DISS. ETH NO. 24957

# MULTI-SENSOR SYSTEM CALIBRATIONS

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

JOERN CHRISTIAN REHDER

Dipl. Ing. in Electrical Engineering,
Hamburg University of Technology

born on 14.06.1984

citizen of Germany

accepted on the recommendation of

Prof. Dr. Roland Y. Siegwart, Examiner
Prof. Dr. Jonathan Kelly, Co-examiner

2018

Institute for Robotics and Intelligent Systems, Autonomous Systems Lab
ETH Zurich
Switzerland

© 2018 Joern Christian Rehder. All rights reserved.

# Abstract

Many applications in robotics require awareness of the state of the robot and its environment. Faced with ambiguities arising in measurements from any single sensor, many applications turn to integrating multiple sensors with often complementary characteristics. This work addresses calibration of multi-sensor systems. It focuses on popular combinations of devices with numerous applications in robotics. Specifically, this work investigates sensor suites comprising cameras and Inertial Measurement Units (IMUs), cameras, an IMU and a Laser Range Finder (LRF), and cameras and an LRF. In this context, it pursues the objectives of providing accurate estimates of the spatial and temporal relations between these sensors and of advancing the understanding of individual measurement models to further improve robustness and accuracy in state estimation.

This thesis builds on a large body of previous work on *continuous-time estimation* and formalizes each calibration problem in terms of probabilistic sensor models. Consequently, each calibration solution lives in the domain of Maximum Likelihood Estimation (MLE), which—under the condition of accurate sensor models—yields the most probable set of parameters to explain the sensor measurements. To this end, it introduces a novel approach to modelling of range measurements recorded by an LRF. This model allows for accurate spatial *and* temporal calibration of the popular combination of cameras and LRFs, yielding precisions in the order of 2 mm, $\frac{1}{10}^\circ$ and $\frac{1}{20}$ms for spatial and temporal parameters respectively. In contrast to established approaches that commonly employ an algebraic error formulation, the probabilistic model is extensible which enables improvements in the understanding of deterministic errors in range measurements. This capability is demonstrated for a deterministic range bias which, if accounted for, improves calibration precision. In many robotic systems, state estimation and low-level controls employ separate IMUs, yielding a need for an accurate estimate of the transformation between these two devices. This thesis proposes a novel estimator that makes use of measurements from all cameras and IMUs in a joint calibration. Building on the same underlying principle, it further advances the model of accelerometers by accounting for different displacements of the sensor structures that perceive specific forces in a single axis. The resulting calibration determines spatial and temporal parameters to precisions of $\frac{1}{5}$mm, $\frac{1}{100}^{\circ}$, and 2 µs respectively—to date the most precise for this class of approaches. Joint calibration is limited in the novel insights it can generate by the least sophisticated sensor model. Consequently, this work explores a more elaborated approach to formulating image sensor measurements. Drawing inspiration from similar approaches in state estimation, it models these directly on intensities rather than on abstracted quantities such as interest point locations. However, currently established direct methods lack a number of important factors. This thesis introduces a chain of models accounting for a number of factors ranging from target illumination over the Point Spread Function (PSF) of the optics to motion blur from camera movements. Results highlight the potential of this approach, for example in rolling-shutter camera calibration, but also the challenges in matching precisions delivered by classical interest point methods.

# Zusammenfassung

Genaue Kenntnisse des Zustands eines Roboters und seiner Umgebung sind die Grundvoraussetzung für den erfolgreichen Betrieb der meisten autonomen Systeme. In vielen Fällen kann es bei der Verwendung eines einzelnen Sensortyps zu Mehrdeutigkeiten bezüglich dieses Zustands kommen.

Diese Arbeit behandelt die Kalibrierung von Systemen, die mehrere Arten von Sensoren verwenden, wobei spezielles Augenmerk auf Kombinationen aus Kameras und Inertialsensoren, Kameras, Inertialsensoren und Laser-Distanzmessern und Kameras und Laser-Distanzmessern gelegt wird. Dabei konzentriert sich die Arbeit auf zwei Hauptziele: Zum einen stellt sie Methoden zur Kalibrierung der räumlichen und zeitlichen Beziehung zwischen den Sensoren vor, zum anderen erweitern die folgenden Kapitel das bestehende Wissen um Sensormodelle mit Mitteln der Kalibrierung und mit dem Ziel, die Zustandsschätzung robuster und genauer zu machen.

Die Arbeit kann dabei auf wichtige Vorleistungen im Bereich der Schätzung kontinuierlicher Zustandsrepräsentationen zurückgreifen. Allen vorgestellten Sensormodellen liegen bewusst gewählte Annahmen über die Verteilung von Messfehlern zugrunde, was “Maximum Likelihood” Kalibrierungen ermöglicht. Durch eine neuartige Formulierung des Messmodells von Laser-Distanzsensoren können diese Messaufnehmer sowohl zeitlich als auch räumlich besser zu anderen Sensoren in Bezug gesetzt werden, wobei Präzisionen von $2\,\mathrm{mm}$, $\frac{1}{10}^\circ$ und $\frac{1}{20}\mathrm{ms}$ für diese räumlichen und zeitlichen Parameter erreicht werden. Gleichzeitig ist das Modell—im Gegensatz zu etablierten Methoden, die nicht auf eine explizite Modellierung der Messfehler setzen—einfach um weitere deterministische Fehlerquellen erweiterbar, wie am Beispiel eines konstanten, additiven Distanzfehlers gezeigt wird.

Viele autonome Systeme sind so aufgebaut, dass Regelung und Zustandsschätzung nicht dieselben Inertialsensoren verwenden, sondern jeweils eigene. Dies macht genaue Kenntnisse der Transformation zwischen den Sensoren zum Betrieb des Systems erforderlich. Diese Arbeit stellt eine Kalibrierungsmethode vor, bei der Messungen aller Inertialsensoren und aller Kameras gleichzeitig verarbeitet werden—der folglich alle Informationen gleichzeitig zur Verfügung stehen. Eine Erweiterung dieses Ansatzes erlaubt es überdies, die genaue Position der Strukturen zu schätzen und fortan zu berücksichtigen, die Beschleunigungen in einzelnen Raumachsen messen. Dieser Ansatz resultiert in Schätzpräzisionen von $\frac{1}{5}$mm und $\frac{1}{100}^{\circ}$ für die räumliche Anordnung und 2 µs für relative zeitliche Fehler in der Zuweisung von Zeitstempeln. Die Kalibrierung eines Systems mit mehreren Sensoren ist ein geeignetes Mittel um tiefere Einblicke in die Modellierung der beteiligten Sensoren zu gewinnen. Die Tiefe dieser Einblicke wird aber wenigstens zum Teil durch das am wenigsten differenzierte Modell begrenzt: Die Suche nach subtileren deterministischen Fehlern in einem Sensor kann durch nicht modellierte Fehler in einem anderen erheblich beeinträchtigt werden. Vor diesem Hintergrund stellt diese Arbeit ein Messmodell für Kameras vor, das anstelle der üblichen Merkmalspositionen direkt Bildintensitäten verwendet. Es stützt sich dabei auf vergleichbare Ansätze der Zustandsschätzung, erweitert diese aber erheblich um den gesteigerten Anforderungen der Kalibrierung nach Modellgenauigkeit gerecht zu werden, wobei speziell die Schätzung der Punktspreizfunktion zur Modellierung der Optik und eine additive Komposition von Bildern zur Nachbildung von Bewegungsunschärfe hervorzuheben sind. Dieser Ansatz hat sich speziell bei der Kalibrierung von Kameras, bei denen individuelle Zeilen sequentiell belichtet werden ("rolling-shutter"), bewährt. Weniger eindeutig ist ihr Vorteil bei der Kalibrierung von Systemen mit Kameras und Inertialsensoren.

# Acknowledgements

First and foremost, I would like to thank my adviser Prof. Roland Siegwart for giving me the opportunity to pursue my doctorate studies at his lab. With the Autonomous Systems Lab, Prof. Siegwart created a truly unique workplace abundant with resources and opportunities, and I consider myself fortunate for having been able to work in such an extraordinary environment. I am deeply grateful for his scientific advice and guidance as well as for his patience and encouragement. I would also like to thank him for opening up opportunities for me (and many others) after completing the doctorate program.

I would further like to thank my second examiner Prof. Jonathan Kelly for his valuable and encouraging feedback that improved this thesis in many ways.

I would like to thank Dr. Paul Furgale for his scientific guidance, honest feedback and advice on how to write scientific papers.

I would like to extend my gratitude to the many outstanding colleagues that I was fortunate to work with over the years. I will always cherish our stimulating discussions during the non-optional coffee breaks.

Finally, I would like to thank my family and Anne for their support and encouragement.

Zurich, 2018 *Joern Rehder*

# Contents

**Abstract** i

**Zusammenfassung** iii

**Acknowledgements** v

**Acronyms** 1

**1 Introduction** 3

- 1.1 Contributions 4
  - 1.1.1 Multi-Camera Synchronization 5
  - 1.1.2 Probabilistic Modelling of Laser Range Finder (LRF) Measurements 6
  - 1.1.3 Multi Inertial Measurement Unit (IMU) Calibration and Estimation of the Displacement of Individual Accelerometer Axes 6
  - 1.1.4 A Direct Camera Measurement Model for Calibration 8
- 1.2 Organization 8

**2 Spatio-Temporal Laser to Visual/Inertial Calibration with Applications to Hand-Held, Large Scale Scanning** 13

- 2.1 Introduction 13
- 2.2 Related Work 14
- 2.3 Methodology 16
  - 2.3.1 Experimental Setup 16
  - 2.3.2 Calibration 16
  - 2.3.3 Model Generation 22
- 2.4 Results 23
- 2.5 Conclusion and Future Work 28

**3 A General Approach to Spatio-Temporal Calibration in Multi-Sensor Systems** 31

3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . 31
3.2 Related Work . . . . . . . . . . . . . . . . . . . . . . . . . . 35
3.3 Theory . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38
3.3.1 Estimating Time Offsets using Basis Functions . . . . . . . . 38
3.3.2 Multi-Sensor Calibration for Combinations of Cameras, an IMU and an LRF . . . . . . . . . . . . . . . . . . . . . . . . 39
3.4 Experiments . . . . . . . . . . . . . . . . . . . . . . . . . . . 50
3.4.1 Equipment . . . . . . . . . . . . . . . . . . . . . . . . . 50
3.4.2 Description of experiments . . . . . . . . . . . . . . . . . . 53
3.4.3 The estimation **J** constitutes a well-posed problem. . . . . . . 54
3.4.4 The temporal offset estimate is accurate to a fraction of the shortest measurement interval . . . . . . . . . . . . . . . . 55
3.4.5 The approach extends to other sensor configuration and synchronization modalities . . . . . . . . . . . . . . . . . . . . 58
3.4.6 The convergence domain suggests practical applicability . . . 62
3.4.7 A closer look at synchronization in software . . . . . . . . . . 65
3.4.8 Delay estimation informs better sensor design . . . . . . . . . 69
3.5 Conclusion and Future Work . . . . . . . . . . . . . . . . . . . 71

**4 Extending *kalibr*: Calibrating the Extrinsics of Multiple IMUs and of Individual Axes 73**
4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . 73
4.2 Related Work . . . . . . . . . . . . . . . . . . . . . . . . . . 75
4.3 Method . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78
4.3.1 Coordinate frame conventions . . . . . . . . . . . . . . . . 78
4.3.2 Accelerometer model . . . . . . . . . . . . . . . . . . . . . 78
4.3.3 Gyroscope model . . . . . . . . . . . . . . . . . . . . . . . 80
4.3.4 The estimator . . . . . . . . . . . . . . . . . . . . . . . . 80
4.4 Experiments . . . . . . . . . . . . . . . . . . . . . . . . . . . 84
4.4.1 Experimental setup and dataset collection . . . . . . . . . . . 84
4.4.2 IMU intrinsics and the extrinsics of multiple IMUs can be precisely inferred in a single estimator . . . . . . . . . . . . 85
4.4.3 Positions of individual accelerometer axes can be discerned . 87
4.4.4 Estimating IMU intrinsics improves camera/IMU extrinsic calibration . . . . . . . . . . . . . . . . . . . . . . . . . . 90
4.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . 90

**5 A Direct Formulation for Camera Calibration 93**
5.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . 93

5.2 Related Work . . . . . . . . . . . . . . . . . . . . . . . . . . 95
5.3 Method . . . . . . . . . . . . . . . . . . . . . . . . . . . . 96
5.3.1 Basic Formulation . . . . . . . . . . . . . . . . . . . . . 97
5.3.2 Applications . . . . . . . . . . . . . . . . . . . . . . . . 101
5.4 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . 104
5.4.1 The direct method exhibits performance similar to established approaches for static images . . . . . . . . . . . . . . . . . . . 106
5.4.2 The approach yields accurate estimates of the line delay in rolling shutter cameras . . . . . . . . . . . . . . . . . . . . . . 106
5.4.3 Exposure time can be accurately inferred from motion blur . . 108
5.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . 111

**6 Camera/IMU Calibration Revisited 113**
6.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . 113
6.2 Method . . . . . . . . . . . . . . . . . . . . . . . . . . . . 115
6.2.1 Problem Statement . . . . . . . . . . . . . . . . . . . . . 115
6.2.2 Coordinate Frame Conventions . . . . . . . . . . . . . . . 116
6.2.3 Estimator Formulation . . . . . . . . . . . . . . . . . . . 116
6.2.4 State Parametrization . . . . . . . . . . . . . . . . . . . . 118
6.2.5 Baseline Camera Measurement Model . . . . . . . . . . . . 119
6.2.6 Direct Camera Measurement Model . . . . . . . . . . . . . 119
6.2.7 IMU Measurement Model . . . . . . . . . . . . . . . . . . 126
6.2.8 Initialization . . . . . . . . . . . . . . . . . . . . . . . . 130
6.3 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . 131
6.3.1 Experimental setup and dataset collection . . . . . . . . . . 131
6.3.2 Appropriate IMU modelling is key to high calibration precision. 134
6.3.3 Exposure time can be accurately inferred from motion blur. . . 137
6.3.4 The direct error formulation yields competitive results. . . . . 137
6.4 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . 139

**7 Conclusion and Future Directions 141**

**List of Publications 145**
Journals . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 145
Conferences . . . . . . . . . . . . . . . . . . . . . . . . . . . 145

**Bibliography 157**

**Curriculum Vitae 159**

# Acronyms

**2D** two-dimensional

**3D** three-dimensional

**ARM** Advanced RISC Machine

**CAD** Computer-Aided Design

**CMOS** Complementary Metal-Oxide-Semiconductor

**EKF** Extended Kalman Filter

**EUROC** European Robotics Challenges

**FPGA** Field Programmable Gate Array

**IA** Accelerometer Input Axes

**IC** Integrated Circuit

**ICP** Iterative Closest Point Algorithm

**IMU** Inertial Measurement Unit

**IRA** Input Reference Axes

**LED** Light-Emitting Diode

**LM** Levenberg-Marquardt

**LRF** Laser Range Finder

**MEMS** Microelectromechanical Systems

**MLE** Maximum Likelihood Estimation

**MSCKF** Multi-State Constraint Kalman Filter

**PCB** Printed Circuit Board

**PDF** Probability Density Function

**PSF** Point Spread Function

**RANSAC** Random Sample Consensus

**RMS** Root Mean Square

**ROS** Robot Operating System

**SLAM** Simultaneous Localization and Mapping

| | |
|---|---|
| **SPKF** | Sigma-Point Kalman Filter |
| **TD-ICP** | Time Delay Iterative Closest Point Algorithm |
| **WVGA** | Wide Video Graphics Array |

# 1

# Introduction

To answer such fundamental question like “*Where am I?*” or “*At which speed am I moving?*”, a robot needs awareness of its own state as well as of that of its surrounding. Consequently, the problem of inferring the underlying state of a setup and its environment from prior knowledge and sensor measurements is central to robotics. In this pursuit, robotic applications commonly deal with the challenge of obtaining sufficient information to unambiguously determine the state. Such ambiguities may for example arise in localization, when a robot equipped with a Laser Range Finder (LRF) travels down a corridor, or in visual odometry, where views of an untextured wall may cause the temporary loss of all feature tracks. In many cases, the problem is efficiently addressed by adding another sensing modality to the platform. Popular combinations of sensors—often coined a *sensor suite*—draw from the complementary strengths of exteroceptive and interoceptive sensors. In the aforementioned examples, the additional sensor adds valuable information about the state. This information could be estimates of the relative motion coming from an Inertial Measurement Unit (IMU) that bridge tracking gaps or ticks from a wheel odometer that provide observations about the travelled distance along the hallway when range measurements only constrain the lateral position of the robot in the corridor.

These examples illustrate that the use of multiple different sensors enables a richer perception of information about the state of the robot. At the same time, their use adds a new level of complexity to state estimation: In addition to modelling the measurement process of the individual sensors, their relationship with respect to each other in the spatial as well as in the temporal domain has to be accounted for—a challenge that does not arise in systems exclusively relying on a single source of measurements. Often, the performance of multi-sensor systems is limited by inadequate modelling of inter-sensor relationships and not by the fidelity of the models of individual sensors. The popular combination of a camera and an IMU is a good example of such a case: Many applications fail to correctly account for the varying delay induced by a fluctuation in exposure time and equally disregard the static temporal offset arising from digital filtering inside the IMU. Yet, best performance can only be achieved in systems where both temporal and spatial relationships between sensors *and* individual measurement processes are accurately modelled. This is where multi-sensor calibration can provide a benefit: By equipping state estimation with accurate insights into the transformation between sensors and the relative offsets in timestamping, calibration can significantly improve the accuracy and robustness of the system.

However, the domain of multi-sensor system calibration is not limited to estimating transformations and delays. Redundancy in the information collected by sensor suites comprising complementary sensing modalities affords the opportunity for an introspection into the sensor suite itself. Accordingly, it can further be a tool to advance the understanding of individual sensor models. Improved modelling of sensor suites is particularly valuable in applications where accuracy is of paramount importance such as inertial-aided photogrammetry or augmented reality applications.

The motivation for this work is rooted in the objective of providing an accurate calibration for multi-sensor systems and advancing the understanding of models of popular sensors. In this pursuit, it investigates specifically sensor suites comprising one or multiple cameras and one or more IMUs as well as combinations of an LRF, cameras, and optionally an IMU.

## 1.1 Contributions

The work compiled in this thesis advances the understanding of multi-sensor systems as well as of individual sensor models in a number of ways.

The contributions are

- a center-exposure synchronization scheme for sensor systems comprising multiple cameras,
- a probabilistic model for laser range measurements and a spatial and temporal maximum likelihood estimator for calibrating sensor systems comprising laser range finders,
- an estimator enabling the calibration of sensor systems comprising multiple IMUs and of the displacement of individual accelerometer axes,
- and a direct formulation for camera calibration.

The following sections will briefly detail on these contributions individually and link to the respective chapters for further reading.

### 1.1.1 Multi-Camera Synchronization

Accurate synchronization is one of the dominating challenges in the design of multi-sensor systems and constitutes a basic prerequisite for correct handling of temporal relations. For sensor suites comprising an IMU and one or multiple cameras, publications commonly propose a strategy that synchronizes cameras through a common trigger signal (e.g. Eling et al. (2015), Grießbach et al. (2012), and Schmid and Hirschmüller (2013)). The time instant of this signal usually determines the timestamp of the respective measurement. Such a synchronization scheme neglects the nature of cameras as a realization of an *integrating* sensor: Instead of perceiving a measurement at an instant, an image sensor integrates irradiance over time. The trigger time relates to the start of the integration process, but it disregards all information about the duration of it. Yet, this duration has a profound impact on the signal, as it leads to motion blur which affects any image, even for short exposure times.

Furgale, Rehder, and Siegwart (2013) confirm that a synchronization approach which assigns the trigger time as measurement instant suffers from an observable delay between camera and IMU that directly depends on camera exposure time. The authors conclude that the mid-exposure time marks a more adequate measurement instant, drawing upon similar observations from the geodesy community (Maune, Photogrammetry, and Sensing, 2007). Nikolic et al. (2014a) implement these findings for a multi-camera system by proposing a novel triggering scheme which compensates for varying exposure times of individual cameras, spacing out *mid-exposure* instants evenly over time and assigning image timestamps accordingly. Such a synchronization yields measurement delays independent of exposure time as demonstrated by Nikolic et al. (2014a) and echoed in Section 3.4.8. It thus enables correct synchronization of multi-camera systems such that measurements of all cameras correspond to the same time instant. Consequently, they can be modelled more correctly as dependent a single, discrete state—an assumption which would be violated by other synchronization schemes.

These efforts in proper design of a multi-sensor system mark the foundation for all subsequent findings in this thesis: The effect of incorrect synchronization can easily eclipse other factors. It would thus severely impact potential insights into more subtle modelling inaccuracies, as for example the displacement of individual accelerometer axes presented in Section 4.3.2.

Finally, confidence in accurate synchronization and spatial and temporal calibration enabled the publication of the European Robotics Challenges (EUROC) visual/inertial dataset (Burri et al., 2016), which has since gained some traction in the robotics community. It also provided the baseline for deriving informed recommendations about synchronization in software as detailed on in Section 3.4.7.

### 1.1.2 Probabilistic Modelling of LRF Measurements

Sensor suites comprising cameras and laser range finders are popular in robotics, and there exist numerous approaches for calibrating the transformation between the two perception modalities (e.g. Bok et al. (2014), Vasconcelos, Barreto, and Nunes (2012), and Zhang and Pless (2004)). However, most of these approaches fall short of the objectives outlined in Chapter 1: Popular methods focus entirely on spatial calibration and completely neglect the temporal relation between sensors. For applications with dynamic motions, a failure to account for temporal offsets between camera and laser range finder measurements can have a profound impact on system performance as highlighted in Fig. 3.1. These effects can easily eclipse those of inaccurately estimated spatial transformations, rendering any calibration efforts focused entirely on spatial relations futile.

Furthermore, established calibration approaches often minimize an algebraic error formulation rather than a quantity informed by a consciously chosen probabilistic model of the sensor. Commonly, these methods employ camera measurements to estimate the parameters of planes in the environment and subsequently model LRF measurement errors as the distances of the transformed range measurements to these planes in the direction of their normals. Such an approach is suboptimal in two respects: It completely disregard uncertainties in camera measurements, and it omits any information about the direction of the laser beam with respect to the plane normal. Consequently, none of the established approaches marks a Maximum Likelihood Estimation (MLE) with respect to any meaningful probabilistic model of the LRF.

Chapters 2 and 3 present a novel method aligned with the idea of providing optimal spatial *and* temporal calibration while advancing the understanding of deterministic errors in range measurements. It treats inputs from camera and LRF consistently and yields transformation estimates that are precise to 2 mm and $\frac{1}{10}$ ° respectively and temporal offset estimates with sub-millisecond precision. The probabilistically motivated LRF model enables further insights into the sensor, suggesting a consistent bias of the range measurements returned by the device under test.

### 1.1.3 Multi IMU Calibration and Estimation of the Displacement of Individual Accelerometer Axes

Numerous robotic systems employ multiple IMUs. The cause for this redundancy is rooted in heterogeneous system design meant to decouple low-level controls from higher level algorithms: For many of these systems, inertial measurements are required–and sufficient–to stabilize the robotic platform. Commonly, a low-level controller with a dedicated IMU is tasked with this stabilization. A secondary IMU is often employed in more advanced estimation of an extended state which serves as input to higher level planning and controls tasks. The spatial relation between these IMUs must be known precisely in order for the low-level controller to correctly execute plans devised in the reference frame of the secondary IMU. Chapter 4 presents a novel estimator for jointly calibrating sensor systems comprising cameras and *multiple* IMUs. The approach makes use of all sensor information at hand in a single MLE.

The significance of the underlying principle is not limited to systems using multiple IMUs, but further informs an advanced model of IMUs in general: Even IMU vendors can be surprisingly ambiguous in reporting the location of the origin of the reference frame for their products. Xsens, for example, specifies the origin as located "*at the accelerometers*" inside their MTi-10 and MTi-100 series (*MTi User Manual*). Given the positioning and dimensions of the accelerometers, this specification constrains the location to about a volume of $1\,\mathrm{cm}^3$, a seemingly unacceptable tolerance for applications where accuracy is crucial and for sensor setups with displacements between cameras and IMU of a few centimeters.

Chapters 4 and 6 highlight the fact that the finite extent of the sensing elements in Microelectromechanical Systems (MEMS) accelerometers invalidates the idea that the specific force in perceived in one spot and, as a consequence, any notion of a well defined origin of the IMU reference frame as a whole. This applies equally to setups featuring individual Integrated Circuits (ICs), such as the aforementioned Xsens device, and to more integrated IMUs which als comprise MEMS structures of finite size. This work advances the prevalent model for accelerometers to further account for the displacement of individual accelerometer axes. It derives a novel estimator that determines these displacements along with IMU scale factor and misalignment corrections. Fig. 6.6a demonstrates that calibration precision increases dramatically when all of these parameters are taken into account. The implications of these findings extend beyond calibration: To date and to our knowledge, all existing approaches to visual/inertial state estimation neglect the individual displacements of the axes in accelerometers. Yet, system designers often integrate large inertial sensor units composed of multiple ICs due to their seemingly superior specifications. Chapters 4 and 6 add a previously neglected perspective on these specifications and raise the question of whether the advantages of these devices in terms of better noise performance and increased bias stability can be leveraged in all applications.

### 1.1.4 A Direct Camera Measurement Model for Calibration

Chapters 4 and 6 as well as the literature (Furgale, Rehder, and Siegwart, 2013; Nikolic et al., 2016b) demonstrate extraordinary precision in estimating constant temporal offsets for camera/IMU systems—often in the order of merely a few microseconds. In indoor settings, exposure time typically ranges in the order of milliseconds and is thus significantly larger than the estimation precision of constant temporal offsets. This discrepancy challenges the universal truth in mid-exposure time being the most adequate measurement instant to assign to an interest point detection: Such an observation is the output of a comparatively complex function of local image gradients. Given the aforementioned integrating characteristics of image sensors, it is not immediately clear why the result of this detection function should always correspond to the projection of a target point at mid-exposure time, independently of the camera motion during image exposure.

To circumvent this issue and ultimately improve camera/IMU calibration, Chapters 5 and 6 explore different routes to directly model intensity measurements–a novelty for geometric calibration. By formulating the measurements as image intensities rather than as abstracted quantities, motion blur can be consistently folded into the model. This capability alleviates the approach from speculatively assigning timestamps at finer granularity than exposure time. As a side product, the direct calibration method enables a novel path to calibrating the line delay in rolling-shutter cameras, resulting in a more intuitive treatment of measurement uncertainties.

However, abolishing the advantages of abstraction provided by interest point detection in favor of a more accurate modelling of motion blur comes at a cost: Predicting accurate intensities entails a host of modelling challenges ranging from uneven illumination of the target over reproducing the camera response function to estimating the Point Spread Function (PSF) of the camera setup. Chapter 6 derives this entire modelling chain and introduces estimation of the PSF as an integral part of the calibration pipeline. It proposes a novel, *joint* estimator for camera poses and PSF. The results raise the question of whether established direct approaches to visual state estimation may equally benefit from modelling these optical effects to higher fidelity .

## 1.2 Organization

This work is organized as a collection of peer-review publications. These map to the aforementioned contributions as follows.

## Chapter 2

This chapter has been published as

> Rehder, J., Beardsley, P., Siegwart, R., and Furgale, P. (2014). „Spatio-Temporal Laser to Visual/Inertial Calibration with Applications to Hand-Held, Large Scale Scanning“. In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Chicago, IL, USA, pp. 459–465

It introduces the probabilistic LRF measurement model and demonstrates joint spatial and temporal calibration of a sensor suite comprising a stereo camera, an IMU, and an LRF. This chapter further highlights the importance of accurate synchronization with examples from three-dimensional (3D) reconstruction.

## Chapter 3

This chapter has been published as

> Rehder, J., Siegwart, R., and Furgale, P. (2016). „A General Approach to Spatiotemporal Calibration in Multisensor Systems“. *IEEE Transactions on Robotics* 32.2, pp. 383–398

It further comprises findings from

> Furgale, P., Rehder, J., and Siegwart, R. (2013). „Unified Temporal and Spatial Calibration for Multi-Sensor Systems“. In: *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*

> Nikolic, J., Rehder, J., Burri, M., Gohl, P., Leutenegger, S., Furgale, P. T., and Siegwart, R. (2014a). „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)*, pp. 431–437

This chapter broadens the use-case of the LRF model to sensor suites consisting only of cameras and an LRF. It further extends the model to account for deterministic biases and contributes a comprehensive experimental analysis. This analysis includes an investigation into the efforts necessary to synchronize devices in software. It also demonstrates that aligning mid-exposure times in multi-camera systems and assigning timestamps accordingly marks an effective strategy to mitigate exposure depended relative time offsets between sensors.

## Chapter 4

This chapter has been published as

> Rehder, J., Nikolic, J., Schneider, T., Hinzmann, T., and Siegwart, R. (2016). „Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes“. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 4304–4311

It extends work by Furgale, Rehder, and Siegwart (2013) to systems comprising multiple IMUs and by augmenting the inertial sensor model with additional intrinsic parameters. This chapter introduces the displacement of individual accelerometer axes as an intrinsic parameter with significant impact on calibration performance.

## Chapter 5

This chapter has been published as

> Rehder, J., Nikolic, J., Schneider, T., and Siegwart, R. (2017). „A Direct Formulation for Camera Calibration“. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 6479–6486

It applies the concept of direct modelling of image measurements to calibration. This approach yields an intuitive formulation of rolling-shutter camera calibration with correct treatment of uncertainties. It further enables the estimation of exposure time from motion blur in images.

## Chapter 6

This chapter has been published as

> Rehder, J. and Siegwart, R. (2017). „Camera/IMU Calibration Revisited“. *IEEE Sensors Journal* 17.11, pp. 3257–3268

It combines and extends findings from

> Rehder, J., Nikolic, J., Schneider, T., Hinzmann, T., and Siegwart, R. (2016). „Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes“. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 4304–4311

> Rehder, J., Nikolic, J., Schneider, T., and Siegwart, R. (2017). „A Direct Formulation for Camera Calibration“. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 6479–6486

To this end, it introduces an improved formulation of the direct camera model, which further accounts for uneven illumination and the PSF of the optics. A comprehensive experimental analysis showcases the impact of high-fidelity IMU models and highlights the complexity of accurate intensity modelling.

# 2

# Spatio-Temporal Laser to Visual/Inertial Calibration with Applications to Hand-Held, Large Scale Scanning

Joern Rehder, Paul Beardsley, Roland Siegwart, and Paul Furgale

## 2.1 Introduction

Time-of-flight laser scanning for three-dimensional (3D) reconstruction is a mature technology with applications in fields ranging from reverse engineering of industrial plants, to architecture, to archaeology (Leica Geosystems, 2014). However, the vast majority of commercially available scanners operates stationarily, and in order to completely capture more complex environments where occlusions are present, the device has to be repositioned multiple times. On the other hand, triangulating, hand-held 3D scanners for small scale objects are widely applied in industry due to their easy deployment (Nikon, 2014).

This work is motivated by the goal of developing a system that can be used to scan large scale structures, but that provides the same ease of deployment of a hand-held sensor, like the one depicted in Fig.2.1, comprised of a Hokuyo UTM-30LX

Published in:
*Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2014,*
*©2014 IEEE*
*DOI:* `10.1109/iros.2014.6942599`

Laser Range Finder (LRF), rigidly connected to a visual/inertial sensor (Nikolic et al., 2014b). Arising from the extended range of dynamic motions of a human operator as compared to a ground vehicle, we employ a novel calibration approach that extends current state of the art by estimating the spatial transformation between the sensors as well as the time offset at which measurements are recorded. Additionally, our framework allows for the calibration of a broader variety of sensor configurations by discarding of the requirement of an overlapping field of view between camera and LRF.

To this end, a continuous-time batch estimation framework (Furgale, Barfoot, and Sibley, 2012) is employed, extending our previous work on camera/Inertial Measurement Unit (IMU) calibration (Furgale, Rehder, and Siegwart, 2013) to integrate laser range measurements. As we perceive the requirement of an overlap in the field of view between the camera and the laser as unnecessarily constraining, ubiquitous planes are identified in the scan data and exploited for modelling the range measurements.

Section 2.4 presents quantitative results for the calibration, demonstrating that the transformation can be estimated to millimeter accuracy and the time delay is determined up to about 5 ms precision. Furthermore, point cloud reconstructions obtained with our system prove the accuracy of the overall system and are well comparable with results reported for similar hand-held scanners (Bok et al., 2011; James and Quinton, 2014).

## 2.2 Related Work

Intensive research has gone into calibrating the transformation between laser scanners and cameras. Many approaches are designed for setups where the scanning plane is rotated (Alismail, Baker, and Browning, 2012; Scaramuzza, Harati, and Siegwart, 2007; Unnikrishnan and Hebert, 2005), or for multi-beam systems (Geiger et al., 2012; Mirzaei, Kottas, and Roumeliotis, 2012; Pandey et al., 2012), and are hence not applicable to our setup. For calibrating a setup of a rigidly connected camera and single-beam laser scanner, Zhang and Pless (2004) proposed an approach, where a set of simultaneously acquired images and static scans of a planar calibration target is used to establish the transformation between the two sensors. Other groups improved upon this algorithm (Li et al., 2007; Vasconcelos, Barreto, and Nunes, 2012), while maintaining the same fundamental principle. Similarly, Núñez et al. (2009) use simultaneous observations of a planar pattern, but additionally employ an inertial measurement unit to further constrain the problem. Mei and Rives (2006) present an algorithm that makes use of the laser trace being visible in the image, which, while

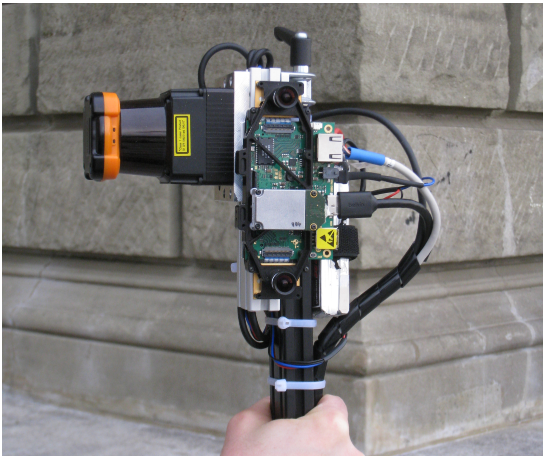


Figure 2.1: The handheld scanning device comprised of a Hokuyo UTM-30LX laser range finder and a visual/inertial sensor (Nikolic et al., 2014b).

not relaxing the requirement of an overlapping field of view, constitutes a different approach to calibration. Bok et al. (2011) generally follow the calibration procedure of Zhang and Pless (2004), but additionally extract measurements of the edges of the calibration pattern from laser data to improve the results. Finally, Moghadam, Bosse, and Zlot (2013) proposes a calibration method that matches edges detected in the image to plane intersections and boundaries in point clouds recorded with their Zebedee system (Bosse, Zlot, and Flick, 2012). The approach is capable of calibrating for devices, where the field of view of the camera does not overlap with the field of view of the range sensor. However, it requires the setup to be able to generate an accurate point cloud irrespectively of the transformation that is calibrated for, and hence is not applicable to our case.

With the exception of Moghadam, Bosse, and Zlot (2013), these approaches have in common that they calibrate the setup based on static scans and completely neglect the temporal relationship between camera and laser scanner. While this might be sufficient for platforms with slow dynamics, it may result in significantly distorted reconstructions for hand-held systems, where angular velocities can reach hundreds of degrees per second. In contrast to these stationary calibration approaches, our calibration is based on continuous-time batch estimation (Furgale, Barfoot, and Sibley, 2012), which allows for a seamless integration of time delays into the calibration framework, and it is an extension of Furgale, Rehder, and Siegwart (2013). The Zebedee system (Bosse, Zlot, and Flick, 2012) continuously estimates the delay between different sensors in operation. While we can see the beauty in this system, our work takes a different approach: In order to increase robustness and decrease the size of the state estimated online, we try to accurately calibrate such quantities beforehand in an offline procedure in a lab environment.

## 2.3 Methodology

### 2.3.1 Experimental Setup

The scanning device is based on the visual/inertial sensor detailed in Nikolic et al. (2014b). This sensor combines two global shutter MT9V034 Wide Video Graphics Array (WVGA) image sensors in a plane-parallel stereo setup with an ADIS16448 inertial measurement unit. The integration of a XILINX Zynq, a combination of a dual core Advanced RISC Machine (ARM) processor with Field Programmable Gate Array (FPGA) fabric, allows for accurate, exposure-compensated triggering of the cameras as well as synchronized polling of inertial data. The sensor has been augmented with a Hokuyo UTM-30LX, which has been rigidly mounted to the visual/inertial sensor.

### 2.3.2 Calibration

Our calibration is based on the continuous-time batch estimation framework proposed in Furgale, Barfoot, and Sibley (2012). In order to estimate the transformation between the laser range finder and the visual/inertial sensor and the inter-sensor delays, we extend our previous work on visual/inertial calibration presented in Furgale, Rehder, and Siegwart (2013). In the following, a brief recapitulation of the visual/inertial calibration framework will be provided, before the contribution to the objective function arising from laser measurements is derived in detail. With this, the description of the algorithm closely follows its processing procedure, as a two step approach is employed, where a smooth sensor path is estimated in a first step, followed by a step that adds laser terms to the estimation. We follow this two step approach, since a sufficiently accurate sensor trajectory is a prerequisite for obtaining an initial point cloud, which in turn is used to obtain a model for the laser measurements. The calibration procedure itself is similar—the setup is waved in front of a checkerboard in a way that excites all degrees of freedom sufficiently to render the calibration parameters observable—but we additionally require the sequence to be recorded in an environment, where a subset of the laser measurements are induced by at least one plane.

### Recapitulation Camera/IMU Calibration

We employ B-splines to represent time-varying states and parametrize the time-varying transformation from the inertial coordinate frame into the world frame as a $6 \times 1$ spline, applying a Euclidean parametrization to translations and an angle/axis representation to orientations. With $\mathbf{C}(\cdot)$ being a function that constructs a rotation matrix from our orientation parametrization $\varphi$ and $\mathbf{t}$ being the translation, the transformation from the body reference frame defined to coincide with the one of the IMU into the world reference frame $\mathbf{T}_{w,i}$ at time $t$ may be expressed as

$$\mathbf{T}_{w,i}(t) = \begin{bmatrix} \mathbf{C}(\varphi(t)) & \mathbf{t}(t) \\ \mathbf{0}^T & 1 \end{bmatrix} \quad . \tag{2.1}$$

Given translations represented as composition of continuously differentiable basis functions, velocities $\mathbf{v}(t)$ and accelerations $\mathbf{a}(t)$ can be obtained by derivation. Angular velocities $\omega(t)$ are obtained similarly with an additional transformation $\mathbf{S}(\cdot)$ relating parameter rates to angular velocities.

With this, the contributions from visual and inertial measurements to the objective function are

$$\mathbf{e}_{y_{mj}} := \mathbf{y}_{mj} - \mathbf{h}\left(\mathbf{T}_{c,i}\mathbf{T}_{w,i}(t_j)^{-1}\mathbf{p}_w^m\right) \tag{2.2a}$$

$$J_y := \frac{1}{2}\sum_{j=1}^{J}\sum_{m=1}^{M} \mathbf{e}_{y_{mj}}^T \mathbf{R}_{y_{mj}}^{-1} \mathbf{e}_{y_{mj}} \tag{2.2b}$$

$$\mathbf{e}_{\alpha_k} := \alpha_k - \mathbf{C}\left(\varphi(t_k)\right)^T \left(\mathbf{a}(t_k) - \mathbf{g}_w\right) + \mathbf{b}_a(t_k) \tag{2.2c}$$

$$J_\alpha := \frac{1}{2}\sum_{k=1}^{K} \mathbf{e}_{\alpha_k}^T \mathbf{R}_{\alpha_k}^{-1} \mathbf{e}_{\alpha_k} \tag{2.2d}$$

$$\mathbf{e}_{\omega_k} := \varpi_k - \mathbf{C}\left(\varphi(t_k)\right)^T \omega(t_k) + \mathbf{b}_\omega(t_k) \tag{2.2e}$$

$$J_\omega := \frac{1}{2}\sum_{k=1}^{K} \mathbf{e}_{\omega_k}^T \mathbf{R}_{\omega_k}^{-1} \mathbf{e}_{\omega_k} \tag{2.2f}$$

$$\mathbf{e}_{b_a}(t) := \dot{\mathbf{b}}_a(t) \tag{2.2g}$$

$$J_{b_a} := \frac{1}{2}\int_{t_1}^{t_K} \mathbf{e}_{b_a}(\tau)^T \mathbf{Q}_a^{-1} \mathbf{e}_{b_a}(\tau)\, d\tau \tag{2.2h}$$

$$\mathbf{e}_{b_\omega}(t) := \dot{\mathbf{b}}_\omega(t) \tag{2.2i}$$

$$J_{b_\omega} := \frac{1}{2}\int_{t_1}^{t_K} \mathbf{e}_{b_\omega}(\tau)^T \mathbf{Q}_\omega^{-1} \mathbf{e}_{b_\omega}(\tau)\, d\tau \tag{2.2j}$$

where $\mathbf{h}(\cdot)$ is an arbitrary projection model, accepting checkerboard corners $\mathbf{p}_w^m$ transformed from the world frame into the camera frame via the transformation $\mathbf{T}_{w,i}^{-1}$ at time $t_j$ and the rigid transformation between inertial and camera frame $\mathbf{T}_{c,i}$. In contrast to Furgale, Rehder, and Siegwart (2013), a time delay between camera and IMU is not considered, since it is compensated for in hardware (Nikolic et al., 2014b). Inertial measurements at times $t_k$ contribute accordingly, with $\mathbf{g}_w$ being the estimated gravity and $\mathbf{b}$ denoting inertial sensor biases, parametrized as B-splines and modelled as driven by zero-mean white Gaussian processes, governed by covariance $\mathbf{Q}$. All measurements are weighted according to the inverse covariances, $\mathbf{R}^{-1}$, of the additive, zero-mean Gaussian distributed perturbation assumed to corrupt the measurement.

With these terms, the initial objective function $J$ for estimating the sensor trajectory is composed as $J := J_y + J_\alpha + J_\omega + J_{b_a} + J_{b_w}$, which is minimized using the Levenberg-Marquardt (LM) algorithm (Marquardt, 1963). Note that with the error term formulation provided above, this constitutes the maximum-likelihood estimation assuming that the perturbation model is sufficiently accurate.

### Incorporating Laser Range Measurements

In the following section, we will derive the approach to modelling laser range measurements for a seamless integration into the objective function $J$. While there exist calibration approaches based on minimizing point cloud entropy (Sheehan, Harrison, and Newman, 2010) that omit assumptions about the scanned structure, we decided to explicitly model the laser range measurements as induced by a structure. For this, some knowledge about the structure is indispensable. Due to their ubiquity, we decided to identify laser measurements induced by planes and model them accordingly. Not relying on a calibration pattern for modelling laser measurements has distinct advantages: Intuitively, the observability of the transformation improves with target size, and manufacturing large targets can quickly become impractical. In contrast to this approach, most calibration approaches rely on an overlapping field of view between camera and laser (Li et al., 2007; Núñez et al., 2009; Vasconcelos, Barreto, and Nunes, 2012; Zhang and Pless, 2004), which limits the applicable sensor configurations. For modelling the laser measurement, we make the following assumptions.

- The visual/inertial sensor is capable of estimating a sufficiently accurate trajectory.
- The initial guess for the transformation between laser and sensor is sufficiently accurate.

- A subset of all individual distances measured by the LRF is induced by planar structures.
- The range measurements are corrupted by additive zero-mean Gaussian distributed noise.
- There is zero error on the beam directions reported by the range finder.

Of these assumptions, the accuracy of the initial estimate of the transformation is the most constraining, as it may be hard to obtain for some setups. In this context, accuracy is sufficient if planes can reliably be identified in the point cloud obtained from the sensor trajectory and with the initial transformation, which is the case when a majority of laser measurements induced by a plane falls within an envelop defined by the chosen Random Sample Consensus (RANSAC) threshold (Fischler and Bolles, 1981).

Given a continuous trajectory of the sensor estimated by minimizing $J$ as defined in Section 2.3.2 and an initial guess of the transformation $\mathbf{T}_{i,l}$ between the inertial sensor and the laser, an initial point cloud can be obtained by transforming the laser measurements $\mathbf{m}_k = [\alpha_k, l_k]^T$ of a single beam, cast at angle $\alpha_k$ and measuring range $l_k$ into a common coordinate frame:

$$[\lambda \mathbf{p}_k, \lambda]^T = \mathbf{T}_{w,i}(t_k + \delta t)\mathbf{T}_{i,l}\,[l_k cos(\alpha_k), l_k sin(\alpha_k), 0, 1]^T \,, \tag{2.3}$$

where $\mathbf{T}_{w,i}$ denotes the transformation from the inertial into the world coordinate frame, $t_k$ is the timestamp of a laser range measurement $l_k$ with corresponding beam angle $\alpha_k$, and $\delta t$ the unknown inter-sensor delay. Fig. 2.2a shows an initial point cloud acquired from a calibration sequence. While some planes are visible, the entire point cloud appears cluttered, with many measurements not stemming from planar structures. In order to identify measurements induced by planes, a RANSAC scheme is applied to the point cloud, with plane hypotheses generated from a minimal set of three non-collinear points and model support being evaluated according to threshold $t$

$$|\mathbf{n}_i^T \mathbf{p}_k - d_i| < t \quad , \tag{2.4}$$

with $\mathbf{n}_i$ being the normal of plane $i$, $d_i$ the distance to the origin, and $\mathbf{p}_k$ being a point evaluated for support. Note that not only measurements induced by this plane satisfy this condition, but points on any structure within the threshold $t$ from the intersection with the plane defined by $\mathbf{n}_i$ and $d_i$. To avoid outliers, the points on each plane are clustered by evenly discretizing them into spatial bins and, starting from the most populated bin, invoking adjacent bins into the plane until the ratio of points in adjacent bins falls below a threshold. As the laser itself scans in a plane, resting the sensor during calibration may result in an accumulation of points that may be detected as a plane in the RANSAC step. To avoid picking such accumulations as planes, we further require the number of points that fall into bins invoked into the plane to represent a certain percentage of all potential plane points and discard of the plane hypothesis otherwise. Having identified the measurements induced by plane $i$ in the environment, and assuming zero error on the beam angle, a prediction of the range $\hat{l}_k$ measured by the LRF can be modelled as

$$\hat{l}_k = \left| \frac{\mathbf{n}_i^T \mathbf{t}_{w,l}(t_k + \delta t) - d_i}{\mathbf{n}_i^T \mathbf{r}_k(t_k + \delta t)} \right| \quad , \tag{2.5}$$

with

$$\mathbf{t}_{w,l}(t_k + \delta t) = \mathbf{C}_{w,i}(t_k + \delta t)\mathbf{t}_{i,l} + \mathbf{t}_{w,i}(t_k + \delta t) \tag{2.6}$$

$$\mathbf{r}_k(t_k + \delta t) = \mathbf{C}_{w,i}(t_k + \delta t)\mathbf{C}_{i,l} \left[cos(\alpha_k), sin(\alpha_k), 0\right]^T . \tag{2.7}$$

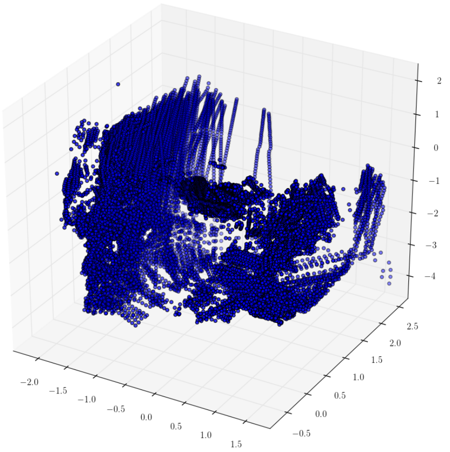
With this, the contribution of the laser range measurements to the objective function $J$ can be determined as

$$\mathbf{e}_l := \hat{l}_k - l_k \tag{2.8a}$$

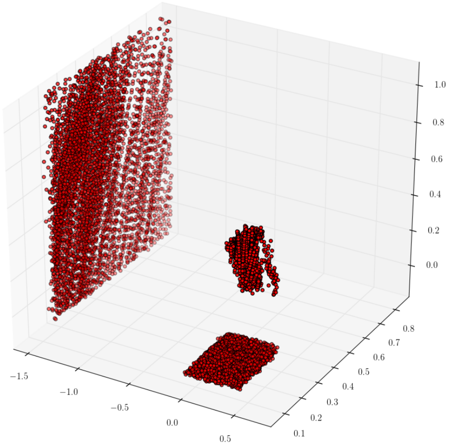
$$J_l := \sum_{i=1}^{I} \sum_{k=1}^{K} \mathbf{e}_l^T \mathbf{R}^{-1} \mathbf{e}_l \tag{2.8b}$$

where $\mathbf{R}$ denotes covariance on the range measurements. The laser calibration quantities $\mathbf{T}_{i,l}$ and $\delta t$—along with the plane parameters $\mathbf{n}_i$ and $d_i$—are then estimated by minimizing the augmented objective function $J := J_y + J_\alpha + J_\omega + J_{b_a} + J_{b_w} + J_l$ analogously to Section 2.3.2, using the result of the trajectory generation as initial guess. To improve robustness to outliers, we further employ the Huber cost function (Hartley and Zisserman, 2000) for the LRF error terms.

Note that the number of planes identified in the RANSAC step is a design parameter that can be adapted to the number of dominant planes in the environment scanned during calibration. Also note that we chose to implement the two different error metrics mostly for convenience: While the distance to the plane in the RANSAC step can be evaluated rapidly with minimal data association, the second metric models the plane induced distance measurement more accurately.



(a)



(b)

Figure 2.2: Different stages of point cloud processing during calibration. Fig. 2.2a shows the initial point cloud generated from a calibration sequence. While planes are visible, the point cloud is cluttered with measurements stemming from non planar objects. Fig. 2.2b depicts the point cloud after identifying the three most dominant planes and clustering.

### 2.3.3 Model Generation

After calibration, the rig can be used to generate 3D models that consist of points from the LRF colored by intensity obtained with the cameras. To estimate the ego-motion of the sensor, the approach presented in Leutenegger et al. (2015) is employed. This algorithm administers a non-linear optimization over a sliding window of key-frames and corresponding inertial measurements. In order to keep computational complexity at bay and allow for real-time operation, entries that correspond to past velocities and inertial sensor biases are continuously marginalized from the estimated state. A detailed description of this framework is outside of the scope of this work, and the interested reader is kindly referred to the original publication.

We use the Robot Operating System (ROS) framework (Quigley et al., 2009) for transforming the laser range measurements from the coordinate frame of the scanner into the global coordinate frame given the static transformation between the laser and the visual/inertial sensor and state estimates from our visual/inertial framework. State estimates are published at the rate of the inertial measurements and interpolated to accommodate for the fact that the range measurements of a single scan are not acquired instantaneously but consecutively over a period of multiple milliseconds.

The laser scan data is augmented with image intensity values using the static transform between the two sensors. To this end, the range measurements are transformed into the camera coordinate frame and projected into the camera using its previously calibrated intrinsics and distortion model. Intensities are sampled at the projections of the laser range respective measurements. This approach constitutes a rather ad-hoc method to laser point cloud coloring, as it neglects the consecutive nature of laser range scans and does not take varying exposure times and gains in successive images into account. There are other approaches that employ more sophisticated coloring schemes, e.g. Chenga et al. (2014). However in that approach, global exposure and gain equalization is performed in post-processing, which makes the method less well suited for immediate model feedback. Furthermore, by considering a single image and scan, the necessity for sophisticated occlusion reasoning is reduced. Although Fig. 2.5 seems to provide a counter argument for that, given that the method would sample incorrect intensity values for the cupboard occluded by the upper left corner of the checkerboard, one has to consider that this problem is induced by the baseline of the laser with respect to the camera, which is mitigated by the distance to the object, and thus less apparent for the majority of scanning use cases, as obvious from the correctly textured building shown in Fig. 2.7.

## 2.4 Results

All following experiments employed the sensor device detailed on in Section 2.3.1, hand guided by an operator. To this end, stereo pairs were recorded on a laptop computer at a rate of 20 Hz, inertial measurements at 200 Hz, and laser range scans at 40 Hz.

### Calibration Results

This section details on results for the spatial and temporal calibration of the LRF with respect to the visual/inertial sensor. Apart from measurement covariances of all sensors and the parameters for the Gaussian processes modelling inertial biases, an initial guess for the transformation between LRF and sensor is required along with a set of free parameters, that can be roughly categorized into parameters inherent to continuous-time batch estimation and parameters governing plane identification and robust estimation. For the continuous time estimation, we employed 6th order B-splines with 120 support points per second. We chose to identify three planes and picked a RANSAC threshold $t$ of 50 mm, allowing for some envelope around the plane to account for errors due to time-delay and the inaccurate initial guess for the inter-sensor calibration. We stopped invoking bins into the plane when the neighbouring bin contained less then half as many points as the current, and discarded a plane hypotheses as potentially having been induced by the scanning plane rather than a structure when the number of clustered points represented less than half of all plane inlier. For down-weighting potential outliers via the Huber cost function, an outlier threshold of 20 mm was chosen in accordance with the measurement accuracies of the Hokuyo as reported by the manufacturer. For all experiments, we assumed zero delay and zero displacement of the laser with respect to the visual/inertial sensor. The relative orientation between the two in Euler angles in XZY convention was approximated by 180°, 0° and 90° Note that the correct selection of these parameters is crucial for achieving accurate calibration results, and we noticed that the approach is sensitive to a correct choice of the RANSAC threshold, which we found to be best set within the range of the accuracy of the LRF.

Fig. 2.3 displays the distribution in errors of the modelled range readings with respect to measured distance for a single plane in one dataset of about 10 s. After initializing the point cloud, but prior to performing the estimation (Fig. 2.3a) errors span an interval of up to 12 cm distance. Optimizing over the augmented cost function and including the spatio-temporal system parameters as well as plane parameters drives down the error to within the tolerance reported by the datasheet of the LRF, with few outliers (Fig. 2.3b). Neglecting the temporal relationship between the sensors results in a broader distribution of errors, with many modelled measurements falling outside of the range specified by the manufacturer (Fig. 2.3c). This suggests that neglecting the temporal relationship between the LRF and other sensors results in impaired reconstruction results.

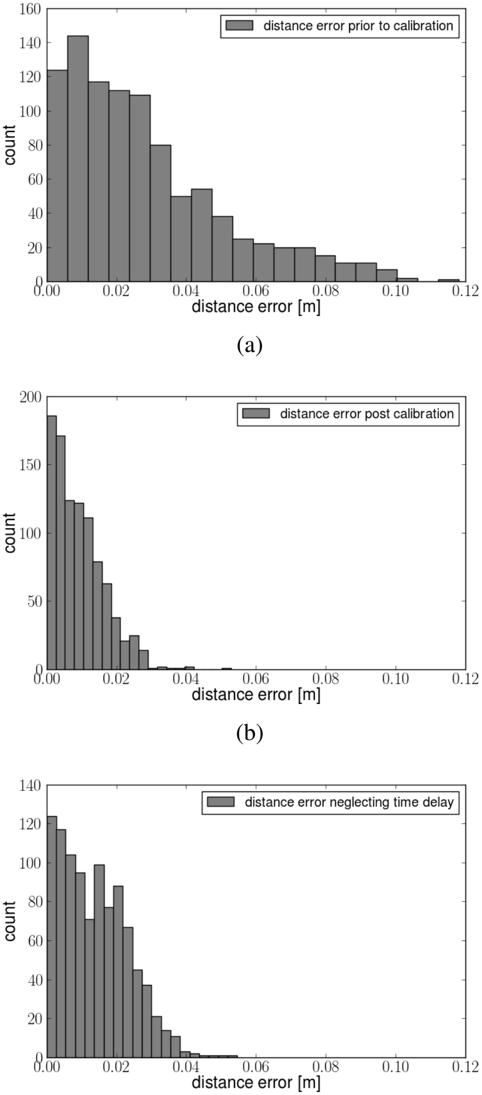
Fig. 2.4 visualizes an experiment for evaluating the accuracy of the temporal calibration. As we lack information on the true delay between the sensors, we simulated different delays by altering the timestamps of the range measurements by $-20\,\mathrm{ms}$, $0\,\mathrm{ms}$, and $20\,\mathrm{ms}$ respectively. We then estimated each time-offset on ten datasets of 4 seconds of data taken from a longer calibration sequence. Of these ten estimates, the mean delays are $47.5\,\mathrm{ms}$, $64.5\,\mathrm{ms}$, and $82.2\,\mathrm{ms}$ with standard deviations $\sigma$ of $4.78\,\mathrm{ms}$, $5.49\,\mathrm{ms}$, and $6.26\,\mathrm{ms}$ respectively. These results show that the approach is capable of estimating the simulated additions, which suggests that it can measure the absolute offset with similar accuracy.

In order to evaluate the spatial calibration, a total of 20 datasets, each of about 40 seconds length, was recorded. The mean and standard deviation of all translation estimates expressed in the coordinate frame of the LRF is $6.34\,\mathrm{cm} \pm 0.84\,\mathrm{cm}$, $-1.59\,\mathrm{cm} \pm 0.64\,\mathrm{cm}$, and $-7.97\,\mathrm{cm} \pm 2.75\,\mathrm{cm}$, as compared to approximate hand measurements of $6.2\,\mathrm{cm}$, $-1.4\,\mathrm{cm}$, and $-9.4\,\mathrm{cm}$. The large uncertainty in estimating the z component of the translation may be explained with a lack of rotational excitement in one axis in the datasets, likely caused both by a conservative operation by the user and by the narrow field of view of the image sensors that causes the sensor to lose track of the calibration pattern more easily when rotating in a certain axis. Orientation in Euler angles in XZY convention was estimated to be $179.73^\circ \pm 0.10^\circ$, $-1.47^\circ \pm 0.47^\circ$, and $90.78^\circ \pm 0.46^\circ$.

Fig. 2.5 provides an visual impression of the accuracy of the spatial calibration. As the image sensors on our device are susceptible to infra-red light, the scan line is visible as illuminated band on table and calibration pattern. The image is superimposed with the scan line modelled using the initial guess (blue) and the transformation as estimated by the framework (red). Although the intersection of the scan plane with two planes leaves unconstrained degrees of freedom, the resemblance of modelled and observed scan line in combination with the other results presented in this work suggest an accurate spatial calibration between the LRF and the visual/inertial sensor.

### Preliminary Reconstruction Results

Fig. 2.6 provides the motivation for this work: In order to illustrate the necessity of a sufficient temporal calibration, the range scan measurements for generating the detail view in Fig. 2.6a were artificially delayed by $60\,\mathrm{ms}$. The resulting point cloud



(a)

(b)

(c)

Figure 2.3: Histogram of laser distance errors prior to (Fig. 2.3a) and after (Fig. 2.3b) estimation of the inter-sensor transformation and delay. Fig. 2.3c depicts results when the time delay between sensors is neglected, resulting in larger overall errors.

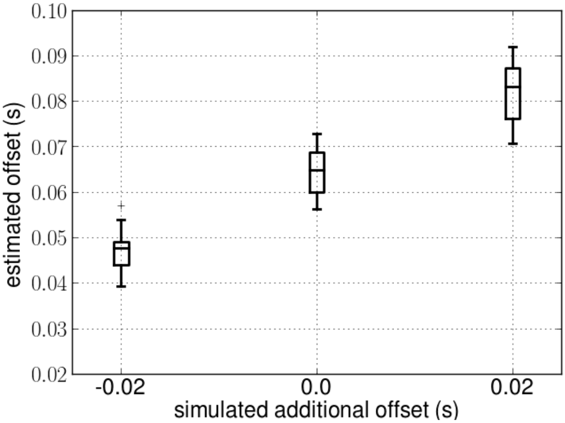


Figure 2.4: Distribution of the intersensor delay. As ground truth was not available for the delay, the accuracy was evaluated by adding a simulated delay of $-20\,\text{ms}$, $0\,\text{ms}$, and $20\,\text{ms}$ respectively to sets of each ten calibration sequences. Results suggest that the proposed approach is capable of accurately estimating the time delay for this sensor combination.

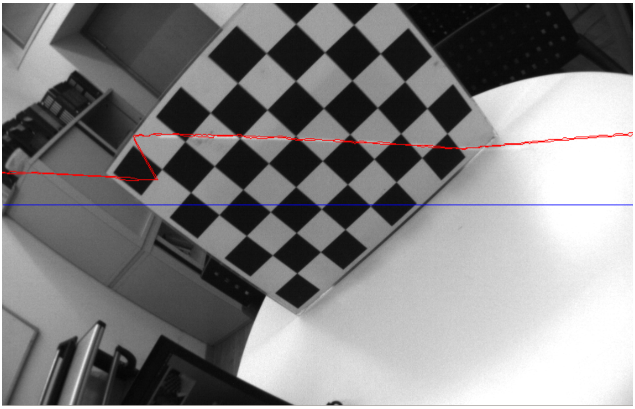
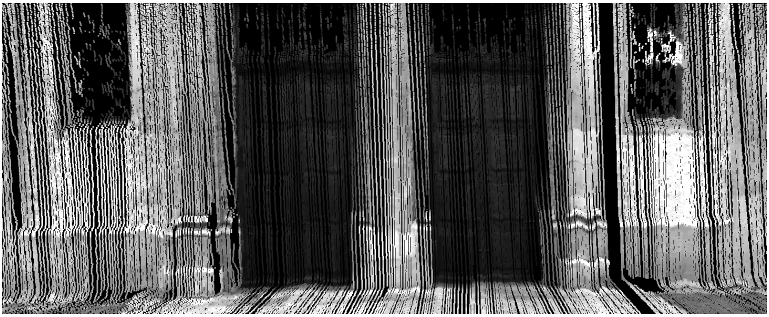
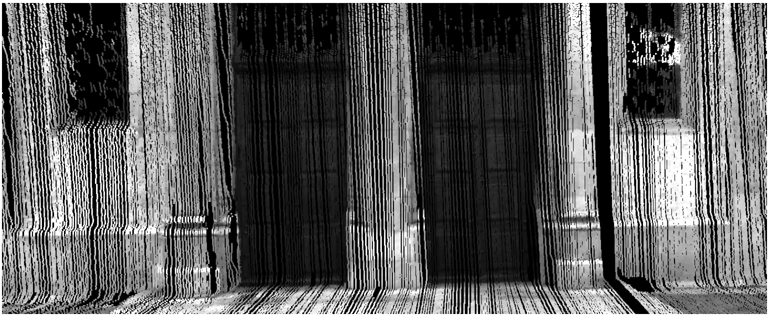


Figure 2.5: Illustration of the calibration accuracy. The laser scanner emits infra-red light that the image sensors are susceptible to, rendering some of the points sampled by the laser range finder visible. The image is superimposed with a projection of the laser measurement into the camera using the initial guess (blue) and the calibrated transformation (red). The degree of coincidence of projected and visible scan line suggests an accurate calibration of the transformation between the sensors.

exhibits noticeable distortions and incorrect coloring, most apparent for the right column, while correctly synchronized measurements allow for a crisp reconstruction(see Fig. 2.6b).



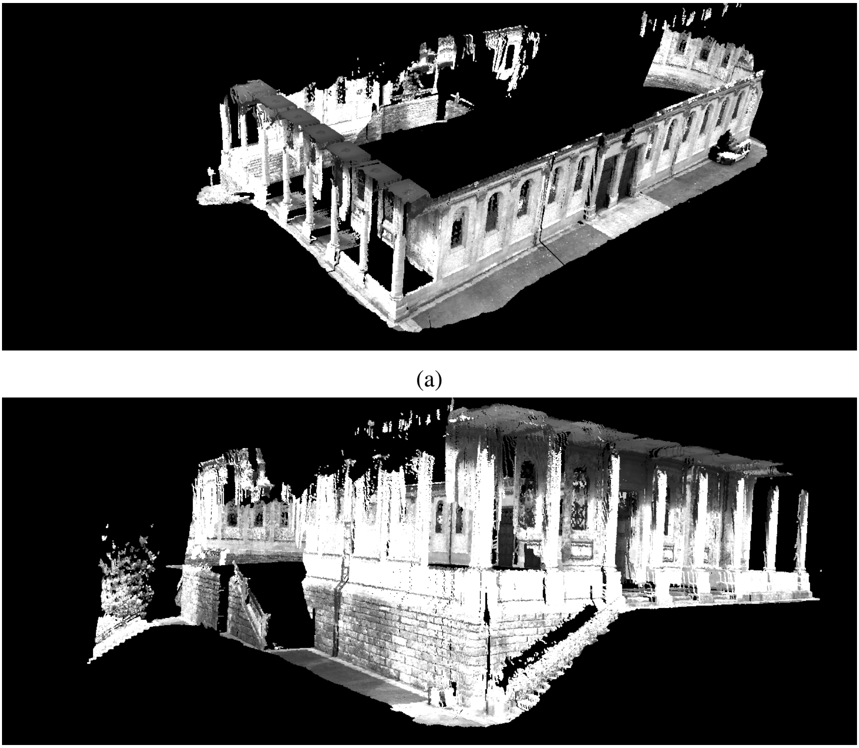
(a)



(b)

Figure 2.6: Detail views of the rendered reconstructions. Laser range data in Fig. 2.6a exhibited a simulated delay of 60 ms with respect to the visual/inertial measurements, while in 2.6b, the sensors were correctly synchronized, resulting in an improved reconstruction of details, which is particularly apparent for the coloring of the column right of the door.

Fig. 2.7 depicts a sample reconstruction result obtained by scanning a church building. The dataset spans roughly 140 s with about 2800 image pairs, 5600 laser range scans and 28 000 inertial measurements. The scanning path followed the contours of the façade. Please note the level of details both in reconstruction and coloring of the model, particularly in the details of the stone wall, which suggests that the calibration is reasonably accurate in estimating time delays as well as the transformation between the LRF and the camera sensor.


(a)

(b)

Figure 2.7: Two views of a 3D reconstruction of a church building. Only laser scan points are visualized for which an intensity value could be retrieved from the camera images. Please note the level of detail, particularly apparent on the stone wall, that suggest—besides accurate state estimation—a precise calibration of the laser with respect to the visual/inertial sensor.

## 2.5 Conclusion and Future Work

This work proposed a spatio-temporal calibration for a combination of laser range finder, camera and inertial measurement unit. Furthermore, it presented preliminary results of colored point cloud reconstructions of buildings, acquired with a hand-held device. Its narrative follows previous work (Furgale, Rehder, and Siegwart, 2013) in that it suggests an accurate and more complete offline-calibration of a multi-sensor setup, and part of its significance lies in the demonstration of unified spatio-temporal calibration applied to a novel combination of sensors. Another contribution lies in the demonstrated application to hand-held, large-scale scanning with results that compare well to other solutions (Bok et al., 2011; James and Quinton, 2014). However, its broader applicability depends on the premise of a fixed time-delay between different sensors, which is not subject to changes on start-up or clock drift. In future work, it remains to be demonstrated that accurate timing can be reproduced over multiple start-ups of the system and that clock drift remains negligible within the time frame of a data collection campaign. Furthermore, by adding regularization terms to the objective function of the estimate, the approach could also be applied to a sensor suite without an IMU, and in the future, we would like to investigate this further.

# 3

# A General Approach to Spatio-Temporal Calibration in Multi-Sensor Systems

Joern Rehder, Roland Siegwart, and Paul Furgale

## 3.1 Introduction

Most methods for state estimation that fuse data from multiple sensors assume and require that the timestamps of all measurements faithfully indicate the measurement instant with respect to a single clock.

Fig. 3.1 depicts reconstructions obtained with our hand-held scanning device shown in Fig. 3.4b. The reconstruction in the middle exhibits distortions arising from a fixed temporal offset corrupting the timestamps of the laser range measurements, while the bottom figure displays a reconstruction where all temporal relations were handled correctly. While Fig. 3.1 provides compelling visual evidence for the importance of considering temporal relationships in multi-sensor systems, we would like to illustrate a core concept of this work with a simpler example: Assume that we would like to add a fixed-exposure camera to our robot. The camera of choice possesses an internal clock and records images at constant rate. The state of this internal clock at the instant at which the camera started exposing is conveyed with the image. The camera further exposes an interrupt line that signals when the image sensor initiates

Published in:
*IEEE Transactions on Robotics 32.2, pp. 383–398, 2016, ©2016 IEEE*
*DOI:* `10.1109/TRO.2016.2529645`

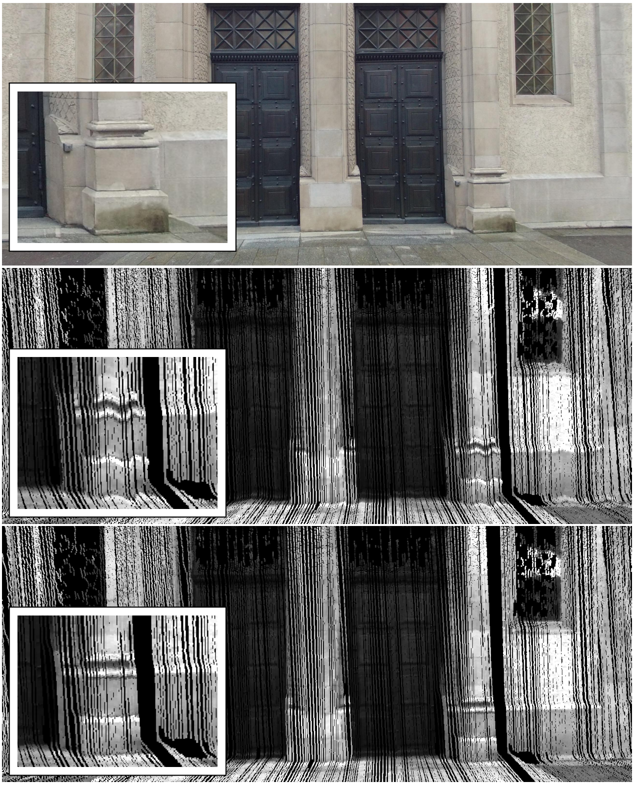


Figure 3.1: Reconstructions obtained from data recorded with our hand-held scanning device (Fig. 3.4b), as well as a photograph for comparison. We used visual/inertial odometry (Leutenegger et al., 2015) for recovering the motion of the sensor and employed the estimated state to transform the range scans into a common coordinate frame. The figure in the middle depicts distortions in the reconstruction caused by a constant temporal offset corrupting the timestamps of the range measurements. In contrast, temporal relations were considered correctly in the figure on the bottom, resulting in a more accurate reconstruction. Image insets highlight features of the rightmost column.

exposure. Finally, the camera implements a protocol that allows the host computer to query the state of the camera's internal clock with minimal latency.

Given these outputs from the camera, it is possible to synchronize the clocks between the robotic system and the camera in a number of different ways. However, the *measurement instant* of the image needed by an estimation framework is the mid-exposure time (see Sections 3.4.4 and 3.4.8 for justification). This implies that, even after clock synchronization, the measurements from the camera exhibit a constant temporal offset. This is typical of many sensors where deterministic temporal offsets arise due to signal integration, transmission times or filter delays.

Consequently, we will distinguish between *clock synchronization* as the process of establishing a single temporal reference for all sensors, and *temporal calibration* as the process of determining constant offsets between measurement instants and timestamps. In the following, we will use the terms *temporal offset* and *delay* interchangeably and will not make any assumptions about the sign of this offset.

Approaches to clock synchronization can be grouped into three different schemes:

1. *Hardware synchronization:* In such a setup, either temporal information about the measurement is conveyed by a change of signal on a dedicated synchronization line or a measurement is initiated by a hardware trigger. In both cases, a central unit records the time at which a trigger was received or generated with respect to a single clock.

2. *Software synchronization with bidirectional communications:* Each sensor assigns timestamps to measurements based on its own internal clock. It further provides a bidirectional communication protocol that allows for synchronization with respect to a common clock. Software synchronization then corrects for jitter (stochastic delay of individual messages), skew (the difference in clock rates), and the communication delay between each sensor and the host system.

3. *Software synchronization with unidirectional communications:* Such a method is usually used when devices do not provide communication protocols for synchronization. These may be sensors that send data at a fixed rate or that are polled regularly. The sensors themselves may have independent clocks according to which they assign device timestamps to their measurements. In this case, software synchronization refers to jitter and skew removal based on an estimate of the measurement rate as perceived by the host system and under the assumption that communication delays, though subject to noise, remain constant over time.

Regardless of the clock synchronization method used, it is still important to account for temporal offsets. These quantities are notoriously hard to obtain from manufacturer's information, and may be dependent on the particulars of integration such as Ethernet switches or other devices that may introduce delays. Even when a device provides a trigger line, it is unclear whether asserting a triggering signal *immediately* initiates a measurement procedure, or if there is some deterministic delay between the trigger and the measurement. Similarly, sensors that allow for clock synchronization through a communication protocol are often comparatively complex black-box systems and it requires faith to assume that all temporal relationships have been considered adequately inside the device.

In this work, we propose an offline calibration technique which estimates the rigid transformation between sensors, while it simultaneously determines constant temporal offsets of sensor timestamps with respect to their measurement instants. The approach is designed for applications with demanding accuracy requirements and can be combined with any of the clock synchronization schemes introduced before.

Conventional discrete-time estimation techniques generally require a state at each measurement time, which makes the estimation of temporal offsets difficult as measurement times shift when the offsets are updated. This has led to the development of specialized algorithms just for estimating time offsets (Kelly, Roy, and Sukhatme, 2014; Mair et al., 2011), that are applied prior to spatial calibration of the sensors.

In contrast, the continuous-time batch estimation algorithm proposed by Furgale, Barfoot, and Sibley (2012) makes it easy to fold time offsets directly into a principled maximum-likelihood estimator. Although we agree with the authors of Mair et al. (2011) that jointly estimating uncorrelated quantities may potentially impair results, we believe that, given accurate measurement models, the optimality implications of maximum likelihood estimation extend to our approach. Consequently, we can achieve the highest accuracy—both for temporal and spatial parameters—by incorporating *all* available information into a unified estimate.

The contributions of this paper are as follows:

1. we propose the first *general* method for spatio-temporal maximum-likelihood estimation that comes as a natural extension to batch, continuous-time estimation (Furgale, Barfoot, and Sibley, 2012) (Section 3.3.1);
2. we derive estimators for sensor suites where temporal calibration has been demonstrated before (e.g. camera/gyroscopes (Kelly, Roy, and Sukhatme, 2014; Mair et al., 2011)) and for novel combinations of sensors (e.g. cam-

era/accelerometer, camera/Laser Range Finder (LRF)) in support of the generality claim (Section 3.3.2.4);

3. we demonstrate temporal calibration, accurate to a fraction of the shortest measurement interval (Section 3.4.4), and spatial calibration of millimeter-precision (Section 3.4.5);
4. we answer questions about the applicability of the approach to different clock synchronization modalities (Section 3.4.5); and
5. we provide a compact account of practical considerations for synchronization in software (Section 3.4.7).

This work compiles findings from our previous contributions (Furgale, Rehder, and Siegwart, 2013; Nikolic et al., 2014b; Rehder et al., 2014), but extends experimental work to specifically answer the question of its applicability to synchronization schemes other than hardware synchronization. It also provides a correction to and extension of the laser range finder measurement model (see Eq. 3.20), which was unfortunately incorrectly stated in Rehder et al. (2014).

## 3.2 Related Work

The focus of this work is on a general approach to incorporating temporal quantities into inter-sensor calibration. To demonstrate this generality, we derive estimators for spatio-temporal camera/Inertial Measurement Unit (IMU), camera/IMU/LRF, and camera/LRF calibration. Accordingly, this section presents an overview of existing approaches for temporal calibration, as well as for spatial camera/IMU and camera/LRF calibration.

Earlier work by Tungadi, Kleeman, et al. (2008) estimates the temporal offset between wheel encoders and a laser range finder by determining the phase shift in the orientation measured independently by both sensors when undergoing a periodic rotational motion. Kelly, Roy, and Sukhatme (2014) propose a procedure, which determines the temporal offset between a camera and an IMU by using a variation of the Iterative Closest Point Algorithm (ICP) to temporally align orientation curves sensed by a camera and gyroscopes individually. This method is intended as a preprocessing step before performing full spatial calibration (Kelly and Sukhatme, 2011). Mair et al. (2011) also recommend independent estimation the time offset by either (a) temporally aligning the independent absolute rotational velocities of the camera and IMU or (b) by determining the phase shift of common frequencies in these signals.

Separating the estimation of temporal and spatial parameters comes at a cost. For example, both (Kelly, Roy, and Sukhatme, 2014) and Mair et al. (2011), neglect temporal evolution of IMU biases over the course of a dataset, which may affect this estimate and in turn bias the subsequent spatial parameter estimation. Furthermore, they all rely on orientation measurements and it is not immediately clear how to extend the approaches to sensor suites where orientation or angular velocity information is not available from every sensor. Unlike these approaches, our method extends to *any calibration problem* that can be cast in terms of maximum-likelihood estimation of the states and parameters—regardless of whether or not orientations are involved. Furthermore, it uses the full model to jointly solve for temporal and spatial parameters, avoiding the compounding of errors as approximations cascade through multiple steps.

There exist online estimation approaches that are capable of estimating the transformation between camera and IMU (Jones and Soatto, 2011; Weiss et al., 2012), but which entirely neglect the temporal relationship. Recently, the integration of temporal offset calibration into a state estimation framework has been demonstrated for sensor setups comprised of a camera and an IMU (Li and Mourikis, 2013; Li and Mourikis, 2014) and for a combination of an IMU and a laser range finder (Bosse, Zlot, and Flick, 2012). Additionally, (Li and Mourikis, 2014) provides an identifiability analysis for the time delay in visual/inertial sensor suites, the implications of which, especially in terms of degenerate calibration motions, also extend to this work. Although these works show that it is possible to add both temporal and spatial calibration parameters to an online estimator, we advocate to separate these steps for two reasons. First, the addition of extra calibration parameters increases the implementation effort and computational complexity of the online filter. Second, the biggest drawback to online self calibration is that the extended estimation problem has several additional *unobservable modes of operation* that the user of the filter has to worry about. In the camera/IMU case, these include planar motion, standing still, and other actions that are clearly very likely for a large class of robots and use cases. While there has been some work addressing unobservability during online calibration (Maye, Furgale, and Siegwart, 2013), the problem remains largely unsolved. In our offline calibration procedure, we can control the motion of the sensor during calibration, ensuring that all parameters are observable. Approaches for online estimation (Bosse, Zlot, and Flick, 2012; Li and Mourikis, 2013; Li and Mourikis, 2014) could equally be employed in a preceding calibration step to determine temporal offsets and inter-sensor transformations from controlled motions. However, we believe that a batch approach, which takes all measurements in the calibration dataset into account and re-linearizes frequently, is likely to yield more accurate results than a real-time method, which operates on a subset of the measurements at any time or which linearizes only once.

Early efforts in spatial camera/IMU calibration required elaborate calibration settings (Alves, Lobo, and Dias, 2003) and estimated orientation and displacement in separate procedures (Lobo and Dias, 2007). More recently, recursive approaches were employed to jointly estimate relative rotation and translation from measurements acquired by dynamically moving a visual/inertial sensor combination in front of a calibration pattern (Kelly and Sukhatme, 2011; Mirzaei and Roumeliotis, 2008). Both studies further address the question of the observability of the calibration, deriving that it can be determined when sufficient rotational velocity is present in the dataset. Other approaches estimated the transformation using batch optimization over a set of inertial measurements and calibration pattern observations (Fleps et al., 2011; Mirzaei and Roumeliotis, 2007). Among those, our algorithm is most similar to Fleps et al. (2011) in that it also employs B-splines to parameterize the motion of the device. In general, we believe that batch approaches (like ours) will always be more accurate than filtering approaches because of the ability to relinearize all model equations during iterative optimization.

For camera/LRF calibration, we will limit our literature overview to work concerning sensor configurations similar to ours with a rigid link between a camera and a two-dimensional (2D) laser range finder. So and Menegatti (2012) provide a comprehensive survey of the state of the art in this field. According to this work, established approaches either minimize the distance from ranges sampled on a plane from its visually perceived estimate (e.g. Mei and Rives (2006), Vasconcelos, Barreto, and Nunes (2012), and Zhang and Pless (2004)) or solve for a set of equations that arises from correspondences of point range measurements to lines detected in the image (e.g. Bok et al. (2011) and Li et al. (2007)). These approaches are mostly practical for setups with overlapping field of view between camera and LRF. The approach by Bok et al. (2014) dropped the requirement that range measurements considered in the calibration had to be induced by a visually perceived entity, making it a more convenient approach for setups with non-overlapping fields of view. In contrast to all aforementioned approaches, our work employs a fully probabilistic model of the range measurements, which allows a maximum likelihood estimate of the calibration parameters. We recently discovered that So and Menegatti (2012) proposed a similar model called “line-of-sight” distance. While the authors remained vague on the computation of this distance, we present a mathematical derivation and develop the model further based on insights from a characterization of the LRF (Demski, Mikulski, and Koteras, 2013). Furthermore, our method overcomes the limitations stated in So and Menegatti (2012) by simultaneously estimating the plane parameters such that uncertainties can be treated consistently. Finally, none of the established techniques addresses the temporal relationship between LRF and camera, despite the fact that this quantity is of practical interest in many applications.

## 3.3 Theory

This section presents the theory for joint estimation of spatial and temporal calibration parameters. Throughout this section, we follow the basis function approach for batch continuous-time state estimation presented in Furgale, Barfoot, and Sibley (2012).

Time-varying states are represented as the weighted sum of a finite number of known analytical basis functions. For example, a $D$-dimensional state, $\mathbf{x}(t)$, may be written as

$$\Phi(t) := \begin{bmatrix} \phi_1(t) & \dots & \phi_B(t) \end{bmatrix}, \quad \mathbf{x}(t) := \Phi(t)\mathbf{c}, \tag{3.1}$$

where each $\phi_b(t)$ is a known $D \times 1$ analytical function of time and $\Phi(t)$ is a $D \times B$ stacked basis matrix. To estimate $\mathbf{x}(t)$, we simply estimate $\mathbf{c}$, a $B \times 1$ column of coefficients.

### 3.3.1 Estimating Time Offsets using Basis Functions

Here we explain the extensions to Furgale, Barfoot, and Sibley (2012) needed to estimate temporal offsets. This section assumes that the system used some form of clock synchronization such that all timestamps are expressed with respect to a common clock.

When estimating time offsets from measurement data, we will encounter error terms such as

$$\mathbf{e}_j := \mathbf{y}_j - \mathbf{h}\big(\mathbf{x}(t_j + d)\big), \tag{3.2}$$

where $\mathbf{y}_j$ is a measurement that arrived with timestamp $t_j$, $\mathbf{h}(\cdot)$ is a measurement model that produces a predicted measurement from $\mathbf{x}(\cdot)$, and $d$ is the unknown time offset. Using basis functions, this becomes

$$\mathbf{e}_j = \mathbf{y}_j - \mathbf{h}\big(\Phi(t_j + d)\mathbf{c}\big), \tag{3.3}$$

which is easy to evaluate for different values of $d$ as it changes during optimization. The analytical Jacobian of the error term, needed for nonlinear least squares estimation, is derived by linearizing (3.3) about a nominal value, $\bar{d}$, with respect to small changes, $\Delta d$. This results in the expression

$$\mathbf{e}_j \approx \mathbf{y}_j - \mathbf{h}\big(\Phi(t_j + \bar{d})\mathbf{c}\big) - \mathbf{H}\dot{\Phi}(t_j + \bar{d})\mathbf{c}\Delta d, \tag{3.4}$$

where the over dot represents a time derivative and

$$\mathbf{H} := \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\mathbf{x}\left(\Phi(t_j+\bar{d})\mathbf{c}\right)} . \tag{3.5}$$

In (3.4), $\Phi(t)$ is a known function and we assume its time derivative, $\dot{\Phi}(t)$, is available analytically.

This approach has two clear benefits. Firstly, it allows us to treat the problem of estimating time offsets within the rigorous theoretical framework of maximum likelihood estimation. Secondly, it allows us to leave the problem in continuous time so that the delayed measurement equations and their Jacobians can be evaluated analytically.

In short, estimating the time offsets in a principled way becomes easy.

### 3.3.2 Multi-Sensor Calibration for Combinations of Cameras, an IMU and an LRF

Rather than delving further into the general case, we will proceed with specific examples. The goal of calibration is to determine the relative rotation, translation, and time offset between the sensors. We suggested that a continuous-time batch formulation of the calibration problem is well-suited for a broad range of sensor combinations and in order to substantiate this claim, we will derive a set of estimators for camera/IMU, camera/IMU/LRF and camera/LRF spatio-temporal calibration.

To perform calibration, we collect a set of data over a short time interval, $T = [t_1, t_K]$ (typically 1–2 minutes), as the sensor head is waved in front of a static calibration pattern, while exciting all rotational degrees of freedom. This procedure forms the basis for calibrating all of the aforementioned sensor suites with the addition that for setups including an LRF, we further require the environment to be at least partly comprised of planes.
Fig. 3.4a shows the basic problem setup. Estimation is performed with respect to an inertial world coordinate frame, $\underrightarrow{\mathcal{F}}_w$. The linear acceleration and angular velocity are measured in the IMU frame, $\underrightarrow{\mathcal{F}}_i$. The camera coordinate frame, $\underrightarrow{\mathcal{F}}_c$, is placed at the camera's optical center with the $z$-axis pointing down the optical axis. Fig. 3.4b depicts the sensor combination employed in LRF calibration. The additional frame $\underrightarrow{\mathcal{F}}_l$ coincides with the center of the LRF with the $x, y$-axes spanning the scanning plane.

#### 3.3.2.1 Parameterization of Time-Varying States

Time-varying states are represented by B-spline functions. B-splines produce simple analytical functions of time with good representational power. Please see (Bartels, Beatty, and Barsky, 1987) for a thorough introduction.

The IMU pose is parameterized as a $6 \times 1$ spline, using three degrees of freedom for orientation and three for translation. The transformation that takes points from the IMU frame $\underrightarrow{\mathcal{F}}_i$ to the world frame $\underrightarrow{\mathcal{F}}_w$ at any time $t$ can be built as

$$\mathbf{T}_{\mathtt{w,i}}(t) := \begin{bmatrix} \mathbf{C}_{\mathtt{w,i}}(t) & \mathbf{t}(t) \\ \mathbf{0}^T & 1 \end{bmatrix}, \tag{3.6}$$

with $\mathbf{C}_{\mathtt{w,i}}(t) := \mathcal{C}\big(\varphi(t)\big)$, where $\mathcal{C}(\cdot)$ is a function that builds a rotation matrix from orientation parameters $\varphi(t) := \Phi_\varphi(t)\mathbf{c}_\varphi$ , and $\mathbf{t}(t) := \Phi_t(t)\mathbf{c}_t$ encodes the translation. The velocity, $\mathbf{v}(t)$, and acceleration, $\mathbf{a}(t)$, of the platform with respect to and expressed in the world frame $\underrightarrow{\mathcal{F}}_w$ are

$$\mathbf{v}(t) = \dot{\mathbf{t}}(t) = \dot{\Phi}_t(t)\mathbf{c}_t\,, \quad \mathbf{a}(t) = \ddot{\mathbf{t}}(t) = \ddot{\Phi}_t(t)\mathbf{c}_t\,. \tag{3.7}$$

For a given rotation parameterization, the relationship to angular velocity is of the form

$$\omega(t) = \mathbf{S}\big(\varphi(t)\big)\dot{\varphi}(t) = \mathbf{S}\big(\Phi(t)\mathbf{c}_\varphi\big)\dot{\Phi}(t)\mathbf{c}_\varphi\,, \tag{3.8}$$

where $\mathbf{S}(\cdot)$ is the standard matrix relating parameter rates to angular velocity (Hughes, 1986). In this paper we used the axis/angle parameterization of rotation where $\varphi(t)$ represents rotation by the angle $\varphi = \sqrt{\varphi(t)^T\varphi(t)}$ about the axis $\varphi(t)/\varphi(t)$.

In our work, the IMU pose is encoded as a sixth-order B-spline (a piecewise fifth-degree polynomial), which allows for encoding accelerations as a cubic polynomial. We found this was necessary to accurately capture the dynamics for the motions exciting the sensor during calibration.

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

$$\mathbf{b}(t) := \Phi_b(t)\mathbf{c}_b \tag{3.9}$$

B-splines are just one possible realization of these basis functions $\Phi(t)$. For more details on the choice of basis functions and their desired properties, please see (Furgale, Barfoot, and Sibley, 2012).

#### 3.3.2.2 Quantities Estimated

The following parameters and states, or a subset thereof, are determined by the estimators proposed in this work:

| | **Time-Invariant** |
|---|---|
| $\mathbf{g}_{\mathtt{w}}$ | direction of gravity expressed in $\underrightarrow{\mathcal{F}}_{w}$ |
| $\mathbf{T}_{\mathtt{c,i}}$ | transformation between IMU and camera |
| $\mathbf{T}_{\mathtt{i,l}}$ | transformation between LRF and IMU |
| $d_c$ | delay in the image timestamps |
| $d_l$ | delay in the range timestamps |
| $\mathbf{n}_{\mathtt{w}}^{h}$ | normal of plane $\pi_h$ expressed in $\underrightarrow{\mathcal{F}}_{w}$ |
| $\mathrm{d}_h$ | distance of plane $\pi_h$ to the origin |
| $b_l$ | cumulative range bias in LRF measurements |
| | **Time-Varying** |
| $\mathbf{T}_{\mathtt{w,i}}(t)$ | pose of the IMU expressed in $\underrightarrow{\mathcal{F}}_{w}$, represented as sixth-order B-spline |
| $\mathbf{b}_a(t)$ | accelerometer bias, represented as cubic B-spline |
| $\mathbf{b}_\omega(t)$ | gyroscope bias, represented as cubic B-spline |

#### 3.3.2.3 Measurement and Process Models

Each accelerometer measurement, $\alpha_k$, and gyroscope measurement, $\varpi_k$, is sampled at time $t_k$, where $k = 1 \ldots K$. We use the standard, discrete-time IMU measurement equations,

$$\alpha_k := \mathbf{C}_{\mathtt{w,i}}\left(t_k\right)^T \left(\mathbf{a}(t_k) - \mathbf{g}_{\mathtt{w}}\right) + \mathbf{b}_a(t_k) + \nu_{a_k}(t_k), \tag{3.10a}$$

$$\varpi_k := \mathbf{C}_{\mathtt{w,i}}\left(t_k\right)^T \omega(t_k) + \mathbf{b}_\omega(t_k) + \nu_{\omega_k}(t_k). \tag{3.10b}$$

In Eq. 3.10a, 3.10b, 3.11 and 3.12, $\nu(t)$ denotes a white Gaussian process $\nu \sim \mathcal{GP}\left(\mathbf{0}, \mathbf{R}\delta(t-t')\right)$ with mean $\mathbf{0}$ and covariance function $\mathbf{R}\delta(t-t')$, where $\delta(\cdot)$ denotes Dirac's delta function. These noise processes are assumed to be statistically independent between sensors.

The pixel location of landmark $\mathbf{p}_{\mathtt{w}}^{m}$ seen at time $t_j + d_c$ is denoted $\mathbf{y}_{mj}$, where $t_j$ is the image timestamp, $d_c$ is the unknown delay of the image timestamps, and $j = 1 \ldots J$ indexes the images. There are $M$ landmarks, $\{\mathbf{p}_{\mathtt{w}}^{m} | m = 1 \ldots M\}$. In this notation, the subscript denotes the frame that the entity is expressed in, here $\underrightarrow{\mathcal{F}}_{w}$, and the superscript marks an identifier, in this case establishing an association with coordinates of a specific landmark on the calibration pattern. With function $\mathbf{h}(\cdot)$ denoting any nonlinear camera model, the projection equation is

$$\mathbf{y}_{mj} := \mathbf{h}\left(\mathbf{T}_{\mathtt{c},\mathtt{i}}\mathbf{T}_{\mathtt{w},\mathtt{i}}(t_j + d_c)^{-1}\mathbf{p}^m_{\mathtt{w}}\right) + \nu_{y_{mj}}. \tag{3.11}$$

Individual laser range measurements, $l_{hi}$, at angle $\psi_i$ in the LRF reference frame are recorded at time $t_i + d_l$ and modelled as induced by plane $\pi_h$, where $h = 1 \ldots H$, with $H$ denoting the number of planes detected in the environment. Plane $\pi_h$ is parameterized by its normal $\mathbf{n}^h_{\mathtt{w}}$, expressed in $\underrightarrow{\mathcal{F}}_w$, and a distance $\mathrm{d}_h \geq 0$, such that ${\mathbf{p}^i_{\mathtt{w}}}^T \mathbf{n}^h_{\mathtt{w}} - \mathrm{d}_h = 0$ holds true for all $\mathbf{p}^i_{\mathtt{w}} \in \pi_h$. The range measurement equation can be expressed as

$$l_{hi} := f\left(\psi_i, \mathbf{T}_{\mathtt{w},\mathtt{i}}(t_i + d_l)\mathbf{T}_{\mathtt{i},\mathtt{l}}, \mathbf{n}^h_{\mathtt{w}}, \mathrm{d}_h\right) + b_l + \nu_{l_{hi}}, \tag{3.12}$$

where $f(\cdot)$ models the range measurement based on planes, $\pi_h$, in the environment and $b_l$ is a constant range bias. This laser model will be explained in more detail in Section 3.3.2.5.

In our formulation of the error terms, the time reference is provided by the IMU. This assumption is only for convenience as it is easier to write out and implement delayed low rate measurements.

We model the IMU biases as driven by white Gaussian processes:

$$\dot{\mathbf{b}}_a(t) = \mathbf{w}_a(t) \qquad \mathbf{w}_a(t) \sim \mathcal{GP}\left(\mathbf{0}, \mathbf{Q}_a\delta(t - t')\right) \tag{3.13a}$$

$$\dot{\mathbf{b}}_\omega(t) = \mathbf{w}_\omega(t) \qquad \mathbf{w}_\omega(t) \sim \mathcal{GP}\left(\mathbf{0}, \mathbf{Q}_\omega\delta(t - t')\right) \tag{3.13b}$$

with mean $\mathbf{0}$ and covariance function $\mathbf{Q}\delta(t - t')$. We assume the bias processes are statistically independent so that $E\left[\mathbf{w}_a(t)\mathbf{w}_\omega(t')^T\right] = \mathbf{0}$ for all $t$, $t'$, where $E[\cdot]$ is the expectation operator.

In some experiments, we present full spatio-temporal calibration between reduced sensor combinations where the measurements do not contain enough information to adequately constrain the trajectory: between a camera and a 3-axis accelerometer as well as between a stereo camera setup and an LRF. We can cope with these cases by making assumptions about the distribution of accelerations exciting the motion of the sensor:

$$\ddot{\varphi}(t) = \mathbf{w}_\varphi(t) \qquad \mathbf{w}_\varphi(t) \sim \mathcal{GP}\left(\mathbf{0}, \mathbf{Q}_\varphi\delta(t - t')\right) \tag{3.14a}$$

$$\ddot{\mathbf{t}}(t) = \mathbf{w}_t(t) \qquad \mathbf{w}_t(t) \sim \mathcal{GP}\left(\mathbf{0}, \mathbf{Q}_t\delta(t - t')\right) \tag{3.14b}$$

Informally, this weak motion prior tells the estimator that, in the absence of other information, it should assume a minimum acceleration path. This setup that fits measurements while minimizing acceleration is typical of spline smoothing used in other disciplines (c.f. (Wahba, 1990)).

#### 3.3.2.4 The Estimators

The previously introduced measurement and process models form the basis from which we compose different estimators. Error terms are constructed as the difference between the measurement and its prediction given the current state estimate. The continuous-time models for IMU biases and motion regularization give rise to integral error terms (refer to Furgale, Barfoot, and Sibley (2012) for more details). In the following, we will introduce the individual components from which the different objective functions will later be composed. These include the IMU cost terms and continuous-time bias models arising from (3.10a) and (3.10b), and (3.13a) and (3.13b),

$$\mathbf{e}_{\alpha_k} := \alpha_k - \left(\mathbf{C}_{\mathtt{w},\mathtt{i}}\left(t_k\right)^T \left(\mathbf{a}(t_k) - \mathbf{g}_{\mathtt{w}}\right) + \mathbf{b}_a(t_k)\right), \tag{3.15a}$$

$$J_\alpha := \frac{1}{2}\sum_{k=1}^{K} \mathbf{e}_{\alpha_k}^T \mathbf{R}_{\alpha_k}^{-1} \mathbf{e}_{\alpha_k}, \tag{3.15b}$$

$$\mathbf{e}_{\omega_k} := \varpi_k - \left(\mathbf{C}_{\mathtt{w},\mathtt{i}}\left(t_k\right)^T \omega(t_k) + \mathbf{b}_\omega(t_k)\right), \tag{3.15c}$$

$$J_\omega := \frac{1}{2}\sum_{k=1}^{K} \mathbf{e}_{\omega_k}^T \mathbf{R}_{\omega_k}^{-1} \mathbf{e}_{\omega_k}, \tag{3.15d}$$

$$J_{b_a} := \frac{1}{2}\int_{t_1}^{t_K} \dot{\mathbf{b}}_a(\tau)^T \mathbf{Q}_a^{-1} \dot{\mathbf{b}}_a(\tau)\, d\tau, \tag{3.15e}$$

$$J_{b_\omega} := \frac{1}{2}\int_{t_1}^{t_K} \dot{\mathbf{b}}_\omega(\tau)^T \mathbf{Q}_\omega^{-1} \dot{\mathbf{b}}_\omega(\tau)\, d\tau, \tag{3.15f}$$

cost terms associated with the camera arising from (3.11),

$$\mathbf{e}_{y_{mj}} := \mathbf{y}_{mj} - \mathbf{h}\left(\mathbf{T}_{\mathtt{c},\mathtt{i}}\mathbf{T}_{\mathtt{w},\mathtt{i}}(t_j + d_c)^{-1}\mathbf{p}_{\mathtt{w}}^m\right), \tag{3.16a}$$

$$J_y := \frac{1}{2}\sum_{j=1}^{J}\sum_{m=1}^{M} \mathbf{e}_{y_{mj}}^T \mathbf{R}_{y_{mj}}^{-1} \mathbf{e}_{y_{mj}}, \tag{3.16b}$$

and cost terms from the LRF measurements based on (3.12),

$$\mathrm{e}_{l_{hi}} := l_{hi} - \left(f\left(\psi_i, \mathbf{T}_{\mathtt{w},\mathtt{i}}(t_i + d_l)\mathbf{T}_{\mathtt{i},\mathtt{l}}, \mathbf{n}_{\mathtt{w}}^h, \mathrm{d}_h\right) + b_l\right), \tag{3.17a}$$

$$J_l := \frac{1}{2} \sum_{(i,h) \in \mathcal{A}} \mathrm{e}_{l_{hi}}^T \mathrm{R}_{l_{hi}}^{-1} \mathrm{e}_{l_{hi}}, \tag{3.17b}$$

where $\mathcal{A}$ is a set of tuples associating range measurement $i$ with plane $\pi_h$. Cost terms from continuous-time motion models based on (3.14) to regularize the estimate when the full IMU is not available result in contributions

$$J_\varphi := \frac{1}{2} \int_{t_1}^{t_K} \ddot{\varphi}(\tau)^T \mathbf{Q}_\omega^{-1} \ddot{\varphi}(\tau)\, d\tau, \tag{3.18a}$$

$$J_t := \frac{1}{2} \int_{t_1}^{t_K} \ddot{\mathbf{t}}(\tau)^T \mathbf{Q}_\omega^{-1} \ddot{\mathbf{t}}(\tau)\, d\tau. \tag{3.18b}$$

Table 3.1 shows the set of estimators composed from these error terms and evaluated in Section 4.4. The Levenberg-Marquardt (LM) algorithm (Nocedal and Wright, 2006) is used to minimize the respective objective functions to find the maximum likelihood estimate of all unknown parameters at once. Estimator **J** has since been released as part of the open-source calibration toolbox *kalibr* (Furgale et al., 2015b).

#### 3.3.2.5 Probabilistic Range Measurement Model

In the following, we will derive a probabilistic model for laser range measurements. It is an extension of the model we proposed in Rehder et al. (2014) and similar to the “line-of-sight” distance mentioned in So and Menegatti (2012). As stated earlier, range measurements are modelled as being induced by a plane, $\pi_h$. Literature on characterizing laser range finders (Demski, Mikulski, and Koteras, 2013) identifies factors such as thermal effects, the angle of incidence at which the beam hits the surface, and surface reflectivity as affecting the distance output. Bosse, Zlot, and Flick (2012) suggest that gyroscopic effects might further impact the device when subjected to dynamic motions. While we acknowledge the existence of all these factors, we chose to not model them individually and instead to make the following simplifying assumptions:

- There is no error on the beam directions reported by the range finder.
- Range measurements are corrupted by additive zero-mean Gaussian noise, the distribution of which is independent of the measured distance.
- The different factors causing an incorrect mean distance to be reported by the LRF can be approximated by a single cumulative range bias.

The results presented in Section 4.4 suggest that this model is valid with respect to our requirements on accuracy.

Given a measurement induced by a plane $\pi_h$, defined by its normal $\mathbf{n}_{\mathtt{w}}^{h}$ and a distance $\mathrm{d}_h$, range $l_{hi}$ can be modelled as

$$l_{hi} := f\left(\psi_i, \mathbf{T}_{\mathtt{w},\mathtt{i}}(t_i + d_l)\mathbf{T}_{\mathtt{i},\mathtt{l}}, \mathbf{n}_{\mathtt{w}}^{h}, \mathrm{d}_h\right) + b_l + \nu_{l_{hi}} \tag{3.19}$$

with $\nu_{l_{hi}} \sim \mathcal{N}(0, R_l)$, and where $f(\cdot)$ is calculated according to

$$f(\cdot) := \left| \frac{{\mathbf{n}_{\mathtt{w}}^{h}}^{T}\, \mathbf{t}_{\mathtt{w}}^{\mathtt{l}}(t_i + d_l) - \mathrm{d}_h}{{\mathbf{n}_{\mathtt{w}}^{h}}^{T}\, \mathbf{r}_{\mathtt{w}}(t_i + d_l)} \right|. \tag{3.20}$$

The position of $\underrightarrow{\mathcal{F}}_l$ in the world coordinate frame $\underrightarrow{\mathcal{F}}_w$ is calculated according to $[\mathbf{t}_{\mathtt{w}}^{\mathtt{l}}(t_i + d_l), 1]^T = \mathbf{T}_{\mathtt{w},\mathtt{i}}(t_i + d_l)[\mathbf{t}_{\mathtt{i}}^{\mathtt{l}}, 1]^T$, where $\mathbf{t}_{\mathtt{i}}^{\mathtt{l}}$ denotes the translational component of transformation $\mathbf{T}_{\mathtt{i},\mathtt{l}}$. The unit vector in beam direction is calculated as

$$\mathbf{r}_{\mathtt{w}}(t_i + d_l) = \mathbf{C}_{\mathtt{w},\mathtt{i}}(t_i + d_l)\mathbf{C}_{\mathtt{i},\mathtt{l}}\left[cos(\psi_i), sin(\psi_i), 0\right]^T, \tag{3.21}$$

Table 3.1: Sensor suites and corresponding estimators considered in this work.

| **Sensor Suite** | **Estimated Quantities** | **Objective Function** |
|---|---|---|
| (**J**) Camera/IMU | $\mathbf{g}_{\mathtt{w}}, \mathbf{T}_{\mathtt{c,i}}, d_c, \mathbf{T}_{\mathtt{w,i}}(t), \mathbf{b}_a(t), \mathbf{b}_\omega(t)$ | $J_\alpha + J_\omega + J_y + J_{b_a} + J_{b_\omega}$ |
| (**G**) Camera/Gyroscopes | $\mathbf{C}_{\mathtt{c,i}}, d_c, \mathbf{T}_{\mathtt{w,i}}(t), \mathbf{b}_\omega(t)$ | $J_\omega + J_y + J_{b_\omega}$ |
| (**A**) Camera/Accelerometers | $\mathbf{g}_{\mathtt{w}}, \mathbf{T}_{\mathtt{c,i}}, d_c, \mathbf{T}_{\mathtt{w,i}}(t), \mathbf{b}_a(t)$ | $J_\alpha + J_y + J_{b_a} + J_\varphi$ |
| (**L**) Camera/IMU/LRF | $\mathbf{g}_{\mathtt{w}}, \mathbf{T}_{\mathtt{c,i}}, \mathbf{T}_{\mathtt{i,l}}, d_c, d_l, \mathbf{n}^h_{\mathtt{w}}, \mathrm{d}_h, b_l, \mathbf{T}_{\mathtt{w,i}}(t), \mathbf{b}_a(t), \mathbf{b}_\omega(t)$ | $J_\alpha + J_\omega + J_y + J_l + J_{b_a} + J_{b_\omega}$ |
| (**C**) Camera/LRF | $\mathbf{T}_{\mathtt{c,l}}, d_l, \mathbf{n}^h_{\mathtt{w}}, \mathrm{d}_h, b_l, \mathbf{T}_{\mathtt{w,i}}(t)$ | $J_y + J_l + J_\varphi + J_t$ |

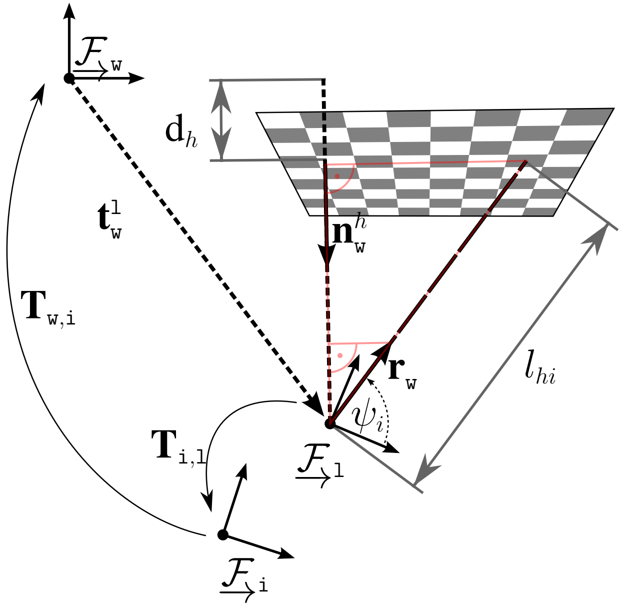

$\underrightarrow{\mathcal{F}}_{\mathtt{w}}$
$\mathrm{d}_h$
$\mathbf{t}^{\mathtt{l}}_{\mathtt{w}}$
$\mathbf{n}^h_{\mathtt{w}}$
$\mathbf{T}_{\mathtt{w,i}}$
$\mathbf{r}_{\mathtt{w}}$
$l_{hi}$
$\psi_i$
$\mathbf{T}_{\mathtt{i,l}}$
$\underrightarrow{\mathcal{F}}_{\mathtt{l}}$
$\underrightarrow{\mathcal{F}}_{\mathtt{i}}$

Figure 3.2: Illustration of the range computation given sensor pose $\mathbf{T}_{\mathtt{w,i}}$ and plane $\pi_h$. The range in beam direction can be computed by means of similar triangles. These relate $l_{hi}$ and the distance of the LRF to the plane $({\mathbf{n}^h_{\mathtt{w}}}^T \mathbf{t}^{\mathtt{l}}_{\mathtt{w}} - \mathrm{d}_h)$ to unit length and the projection of the unit vector in beam direction onto the normal direction ${\mathbf{r}_{\mathtt{w}}}^T \mathbf{n}^h_{\mathtt{w}}$.

where $\mathbf{C}_{\mathtt{i,l}}$ marks the rotational component of $\mathbf{T}_{\mathtt{i,l}}$. Fig. 3.2 visualizes these quantities and provides an intuition for the range computation. The cumulative range bias, $b_l$, is representative of properties of the environment the calibration was conducted in, such as reflectivity and mean depth and slant of identified planes. Hence, it is not genuinely suited to correct measurements in other environments.

### 3.3.2.6 Automatic Plane Detection

The previously introduced probabilistic model is only valid for measurements induced by planes and the problem remains of identifying those measurements as well as the corresponding planes, $\pi_h$.

With the sensor trajectory recovered in a previous step (either through estimator $\mathbf{J}$ or, in the absence of an IMU, by minimizing the objective function $J_y + J_\varphi + J_t$), and assuming a sufficiently accurate initial estimate of the transformation, $\mathbf{T}_{\mathtt{i,l}}$, an initial point cloud can be obtained by transforming the laser measurements of a beam cast at angle, $\psi_i$, and measuring range, $l_i$, into the world coordinate frame, $\underset{\rightarrow}{\mathcal{F}}_w$:

$$\begin{bmatrix} \mathbf{p}^i_{\mathtt{w}} \\ 1 \end{bmatrix} = \mathbf{T}_{\mathtt{w,i}}(t_i + d_l)\mathbf{T}_{\mathtt{i,l}}\left[l_i cos(\psi_i), l_i sin(\psi_i), 0, 1\right]^T . \tag{3.22}$$

Fig. 3.3 depicts such an initial point cloud acquired from a calibration sequence. As a result of the errors in the initial estimate for the transformation, $\mathbf{T}_{\mathtt{i,l}}$, and the temporal calibration parameter, $d_l$, the point cloud appears fuzzy despite most features of the room being visible. In order to automatically identify candidates for measurements having been induced by planes, a Random Sample Consensus (RANSAC) scheme (Fischler and Bolles, 1981) is applied to the point cloud, with plane hypotheses generated from a minimal set of three non-collinear points and model support being evaluated according to threshold, $\varepsilon$,

$$|\mathbf{n}^{h^T}_{\mathtt{w}} \mathbf{p}^i_{\mathtt{w}} - \mathrm{d}_h| < \varepsilon \quad , \tag{3.23}$$

with $\mathbf{p}^i_{\mathtt{w}}$ being a point evaluated for support.

Finding planes in this way is not easy. Not all inliers are necessarily induced by the identified plane and the parameters determined by RANSAC might not even correspond to a physically meaningful entity at all; points on any structure within the distance, $\varepsilon$, from plane, $\pi_h$, will be included. This complication can be illustrated with the table shown in Fig. 3.3: Using the same parameters, $\pi_h$, (3.23) would hold true for the table top, but also for any point on the wall that has the same distance to the ground as the surface of the table. Similarly, the range finder samples points in the scanning plane and a lack of dynamics in the calibration motion or frequent revisiting of similar sensor poses results in accumulations of points that satisfy (3.23) even though they are only induced by the scanning plane of the laser rather than a physical structure. Finally, while the plane identification is based on an error measuring the distance of a point to a plane in normal direction, our probabilistic measurement model considers the distance of the point to the plane in beam direction, and intuitively these two errors can be vastly different[1]. For these reasons, a subsequent processing step aims at omitting all spurious data points.

A simple region growing is employed to remove outliers that stem from the intersection of plane $\pi_h$ with other structures in the environment as well as incorrect associations arising at the intersection of two physical planes. Starting from a set of

---

[1] A RANSAC support measure based on the distance in beam direction is equally feasible. However, evaluating support according to (3.23) is more efficient and does not require to maintain associations between points and beam directions during plane detection.

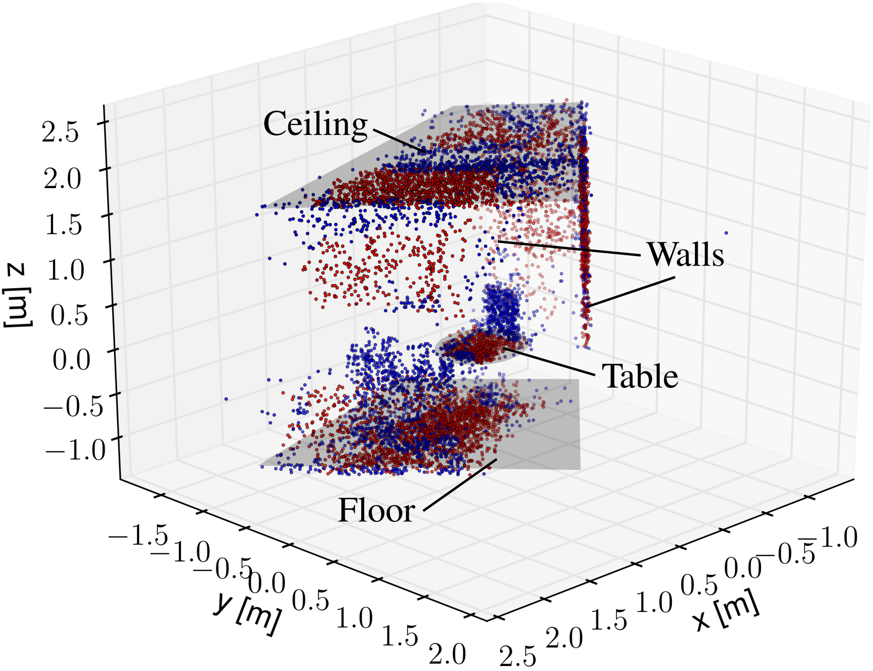


Figure 3.3: An illustration of the automatic plane identification. Measurements induced by planes are marked in red, while blue points mark range data not considered in the calibration process. The majority of planes in the environment are detected correctly, while points at the intersection of two planes are excluded to avoid issues arising from incorrect associations. The lack of crispness in the point cloud is due to discrepancies in the initial guess for the inter-sensor transformation $\mathbf{T}_{\mathtt{i},\mathtt{l}}$ and $d_l$ to their true values.

seed points, neighbouring measurements are incorporated into the plane hypothesis if they (a) are considered a RANSAC inlier, (b) are within a certain distance, and (c) not a single point in the same neighbourhood is considered an outlier. Of all seeded regions, the most populated is selected and subjected to a heuristic test based on the ratio of the eigenvalues of the matrix expressing the second central moment of its points as well as their absolute number in order to exclude regions induced by the scanning plane or of insufficient extent to reliably constrain the normal estimation. Finally, all measurements within the region are filtered with respect to their range error in beam direction. Subtle disruptions of planes, such as flat ceiling lamps, may be eclipsed by an inaccurate initial $\mathbf{T}_{\mathtt{i},\mathtt{l}}$. We mitigate their effect on the estimation by employing the Blake-Zisserman robust cost function (Blake and Zisserman, 1987) on range residuals.

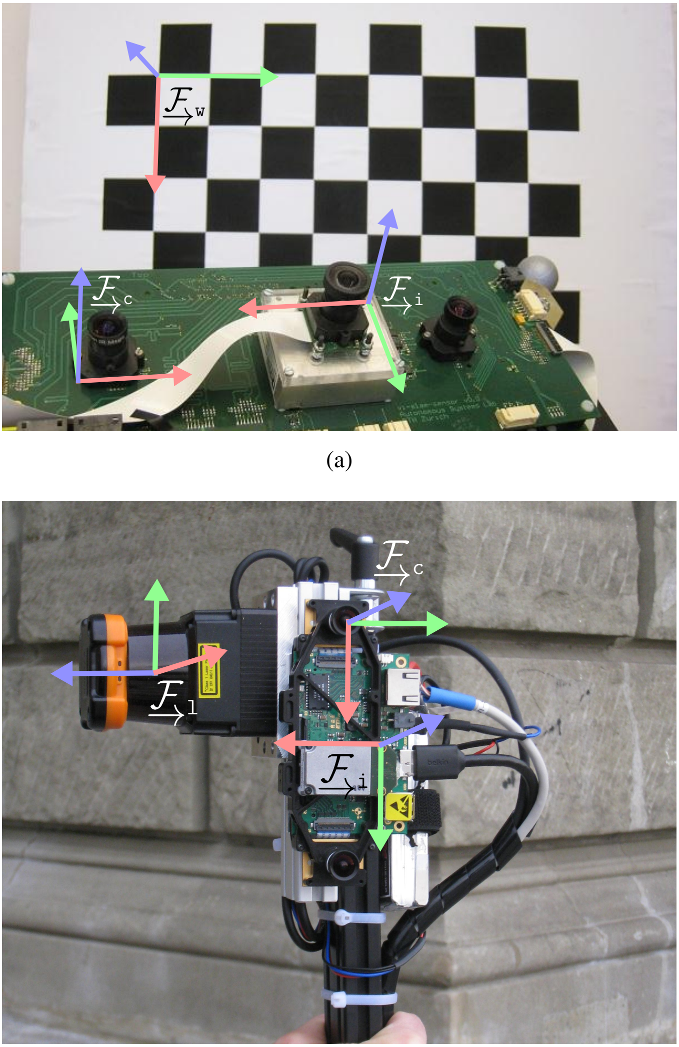
We iterate the plane detection while removing measurements corresponding to verified detections until the number of remaining points falls below the threshold of what we accept as a plane hypothesis. In cases where a plane hypothesis is rejected by our heuristic as presumably induced by the scanning plane, a certain percentage of randomly selected inliers is excluded from the data to avoid selecting similar configurations in future iterations. A result of automatic plane identification is depicted in Fig. 3.3, where red points mark measurements identified as being induced by planar structures. Ceiling, floor and most walls are reliably detected, while the transition from ceiling to wall is not considered to avoid incorrect associations.

## 3.4 Experiments

In this section, we will demonstrate the stability and accuracy of the estimators proposed in Section 3.3.2.4 and highlight the applicability of our temporal calibration to the different synchronization modalities introduced in Section 3.1.

### 3.4.1 Equipment

Our experiments used two hardware setups: *Setup I*, depicted in Fig. 3.4a, and *Setup II*, shown in Fig. 3.4b. Central to both are different generations of a visual/inertial sensor (Nikolic et al., 2014b) comprised of multiple Aptina MT9V034 Wide Video Graphics Array (WVGA) global shutter image sensors and an Analog Devices ADIS16488/ADIS16448 IMU respectively. In addition, the setup shown in Fig. 3.4b features a Hokuyo UTM-30LX laser range finder. For experiments conducted with Setup I, only a single camera was employed. For evaluations with Setup II, error term contributions from both cameras were considered. In both setups, data inside the visual/inertial sensor was routed through a Field Programmable Gate Array (FPGA) so timestamps could be assigned in hardware with true concurrency to all sensor readings including the LRF trigger output. Datasets were collected by dynamically moving the sensor suite in front of a checkerboard while exciting all rotational degrees of freedom to render all calibration parameters observable (Li and Mourikis, 2014; Mirzaei and Roumeliotis, 2008). The geometry of the checkerboard was known a priori, and the pattern was approximately aligned with gravity to facilitate an initial estimate of $\mathbf{g}_{\mathtt{w}}$.


$\underrightarrow{\mathcal{F}}_{w}$
$\underrightarrow{\mathcal{F}}_{c}$
$\underrightarrow{\mathcal{F}}_{i}$
$\underrightarrow{\mathcal{F}}_{c}$
$\underrightarrow{\mathcal{F}}_{l}$
$\underrightarrow{\mathcal{F}}_{i}$

(a)

(b)

Figure 3.4: In *Setup I* (Fig. 3.4a) $\underrightarrow{\mathcal{F}}_c$ marks the coordinate frame of the single MT9V034 global shutter image sensor used in the experiments. $\underrightarrow{\mathcal{F}}_i$ shows the IMU coordinate frame (Analog Devices ADIS16488), while $\underrightarrow{\mathcal{F}}_w$ marks the inertial frame attached to a static calibration pattern. *Setup II* (Fig. 3.4b) additionally employs an LRF, whose coordinate frame is denoted with $\underrightarrow{\mathcal{F}}_l$. In this setup, both cameras were employed and the IMU used was an ADIS16448.

Table 3.2: Experiments

| **Section** | **Sensor Setup** | **Dataset** | **Est.** | **Objective** |
|---|---|---|---|---|
| 3.4.3 | simulated | 500 sinusoidal motions, 90 second each, 5 temporal offsets | **J** | show accuracy, lack of bias for spatio-temporal calibration with accurately known references |
| 3.4.4 | Setup I (Fig. 3.4a) | 40 hand-guided motions, ∼90 second each, 4 fixed camera exposures | **J**, **G**, **A** | show accuracy in recovering exposure-dependent temporal offsets, general applicability through novel estimator (**A**) |
| 3.4.5 | Setup II (Fig. 3.4b) | 30 hand-guided motions, 60 second each, 3 simulated temporal offsets | **L**, **C** | demonstrate general applicability through novel estimators (**L**, **C**), accuracy of spatial estimates through visual overlay |
| 3.4.6 | Setup II | 100 hand-guided motions, 30 second each, varying initial estimates | **L**, **J** | provide a general intuition about the convergence domain |
| 3.4.7 | Setup II | 30 hand-guided motions, 60 second each, 3 simulated temporal offsets, ranges timestamped on arrival | **L** | highlight impaired calibration caused by non-deterministic corruptions of timestamps, evaluate remedies |
| 3.4.8 | Setup II | 30 hand-guided motions, ∼60 second each, 3 fixed camera exposures, delay compensation | **J** | demonstrate temporal offset removal through filter delay and exposure compensation |

### 3.4.2 Description of experiments

We will adhere to the naming conventions for estimators established in table 3.1. Table 3.2 provides an overview of the experiments covered in the following sections.

For estimators **J**, **G** and **A**, 4 sets of 10 continuous motions were recorded with Setup I. Each run was of about 90 s in length, and different sets exhibited different, fixed exposure times. In order to evaluate estimators **L** and **C**, Setup II was employed to record a single dataset of about 30 minutes length in a mostly empty room with uncluttered walls, floor and ceiling. The dataset was subsequently split into 3 sets of 10 one minute recordings. The timestamps of range measurements within one set were artificially offset by $-5$ ms, 0 ms and 5 ms respectively. For this dataset, we fixed the exposure time and allowed the LRF to warm up to avoid range bias drift (Demski, Mikulski, and Koteras, 2013). With both platforms, inertial measurements were recorded at a rate of 200 Hz, while the frame rate of the cameras was fixed to 20 Hz. The Hokuyo UTM-30LX provided full 270° range scans at a rate of 40 Hz. In order to limit run time of the calibration, range measurements were sub-sampled to about 15 % of their initial count. We performed an Allan Variance analysis (IEEE Aerospace and Electronic Systems Society. Gyro and Accelerometer Panel and Institute of Electrical and Electronics Engineers, 1998, 1999) for both types of IMU to identify the parameters of our noise model. Using the well established camera calibration toolbox (Bouguet, 2004), lens distortion was modeled according to the equidistant model (Kannala and Brandt, 2006), and it was assumed that the the projections of landmarks into the images were subject to isotropic, zero-mean Gaussian distributed noise with a standard deviation of 0.5 pixels. Based on the findings in Demski, Mikulski, and Koteras (2013), the standard deviation of additive zero-mean Gaussian distributed noise corrupting range measurements was set to 7.5 mm.

To initialize the pose spline, we first employed the perspective n-point algorithm from Bouguet (2004) to obtain initial estimates of the pose of the sensor at times when images had been taken. These estimates were transformed into the IMU frame using the initial estimate of the inter-sensor transformation. Finally, the IMU pose spline was initialized using the linear solution of Schoenberg and Reinsch (Chapter XIV of de Boor (2001)).

In addition to experiments on real data, we conducted a study of estimator **J** on simulated data, using noise characteristics identical to those assumed for Setup I and a similar sensor configuration with a slightly longer lever arm between camera and IMU. Five sets of 100 realizations of a 90 s experiment were simulated, where the sensor moved in front of landmarks distributed in a planar, regular grid. The motion in position and orientation followed compositions of sinusoidal functions. Each set exhibited a different simulated delay of $-8\,\mathrm{ms}$, $-4\,\mathrm{ms}$, $0\,\mathrm{ms}$, $4\,\mathrm{ms}$, and $8\,\mathrm{ms}$ respectively.

### 3.4.3 The estimation J constitutes a well-posed problem.

We simulated a camera rotated by 180° about the optical axis and displaced by $\mathbf{t}^{\mathtt{i}}_{\mathtt{c}\mathrm{sim}} = \begin{bmatrix}103, & -15, & -10\end{bmatrix}^T$ mm, where $\mathbf{t}^{\mathtt{i}}_{\mathtt{c}}$ marks the translational component of the transformation $\mathbf{T}_{\mathtt{c},\mathtt{i}}$. The initial estimate for the relative orientation was accurate up to a few of degrees, while an initial displacement estimate of $\mathbf{t}^{\mathtt{i}}_{\mathtt{c}\mathrm{init}} = \begin{bmatrix}0, & 0, & 0\end{bmatrix}^T$ mm was provided to the estimator. The initial estimate of the delay was set to $0\,\mathrm{ms}$.

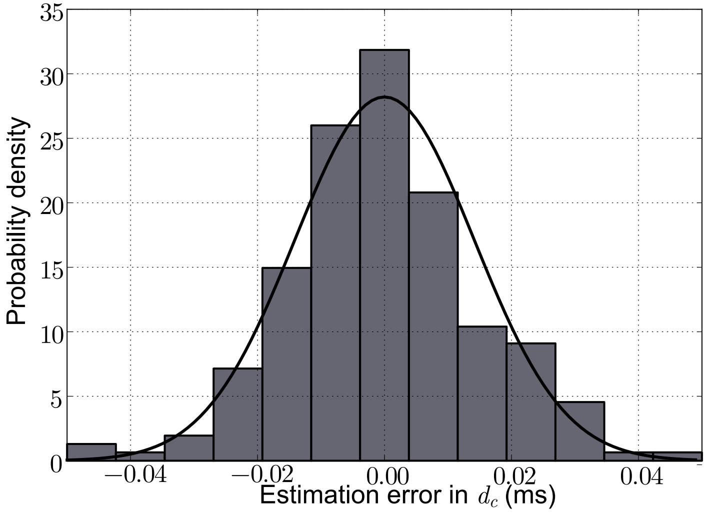


Figure 3.5: A histogram of the error in the estimated time offset over 500 simulation trials with the time offset varying between $-8\,\mathrm{ms}$ and $8\,\mathrm{ms}$. The marginal uncertainty returned by the estimator is plotted as a Gaussian probability density function (solid black line). The results clearly show that, if the correct noise models are known, this method is able to estimate the time offset between the two devices and return a reasonable uncertainty of the estimate.

Figure 3.5 depicts a histogram of errors in time offset estimation overlaid with the marginal uncertainty returned by estimator **J** and plotted as a Gaussia Probability Density Function (PDF). The plot shows that, given the correct noise models, the approach is capable of accurately estimating the time offset and returning a reasonable

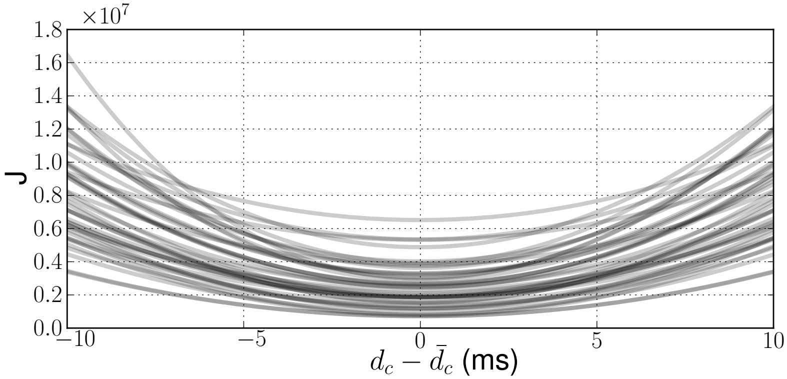


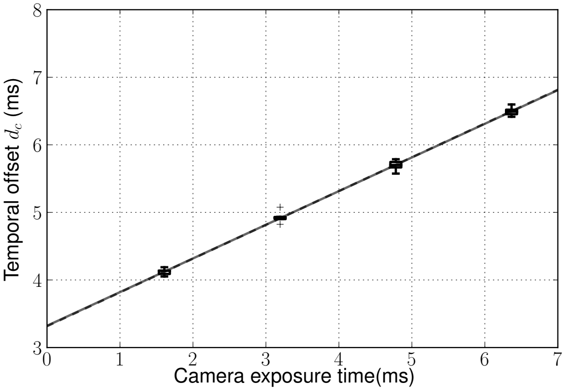
Figure 3.6: The cost function evaluated for different values of the time offset, $d_c$, in the neighborhood of the minimum, $\bar{d}_c$, on 40 real datasets from Section 3.4.4. In this neighborhood, the cost function is convex with respect to $d_c$.

uncertainty of the estimate. The mean and standard deviation of all estimates of the displacement between camera and IMU was $\bar{\mathbf{t}}^{\mathtt{i}}_{\mathtt{c}\mathrm{est}} = [103.73, -15.18, -9.98]^T \pm [0.38, 0.98, 0.17]^T$ mm. Yaw, pitch and roll were estimated as $\overline{\varphi}_{\mathrm{est}} = [179.999°, -0.010°, 0.000°]^T \pm [0.003°, 0.009°, 0.007°]^T$.

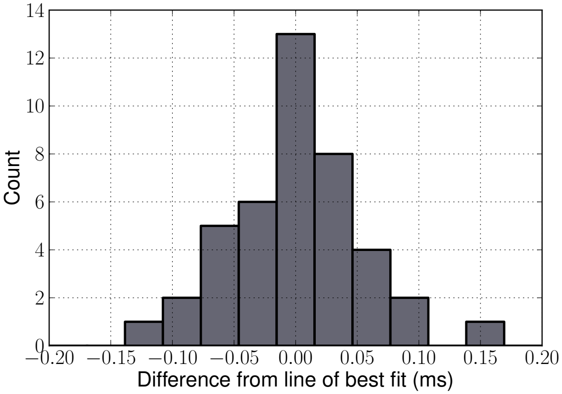
Figure 3.6 shows the cost function evaluated in the neighborhood of the minimum on the 40 real datasets also employed in Section 3.4.4. The figure clearly shows that the cost function in the neighborhood of the minimum is convex and steep with respect to changes in $d_c$. This further suggests that the optimization problem is well-posed. While these experiments do not constitute a formal proof, they suggest that, given an accurate noise model, the estimator faithfully recovers the calibration quantities.

### 3.4.4 The temporal offset estimate is accurate to a fraction of the shortest measurement interval

This section describes the evaluation of estimators **J**, **G** and **A** and compares our approach to temporal calibration with the existing state of the art. We used a dataset consisting of 4 sets of 10 recordings collected with Setup I. For each set of 10, the camera had a different fixed exposure time. Hardware clock synchronization (see scheme 1, Section 3.1) was used; timestamps were assigned to images according to their triggering instant and inertial measurements were timestamped when a polling operation was initialized. For all experiments, we used an initial guess of 0 ms temporal offset, no translation between IMU and camera, and a relative orientation of 180° rotation about the optical axis of the camera.



(a) Estimation incorporating camera, gyroscopes and accelerometer



(b) Difference of offset estimation to the line of best fit.

Figure 3.7: Results from Section 3.4.4 on camera to IMU calibration. Figure 3.7a depicts the time offset estimated for four different, fixed exposure times and ten datasets per exposure setting. The estimation made use of all inertial sensors available in the IMU. The slope of the line of best fit (drawn as a solid gray line) is estimated as 0.498, which compares well with its theoretical value of 0.5 (marked by the dashed gray line). Figure 3.7b shows a histogram of the difference between the estimates and the line of best fit for all 40 experiments. For all datasets, the difference stays within a domain of ±0.2 m, which constitutes just 4 % of the measurement interval of the IMU.

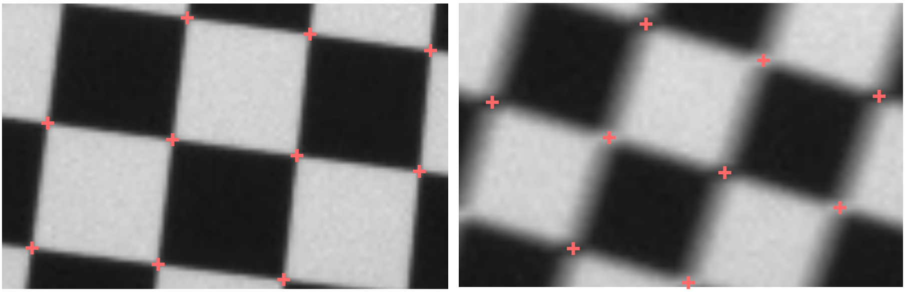


Figure 3.8: Images of the checkerboard may be blurred due to the motion of the camera. This figure shows details from two images taken from one of the datasets. The corner finding algorithm used in this paper performs well for images taken with a static camera (left) as well as under motion blur (right), returning the location of the corner near the middle of the exposure time for the vast majority of motions.

Figure 3.7 depicts the key results for the temporal calibration as a comparison between estimated time offsets and fixed exposure times. The middle of the exposure time constitutes the ideal point to timestamp an image (Maune, Photogrammetry, and Sensing, 2007) and Fig. 5.5 illustrates this; each corner point extracted from an image in the presence of motion blur resembles the position of the projection of the corresponding world point in the middle of the exposure time. In this experiment, we expected the time offset to account for fixed communication and filter delays *plus half the exposure time* as the images were timestamped at the start of the exposure time. Hence, we expect that a plot of temporal offset versus exposure time should show a linear relationship with slope 0.5. As the true delays are unavailable for our experiments, we evaluate the results using (a) the deviation in slope of the line of best fit from the theoretical value and (b) the Root Mean Square (RMS) error with respect to a line of slope 0.5, fitted in a least squares sense.

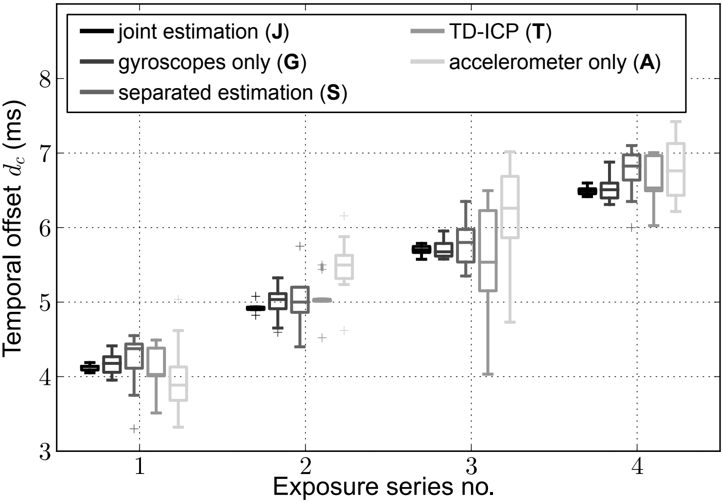
Figure 3.7a shows that our framework is capable of reproducing the inter-sensor time delay up to high accuracy, estimating a slope of 0.498. Figure 3.7b shows the differences of the estimates to the line of best fit, which are all below 0.2 ms. This suggests that the method is accurate to a fraction of the IMU sampling period of 5 ms. Over all experiments, mean and standard deviation of the spatial calibration between camera and IMU were determined as $\bar{\mathbf{t}}^{\mathtt{c}}_{\mathtt{i}\,\mathrm{est}} = [74.54, -8.68, 12.39]^T \pm [1.61, 0.91, 0.76]^T$ mm for displacement and $\overline{\varphi}_{\mathrm{est}} = [180.753^\circ, 0.178^\circ, -0.165^\circ]^T \pm [0.021^\circ, 0.060^\circ, 0.042^\circ]^T$ for yaw, pitch and roll.

We compared our approach to reference implementations of two established approaches based on distinctively different principles. The approach proposed by Mair et al. (2011) completely separates spatial and temporal calibration using the frame independent absolute angular velocity (labeled **S**). Our measurement models will rarely be an absolute faithful representation of all processes involved when assigning a numeric value to a physical quantity. Whether a separation of independent entities in the estimator is advantageous depends on the degree of discrepancy between model and reality and the amount of information omitted in the separation. The Time Delay Iterative Closest Point Algorithm (TD-ICP) algorithm by Kelly, Roy, and Sukhatme (2014) estimates the relative orientation between camera and gyroscopes along with a temporal offset by aligning orientation curves in a fashion resembling the iterative closest point algorithm (labeled **T**). Our comparison further included subsets of the sensor suite to enable an evaluation of the gain from considering richer information, but also to allow a direct comparison to established approaches and even to highlight the broad applicability of our approach by demonstrating spatio-temporal calibration for a combination of a camera and accelerometers (labeled **A**) or camera and gyroscopes (labeled **G**).

Figure 3.9 visualizes this comparison and provides a table comparing the estimated slope and RMS error for each estimator. The results suggest that incorporating measurements from all available sensors into a continuous-time batch optimization yields significantly more accurate and consistent results compared to algorithms that only make use of a subset of the measurements at hand. In our experiments, the gain from the additional information comprised in the accelerometer readings appears to outweigh possible drawbacks of jointly estimating parameters that could be separated otherwise. Under the assumption that the distribution of accelerations approximates a Gaussian distribution, $\mathbf{T}_{\mathtt{c,i}}$ is fully observable for estimator **A**. We empirically determined the parameters of this distribution using a single dataset and applied them to all subsequent evaluations.

### 3.4.5 The approach extends to other sensor configuration and synchronization modalities

This section compares spatio-temporal calibration results using different clock synchronization methods, either (a) hardware synchronization (scheme 1), or (b) unidirectional software synchronization (scheme 3). To correct the host timestamps in the second case, the algorithm of Zhang, Liu, and Honghui Xia (2002) was employed to remove jitter and skew (see Section 3.4.7 for more details on synchronization in software).



| | J | G | S | T | A |
|---|---|---|---|---|---|
| **slope** | 0.498 | 0.493 | 0.531 | 0.515 | 0.553 |
| **RMS** | 0.054 ms | 0.165 ms | 0.344 ms | 0.467 ms | 0.572 ms |

Figure 3.9: Comparison of approaches for determining the time offset between a camera and IMU. The joint estimation (**J**) incorporating all sensor information available results in significantly increased precision in the estimates and the most consistent results. Using a subset of all sensors—either only the gyroscopes (estimator **G**) or only the accelerometer (**A**) in addition to the camera—yields less precise estimates. A separation of temporal and spatial calibration (**S**) as proposed in Mair et al. (2011) resulted in less precise estimates, suggesting that the calibration may benefit more from additional measurements than from the separation of uncorrelated parameters. Employing the same sensors as estimator **G**, the TD-ICP (**T**) approach (Kelly, Roy, and Sukhatme, 2014) produced less accurate estimates.

The experiments in this section use the dataset recorded with Setup II. In order to quantify the performance of the temporal calibration, simulated offsets of $-5\,\mathrm{ms}$, $0\,\mathrm{ms}$, and $5\,\mathrm{ms}$ were applied to three disjoint sets of 10 recordings. Each recording was one minute long. Using this data, we ran the following experiments:

1. Estimator **L** (Camera/IMU/LRF) combined with hardware synchronization, outlined as synchronization scheme 1 in Section 3.1.
2. Estimator **L** (Camera/IMU/LRF) with synchronization scheme 3.
3. Estimator **C** (Camera/LRF) with synchronization scheme 3.

The approach can be tuned through a small set of parameters, and we employed an independent dataset recorded in a different environment to adapt the RANSAC threshold, $\varepsilon$, to $60\,\mathrm{mm}$ and the M-Estimator parameter to $15\,\mathrm{mm}$, twice the standard deviation of the noise assumed to corrupt the range measurements. Plane hypotheses were rejected when they were supported by less that 10 points or the ratio of the two largest eigenvalues of the matrix expressing the second central moment of the supporting points with respect to the smallest eigenvalue fell below a threshold 10 and 5 respectively. The same preliminary experiment was used to determine the parameters of the Gaussian processes, (3.14), governing the accelerations in estimator **C**. The initial estimate was set to $0\,\mathrm{ms}$ temporal offset and no translation between $\underrightarrow{\mathcal{F}}_{i}$ and $\underrightarrow{\mathcal{F}}_{l}$. The initial orientation estimate was correct to about a degree.

Fig. 3.10 depicts the results of recovering the simulated delays. For the Hokuyo UTM-30LX, the device timestamp and hardware trigger mark different instants within one scanning cycle. For improved comparability, we compensated for this in Fig. 3.10 by removing a fixed offset from the estimates obtained using hardware synchronization. The accuracy of these results suggests that the applicability of our approach is not limited to camera/IMU calibration, but that it extends to other sensor suites as well. Table 3.3 shows results for different quantities estimated in this experiment. For estimator **L**, both hardware and software synchronization yielded similar mean displacements with comparable $1\sigma$ precision. Omitting inertial measurements in estimator **C** resulted in a similar spatial offset and only slightly increased variance in the estimates over all datasets.[2] These values compare well with the displacement, $\mathbf{t}^{\mathtt{l}}_{\mathtt{i}\,\mathrm{meas}} = [69, -42, -65]^T\,\mathrm{mm}$, determined by hand measurement.[3] The temporal off-

---

[2] In this experiment, the transformation between $\underrightarrow{\mathcal{F}}_{l}$ and $\underrightarrow{\mathcal{F}}_{c}$ is estimated. For improved comparability, results were transformed with $\mathbf{T}_{\mathtt{c},\mathtt{i}}$ determined in the second experiment.

[3] The discrepancy in this measurement to previous work (Rehder et al., 2014) can be attributed to a different displacement in the movable LRF mount as well as to an improved understanding of the location of $\underrightarrow{\mathcal{F}}_{i}$ with respect to the IMU package. These measurements are only accurate to a few millimeters.

set, $d_l$, estimated for different synchronization modalities, comes close to the respective value stated in the Hokuyo UTM-30LX product and communication protocol specifications.[4] These documents state an expected offset of 2.725 ms for the synchronization trigger signal and of 0 ms with respect to the timestamps assigned by the device. The results further suggest that the estimated range bias was reproducible across different synchronization modalities and individual 60 s recordings.

Arguably, misalignments in the orientation of the LRF with respect to $\underrightarrow{\mathcal{F}}_i$ will have the most impact in the majority of applications. Given that the alignment of $\underrightarrow{\mathcal{F}}_i$ with respect to the physical IMU package was not accurately known, there was no way of obtaining an accurate reference for the relative orientation for Setup II. To assess the repeatability of the orientation estimate, we evaluated the square root of the variance with respect to the Fréchet expectation (Pennec, 1999) over all experiments. Values of 0.096° for estimator **L** and 0.121° for estimator **C** suggest precise orientation estimates.

Table 3.3: Results for LRF Spatio-Temporal Calibration

| | estimator **L**, hardware synchronized | estimator **L**, software synchronized | estimator **C**, software synchronized |
|---|---|---|---|
| spatial displacement $\bar{\mathbf{t}}^{\mathtt{l}}_{\mathtt{i}\,\mathrm{est}}$ [mm] | $\begin{bmatrix} 70.0 \pm 1.4 \\ -40.9 \pm 1.5 \\ -67.4 \pm 1.0 \end{bmatrix}$ | $\begin{bmatrix} 69.8 \pm 1.4 \\ -40.8 \pm 1.7 \\ -67.3 \pm 0.9 \end{bmatrix}$ | $\begin{bmatrix} 71.2 \pm 2.0 \\ -41.8 \pm 1.4 \\ -66.9 \pm 1.1 \end{bmatrix}$ |
| temporal offset $\bar{d}_l$ [ms] | $2.603 \pm 0.045$ | $-0.023 \pm 0.106$ | $0.106 \pm 0.088$ |
| cumulative bias [mm] $\bar{b}_l$ | $-16.6 \pm 4.0$ | $-17.4 \pm 4.2$ | $-21.8 \pm 7.1$ |

Fig. 3.11 provides a visual validation of the accuracy of both spatial and temporal calibration of the LRF with respect to the camera. The hardware synchronization scheme 1 was employed to produce this plot. The points sampled by the LRF were observable in the images as a bright line. The resemblance of this line with the simulated projection of the sampled ranges is representative of the entire calibration chain: Both spatial transformations $\mathbf{T}_{\mathtt{c},\mathtt{i}}$ and $\mathbf{T}_{\mathtt{i},\mathtt{l}}$ were expressed with respect to $\underrightarrow{\mathcal{F}}_i$ and chained to form $\mathbf{T}_{\mathtt{c},\mathtt{l}}$. In addition, the projection employed the camera intrinsics and distortion parameters obtained in a separate calibration procedure (using the calibration

---

[4] Accessible at `http://www.hokuyo-aut.jp/02sensor/07scanner/download/pdf/UTM-30LX_spec_en.pdf` and `http://www.hokuyo-aut.jp/02sensor/07scanner/download/pdf/URG_SCIP20.pdf` (Jan. 2015)

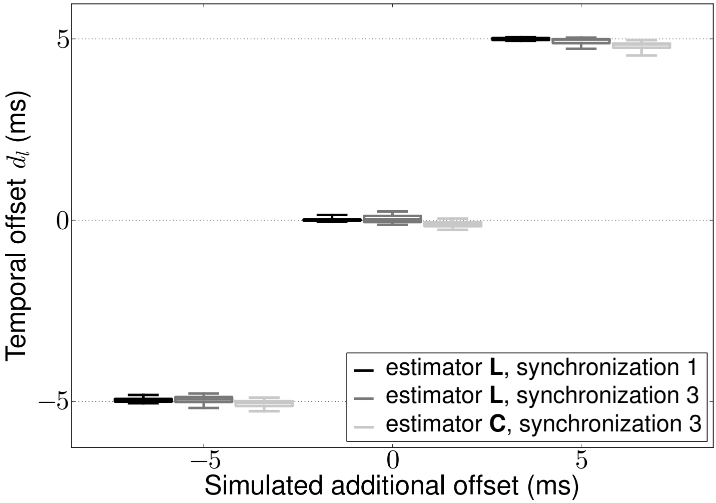
procedure for fisheye lenses provided in Bouguet (2004)). Finally, the timestamps of the image and the range scan were corrected according to $d_c$ and $d_l$. In the figure, the green line corresponds to a projection modelled according to the initial estimate, while blue and red plots mark the projection for these quantities determined by measuring by hand and through calibration respectively. The top Fig. 3.11 depicts a superimposition for an extended exposure time and clearly shows the improvement over the initial estimate as well as over the measured transformation. The bottom Fig. 3.11 visualizes an overlay for a short exposure time of about 2.5 ms, which allows for an evaluation of the temporal calibration. The close resemblance of the red line with the visible scan suggests that the temporal offset was estimated accurately. The green and blue plots are based on delays of 0 ms and 2.725 ms respectively.

Fig. 3.11 allows for a visual assessment of the accuracy of the approach, but it provides little intuition about the impact of spatio-temporal calibration on real applications. Fig. 3.1 provides this context, demonstrating the improvement in reconstruction results produced by correctly handling temporal relationships between sensors.

### 3.4.6 The convergence domain suggests practical applicability

To assess the applicability of the approach to a particular calibration problem, the accuracy to which the calibration parameters have to be known a priori is of vital importance. Following Kelly, Roy, and Sukhatme (2014), we will label this region of initial estimates for which the calibration produces “correct” parameters the *domain of convergence*. Here, we will present on a brief study conducted for estimators **J** and **L**. The shape of this domain is highly dependent on factors that are difficult to quantify, such as the precise motion of the setup during calibration. Consequently, this study is qualitative in nature and merely provides an intuition about the accuracy to which the calibration parameters have to be known a priori.

We used 100 thirty seconds chunks of the 30 one minute datasets recorded with Setup II. For each realization of the experiment, the mean estimate of transformations $\mathbf{T}_{c,i}$ and $\mathbf{T}_{i,l}$ from Section 3.4.5 was perturbed by a random translation and rotation. We used an axis-angle representation for the rotation and sampled the axis uniformly on the sphere, while the angle was sampled from a uniform distribution within some bounds. Analogously, the direction of the translatory perturbation was sampled on the entire sphere, and the magnitude was sampled from a uniform distribution. A corruption to the initial estimate of the temporal offsets $d_c$ and $d_l$ was sampled from a uniform distribution. Automatic plane detection almost certainly fails for vastly incorrect initial estimates, and hence we limited their range to accuracies that are realistically achievable by careful system design and measurements by hand. The



| | **L/1** | **L/3** | **C/3** |
|---|---|---|---|
| **slope** | 0.997 | 0.988 | 0.988 |
| **RMS** | 0.057 ms | 0.114 ms | 0.110 ms |

Figure 3.10: Estimated temporal offsets between an LRF and additional sensors using either hardware clock synchronization (L/1) or unidirectional timestamp correction (L/3, C/3). In order to quantify the performance of the temporal calibration, different simulated offsets were applied to three disjoint sets of 10 recordings, each of one minute length. These results are consistent with values determined from the Hokuyo UTM-30LX product specification and they suggest that the approach is capable of accurately estimating delays. They also show that careful software clock synchronization has precision and accuracy comparable to hardware clock synchronization.

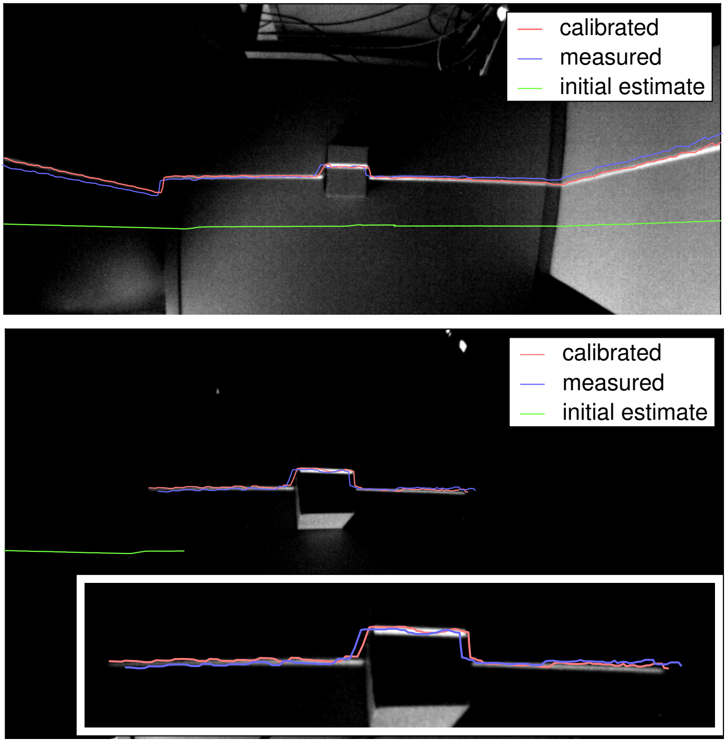


Figure 3.11: Superimposition of camera images and modelled projections of laser scans. The green line depicts a projection using the initial estimate for $\mathbf{T}_{\mathrm{c},\mathrm{l}}$ and $d_l$ prior to calibration. Blue and red lines visualize projections based on these quantities determined by measuring by hand and by our calibration procedure respectively. The top figure shows an image recorded with extended camera exposure time. In the bottom figure, the camera exposure was set to 2.5 ms, capturing the fraction of the scan that falls within the shutter time as a bright line. The close resemblance between the visible band of points sampled by the LRF and its modelled projection suggest an accurate spatial and temporal calibration. Note the significant improvement over the initial estimate provided to our calibration approach.

corruption to $d_l$ was sampled in the bounds of ±50 ms, while translations and rotations were sampled in the bounds of 0.2 m and 10° respectively. For estimator **J**, we sampled from a much wider range, allowing initial perturbations in the bounds of ±100 ms, 1.0 m and 90° respectively. Based on the precision of the estimates observed in Section 3.4.4 and 3.4.5, we defined a metric for classifying calibration results as correct. Estimator **J** was considered successful, if the error in $d_c$ was smaller than 100 µs and the magnitude of the error translation and rotation did not exceed 5 mm and 0.5° respectively. For estimator **L**, these bounds were set to 2.0 ms, 10 mm and 1°. Additionally, calibration with estimator **L** could fail at the plane detection stage, and it was considered unsuccessful when a fewer than 2 planes were detected.

By this metric, estimator **J** provided correct calibrations in 92 experiments, while estimator **L** was successful in 79 cases. We observed a weak correlation between the magnitude of the perturbation and the success rate of both estimators, with the probability of an unsuccessful calibration increasing with larger errors in the initial estimate. Nevertheless, successful runs were observed over the entire range of perturbations. A stronger correlation could be observed between the initial temporal offset $d_l$ and its post calibration error, with estimates biased in the direction of the perturbation and in the order of around 2 ms at 50 ms initial corruption. Presumably, this deterioration in accuracy is rooted in the automatic plane detection returning a biased set of measurements in support of the initial temporal offset. This effect could potentially be mitigated by iterating the calibration multiple times, initializing each plane detection with the previously calibrated quantities. For the temporal offset $d_c$, this effect could not be observed.

While the significance of our quantitative findings is limited to the very realisation of the experiment, the results generally suggest that the approach is applicable to initial estimates within the range of accuracy that can be realistically achieved by measuring by hand.

### 3.4.7 A closer look at synchronization in software

This section describes the effort that is required to get good host timestamps from a device that only provides measurement timestamps based on its internal clock. Many off-the-shelf devices operate this way and we believe it is useful to present a full example, describe the tools needed to clean up the host timestamps, highlight some common pitfalls of timestamp processing, and show what happens when these effects are not accounted for.

Mair et al. (2011) propose a procedure to recover host timestamps for devices that *do not* provide internal timing information. Their method removes the effects of stochastic corruptions and corrects "data jams" where several messages are delayed and then arrive together. While their work can serve as guidance for such systems, here we will take the Hokuyo UTM-30LX as an example to focus on synchronization with sensors that do not allow for bidirectional communication, but provide device timestamps along with their measurements. Although the Hokuyo UTM-30LX does implement a communication protocol for synchronization, we chose to use it in this section because it exhibits timestamp quantization. Hence, treating the device as a one-way communicator allows us to examine all timestamp issues with a single example. Specifically, we will look at (a) clock offsets, (b) jitter, (c) clock skew, and (d) timestamp quantization in systems with multiple clocks and exclusively unidirectional communication. As before, we require communication delays, though subject to noise, to remain constant.

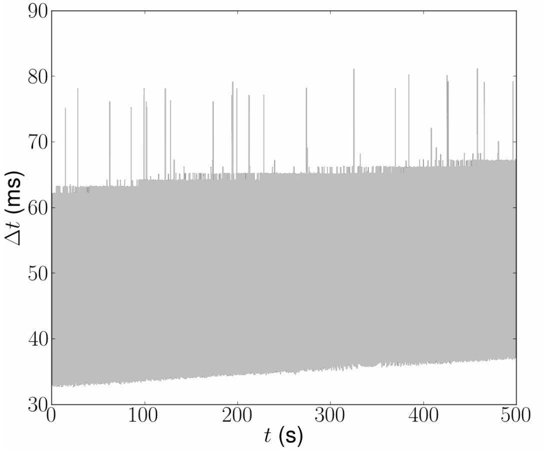
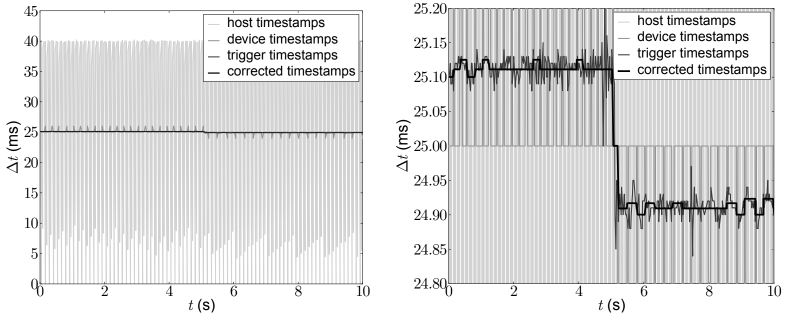


Figure 3.12: The difference between host time and device time for the example of Hokuyo UTM-30LX, plotted over host time. Slightly different rates in individual clocks result in skew, which can be significant even for short datasets. At the same time, various effects in the host system corrupt the timestamps non-deterministically.

Fig. 3.12 depicts the difference between device timestamps and host timestamps for a period of about 8 minutes of the 30 minutes dataset recorded with Setup II. The plot highlights all the previously listed effects: First, there is a distinct lower bound to the difference in timestamps, resulting from a combination of the offset between the clocks and the communication delay between sensor and host system. Second, the delay exhibits significant randomness, often referred to as jitter. Third, even over this short period, the clocks diverge, caused by slightly differing rates and apparent as an increasing lower bound. Forth and more subtly, the upper bound of the plot shows a



(a) (b)

Figure 3.13: Difference in timestamps of successive scans as perceived by different modalities. Timestamping on arrival at the host results in excessive jitter. The Hokuyo UTM-30LX provides device timestamps with 1 ms granularity, which results in frequent spikes in the perceived measurement intervals. Improving timestamp resolution by means of interpolation increases accuracy significantly and corresponds well to timings independently measured in hardware by recoding the trigger instant as highlighted in the scaled view, Fig 3.13b.

staircase profile with 1 ms steps, which hints to the coarse quantization of the device timestamps.

Fig. 3.13 depicts the difference in timestamps of successive measurements and identifies the host timestamps as the obvious source of jitter. In contrast, differences in device timestamps are significantly closer to the nominal value of 25 ms. In Sections 3.4.4 and 3.4.5, we demonstrated estimation of delays well below the shortest measurement period and with sub-millisecond precision. The jitter eclipses these delays by far.

Relying solely on device timestamps is equally undesirable, since Fig. 3.12 suggests that even for short datasets, clock skew would noticeably affect the estimated offset. Finally, Fig. 3.13b shows a 200 µs step in the scan rate occurring in the dataset and independently recorded by timestamping the hardware trigger output. Omitting timing information provided by the sensor and following the host timestamp correction approach outlined in Mair et al. (2011) would render the synchronization incapable of perceiving such a change and in turn affect accuracy.

The synchronization approach proposed by Zhang, Liu, and Honghui Xia (2002)—an extension of previous work by Moon, Skelly, and Towsley (1999)—exploits the absence of randomness in device timestamps. It estimates a lower convex hull to the delays in order to determine the skew between host and device clock. Subsequently, the approach administers appropriate corrections to the device timestamps to obtain skew- and jitter-corrected host timestamps. Zhang's correction algorithm assumes that device timestamps faithfully represent the state of the device clock. For the Hokuyo UTM-30LX, this does not strictly hold true, since device timestamps are quantized with 1 ms, while measurement intervals do not strictly adhere to millisecond boundaries. As a result of this coarse quantization, the difference in consecutive timestamps exhibits spikes, as depicted in Fig. 3.13. In our experiments, the coarse quantization resulted in reduced precision in the estimates of the temporal offset as well as in biased results.

In order to counter quantization effects, the resolution of the device timestamps can be increased. Assuming that the quantized timestamps are the result of a rounding operation, we assign the respective timestamp ±0.5 ms to the spikes and interpolate linearly to obtain timestamps for measurements in between. We do not know that rounding was used internally; it could just as easily be a ceiling or floor operation. But, at most, this would introduce a ±0.5 ms temporal offset that would be estimated by our calibration routine. It is important that, after calibration, we use the same timestamp correction and upsampling methods in the device driver so that any delay introduced is still present.

Fig. 3.13b shows the corrected differences in black. While the original timestamps result in discrete steps, the correction yields timings closely resembling those independently recorded in hardware. Fig. 3.13b also highlights the importance of having access to device timestamps in applications with high accuracy demands: Albeit on the scale of a few 100 µs, the measurement period is clearly non-constant and exhibits rapid changes. The varying scanning period has a direct effect on the temporal spacing of the individual range samples, and correcting for this yielded slightly increased precision in the spatial calibration over all realizations of the experiment.

To highlight the importance of removing jitter from measurement timestamps, we repeated the spatio-temporal calibration with estimator $\mathbf{L}$ on the same dataset, but using the timestamps assigned to the range measurements on arrival. Fig. 3.14 depicts the results as compared to estimates obtained with corrected timestamps. The timestamp correction (Zhang, Liu, and Honghui Xia, 2002) fits a lower convex hull based on messages of *minimal* delay, while timestamps assigned on arrival exhibit a significantly different distribution. Consequently, the estimated temporal offset is (and is expected to be) different in both cases, and we removed this fixed difference in Fig. 3.14 to improve comparability.

The distribution of the estimated offset over multiple trials exhibits a significantly increased standard deviation, despite the fact that the simulated additional offsets were

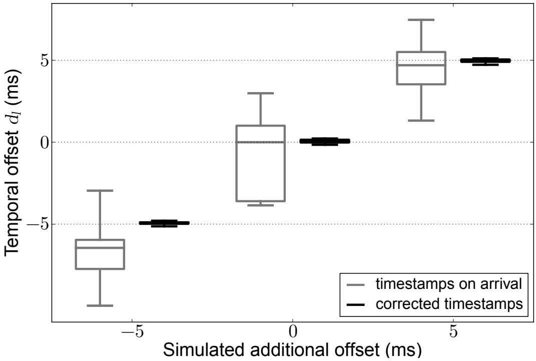


Figure 3.14: Temporal calibration (estimator **L**) using timestamps assigned on arrival as compared to one using timestamps corrected according to Zhang, Liu, and Honghui Xia (2002). The jitter present in timestamps assigned on arrival results in significantly increased standard deviations of the delays estimated for 30 recordings. Nevertheless, the simulated additional delays are approximately recovered.

approximately recovered. The impairment in estimating the temporal relationship reflected in reduced precision in the spatial calibration, with the translation being determined as $\bar{\mathbf{t}}^{\mathtt{l}}_{\mathtt{i}\,\mathrm{est}} = [64.1, -40.3, -67.3]^T \pm [12.1, 6.6, 3.2]^T$ mm. In this case, the square root of the orientation variance with respect to the Fréchet expectation was 0.854° and thus significantly increased over the calibration experiment with corrected timestamps reported in Section 3.4.5.

### 3.4.8 Delay estimation informs better sensor design

Fig. 3.7 suggests that the proposed method is capable of estimating static delays with very high precision. We advocate for making temporal calibration an essential part of any sensor design and for taking any deterministic delays into account at the design stage. This measure reduces the need for post-processing of sensor data and eases the implementation of discrete-time state estimation algorithms by ensuring that measurements are precisely synchronized.

In this section, we highlight some design considerations using Setup II (Fig. 3.4b) as an example. In our work, these insights directly informed the hardware description specifying the configuration of an FPGA, which can be considered as delay compensation in hardware. Implementing the same compensation mechanisms in software on a dedicated processor is equally feasible.

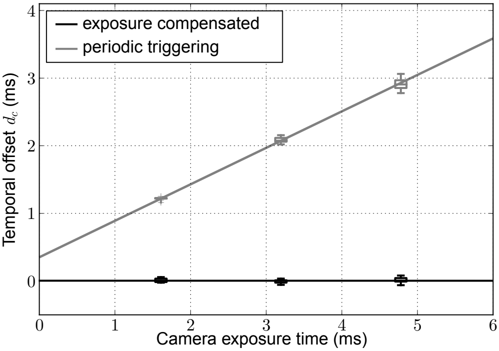


Figure 3.15: Estimated time delays for a compensated sensor setup. In this experiment, the exposure-dependent time delay has been mitigated by bringing the trigger forward by half the exposure time, while filtering and communication delays in the IMU have been compensated for by offsetting the polling of the sensor with respect to the timestamp assigned to the measurements. As a consequence, the delay virtually vanishes.

In order to remove the need to interpolate measurements in discrete-time estimators, we spaced mid-exposure times equally in time and aligned them with time instants at which inertial measurements were sampled. To this end, individual camera trigger instants were brought forward in time by half the exposure time of the image. In contrast, inertial measurements traverse a cascade of buffers and analog and digital filters, thus their output appears delayed. This filter offset is notoriously hard to obtain from manufacturer's information, and the overall delay is further impacted by communication delays that may be specific to the sensor hardware. In order to estimate the IMU offset, we recorded about 30 one minute datasets for 3 fixed exposure times with an uncompensated setup (resembling the experiment proposed in Section 3.4.4). The gray line in Fig. 3.15 depicts the estimated delays. Using this data, the combined filter and communication delay was determined from extrapolating the time delay curve for zero exposure time. Subsequently, we postponed inertial sensor polling with respect to the timestamp by this value. The black line in Fig. 3.15 marks the result of delay estimation for repeating the experiment, but using a sensor employing the proposed delay compensation. After compensating for exposure time and offsetting the IMU polling, the discrepancy in the timestamps of the two measurement modalities virtually vanished. Without additional information, only the *relative* time delay is observable. However, given information from the manufacturer of this particular image sensor that specifies the offset between receiving a trigger pulse and starting exposure of an image at around 10 µs, we could take camera timing as reference. For additional details on the design of the visual/inertial sensor, please see (Nikolic et al., 2014b).

## 3.5 Conclusion and Future Work

In this study, we presented the first general approach to jointly calibrate for temporal offsets and spatial transformations between multiple sensors. Using a continuous state representation allows us to treat the problem of estimating temporal offsets within the rigorous theoretical framework of maximum likelihood estimation.

Established approaches use a two-stage procedure, exploit domain-specific properties and often require making overly simplifying assumptions. In contrast, our approach does not suffer from any of these shortcomings and we believe that the range of estimators presented in this work supports our claim of its general applicability.

For the case of camera/IMU calibration, we showed that it was beneficial to calibrate for time offsets and inter-sensor transformations in a single estimator based on all sensor measurements available. The same estimator then formed the basis for a novel spatio-temporal calibration approach for camera/LRF setups.

Furthermore, this work answered the question about applicable synchronization schemes raised in Furgale, Rehder, and Siegwart (2013) and Rehder et al. (2014), providing experimental results that suggest that the accuracy of calibrations using timestamps corrected for jitter and skew in software, resembles that of hardware synchronized systems. In this context, we also highlighted that this correction can potentially be complex.

Future work will investigate improvements in the model governing range measurements and might extend the IMU model to reflect our improved understanding of accelerations not being perceived in a single location.

# 4

# Extending *kalibr*: Calibrating the Extrinsics of Multiple Inertial Measurement Units (IMUs) and of Individual Axes

Joern Rehder, Janosch Nikolic, Thomas Schneider, Timo Hinzmann, and Roland Siegwart

## 4.1 Introduction

With the costs for inertial measurement units steadily declining and the emergence of integrated visual/inertial sensors, an increasing number of robotics platforms feature multiple inertial measurement units. An example for such a system is the Boston Dynamics quadrupedal platform (Ma et al., 2015) equipped with a tactical grade IMU rigidly mounted to a stereo camera setup and used for visual/inertial odometry and with a navigation grade IMU positioned inside the body of the robot. Another example is the quadrotor platform by Shen, Michael, and Kumar (2015), which employs a low-cost IMU for low-level controls in the autopilot and an additional, high-performance IMU for visual/inertial motion estimation. These platforms have in common that they employ a main IMU positioned and aligned in a way meaningful for locomotion (i.e. mounted close the center of gravity and aligned with the main

Published in:
*2016 IEEE International Conference on Robotics and Automation (ICRA), 2016, ©2016 IEEE*
*DOI:* `10.1109/icra.2016.7487628`

axes of the platform) and a second, auxiliary IMU mounted in the vicinity of some exteroceptive sensors in a location with minimal obstruction by the platform itself. For most platforms, these two locations will be vastly different. In order to make sense of ego-motion estimates from the auxiliary sensor suite for controls and locomotion, they will have to be transformed to the coordinate frame of the main IMU. To this end, an accurate estimate of the transformation between the two coordinate frames is required.

While it is possible to estimate the transformation of both IMUs with respect to the exteroceptive sensor and subsequently chain them, little work presents on fusion of measurements from multiple IMUs inside a single estimator. We suspect one of the reasons for this to lie in the fact that angular accelerations are required to model accelerations perceived in any location outside the Accelerometer Input Axes (IA)—a quantity that is often not measured directly.[1] While it would be possible to derive an estimate of angular acceleration from numerically differentiating angular velocity measurements perceived by the gyroscopes, we pursued a different approach here: The well-established continuous-time batch estimation framework presented by Furgale, Barfoot, and Sibley (2012) fits a spline representing the evolution of the relative orientation of two coordinate frames over time to a series of orientation and angular velocity measurements. Assuming that angular velocity varies smoothly, an estimate of angular acceleration can be directly derived from this orientation curve.

The same estimator enables further applications: High-end IMUs often employ one Integrated Circuit (IC) per axis for acceleration measurements rather than a single IC that combines all axes on a single die. Individual axes may be multiple centimeters apart, which violates the assumption that they are subject to the same acceleration under general motion. If unaccounted for, this introduces errors which are sometimes referred to as the "size effect" in the navigation literature (Hung et al., 1979). Consequently, the offsets of individual axes to the origin of the Input Reference Axes (IRA) should be considered for maximum calibration performance.

The contributions of this work are the following:

- We derive an estimator for simultaneous intrinsic and extrinsic calibration of multiple IMUs with respect to one or multiple exteroceptive sensors.

---

[1] There exist different approaches for measuring angular accelerations, a not so recent review of which is provided in Ovaska and Valiviita (1998). More recently, consumer grade Microelectromechanical Systems (MEMS) angular acceleration sensors have been announced (*Murata announces world's first surface mount MEMS angular acceleration sensor*). However, these devices are currently not widely employed.

- We generalize this estimator to additionally determine the location of individual accelerometer axes.
- We present a comprehensive experimental study demonstrating precise intrinsic calibration and showing that it is possible to locate individual accelerometer axes inside a commercial grade IMU.

The approach was implemented as an extension to the open-source camera/IMU calibration toolbox *kalibr*[2] (Furgale, Rehder, and Siegwart, 2013) and will be released as an update to it.

## 4.2 Related Work

This work is concerned with calibrating a sensor suite comprising one or multiple IMUs and one or multiple exteroceptive sensors. The goal of the calibration is to improve state estimation results obtained from fusing measurements from *all* sensors available. Accordingly, estimating the extrinsics of the IMUs with respect to an exteroceptive sensor is an integral part of the approach, and we will limit the review of related work to approaches similar in scope.

Nevertheless, there exists a large body of work addressing the problem of calibrating redundant IMUs for applications where fusion with additional sensors is not a focus. Possible starting points for further literature review in this direction could be the work by Pittelkau (2005), Hwangbo, Kim, and Kanade (2013),Nilsson, Skog, and Handel (2014).

Mirzaei and Roumeliotis (2008) and Kelly and Sukhatme (2009) proposed an extended Extended Kalman Filter (EKF)-based framework that estimated the transformation between an IMU and a camera from a calibration sequence recorded by moving the setup in front of a visual target. Using a similar calibration procedure, Fleps et al. (2011) determined these quantities by means of batch optimization. Their approach estimated a continuous trajectory encoded as a spline rather than representing the motion as a discrete sequence of states. Furgale, Rehder, and Siegwart (2013) pursued a similar continuous-time approach, but additionally folded the estimation of a temporal offset between camera and IMU into the estimator—a parameter that had previously been estimated in a separate procedure (Kelly, Roy, and Sukhatme, 2014; Mair et al., 2011). Krebs extended the approach by IMU intrinsics (Krebs, 2012). Similarly, Zachariah and Jansson (2010) incorporated intrinsic parameters

---

[2] `https://github.com/ethz-asl/kalibr`

Table 4.1

| | | **Estimated Quantities** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Estimation Approach** | $\mathbf{S}_{\alpha,\omega}$ | $\mathbf{M}_{\alpha,\omega}$ | $\mathbf{A}_{\omega}$ | ${}_{B}\mathbf{r}_{BA}$ | $\mathbf{C}_{BC}, {}_{B}\mathbf{r}_{BC}$ | $d_c$ | $\mathbf{fc}, \mathbf{cc}, \mathbf{kc}$ | $\mathbf{b}_{\alpha,\omega}$ |
| (Mirzaei and Roumeliotis, 2008), (Kelly and Sukhatme, 2009), (Kelly, Roy, and Sukhatme, 2014) | EKF | | | | | • | •* | | • |
| (Fleps et al., 2011), (Mair et al., 2011) | cont.-time batch optimization | | | | | • | •* | | • |
| (Furgale, Rehder, and Siegwart, 2013) | cont.-time batch optimization | | | | | • | • | | • |
| (Zachariah and Jansson, 2010) | SPKF | • | • | • | | • | | | • |
| (Li et al., 2014) | MSCKF | • | • | • | | • | • | • | • |
| (Krebs, 2012) | cont.-time batch optimization | • | • | • | | • | • | | • |
| **Ours** | cont.-time batch optimization | • | • | • | • | • | • | | • |

into a discrete-time Sigma-Point Kalman Filter (SPKF) estimation framework. Recently, Li et al. (2014) demonstrated the estimation of camera/IMU extrinsics, a time delay and IMU intrinsics as an integral part of an online state estimation framework using a Multi-State Constraint Kalman Filter (MSCKF). In contrast to other methods reviewed here, their approach uses natural visual landmarks rather than a dedicated calibration pattern and additionally estimates the camera intrinsic parameters focal length **fc**, principle point **cc** and distortion parameters **kc**.

Our approach is based on (Furgale, Rehder, and Siegwart, 2013) and extends that method to incorporate multiple IMUs into a single estimator. The same formulation can be employed to determine the displacement of individual accelerometer axes, arriving at a more complete model even in sensor suites comprising only a single IMU. Borrowing from Krebs (2012), IMU intrisics were added to the calibration parameters to improve results.

Table 4.1 summarizes these approaches using the notation that will be introduced in Section 4.3.4. Asterisks mark approaches where temporal calibration is performed in a separate, preceding step.

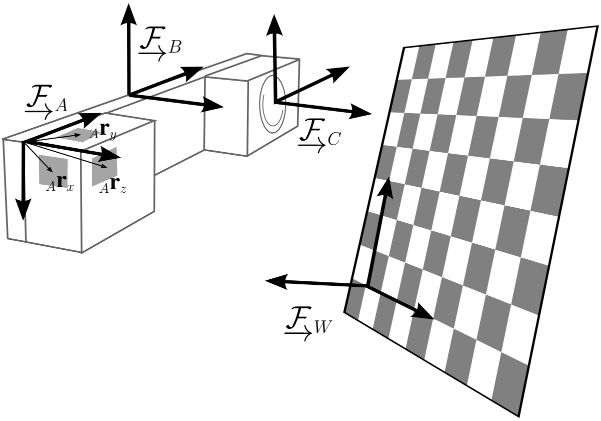

$\underrightarrow{\mathcal{F}}_B$
$\underrightarrow{\mathcal{F}}_A$
${}_A\mathbf{r}_y$
${}_A\mathbf{r}_x$
${}_A\mathbf{r}_z$
$\underrightarrow{\mathcal{F}}_C$
$\underrightarrow{\mathcal{F}}_W$

Figure 4.1: Coordinate frame convention. $\underrightarrow{\mathcal{F}}_W$ denotes the world reference frame, and $\underrightarrow{\mathcal{F}}_B$ and $\underrightarrow{\mathcal{F}}_A$ mark the input reference axes (IRA) and the eccentric IMU frame respectively. A camera was used as exteroceptive sensor, denoted here with $\underrightarrow{\mathcal{F}}_C$. $\underrightarrow{\mathcal{F}}_B$, $\underrightarrow{\mathcal{F}}_C$ and $\underrightarrow{\mathcal{F}}_A$ are connected through a rigid mechanical link. Gray boxes mark the locations of individual accelerometer ICs, which are displaced with respect to $\underrightarrow{\mathcal{F}}_A$ by individual lever arms ${}_A\mathbf{r}_{x,y,z}$. For simplicity, we will assume ${}_A\mathbf{r}_x = \mathbf{0}$ in the following.

# 4.3 Method

## 4.3.1 Coordinate frame conventions

Fig. 4.1 visualizes the different coordinate frames used in this work. The inertial frame $\underrightarrow{\mathcal{F}}_W$ is attached to the calibration pattern. $\underrightarrow{\mathcal{F}}_B$ marks the body IRA, while $\underrightarrow{\mathcal{F}}_A$ and $\underrightarrow{\mathcal{F}}_C$ denote the IMU frame and the frame of the exteroceptive sensor respectively. $\underrightarrow{\mathcal{F}}_B$, $\underrightarrow{\mathcal{F}}_C$ and $\underrightarrow{\mathcal{F}}_A$ are rigidly connected through a mechanical link. We will estimate the time-varying relative orientation and position of $\underrightarrow{\mathcal{F}}_B$ with respect to $\underrightarrow{\mathcal{F}}_W$.

In order for the IRA to be well defined in practice, the spatial offset to at least one axis of at least one IMU needs to be fixated. This could be an arbitrary displacement, but for convenience, we choose to align the IRA with one IMU when calibrating with multiple devices, and the $x$-axis of the IRA with the accelerometer sensing the specific force in $x$ direction when calibrating for displacements of individual axes (i.e. ${}_A\mathbf{r}_x = \mathbf{0}$).

## 4.3.2 Accelerometer model

Here, we will derive the model of inertial measurements used for calibration in the estimator.

Let ${}_W\mathbf{t}_{WB}(t)$ denote the time-varying vector from the origin of coordinate frame $\underrightarrow{\mathcal{F}}_W$ to coordinate frame $\underrightarrow{\mathcal{F}}_B$ expressed in $\underrightarrow{\mathcal{F}}_W$. With this, the acceleration of the origin of coordinate frame $\underrightarrow{\mathcal{F}}_B$ expressed in $\underrightarrow{\mathcal{F}}_W$ is given by ${}_W\ddot{\mathbf{t}}_{WB}(t)$.

The measurements of an ideal accelerometer (i.e. the specific force) at coordinate frame $\underrightarrow{\mathcal{F}}_B$ can be expressed as

$$ {}_B\mathbf{a}_{WB}(t) = \mathbf{C}_{BW}(t)\left({}_W\ddot{\mathbf{t}}_{WB}(t) - {}_W\mathbf{g}\right) \tag{4.1} $$

where $\mathbf{C}_{BW}(t)$ denotes the time-varying direction cosine matrix that transforms a vector from $\underrightarrow{\mathcal{F}}_W$ to $\underrightarrow{\mathcal{F}}_B$, and ${}_W\mathbf{g}$ marks the gravitational force.

Now assume that we would like to model the acceleration of a coordinate frame $\underrightarrow{\mathcal{F}}_A$, rigidly attached to $\underrightarrow{\mathcal{F}}_B$ with constant displacement ${}_B\mathbf{r}_{BA}$. The temporal evolution of the origin of this coordinate frame, expressed in $\underrightarrow{\mathcal{F}}_W$, is given by ${}_W\mathbf{t}_{WA} = {}_W\mathbf{t}_{WB}(t) + \mathbf{C}_{BW}^T(t){}_B\mathbf{r}_{BA}$. Accordingly, ${}_A\mathbf{a}_{WA}(t)$ is given by

$$ \begin{aligned} {}_A\mathbf{a}_{WA}(t) = \mathbf{C}_{AB}^{\alpha}(&\mathbf{C}_{BW}(t)({}_W\ddot{\mathbf{t}}_{WB}(t) - {}_W\mathbf{g}) \\ &+ \lfloor {}_B\dot{\omega}_{WB}(t) \rfloor_{\times} {}_B\mathbf{r}_{BA} \\ &+ \lfloor {}_B\omega_{WB}(t) \rfloor_{\times}^{2} {}_B\mathbf{r}_{BA}) \end{aligned} \tag{4.2} $$

where $\mathbf{C}^{\alpha}_{AB}$ marks the rotation matrix relating $\underrightarrow{\mathcal{F}}_A$ and $\underrightarrow{\mathcal{F}}_B$, ${}_B\omega_{WB}(t)$ denotes the angular velocity of $\underrightarrow{\mathcal{F}}_W$ with respect to $\underrightarrow{\mathcal{F}}_B$ and ${}_B\dot{\omega}_{WB}(t)$ denotes the angular acceleration. The operator $\lfloor\cdot\rfloor_\times$ denotes the skew-symmetric matrix expressing the cross products.

In the most simplistic accelerometer model, we assume that the IA are aligned with $\underrightarrow{\mathcal{F}}_A$ and that the accelerometer measurements $\alpha(t)$ are only affected by noise:

$$\alpha(t) = {}_A\mathbf{a}_{WA}(t) + \mathbf{b}_\alpha(t) + \nu_\alpha \tag{4.3a}$$

$$\dot{\mathbf{b}}_\alpha(t) = \nu_{b\alpha} \tag{4.3b}$$

where $\nu_\alpha$ and $\nu_{b\alpha}$ are zero-mean, white Gaussian noise processes of strength $\sigma^2_\alpha\mathbf{I}$ and $\sigma^2_{b\alpha}\mathbf{I}$. In other words, the accelerometer measurements are independently affected by white noise $\nu_\alpha$ and a slowly varying random walk process of diffusion $\sigma^2_{b\alpha}\mathbf{I}$, $\mathbf{b}_\alpha(t)$.

This model is a good approximation for devices with factory calibrated intrinsics, but may produce impaired calibration results for low-cost, consumer grade inertial sensors which exhibit significant axis misalignment and scale factor errors. Hence, for these sensors, the model is augmented to include misalignment and incorrect scales:

$$\alpha(t) = \mathbf{S}_\alpha\mathbf{M}_{\alpha A}\mathbf{a}_{WA}(t) + \mathbf{b}_\alpha(t) + \nu_\alpha \tag{4.4}$$

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

Equation 6.35 can be extended to accommodate a design trait common to many high-end IMUs: These often employ an individual sensor IC per measurement axis, and there are physical limits on the proximity in which the sensors can be mounted. Consequently, *each axis* is displaced differently from the IRA. With $\omega(t) := {}_B\omega_{WB}(t)$ and $\dot{\omega}(t) := {}_B\dot{\omega}_{WB}(t)$, the complete model with individually displaced accelerometer axes amounts to

$$\begin{aligned} {}_A\mathbf{a}_{WA}(t) =& \mathbf{C}^{\alpha}_{AB}(\mathbf{C}_{BW}(t)({}_W\ddot{\mathbf{t}}_{WB}(t) - {}_W\mathbf{g}) \\ &+ \mathrm{diag}(\lfloor\dot{\omega}(t)\rfloor_\times\mathbf{R}_\alpha + \lfloor\omega(t)\rfloor^2_\times\mathbf{R}_\alpha)), \end{aligned} \tag{4.5}$$

where $\mathrm{diag}(\cdot)$ extracts the $N \times 1$ vector from the diagonal of a matrix and $\mathbf{R}_\alpha$ is composed of the lever arms of individual accelerometers (identified by the subscripts) according to

$$\mathbf{R}_\alpha = \begin{bmatrix} {}_B\mathbf{r}_{BA_x} & {}_B\mathbf{r}_{BA_y} & {}_B\mathbf{r}_{BA_z} \end{bmatrix}. \tag{4.6}$$

### 4.3.3 Gyroscope model

Analogously, given the angular velocity ${}_B\omega_{WB}$ governing the time-varying change in orientation between $\underrightarrow{\mathcal{F}}_B$ and $\underrightarrow{\mathcal{F}}_W$ expressed in $\underrightarrow{\mathcal{F}}_B$, the angular velocity expressed in $\underrightarrow{\mathcal{F}}_A$ is given as

$$ {}_A\omega_{WB}(t) = \mathbf{C}^{\omega}_{AB}\,{}_B\omega_{WB}(t) \tag{4.7} $$

The rationale behind estimating $\mathbf{C}^{\omega}_{AB}$ and $\mathbf{C}^{\alpha}_{AB}$ separately lies in sensor imperfections: The gyroscopes may not be perfectly aligned with the accelerometers, and estimating a single $\mathbf{C}_{AB}$ would in turn be a source of deterministic errors in the model.

Again, a properly factory calibrated gyroscope can be modelled as

$$ \varpi(t) = {}_A\omega_{WB}(t) + \mathbf{b}_{\omega}(t) + \nu_{\omega} \tag{4.8a} $$

$$ \dot{\mathbf{b}}_{\omega}(t) = \nu_{b\omega} \tag{4.8b} $$

where $\nu_{\omega}$ and $\nu_{b\omega}$ are zero-mean, white Gaussian noise processes of strength $\sigma^2_{\omega}\mathbf{I}$ and $\sigma^2_{b\omega}\mathbf{I}$, i.e. the gyroscopes are independently affected by white noise and a random walk process, analogously to the accelerometers.

For consumer grade devices, the influence of axis misalignment and incorrect measurement scaling as well as of linear accelerations on gyroscope measurements (“g-sensitivity”) can be modelled as

$$ \varpi(t) = \mathbf{S}_{\omega}\mathbf{M}_{\omega A}\,\omega_{WB}(t) + \mathbf{A}_{\omega A}\mathbf{a}_{WA}(t) + \mathbf{b}_{\omega}(t) + \nu_{\omega} \tag{4.9} $$

where $\mathbf{S}_{\omega}$ and $\mathbf{M}_{\omega}$ are defined analogously to $\mathbf{S}_{\alpha}$ and $\mathbf{M}_{\alpha}$ in (4.4), and $\mathbf{A}_{\omega}$ is a fully populated matrix. Despite presumably having different displacements from the IRA, only a single lever arm is considered in the calculation of ${}_A\mathbf{a}_{WA}(t)$. In general, the effect of linear accelerations on gyroscope measurements is small and insufficient to properly constrain the estimate of a spatial displacement. In this work, we assume the accelerometer and gyroscopes to be sufficiently close and employ the lever arm estimated for the accelerometers. In cases where an individual lever arm per accelerometer axis is determined, the estimate of one axis is employed for all axes of the gyroscope.

### 4.3.4 The estimator

So far, we established the basis for modelling accelerometer and gyroscope measurements from devices mounted with an offset to the IRA. Generally, these models could be employed in any estimator. However, both, (6.35) and, to a lesser extent, (4.9), depend on angular accelerations, and this quantity is not measured directly in most sensor suites. Accordingly, it has to be inferred, and we employ the continuous-time batch optimization paradigm (Furgale, Barfoot, and Sibley, 2012), which estimates a continuously differentiable sensor trajectory, yielding a smooth estimate of angular accelerations.

In the following, we will give a brief introduction to continuous-time estimation, which will follow Furgale, Rehder, and Siegwart (2013) very closely. For a more thorough derivation, please see the original publication (Furgale, Barfoot, and Sibley, 2012).

Time-varying states are represented as the weighted sum of a finite number of known analytical basis functions. For example, a $D$-dimensional state, $\mathbf{x}(t)$, may be written as

$$\Phi(t) := \begin{bmatrix} \phi_1(t) & \dots & \phi_B(t) \end{bmatrix}, \quad \mathbf{x}(t) := \Phi(t)\mathbf{c}, \tag{4.10}$$

where each $\phi_b(t)$ is a known $D \times 1$ analytical function of time and $\Phi(t)$ is a $D \times B$ stacked basis matrix. We estimate $\mathbf{x}(t)$ by determining $\mathbf{c}$, a $B \times 1$ vector of coefficients.

While various basis functions are feasible, we employ B-splines due to their simple analytical derivatives, good representational power and finite temporal support, yielding a sparse system of equations in the estimator that can be solved efficiently.

The pose of $\underrightarrow{\mathcal{F}}_B$ is parameterized as a $6 \times 1$ spline with 3 degrees of freedom for relative translation and 3 degrees of freedom for relative orientation:

$$_W\mathbf{t}_{WB}(t) := \Phi_t(t)\mathbf{c}_t \tag{4.11}$$

$$\varphi(t) := \Phi_\varphi(t)\mathbf{c}_\varphi\,. \tag{4.12}$$

In this paper, we use the axis/angle parameterization for rotations, where $\varphi(t)$ represents rotation by the angle $\varphi = \sqrt{\varphi(t)^T\varphi(t)}$ about the axis $\varphi(t)/\varphi(t)$. The orientation of $\underrightarrow{\mathcal{F}}_W$ with respect to $\underrightarrow{\mathcal{F}}_B$ at time $t$ is given by $\mathbf{C}_{BW}(t) := \mathcal{C}\big(\varphi(t)\big)^T$, where $\mathcal{C}(\cdot)$ is a function that builds a direction cosine matrix from the orientation parameters $\varphi(t)$.

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

$$_W\ddot{\mathbf{t}}_{WB}(t) = \ddot{\Phi}_t(t)\mathbf{c}_t \tag{4.13}$$

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

Angular velocity and angular acceleration as perceived in $\underrightarrow{\mathcal{F}}_B$ are computed as

$$
\begin{aligned}
{}_B\omega_{WB}(t) &= \mathbf{C}_{BW}(t)\; {}_W\omega_{WB}(t) && (4.14)\\
{}_B\dot{\omega}_{WB}(t) &= \mathbf{C}_{BW}(t)\; {}_W\dot{\omega}_{WB}(t) && (4.15)
\end{aligned}
$$

with

$$
\begin{aligned}
{}_W\omega_{WB}(t) &= \mathbf{S}\big(\varphi(t)\big)\dot{\varphi}(t) = \mathbf{S}\big(\Phi(t)\mathbf{c}_\varphi\big)\dot{\Phi}(t)\mathbf{c}_\varphi && (4.16)\\
{}_W\dot{\omega}_{WB}(t) &= \mathbf{S}\big(\varphi(t)\big)\ddot{\varphi}(t) = \mathbf{S}\big(\Phi(t)\mathbf{c}_\varphi\big)\ddot{\Phi}(t)\mathbf{c}_\varphi && (4.17)
\end{aligned}
$$

where $\mathbf{S}(\cdot)$ is the matrix relating parameter rates to angular velocities and accelerations (Hughes, 1986).

For both, orientation and translation, a sixth-order B-spline is employed, which encodes linear and angular acceleration as a cubic polynomial.

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

$$
\mathbf{b}(t) := \Phi_b(t)\mathbf{c}_b \tag{4.18}
$$

The estimator further requires inputs from exteroceptive sensors to sufficiently constrain the trajectory. This can be any sensor that acquires measurements sufficient to render all quantities of interest observable. Since this work is an extension of *kalibr*, we employed a global shutter camera with a static calibration pattern for this purpose.

Projections of reference points on the calibration pattern ${}_W\mathbf{p}^i$ are modelled according to the well-established pinhole camera model

$$
\begin{aligned}
\mathbf{y}_k^i = f(\mathbf{C}_{BC}^T(&\mathbf{C}_{BW}(t_k + d_c)\big({}_W\mathbf{p}^i + {}_W\mathbf{t}_{WB}(t_k + d_c)\big)\\
&+ {}_B\mathbf{r}_{BC})) + \nu_y \quad ,
\end{aligned} \tag{4.19}
$$

where the function $f(\cdot)$ denotes a perspective projection. $d_c$ is an unknown relative temporal offset that compensates for either the IMU or the camera assigning timestamps with a fixed offset with respect to their measurement instant. We assume that the projections are corrupted in the image plane by a zero-mean, discrete-time, white Gaussian noise process of variance $\sigma_y^2\mathbf{I}$.

The estimator is formulated as a non-linear least-square optimization problem. Our previously introduced measurement models ((4.4), (4.9), and (4.19)) are all of the form $\mathbf{m}(t) := h(\Theta, t) + \nu$, where $\Theta$ is a vector containing all estimated quantities, $t$ denotes the instant at which the measurement was recorded and the model is evaluated, and $\nu$ is a zero-mean, white Gaussian noise process of strength $\sigma^2\mathbf{I}$. Accordingly, the contribution of measurements $\tilde{\mathbf{m}}_k^i$ recorded with sensor $i$ at times $[t_1, \ldots, t_N]$ to the objective function $J$ can be formulated as

$$J_i := \sum_{k=1}^{N} \frac{1}{\sigma_i^2} \left| \tilde{\mathbf{m}}_k^i - h^i(\Theta, t_k) \right|^2 . \tag{4.20}$$

Contributions from bias terms are evaluated according to

$$J_b := \int_{t_1}^{t_N} \frac{1}{\sigma_b^2} \left| \dot{\mathbf{b}}(\tau) \right|^2 \mathrm{d}\tau. \tag{4.21}$$

The objective function is composed from these sensor and bias terms, and the estimate is determined as the $\Theta$ that minimizes $J$:

$$\Theta = \underset{\Theta}{\operatorname{argmin}} \left( J_\alpha + J_\varpi + J_y + J_{b_\alpha} + J_{b_\omega} \right). \tag{4.22}$$

We employ the Levenberg-Marquardt algorithm (Nocedal and Wright, 2006) for non-linear optimization.

The following table lists the parameters and states comprised in $\Theta$ and partitions them into time-varying and time-invariant, and IMU and "auxiliary" parameters.

| | **Time-Invariant** |
|---|---|
| $\mathbf{C}_{AB}^{\alpha}$ | orientation of the accelerometers |
| $\mathbf{C}_{AB}^{\omega}$ | orientation of the gyroscopes |
| ${}_B\mathbf{r}_{BA}$ | displacement of the IMU |
| $\mathbf{S}_\alpha$ | accelerometer scale factors |
| $\mathbf{M}_\alpha$ | accelerometer misalignment |
| $\mathbf{S}_\omega$ | gyroscope scale factors |
| $\mathbf{M}_\omega$ | gyroscope misalignment |
| $\mathbf{A}_\omega$ | effect of linear accelerations on gyroscopes |
| $\mathbf{C}_{BC}$ | orientation of the camera |
| ${}_B\mathbf{r}_{BC}$ | displacement of the camera |
| $d_c$ | temporal offset between IMU and camera |
| ${}_W\mathbf{g}$ | direction of gravity |
| | **Time-Varying** |
| ${}_W\mathbf{t}_{WB}$ | position of the IRA expressed in $\underset{\rightarrow}{\mathcal{F}}_W$ |

| | |
|---|---|
| $\varphi$ | orientation parameters of the IRA |
| $\mathbf{b}_a$ | accelerometer bias |
| $\mathbf{b}_\omega$ | gyroscope bias |

# 4.4 Experiments

## 4.4.1 Experimental setup and dataset collection

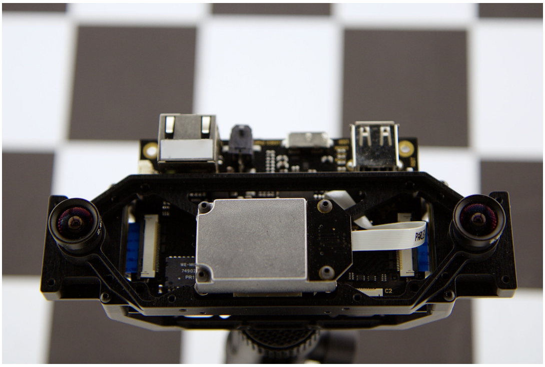


Figure 4.2: The experimental setup used in this work. The integrated visual/inertial sensor(Nikolic et al., 2014b) is equipped with two global shutter image sensors, as well as three MEMS IMU. A factory calibrated Analog Devices ADIS16448 is mounted centrally on the sensor frame; the two consumer grade Invensense MPU9150 IMUs are located on the back of each image sensor board.

For our experiments, we employed a visual/inertial sensor (Nikolic et al., 2014b), which was equipped with an Analog Devices ADIS16448 and two Invensense MPU9150 IMUs. The latter fall into the class of consumer grade devices, while the ADIS16448 was calibrated intrinsically by the manufacturer. The sensor unit was manufactured by Skybotix, but retrofitted with custom firmware to provide control over the filtering of inertial measurements. Both IMUs were sampled at a rate of 800 Hz. For the ADIS16448, a 2-tap filter was enabled; for the MPU9150, we chose a cut-off frequency of about 190 Hz. For exteroceptive perception, two MT9V034 Wide Video Graphics Array (WVGA) global shutter image sensors were employed. The cameras were triggered at a rate of 20 Hz and set to a constant, low exposure time. We calibrated the cameras intrinsically and the stereo extrinsics using the camera calibration functionality of *kalibr* on a separate dataset beforehand.

This setup was dynamically moved by hand in front of a checkerboard of known dimensions. Subsequently, the recorded dataset was split into 20 chunks of *only* 10 s length. We ensured that all rotational degrees of freedom were excited sufficiently.

The parameters of the accelerometer and gyroscope noise models (4.3) and (4.8) were determined from static sensor data, i.e. from measurements where the IMUs were at rest. For this purpose, the sensors were mechanically fixated, and raw sensor measurements were captured at a rate of 800 Hz for a duration of 5 h. The sensor filter and range settings were identical to those used during the experiments. Table 4.3 lists the parameters that were identified and used for the experimental evaluation. Fig. 4.3 shows the sample Allan deviation of the gyroscopes and accelerometers, and the Allan deviation that corresponds to the selected noise model parameters.

For all experiments, we used 50 knots per second for the B-spline representing biases and 250 knots per second for the spline encoding the sensor trajectory.

### 4.4.2 IMU intrinsics and the extrinsics of multiple IMUs can be precisely inferred in a single estimator

For the experiment on extrinsic calibration of multiple IMUs, measurements of the two MPU9150 devices were employed. We defined $\underrightarrow{\mathcal{F}}_B$ to align with one of the devices and included IMU intrinsics—$\mathbf{S}_{\alpha,\omega}$,$\mathbf{M}_{\alpha,\omega}$ and $\mathbf{A}_{\omega}$—for both sensors, as well as extrinsics—$\mathbf{C}_{AB}^{\alpha,\omega}$ and ${}_B\mathbf{r}_{BA}$—into the estimator. The displacement between the IMUs was estimated as ${}_B\mathbf{r}_{BA} = [-5.98, 120.4, -1.02]^T$ mm with standard deviation of $\sigma = [1.44, 0.67, 1.20]$ mm. These values compare well with those determined through measuring by hand ($[-10.0, 121.0, 0.0] \pm [8.0, 8.0, 0.0]$ mm). Note that the displacement between the centroids of the packages was measured, and that the uncertainty bounds (given by the package dimensions of $4 \times 4$ mm and the relative orientations of the two devices) reflect our lack of knowledge about the accurate position of the accelerometer axes inside the device. Due to imperfections in soldering devices to the Printed Circuit Board (PCB) and in the mechanical mount connecting the PCBs holding both MPU9150 IMUs, it is impossible to acquire accurate reference measurements for the relative orientation of the two devices. Instead, we assessed the precision as the square root of the orientation variance with respect to the Fréchet expectation (Pennec, 1999). For $\mathbf{C}_{AB}^{\omega}$, this was evaluated to about $0.01°$, for $\mathbf{C}_{AB}^{\alpha}$ to $1.95°$. Note that absolute accuracy cannot be inferred from this assessment. While the relative orientation of the gyroscopes to the IRA exhibits a small variance, the estimate of the orientation of the accelerometers is noticeably less precise.

To ensure comparability between devices, we repeated the experiment to demonstrate intrinsic calibration for a single MPU9150 and for the ADIS16448. The estimation

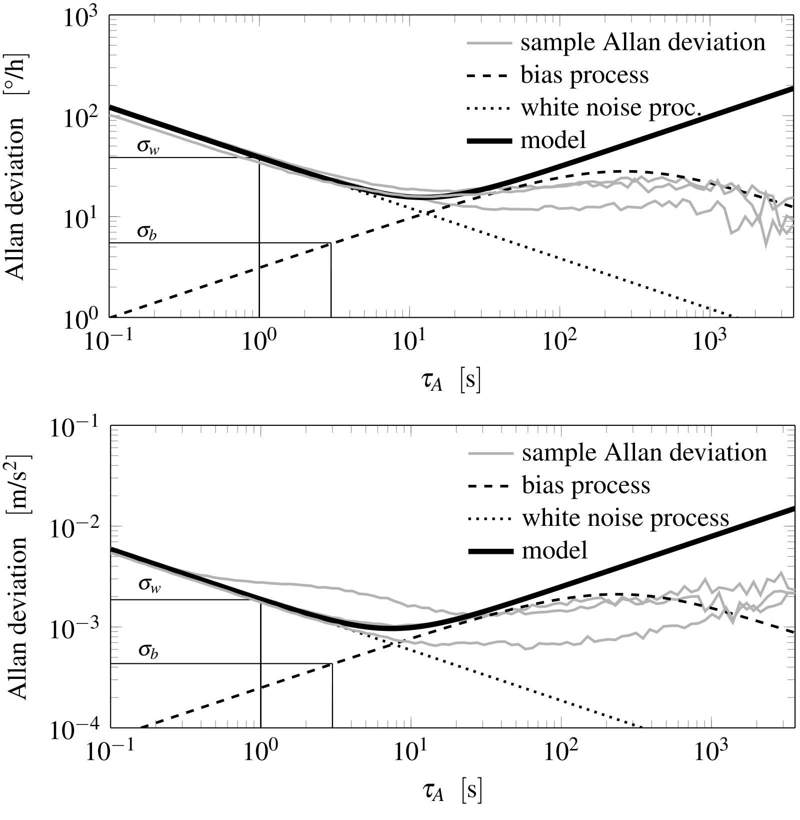


Figure 4.3: Allan deviation of the MPU9150 gyroscopes (top) and accelerometers (bottom). The sample Allan deviations are shown in grey, and the Allan deviations corresponding to the noise model parameters used during the experiments are shown in black (solid).

Table 4.3

| | sym. | unit | ADIS16448 | MPU9150 |
|---|---|---|---|---|
| **Gyroscopes** | | | | |
| White noise str. | $\sigma_\omega$ | $°/(\mathrm{h}\sqrt{\mathrm{Hz}})$ | $3.85 \times 10^{1}$ | $1.84 \times 10^{1}$ |
| Bias diffusion | $\sigma_{b\omega}$ | $\mathrm{rad}/(\mathrm{s}^2\sqrt{\mathrm{Hz}})$ | $2.66 \times 10^{-5}$ | $1.08 \times 10^{-5}$ |
| **Accelerometers** | | | | |
| White noise str. | $\sigma_\alpha$ | $\mathrm{m}/(\mathrm{s}^2\sqrt{\mathrm{Hz}})$ | $1.86 \times 10^{-3}$ | $2.24 \times 10^{-3}$ |
| Bias diffusion | $\sigma_{b\alpha}$ | $\mathrm{m}/(\mathrm{s}^3\sqrt{\mathrm{Hz}})$ | $4.33 \times 10^{-4}$ | $7.53 \times 10^{-5}$ |

results for $\mathbf{S}_{\alpha,\omega}$,$\mathbf{M}_{\alpha,\omega}$ and $d_c$ are summarized in Table 4.4. While the ADIS16448 appears to be well calibrated by the manufacturer, significant gyroscope scale factor errors and axes misalignments were estimated for the MPU9150 (up to 1 % and 1°). For both IMUs, a device intrinsic time delay of approximately 3 ms was estimated. Over the 20 datasets, the standard deviation in the estimates were only about 15 µs and 20 µs respectively.

### 4.4.3 Positions of individual accelerometer axes can be discerned

In this experiment, three different calibrations were performed for the ADIS16448 and one of the MPU9150: A) Assuming that scaling and misalignment errors are compensated for or negligible, a standard IMU/camera calibration was performed, B) misalignment, scale errors and the effect of linear accelerations on gyroscopes were estimated, but individually different accelerometer axis displacements were neglected, and C) a full calibration including individual axis offsets was performed.

Fig. 4.4 depicts the estimated accelerometer positions expressed in $\underrightarrow{\mathcal{F}}_C$ for both IMUs. We determined a rough estimate of the sensor package dimensions by hand and visualized it as gray wire frames in the figures. Crosses (×) mark the estimated position of $\underrightarrow{\mathcal{F}}_A$ for calibration A. For the ADIS16448, the estimates lie clearly within the sensor package and shows a comparatively small dispersion. For the MPU9150—which is not factory calibrated—this estimate exhibits a bias and is located outside the approximate sensor dimensions. Consequently, estimating misalignment, g-sensitivity and scale in calibration B yields improved results for this device as depicted as pluses (+) in Fig. 4.4b. In Fig. 4.4b, the measured footprint does not align

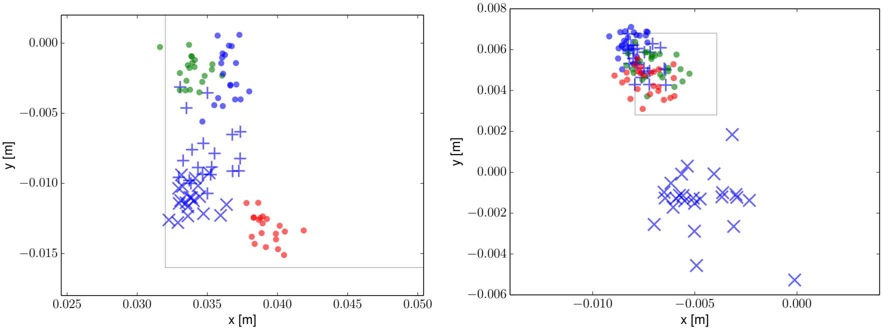
Table 4.4

| symbol | unit | ADIS16448 | MPU9150 |
|---|---|---:|---:|
| **Accelerometer** | | | |
| $\mathbf{S}_\alpha - \mathbf{I}$ | ppm | $1.73 \times 10^3 \pm 1.8 \times 10^3$ | $-4.20 \times 10^3 \pm 2.9 \times 10^3$ |
| | ppm | $-6.37 \times 10^3 \pm 4.1 \times 10^3$ | $7.76 \times 10^3 \pm 1.2 \times 10^3$ |
| | ppm | $6.60 \times 10^3 \pm 6.2 \times 10^3$ | $7.92 \times 10^3 \pm 2.7 \times 10^3$ |
| $\mathbf{M}_\alpha$ | ″ | $-0.80 \times 10^3 \pm 0.81 \times 10^3$ | $-0.17 \times 10^3 \pm 0.53 \times 10^3$ |
| | ″ | $-1.77 \times 10^3 \pm 1.11 \times 10^3$ | $0.85 \times 10^3 \pm 0.54 \times 10^3$ |
| | ″ | $-0.11 \times 10^3 \pm 1.74 \times 10^3$ | $-0.36 \times 10^3 \pm 0.44 \times 10^3$ |
| **Gyroscope** | | | |
| $\mathbf{S}_\omega - \mathbf{I}$ | ppm | $-1.85 \times 10^3 \pm 1.6 \times 10^3$ | $3.01 \times 10^3 \pm 0.6 \times 10^3$ |
| | ppm | $1.85 \times 10^3 \pm 1.1 \times 10^3$ | $-9.29 \times 10^3 \pm 0.5 \times 10^3$ |
| | ppm | $0.99 \times 10^3 \pm 0.3 \times 10^3$ | $2.74 \times 10^3 \pm 0.2 \times 10^3$ |
| $\mathbf{M}_\omega$ | ″ | $-0.14 \times 10^3 \pm 0.16 \times 10^3$ | $-0.31 \times 10^3 \pm 0.09 \times 10^3$ |
| | ″ | $-0.14 \times 10^3 \pm 0.17 \times 10^3$ | $-1.71 \times 10^3 \pm 0.18 \times 10^3$ |
| | ″ | $0.03 \times 10^3 \pm 0.25 \times 10^3$ | $3.13 \times 10^3 \pm 0.13 \times 10^3$ |
| $d_C$ | µs | $2.94 \times 10^3 \pm 20$ | $3.03 \times 10^3 \pm 15$ |

perfectly with our estimates. While the reason for this could lie in biased calibrations, it is similarly plausible that the device footprint measurements are inaccurate, particularly given the complications associated with determining the origin of $\underset{\rightarrow}{\mathcal{F}}_C$. For the Analog Devices product, estimating IMU intrinsics did not yield improved precision, again suggesting that the factory calibration accurately compensates for these effects. Calibration C produced clearly separated estimates for the location of individual accelerometer axes in Fig. 4.4a. The $x$-axis was estimated to be located about a centimeter apart from the $y$- and $z$- axis. This may suggest that it is housed inside a different IC, while the other two axes may share the same die. This separation of axes is less pronounced for the Invensense product, since this IMU is a single $4 \times 4$ mm chip. Nevertheless, the order of axes on the device can be discerned and their approximate location can be inferred. Table 4.5 compiles the estimates of the $y$- and $z$-axis position expressed in $\underset{\rightarrow}{\mathcal{F}}_A$. Note that $\underset{\rightarrow}{\mathcal{F}}_A$ is defined in a way that its $x$-axis aligns with the $x$-axis of the device. Estimates for both devices yielded similar precision.

Table 4.5

| | | x | y | z |
|---|---|---|---|---|
| **ADIS16448** | $y$-axis [mm] | $5.27 \pm 0.53$ | $-11.19 \pm 0.31$ | $0.67 \pm 2.00$ |
| | $z$-axis [mm] | $2.91 \pm 1.12$ | $-11.14 \pm 1.36$ | $-1.15 \pm 2.59$ |
| **MPU9150** | $y$-axis [mm] | $0.71 \pm 0.20$ | $0.33 \pm 0.33$ | $0.32 \pm 0.72$ |
| | $z$-axis [mm] | $1.76 \pm 0.49$ | $-0.80 \pm 0.64$ | $0.71 \pm 1.27$ |



(a) ADIS16448

(b) MPU9150

Figure 4.4: Visualization of ${}_{C}\mathbf{p}_{CA}$, the estimated displacement between camera and IMU. Crosses (×) mark the estimated position of $\underrightarrow{\mathcal{F}}_{A}$ when intrinsics are neglected (estimator A). Each cross indicates the result from one experiment. Pluses (+) visualize the position when IMU intrinsics are included in the estimation (estimator B). The results of the full-scale estimator C are indicated with dots marking the estimated position of each individual accelerometer axis (red: x-axis, green: y-axis, blue: z-axis). Results suggest that the approach is capable of discerning the positions of individual axes. The results are less pronounced for the MPU9150, where all axes are integrated in a package of $4 \times 4$ mm.

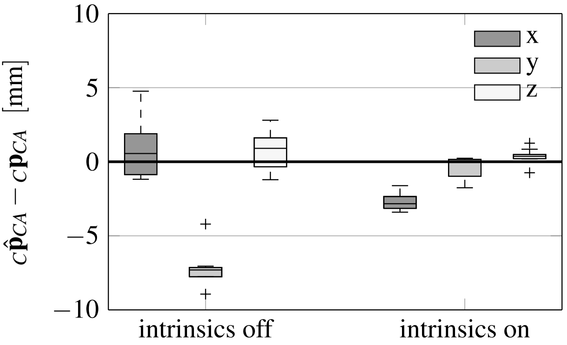


Figure 4.5: Camera-IMU extrinsic translation estimation errors for experiments with the MPU9150, with and without estimating the IMU intrinsic calibration parameters. These results indicate that incorporating IMU intrinsic calibration terms improves the accuracy of the extrinsic calibration parameter estimates.

### 4.4.4 Estimating IMU intrinsics improves camera/IMU extrinsic calibration

Fig. 4.4b suggests that IMU scale errors and misalignments do not only result in increased variance in the estimates, but even in noticeably biased quantities. Accordingly, intrinsic calibration should be an integral part of calibration involving low-cost devices. Fig. 4.5 visualizes the results of Section 4.4.3 quantitatively for the MPU9150: Using reference measurements extracted from Computer-Aided Design (CAD) data, we determined the accuracy of the approach. Note that the estimate of the $y$-axis is significantly biased when IMU intrinsics are not incorporated into the estimator. Including them yields both, higher precision and greater accuracy. Please note that the aforementioned problems regarding the acquisition of reference measurements apply here as well.

## 4.5 Conclusion

In this work, we presented an extension to the open-source calibration toolbox *kalibr* that allows for determining the extrinsics and intrinsics of multiple IMUs in a single estimator. We further demonstrated that it is feasible to infer the location of individual accelerometer axes to millimeter precision.

We believe that the significance of this contribution extends beyond the application of calibrating multiple IMUs, and we intend to further investigate this in future work:

- Neglecting the physical displacement of individual accelerometer axes in high-end IMUs yields a source of deterministic error, which might be worth addressing in applications where accuracy is crucial.
- Where multiple IMUs are available, state estimation may benefit from incorporating measurements from all devices. Recently, different approaches to continuous-time Simultaneous Localization and Mapping (SLAM) have been proposed (e.g. Anderson, MacTavish, and Barfoot (2015) and Patron-Perez, Lovegrove, and Sibley (2015)), and it would be straight-forward to extend these to consider inputs from more than one IMU.

The biggest drawback of fusing data from multiple IMUs or individually displaced accelerometer axes lies in the dependence on angular accelerations, which in most sensor suites are not sensed directly. While this work showed that the continuous-time batch estimation framework is capable of inferring reasonable angular accelerations for our calibration use-case, it remains future work to demonstrate this fact for other applications and estimation frameworks. Furthermore, we did not include temporal offsets (apart from an image delay $d_c$) in the estimator, and future work will estimate individual delays for different IMUs as well as for accelerometers and gyroscopes, acknowledging the fact that these may not employ filters with identical characteristics.

# 5

# A Direct Formulation for Camera Calibration

Joern Rehder, Janosch Nikolic, Thomas Schneider, and Roland Siegwart

## 5.1 Introduction

Camera intrinsic calibration is a mature technology with a growing number of calibration toolboxes freely available (Bouguet, 2004; Furgale et al., 2015b; Scaramuzza, 2006). The majority of these toolboxes relies on some sort of a calibration pattern composed of visual identifiers for known three-dimensional (3D) points. Their calibration routines are based on extracting the position of these identifiers in the calibration images. Subsequently, the positions in the images and the location of the associated points in the world are used to estimate the parameters of a projection model. Commonly, this is achieved via an initialization step which minimizes some algebraic constraints arising from multiple view geometry. The initialization is followed by a probabilistically motivated optimization step which typically minimizes the reprojection error.

This procedure has clearly stood the test of time and is applied widely, both in industry and in research.

However, reducing the richness of image measurements to mere interest point positions as the *very first* step of the calibration marks a substantial abstraction. We argue that for certain use-cases, valuable information is omitted that could either yield

Published in:
*2017 IEEE International Conference on Robotics and Automation (ICRA), 2017, ©2017 IEEE*
*DOI:* `10.1109/icra.2017.7989765`

higher precision in calibration or that would otherwise allow for a more intuitive formulation of the parameter estimation problem.

Throughout this work, we will highlight motion blur and the rolling shutter effect as two examples of such use-cases: Motion blur results from camera motion during image exposure. Since the camera pose varies over the course of the exposure, what is the timestamp that should be assigned to the interest point observations? In rolling shutter cameras, image lines are exposed individually and consecutively. This property leads to distortions in the image when the camera is subjected to motion during the acquisition of a frame. How is uncertainty in the localization of the visual identifiers modelled correctly in the presence of these complex distortions?

To avoid premature abstraction, we drew inspiration from a growing number of approaches to visual state estimation referred to as either *direct* (Forster, Pizzoli, and Scaramuzza, 2014; Tykkälä, Audras, and Comport, 2011) or *(semi-)dense* (Comport, Meilland, and Rives, 2011; Engel, Sturm, and Cremers, 2013). Rather than formulating the camera measurement model on discrete interest point positions, these methods employ image intensities.

Using intensities allows for modelling the image exposure process. Accordingly, the aforementioned use-cases become more intuitive: Rather than assigning an arbitrary timestamp inside the bounds given by the exposure time to an interest point, intensity can be modelled as a function of the exposure time and the camera trajectory during this period. Similarly, instead of modelling the effect that rolling shutter distortion exerts on the uncertainty of corner detections, camera frames can be treated as a composition of individually exposed rows which reduces the modelling of measurement uncertainty to the noise corrupting image intensities.

The scope of this work is *not* to remove the need for interest point extraction altogether. We acknowledge their necessity for proper initialization of all calibration parameters, and we employ the Perspective-n-Point algorithm(Lepetit, Moreno-Noguer, and Fua, 2009) based on corner detections to initialize camera poses in our optimization. Instead, we advocate for replacing the reprojection error with a direct error formulation in the optimization step—particularly in applications where discrete interest point positions mark an inadequate abstraction.

This idea may appear incremental, but we believe that it constitutes an (albeit small) paradigm shift from abstracted quantities with often arbitrarily chosen uncertainties to probabilistically more correct modelling of measurements acquired by cameras. Furthermore, this approach will enable future estimators to account for properties that are currently commonly neglected, such as defocus blur, as well as calibrations directly on Bayer pattern images, circumventing inaccuracies arising from interpolation in demosaicing.

The contributions of this work are as follows:

- We highlight cases where modelling the image acquisition process yields distinct advantages over abstraction to discrete interest points.
- We derive a direct error formulation analogously to the state estimation literature and present a complete calibration pipeline comprising adaptive point selection and dimensionality reduction.
- We extend this basic formulation in different ways to cover the use-cases of line delays in rolling shutter cameras and motion blur in camera/Inertial Measurement Unit (IMU) calibration.
- We present a comprehensive evaluation suggesting that the direct formulation yields competitive results for estimating camera intrinsics from static images, but further allows for estimating exposure time from motion blur and line delays for rolling shutter cameras, both to high accuracy.

## 5.2 Related Work

This work is motivated by what we perceive as a gap in the camera calibration literature: Recent advances in state estimation based on a direct formulation of the camera measurement model have not been matched by similar methods for calibration—despite some potential advantages of this formulation.

Many prevailing methods in camera intrinsic and multi-camera extrinsic calibration (Mei and Rives, 2007; Scaramuzza, Martinelli, and Siegwart, 2006; Zhang, 2000) are based on extracting the position of visual identifiers of known world points from the images. Subsequently, these approaches operate exclusively on the positions and omit all image data. In most implementations, noise on keypoints is modelled as isotropic zero-mean Gaussian corruptions. This choice may be an inaccurate model of the detection uncertainty for views where the target is significantly distorted due to projective foreshortening. In contrast, the direct model is formulated on the lowest level of abstraction from the sensor measurement and hence facilitates a correct treatment of measurement uncertainties. Furthermore, certain calibration use-cases require camera motion during calibration (e.g. camera/IMU calibration (Furgale, Rehder, and Siegwart, 2013; Mirzaei and Roumeliotis, 2008), rolling shutter calibration (Oth et al., 2013)). In this context, (Furgale, Rehder, and Siegwart, 2013) observed that a finite exposure time—and the resulting motion blur—have a profound impact on delay estimation. The authors concluded that the mid-exposure time should be assigned as a timestamp to keypoint observations. Under general motion however, it is not obvious why the corner position returned by the detector should always correspond to the projection at mid-exposure time. Instead, we suspect that the mid-exposure time yields the smallest average error in timestamping. Nevertheless, these errors likely correlate with the motion, which violates fundamental assumptions about their distributions. In applications where camera measurements are leveraged to estimate intrinsic parameters of other sensors (e.g. Rehder et al. (2016) and Zachariah and Jansson (2010)), these correlations and the resulting inaccuracy in timing may severely limit possible insights. In contrast, (Meilland, Drummond, and Comport, 2013) proposed a direct model which incorporates motion blur and would remove the need to assign timestamps at finer granularity than exposure time. Similarly, modelling the uncertainty of interest points detected in images from rolling shutter cameras is highly involved (Oth et al., 2013). Employing a measurement model directly based on intensities yields a more intuitive formulation (Kim, Cadena, and Reid, 2016; Meilland, Drummond, and Comport, 2013).

The direct formulation will introduce radiometric entities into the calibration procedure. In this work, many of the complications arising from this fact are mitigated by careful experimental design and simplifying assumptions. We adopted naming conventions and concepts from Grossberg and Nayar (2004) and Debevec and Malik (2008), and we will incorporate ideas from Kim and Pollefeys (2008) for folding geometric and radiometric calibration into a single estimator in future work.

## 5.3 Method

In this work, we employ non-linear weighted least-squares optimization over a batch of measurements for parameter estimation. A set of parameters $\Theta$ is determined by minimizing an objective function based on some sensor model $\mathbf{f}(\cdot)$, measurements $\tilde{\mathbf{m}}$, and the covariance $\Sigma$ of the error corrupting the measurements:

$$\Theta = \underset{\Theta}{\operatorname{argmin}} \left(\mathbf{f}(\Theta) - \tilde{\mathbf{m}}\right)^T \Sigma^{-1} \left(\mathbf{f}(\Theta) - \tilde{\mathbf{m}}\right). \tag{5.1}$$

In a direct formulation of camera measurements, the measurement vector $\tilde{\mathbf{m}}$ is composed of image intensities at individual pixels. The sensor model $\mathbf{f}(\cdot)$ corresponds to a rendering of the respective camera view based on the parameters $\Theta$. Next, we will introduce the sensor model in general terms, followed by the more concrete example of a pinhole camera observing an evenly lit checkerboard. For this setup, further considerations concerning motion blur and rolling shutter sensors will be presented.

## 5.3.1 Basic Formulation

### 5.3.1.1 Direct error formulation

For the direct error formulation, we are interested in the mapping from target radiance to image intensities:

$$\mathrm{L}({}_W\mathbf{p}) \mapsto \mathrm{B}({}_I\mathbf{p}). \tag{5.2}$$

We will make the assumption that the target behaves perfectly Lambertian. Hence, the radiance function of the target is formulated as dependent on its coordinates ${}_W\mathbf{p}$ expressed in the world coordinate frame $\underrightarrow{\mathcal{F}}_W$, but not as dependent on the viewing direction. The function governing image intensities is expressed in terms of image coordinates ${}_I\mathbf{p}$. Mapping (5.2) can be decomposed into a *geometric* component that concerns the mapping from $\underrightarrow{\mathcal{F}}_W$ into the image (${}_W\mathbf{p} \mapsto {}_I\mathbf{p}$) and a *radiometric* component that concerns the mapping from scene radiance to image intensity.

The geometric mapping is given by a coordinate transformation and a subsequent projection $\pi(\cdot)$:

$${}_W\mathbf{p} \xmapsto{\mathbf{T}_{CW}} {}_C\mathbf{p} \xmapsto{\pi(\cdot)} {}_I\mathbf{p}. \tag{5.3}$$

Given the camera pose $\mathbf{T}_{CW}$, a world point ${}_W\mathbf{p}$ is transformed into the coordinate frame of the camera, $\underrightarrow{\mathcal{F}}_C$, via

$${}_C\mathbf{p} = \mathbf{T}_{CW}({}_W\mathbf{p}). \tag{5.4}$$

The camera is characterized by a projection function ${}_I\mathbf{p} = \pi({}_C\mathbf{p})$ that maps points from $\underrightarrow{\mathcal{F}}_C$ onto the image plane.

The radiometric part is given by

$$\mathrm{L} \xmapsto{\mathrm{S}(\cdot)} \mathrm{E} \xmapsto{\int \mathrm{d}t} \mathrm{X} \xmapsto{\mathrm{R}(\cdot)} \mathrm{B} \tag{5.5}$$

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.

Here, we adhere loosely to the naming conventions introduced in Grossberg and Nayar (2004) and Debevec and Malik (2008)—work which is recommended for more detailed insights into the radiometric aspects of the image forming process. We further follow Debevec and Malik (2008) in justifying the use of the term *irradiance* while echoing the authors' note that this use neglects the weighting with the spectral response function of the sensor.

Image intensity $\mathrm{B}(\cdot)$ at an image point ${}_I\mathbf{p}$ is given by the target radiance at the corresponding point ${}_W\mathbf{p}$ and the radiometric mapping of the optical setup as well as the characteristics of the image sensor used:

$$\mathrm{B}({}_I\mathbf{p}) = \mathrm{R}\left(\int_{t_0}^{t_0+t_e} \mathrm{S}\left(\mathrm{L}({}_W\mathbf{p})\right)\mathrm{d}t\right); {}_I\mathbf{p} = \pi\left(\mathbf{T}_{CW}({}_W\mathbf{p})\right) \tag{5.6}$$

where $t_0$ marks the start of exposure and $t_e$ is the exposure time.

In practice, two more aspects are of interest: Image sensors are composed of many sensitive elements (“pixels”) which have a finite area. Accordingly, the intensity perceived at a discrete pixel location ${}_I\hat{\mathbf{p}}$ is given by the integral of the sensor irradiance $\mathrm{E}(\cdot)$ over the area of the respective sensitive element and over exposure time $t_e$. The finite size of the sensor element is *neglected* here in order to reduce the computational complexity. However, the sensor might not be at rest during image exposure. In this case, different ${}_W\mathbf{p}(t)$ will map to a single image location ${}_I\mathbf{p}$ during image formation; an effect we exploit to estimate exposure time.

The predicted measurement error $e_\mathrm{B}({}_I\hat{\mathbf{p}})$ for a single pixel is given as the difference between the predicted intensity $\mathrm{B}({}_I\hat{\mathbf{p}})$ and the measured intensity $\tilde{\mathrm{B}}_{{}_I\hat{\mathbf{p}}}$ at this position:

$$e_\mathrm{B}({}_I\hat{\mathbf{p}}) := \mathrm{B}({}_I\hat{\mathbf{p}}) - \tilde{\mathrm{B}}_{{}_I\hat{\mathbf{p}}}. \tag{5.7}$$

The error for an entire image or a subset of an image is composed of the errors of all the individual pixels:

$$\mathbf{e}_\mathrm{B} := \begin{bmatrix} e_\mathrm{B}({}_I\hat{\mathbf{p}}^1) \\ \vdots \\ e_\mathrm{B}({}_I\hat{\mathbf{p}}^N) \end{bmatrix}; {}_I\hat{\mathbf{p}}^1, \ldots, {}_I\hat{\mathbf{p}}^N \in \mathrm{P}, \tag{5.8}$$

where P denotes the set of all pixels active in the error term.

The intensity error (5.8) depends on a set of different parameters:

- The unknown pose of the camera with respect to $\underset{\rightarrow}{\mathcal{F}}_W$, $\mathbf{T}_{CW}(t)$.
- The result of $\pi(\cdot)$ depends on parameters such as focal length, the location of the principal point, and parameters of the lens distortion model.
- $\mathrm{L}(\cdot)$ as well as $\mathrm{S}(\cdot)$ and $\mathrm{R}(\cdot)$ may be described by models depending on unknown parameters. These parameters may be coefficients of a polynomial describing vignetting or factors weighing basis functions that form the response function (Grossberg and Nayar, 2004; Kim and Pollefeys, 2008).

– For certain applications, the exposure time $t_e$ may be of interest but unknown.

The optimization problem (5.1) is solved iteratively for these parameters. In each iteration, a set of error functions (5.8) is linearized around the current parameters estimates. With $\mathbf{J}_\chi := \frac{\delta \mathbf{e}_\mathrm{B}}{\delta \chi}$ and using the function names as an identifier for their parameter set, the linearization is given as

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_\mathrm{R} & \mathbf{J}_{t_e} & \mathbf{J}_\mathrm{S} & \mathbf{J}_\mathrm{E} & \mathbf{J}_\pi & \mathbf{J}_{\mathbf{T}_{CW}} \end{bmatrix}. \tag{5.9}$$

To implement (5.9), we render exposure images $\mathrm{X}({}_I\hat{\mathbf{p}})$ and compute their spatial derivatives numerically by means of an image gradient operation, using a Sobel kernel for increased robustness. The individual Jacobians can then be computed through repeated application of the chain rule (Tykkälä, Audras, and Comport, 2011).

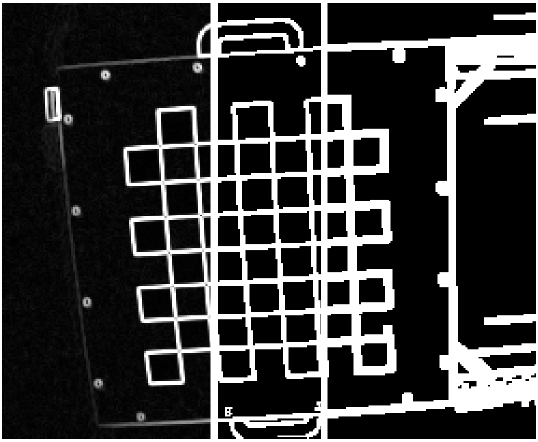
#### 5.3.1.2 Point selection

For rendering $\mathrm{X}(\cdot)$ with the aforementioned assumptions, $\mathrm{L}(\cdot)$ must be evaluated at discrete positions ${}_W\hat{\mathbf{p}}$. To reduce the amount of computation, $\mathrm{L}(\cdot)$ should be queried as sparsely as possible. Hence, we focus on the most informative areas, which correspond to pixel positions ${}_I\hat{\mathbf{p}}$ where $\mathrm{X}({}_I\hat{\mathbf{p}})$ exhibits large gradients. To identify these, a reverse approach is taken: Assuming that the estimated parameters are reasonably close to the true values, the rendered exposure image will be similar to the exposure that induced the measured image $\tilde{\mathrm{B}}_{{}_I\hat{\mathbf{p}}}$. Accordingly, pixel locations with large gradients in the *measured* image $\tilde{\mathrm{B}}$ are likely to also yield informative regions in $\mathrm{X}(\cdot)$. This is formalized as a classification function $\mathrm{C}(\cdot)$ dependent on threshold t:

$$\mathrm{C}({}_I\hat{\mathbf{p}}) := \begin{cases} 1 & \text{if } \left|\nabla \tilde{\mathrm{B}}_{{}_I\hat{\mathbf{p}}}\right| > \mathrm{t} \\ 0 & \text{else} \end{cases} \tag{5.10}$$

The set of active pixels ${}_I\hat{\mathbf{p}} \in \mathrm{P}$ is composed of all locations with $\mathrm{C}({}_I\hat{\mathbf{p}}) = 1$. These locations are projected back onto the target given the current parameter estimates. Linearization (5.9) requires the gradient of the rendered exposure to be defined for all ${}_I\hat{\mathbf{p}}$ for which error terms are considered. To achieve this, the exposure image is additionally rendered for all pixels adjacent to pixels with large gradients. This functionality is implemented as a morphological dilation operation on $\mathrm{C}(\cdot)$. Fig. 5.1 visualizes the individual steps of the process.

For this approach to be correct, the parameters *do not* need to be known precisely. In the worst case, the radiance function is sampled in less informative regions, but since it is defined in terms of ${}_W\mathbf{p}$, the association between $\mathrm{L}(\cdot)$ and ${}_W\mathbf{p}$ will always be correct.



(a) (b) (c)

Figure 5.1: Visualization of the point selection process. Fig. 5.1a depicts the magnitude of the result of the Sobel operation on the input image. Fig. 5.1b and Fig. 5.1c show the result of the subsequent classification $\mathrm{C}(\cdot)$ and the dilation operations.

#### 5.3.1.3 Error compression

The aforementioned point selection yields a varying number of equations of the form (5.8), where the number mostly depends on the distance of the camera to the calibration target. This number is usually large compared to the number of parameters on which the error terms for a single image (or an image line in the rolling shutter case 5.3.2.2) depend. Accordingly, the system is generally overdetermined. We deem a vast and varying number of our 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 governing the entire measurement sequence (i.e. all camera frames) could be precomputed which would allow for more efficient solving (Furgale, Rehder, and Siegwart, 2013).

Fixating the number of error terms is accomplished through QR decompositions (Golub and Loan, 1996). The Jacobian $\mathbf{J}$ of all intensity errors associated with a single image (or an image line) can be decomposed as

$$\mathbf{J}\mathbf{P} = \mathbf{Q}\mathbf{R} = \begin{bmatrix}\mathbf{Q}_1 & \mathbf{Q}_2\end{bmatrix}\begin{bmatrix}\mathbf{R}_1 \\ \mathbf{0}\end{bmatrix} \tag{5.11}$$

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 compressed Jacobian can then be computed as

$$\hat{\mathbf{J}} := \mathbf{R}_1 \mathbf{P}^T \tag{5.12}$$

and the corresponding intensity errors are given as

$$\hat{\mathbf{e}}_I := \mathbf{Q}_1^T \mathbf{e}_I. \tag{5.13}$$

### 5.3.2 Applications

So far, we introduced the concept in a broader scope. In this section, concrete manifestations will be presented. For the following, we will make some simplifying assumptions:

- A simple radiance model is a sufficiently faithful representation of the reflectance properties of our target; target illumination does not change over the course of a dataset collection.
- Camera gain and exposure are fixated during acquisition of the calibration dataset; the optical transmission function is approximated by an attenuation factor that affects the image uniformly (i.e. no vignetting); the camera response function is assumed to be linear.
- The sensor motion during calibration is sufficiently smooth.

We used a checkerboard pattern due to its simplicity. For points ${}_W\mathbf{p}$ on the target, the radiance function is modelled as

$$\mathrm{L}({}_W\mathbf{p}) := \begin{cases} L_0 & \text{if } \left(\lfloor \frac{{}_W\mathbf{p}_x}{d_x} \rfloor + \lfloor \frac{{}_W\mathbf{p}_y}{d_y} \rfloor\right) \bmod 2 = 0 \\ 0 & \text{else} \end{cases} \tag{5.14}$$

where the operator $\lfloor\cdot\rfloor$ denotes a floor operation and $d_x$ and $d_y$ the extent of individual checkerboard tiles in $x$ and $y$ direction. $L_0$ was chosen arbitrarily, and is not observable without additional information as will be highlighted in (5.18).

The projection function uses a pinhole camera model with distortion modelled as a 4th degree polynomial of the incidence angle (Kannala and Brandt, 2006):

$$\pi({}_C\mathbf{p}, \mathbf{k}) := \mathrm{d}\left(\arctan\left(\sqrt{\left(\frac{{}_C\mathbf{p}_x}{{}_C\mathbf{p}_z}\right)^2 + \left(\frac{{}_C\mathbf{p}_y}{{}_C\mathbf{p}_z}\right)^2}\right)\right) {}_I\check{\mathbf{p}} \tag{5.15}$$

where ${}_{I}\check{\mathbf{p}}$ is the ideal pinhole projection

$$ {}_{I}\check{\mathbf{p}} := \begin{bmatrix} k_0 \frac{{}_{C}\mathbf{p}_x}{{}_{C}\mathbf{p}_z} + k_2 \\ k_1 \frac{{}_{C}\mathbf{p}_y}{{}_{C}\mathbf{p}_z} + k_3 \end{bmatrix} \tag{5.16} $$

and $\mathrm{d}(\cdot)$ denotes the distortion function

$$ \mathrm{d}(\alpha) := \alpha + k_4\alpha^3 + k_5\alpha^5 + k_6\alpha^7 + k_7\alpha^9. \tag{5.17} $$

With the aforementioned assumption about the transmission function $\mathrm{S}(\cdot)$ and the target radiance, all radiometric relations can be collapsed into a single linear mapping. For convenience, we will nevertheless refer to this mapping as the response function $\mathrm{R}(\cdot)$ depending on exposure X:

$$ \mathrm{R}(\mathrm{X}) := s\mathrm{X} + o. \tag{5.18} $$

where $s$ denotes a scale factor and $o$ an offset. There exist ambiguities in the mapping from target radiance to exposure and image intensity. In absence of additional information, the functions $\mathrm{R}(\cdot)$ and $\mathrm{X}(\cdot)$ as well as the parameters $s$ and $o$ elude any physically meaningful interpretation. The parameters even transcend the domains of scene characteristics and camera properties: As an example for image sensors with linear response, offset $o$ will correspond to an inseparable amalgamation of the dark current of the sensor and the radiance of the dark patches in the target weighted by the attenuation induced by the optics. This model marks a vast simplification. Whether the effects of this reduction of complexity are negligible in practice depends on such factors as the shape of the true camera response function and the amount of vignetting induced by the optics. Our results for two different sets of camera makes and optics presented in Section 5.4 suggest that these assumptions do not necessarily constitute an over-simplification that renders the approach inapplicable in practice. To further broaden the applicability of the approach, we are currently working on additionally modelling uneven illumination of the target and vignetting effects.

In the following, camera specific models will be introduced.

#### 5.3.2.1 Global shutter

In the global shutter case with negligible exposure time, the error term $\mathbf{e}_{\mathrm{B}}$ corresponding to a single image is composed of the set of pixels ${}_{I}\hat{\mathbf{p}}^{i} \in \mathrm{P}$ in the frame for which the classification (6.24) yields $\mathrm{C}({}_{I}\hat{\mathbf{p}}^{i}) = 1$. Since the entire image is exposed simultaneously $\mathbf{e}_{\mathrm{B}}$ depends only on a single camera pose $\mathbf{T}_{CW}(t_j)$ at some time $t_j$.

Table 5.1: Estimated parameters

| Parameter | Symbol | |
|---|---|---|
| Camera poses | $\mathbf{T}_{CW}$ | |
| Camera intrinsics | $\mathbf{k}$ | |
| Radiometric properties | $s$, $o$ | |
| Exposure time | $t_e$ | see 5.3.2.3 |
| Temporal offset between lines | $t_l$ | see 5.3.2.2 |

#### 5.3.2.2 Rolling shutter

In the rolling shutter case—again for negligible exposure time—each *image line* is exposed individually and consecutively. According to *Flea3 USB 3.0 Digital Camera Technical Reference* (2016), the temporal offset $t_l$ between lines is further constant and corresponds to the line readout time. To account for this behavior, the intensity error is not formulated on the entire image but on individual rows. The set of pixels ${}_I\hat{\mathbf{p}}^i \in \mathrm{P}$ contributing to the error at line $k$ is determined by $(\mathrm{C}({}_I\hat{\mathbf{p}}^i) = 1) \cap ({}_I\hat{\mathbf{p}}^i_y = k)$. Accordingly, each error term $\mathbf{e}^k_{\mathrm{B}}$ depends on a different camera pose $\mathbf{T}_{CW}(t_j + kt_l)$ where $t_j$ marks the time at which line 0 is exposed.

Contrary to other rolling shutter calibration approaches (see (Oth et al., 2013) for examples), no specific accommodation were made as compared to the global shutter use-case beyond formulating the error on subsets of the entire image. This highlights that the direct error formulation lends itself well to extending existing frameworks designed for global shutter cameras to rolling shutter use-cases.

#### 5.3.2.3 Motion blur

Analogously to Meilland, Drummond, and Comport (2013), motion blur is implemented as discretization of (5.6). With the simplifications introduced in Section 5.3.2, this is achieved as discretization of $\mathrm{X}(\cdot)$ in time:

$$\mathrm{X}({}_I\hat{\mathbf{p}}) := \sum_{k=0}^{K} \mathrm{L}\left({}_W\mathbf{p}\left(t_0 + \frac{k}{K}t_e\right)\right)\frac{1}{K}t_e \tag{5.19}$$

where $t_0$ marks the start of the exposure time and $t_e$ is the estimated exposure time. The term ${}_W\mathbf{p}^{t_k} := {}_W\mathbf{p}\left(t_0 + \frac{k}{K}t_e\right)$ denotes the surface point that projects onto ${}_I\hat{\mathbf{p}}$ at time $t_0 + \frac{k}{K}t_e$. The mapping is given by

$$ {}_I\mathbf{p} = \pi\left(\mathbf{T}_{CW}\left(t_0 + \frac{k}{K}t_e\right){}_W\mathbf{p}^{t_k}\right). \tag{5.20}$$

The above formulation is similar in function to the one presented in Section 5.3.2.1, but it addresses the case when exposure time *cannot* be neglected. Accordingly, a model for rolling shutter cameras in the presence of motion blur can be derived from it analogously to Section 5.3.2.2.

## 5.4 Results

We conducted experiments using the setup depicted in Fig. 6.5. The sensor suite comprises an MT9V034 Complementary Metal-Oxide-Semiconductor (CMOS) global shutter camera, a Point Grey Flea3 FL3-U3-32S2C rolling shutter camera and an ADIS16448 IMU. The global shutter cameras and the IMU were synchronized in hardware and ran at 20 Hz and 800 Hz respectively. The rolling shutter camera was later synchronized in software. During the collection of a single dataset, exposure times and gains were fixated. In all experiments, we assumed intensity measurements to be corrupted by additive zero-mean Gaussian-distributed noise with $\sigma = 5.0$. The calibration target were static checkerboards with $8\times7$ quadratic tiles of size $50\times50$ mm and $70\times70$ mm. We illuminated the targets evenly using two large Light-Emitting Diode (LED) light bar modules powered by a direct current power supply. We further disabled any fluorescent lighting in the vicinity to avoid flickering. While the experiments were set up with particular consideration on even illumination, we observed comparable performance with regular office lighting in preliminary experiments. For data collection, the setup was moved in front of the visual target while ensuring that a) the visual target was in the field of view of the camera for most of the dataset, b) all rotational degrees of freedom were sufficiently excited and c) that the maximum velocity of the motion was limited to values such that corner detection and initial parameter estimation still yielded acceptable results despite the presence of motion blur.

In the estimators determining line delay (Section 5.4.2) and exposure time (Section 5.4.3), we represented the sensor pose $\mathbf{T}_{CW}(t)$ as continuous quantity and further fused IMU measurements to constrain the sensor trajectory, drawing inspiration from Furgale, Barfoot, and Sibley (2012). An IMU is not strictly required in either application: For the approach to yield reasonable results, estimates of the camera pose during image acquisition have to follow realistic trajectories. Such estimator behavior can be enforced by penalizing physically infeasible temporal changes in the sensor pose. While fusion with inertial measurements is one realization of this objective, it can also be achieved via a motion prior, trading measurements for models of the distribution of linear and angular accelerations (Oth et al., 2013).

Table 5.1 compiles a list of all parameters relevant to the different use-cases presented in this section. For each estimator, we will identify those parameters that are estimated *in addition to* the camera poses $\mathbf{T}_{CW}$ and the parameters that govern the radiometric model, $s$ and $o$. Any quantity that is not estimated is either not present in the model (as for example the line delay $t_l$ when a global shutter camera is employed) or is known and constant. In either case, (5.9) is adjusted to only reflect the estimated parameters.

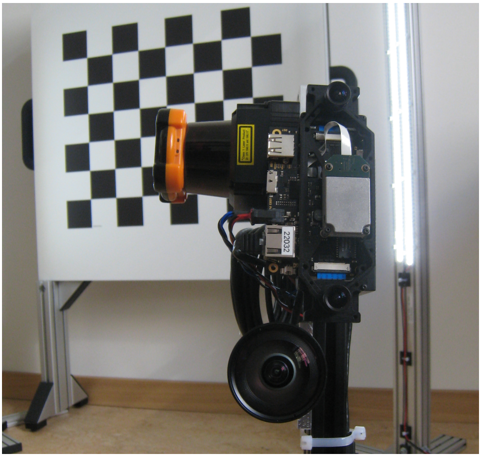


Figure 5.2: The experimental setup and visual target employed in this work. The sensor suite features a visual/inertial sensor comprising two MT9V034 global shutter image sensors and an ADIS16448 IMU (Nikolic et al., 2014b). In addition, a Point Grey Flea3 FL3-U3-32S2C rolling shutter camera with C-mount optics was rigidly attached to the other sensors. The laser range finder present in the setup was not used in this work.

### 5.4.1 The direct method exhibits performance similar to established approaches for static images

In this experiment, we used the FL3-U3-32S2C camera to record roughly 500 static images of the larger calibration target. The camera was operated at its native resolution of 2080×1552 px. Images were converted to grayscale and subsequently down-sampled by a factor of eight.

At this resolution, calibration using a single set of images took approximately 5 minutes on an Intel Core i7-2720QM at 2.2 GHz. Calculating the direct error is only marginally more expensive than computing a reprojection error. However, the error is evaluated for a significantly larger number of points, yielding a runtime increase in the order of the ratio of evaluated intensities to interest points in the target. Accordingly and in contrast to approaches based on interest point locations, the cost accruing in the estimation scales with image resolution. We believe that the benefits of a more correct treatment of uncertainties and timestamps in the rolling shutter and motion blur use case outweigh the drawback of increased processing time in offline calibration. Runtime improvements currently under development will offset some of the additional computation, which will make the advantages of this approach more distinct in the future.

In this experiment, camera intrinsics $\mathbf{k}$ were calibrated. For 30 random subsets of 20 images and using corner position extracted from the images, a vanishing point based initialization of camera intrinsics $\mathbf{k}_{0-3}$ (Hughes et al., 2010) and the Perspective-n-Point algorithm for pose initialization (Lepetit, Moreno-Noguer, and Fua, 2009) were employed. The distortion estimate $\mathbf{k}_{4-7}$ was initialized to zero. Subsequently, the initial estimate was refined on these subsets using an implementation of the optimization step a) based on the reprojection error (using the implementation provided in Furgale et al. (2015b)) and b) based on the direct formulation as described in Section 5.3.2.1.

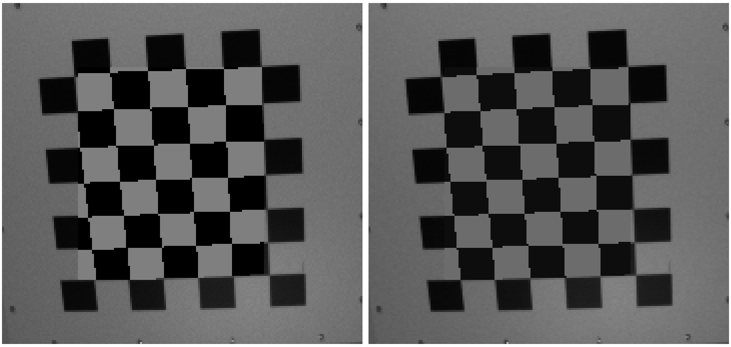
Table 5.2 lists the results in terms of mean and standard deviation for both approaches. They suggest that the direct formulation exhibits performance similar to the reprojection error in static cases. In our experiments, the interest point based approach demonstrated greater precision. This can likely be attributed to our simplifying assumption that a single radiance sample is sufficient to determine the intensity for a pixel.

### 5.4.2 The approach yields accurate estimates of the line delay in rolling shutter cameras

In this experiment, the FL3-U3-32S2C rolling shutter camera as well as the ADIS16448 IMU were used. Again, images were converted to grayscale and down-sampled by a factor of eight. Exposure time was set to 2 ms. The camera did not provide an interface to directly manipulate the line delay $t_l$ (or its reciprocal, the *line frequency* $f_l$), but it allowed for setting different video modes with two distinctively different line frequencies of 24 193.6 Hz and 96 774.2 Hz respectively (*Flea3 USB 3.0 Digital Camera Technical Reference* 2016). With these settings, the camera provided images at a rate of 15 Hz and 20 Hz with a resolution of 1920×1080 px. For each line de-

Table 5.2: Standard vs. direct method for estimation of the camera intrinsics.

| | **Standard method** | **Direct method** | **Unit** |
|---|---|---|---|
| $k_0$ | $176.72 \pm 0.85$ | $175.43 \pm 0.90$ | px |
| $k_1$ | $176.76 \pm 0.85$ | $175.32 \pm 1.02$ | px |
| $k_2$ | $129.91 \pm 0.54$ | $129.92 \pm 0.81$ | px |
| $k_3$ | $96.07 \pm 0.87$ | $96.08 \pm 0.71$ | px |
| $k_4$ | $0.11 \pm 0.01$ | $0.10 \pm 0.02$ | |
| $k_5$ | $0.08 \pm 0.07$ | $0.06 \pm 0.13$ | |
| $k_6$ | $-0.08 \pm 0.25$ | $-0.03 \pm 0.41$ | |
| $k_7$ | $0.20 \pm 0.32$ | $0.25 \pm 0.59$ | |



(a) (b)

Figure 5.3: A comparison of the modelled target given the initial estimate based on a global shutter approximation (Fig. 5.3a) and after estimating the calibration parameters (Fig. 5.3b). Albeit subtle, the effect of consecutively exposed image lines is visible as a distortion of the pattern, resulting in an incorrect fit of the initial rendering of the target. This effect is most pronounced in the lower left corner. After calibration, the rolling shutter distortion is correctly accounted for, resulting in a seamless superimposition of camera image and rendered target.

lay, a dataset of approximately 60 s length was recorded and subsequently split into chunks of 10 s. In this experiment, the line delay $t_l$ as well as the camera intrinsics and distortion **k** were estimated. As initial estimate, we set $t_l$ to 0 s and initialized the intrinsics with the values calibrated in Section 5.4.1, despite the different resolutions in the two experiments.

Table 5.3 shows the results of estimating the line frequency $1/t_l$. The results suggest that the approach is capable of accurately estimating the line frequency with mean absolute estimation errors below 1 %.

Table 5.3: True vs. estimated line frequency.

| **line frequency** [Hz] | $\overline{f_l}$ [Hz] | $\sigma_{f_l}$ [Hz] | $\bar{e}_{f_l}$ [Hz] |
|---|---|---|---|
| 24193.6 | 24147.3 | 138.5 | 133.4 |
| 96774.2 | 96401.0 | 929.6 | 751.4 |

Camera intrinsics were determined as $\overline{\mathbf{k}}_{0-3} = [176.56, 176.33, 120.02, 67.21]$ px with standard deviations $\sigma_{0-3} = [1.65, 1.577, 0.735, 1.169]$ px over all estimates. The focal length is similar to the value estimated in Section 5.4.1, suggesting that image resolution was reduced through a cropping operation. Distortion was estimated as $\overline{\mathbf{k}}_{4-7} = [0.106, 0.058, 0.122, -0.319]$ with standard deviation $\sigma_{4-7} = [0.01, 0.129, 0.65, 1.238]$, which compares well with the values determined in the static experiment. The parameters governing mapping (5.18), $s$ and $o$, were determined as $\bar{s} = 0.8$ with $\sigma_s = 0.05$ and $\bar{o} = 13.35$ with $\sigma_o = 1.61$ respectively. While these values are representative exclusively of the conditions present during data collection (e.g. the target, lighting, optics, camera make and settings, and the specific choice of $L_0$ in (5.14)), they suggest that $s$ and $o$ can be estimated in a repeatable fashion over multiple datasets—a prerequisite for our direct approach to yield meaningful results.

Fig. 5.3 is representative of the magnitude of the rolling shutter effect in the dataset. Assuming a zero line delay $t_l$ as initial estimate yields subtle inconsistencies when superimposing the target rendering onto the image, most apparent in the lower left corner of Fig. 5.3a. After calibration, these are resolved by accounting for the consecutive nature of rolling shutter frame exposure (Fig. 5.3b).

### 5.4.3 Exposure time can be accurately inferred from motion blur

In this experiment, we used the MT9V034 Wide Video Graphics Array (WVGA) global shutter camera in combination with the ADIS16448 IMU to estimate exposure time $t_e$. Using the smaller calibration target, three datasets of approximately 60 s each were recorded with 2, 4, and 6 ms exposure time. These datasets were subsequently split into chunks of 5 s length. For estimating the exposure time, the blurred image was additively composed of images rendered for 10 discrete poses. Our camera setup assigned timestamps at mid-exposure time, which was accounted for by extending (5.19) by additionally subtracting half the exposure time $t_e$. This adjustment gave rise to an ambiguity in the sign of the estimate of the exposure time. Accordingly, the evaluation was performed on the absolute value of the estimate.

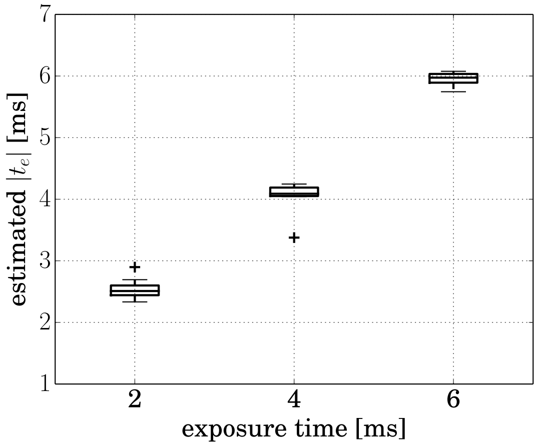
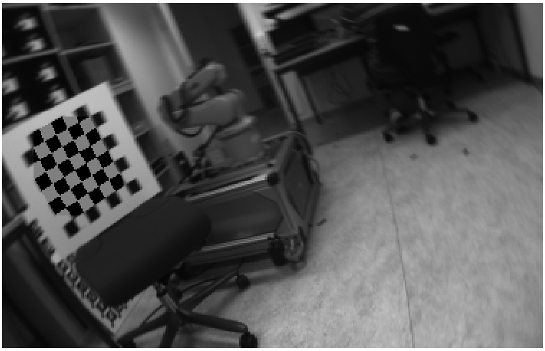
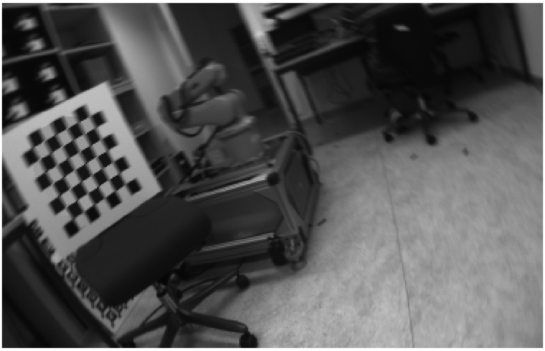


Figure 5.4: Exposure time estimates inferred from motion blur in global shutter images. The estimator accurately determined exposure times of 4 ms and 6 ms, while returning biased results for 2 ms. In the 2 ms case, the effect of motion blur is subtle, which—in combination with our discretization approach—could potentially explain the bias.

Fig. 5.4 depicts a box plot of the estimated exposure times. The median values of the estimates are 2.50 ms, 4.09 ms, and 5.97 ms. Longer exposure times (4 ms and 6 ms) could be accurately inferred from motion blur *alone*. Note that this result is different from Furgale, Rehder, and Siegwart (2013) where conceptually exposure time could be observed by reconciling sensor trajectories perceived by the camera and the IMU by means of a fixed, relative temporal offset. For sensor setups that assign image timestamps at mid-exposure time, such an approach is not capable of inferring exposure as demonstrated in Nikolic et al. (2014b). For short exposure times (2 ms), our approach returned a biased estimate. From examining the images in the dataset, we concluded that the effect of motion blur was rather subtle, and our temporal discretization—in addition to the current lack of support for sub-pixel effect in rendering—may have made it difficult to perceive short exposure times.



(a)



(b)

Figure 5.5: A comparison of the target rendered onto the real image. Fig. 5.5a depicts a rendering based on the initial guess of $o$, $s$ and $t_e$, while Fig. 5.5b shows the rendering after calibration. The effect of motion blur is realistically captured as apparent from the close resemblance of the rendered checkerboard squares with the squares in the right column of the target.

We use the estimation of exposure time as a proxy to shed light on an underlying concept: There exists a number of approaches (e.g. Rehder et al. (2016) and Zachariah and Jansson (2010)) that employ camera measurements to improve the models of other devices comprised in the same sensor suite. However, these approaches are likely limited by the least sophisticated model contributing to the estimator. Formulating the camera model directly on image intensities paves the way for more comprehensive modelling of image measurements and in turn for further advances in the models of the additional sensor comprised in the setup. Fig. 5.5 highlights the accuracy of the currently implemented model for motion blur. Starting from an initial estimate (Fig. 5.5a) of negligible exposure time and with inaccurate estimates of the parameters $s$ and $o$, the approach converges to estimates that yield a realistic impression of the target recorded under camera motion (Fig. 5.5b), as obvious from the close resemblance of the rendered part of the checkerboard with its right column. Fig. 5.5b also highlights a limitation of our simplified radiance model: In contrast to our assumption, the checkerboard is not perfectly evenly lit, resulting in an intensity gradient. This mismatch between model and reality yields subtle seams, especially visible in the top row of the target.

## 5.5 Conclusion

Our results suggest that camera calibration benefits from a direct formulation of the camera measurement model—an approach that increasingly gains traction in visual state estimation. The direct method is computationally more involved, but it exhibits some distinct advantages: It marks the lowest level of abstraction from image sensor readings which facilitates a correct treatment of measurement uncertainties. It further allows for more comprehensive measurement models that bypass complex issues like assigning a timestamp to an interest point extracted from an image blurred by motion or determining the uncertainty of a corner detection in an image distorted by the rolling shutter effect.

However, this comes at the cost of incorporating radiometric aspects into the calibration even in cases where we are exclusively interested in temporal or geometric parameters.

Currently, we are working on relaxing some of the assumptions made in this work by extending the model to incorporate uneven illumination and optical vignetting. We are further addressing defocus by estimating the point spread function of the optical system. Reducing computation time is an additional concern of ours.

# 6

# Camera/Inertial Measurement Unit (IMU) Calibration Revisited

Joern Rehder and Roland Siegwart

## 6.1 Introduction

With an increasing number of approaches emerging which leverage the complementary strengths of IMUs and cameras (e.g. Jones and Soatto (2011), Leutenegger et al. (2015), and Mourikis and Roumeliotis (2007)), 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 (Fleps et al., 2011; Kelly and Sukhatme, 2011; Mirzaei and Roumeliotis, 2008). Early on, temporal offsets have been identified as a significant source of deterministic error (Kelly, Roy, and Sukhatme, 2014; Mair et al., 2011). Consequently, the estimation of temporal quantities has been incorporated as an integral part into camera/IMU calibration (Furgale, Rehder, and Siegwart, 2013; Nikolic et al., 2016b).

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 (Krebs, 2012; Nikolic et al., 2016b; Zachariah and Jansson, 2010).

Published in:
*IEEE Sensors Journal 17.11, pp. 3257–3268, 2017, ©2017 IEEE*
*DOI:* `10.1109/JSEN.2017.2674307`

Recently, different approaches for online calibration of camera/IMU systems have been proposed. Some methods are limited to the transformation between camera and IMU (Jones and Soatto, 2011; Leutenegger et al., 2015) while others additionally also determine the time offset (Li and Mourikis, 2014) and IMU and camera intrinsics (Li et al., 2014). 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*(Hung et al., 1979), 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, Drummond, and Comport (2013). 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 (2005), and it is able to treat defocus blur explicitly. The later incorporates insights from Joshi, Szeliski, and Kriegman (2008) into estimating the point spread function for modelling blur. The approach differs from similar modelling proposed by Meilland and Comport (2013) 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 (Rehder et al., 2016) with a direct image measurement formulation (Rehder et al., 2017). It extends this work by a novel formulation of the direct image error that facilitates modelling uneven target illumination.

# 6.2 Method

## 6.2.1 Problem Statement

Most fundamentally, calibration aims at establishing a set of parameters $\Theta$ 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.

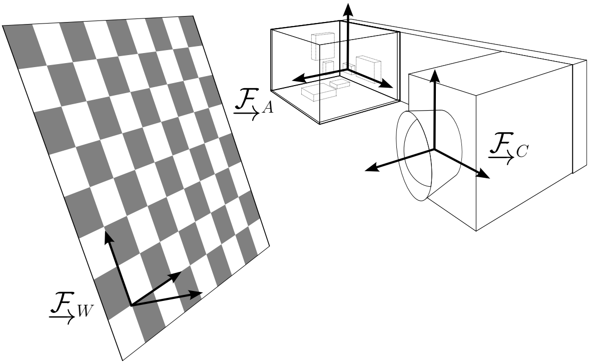

$\underrightarrow{\mathcal{F}}_A$
$\underrightarrow{\mathcal{F}}_C$
$\underrightarrow{\mathcal{F}}_W$

Figure 6.1: 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. 6.4 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. 6.1 depicts the general calibration setup: A sensor suite, comprising an IMU rigidly attached to an intrinsically 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 6.1 compiles all states and parameters estimated in this work.

### 6.2.2 Coordinate Frame Conventions

The different coordinate frames used in this work are shown in Fig. 6.1. We will refer to the target frame as $\underrightarrow{\mathcal{F}}_W$, while the camera 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 $4 \times 4$ transformation matrix $\mathbf{T}_{CW}$ denoting the mapping ${}_C\mathbf{p} = (\mathbf{T}_{CW})_W\mathbf{p}$ of points expressed in homogeneous coordinates:

$$\mathbf{T}_{CW} := \begin{bmatrix} \mathbf{R}_{CW} & {}_C\mathbf{t}_{CW} \\ \mathbf{0} & 1 \end{bmatrix}, \tag{6.1}$$

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 two-dimensional (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 (6.14).

### 6.2.3 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 $\mathbf{e}_h := \mathbf{h}(\mathbf{x}(t), \Theta) - \tilde{\mathbf{m}}$ 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

Table 6.1: States $\mathbf{x}(t)$ and parameters $\Theta$ estimated in this work.

| **Symbol** | **Description** | **Section** |
|---|---|---|
| | **States** $\mathbf{x}(t)$ | |
| $\mathbf{T}_{AW}(t)$ | Time-varying pose of the IMU | 6.2.4 |
| $\mathbf{b}_{\alpha}(t)$ | Time-varying accelerometer bias | 6.2.7.1 |
| $\mathbf{b}_{\omega}(t)$ | Time-varying gyroscope bias | 6.2.7.2 |
| | **Parameters** $\Theta$ | |
| $\mathbf{T}_{CA}$ | Fixed relative transformation between $\underrightarrow{\mathcal{F}}_A$ and $\underrightarrow{\mathcal{F}}_C$ | 6.2.6.1 |
| $d_C$ | Fixed temporal offset of image timestamps with respect to accelerometer timestamps | 6.2.6.1 |
| $\mathbf{a}^k$ | Coefficients of a polynomial illumination model for image $k$ | 6.2.6.2 |
| $\rho_b^k$ | Reflectance of the black tiles of the calibration pattern for image $k$ | 6.2.6.2 |
| $t_e$ | Exposure time of the camera | 6.2.6.2 |
| $o$ | offset of the linear camera response function | 6.2.6.2 |
| $\mathbf{S}_{\alpha,\omega}$ | Accelerometer and gyroscope scaling factors | 6.2.7.1,6.2.7.2 |
| $\mathbf{M}_{\alpha,\omega}$ | Accelerometer and gyroscope misalignments | 6.2.7.1,6.2.7.2 |
| ${}_A\mathbf{r}_{A\alpha_{y,z}}$ | Displacement of accelerometer axes $y$ and $z$ from $\underrightarrow{\mathcal{F}}_A$ | 6.2.7.1 |
| $\mathbf{C}_{A\omega}$ | Relative rotation between $\underrightarrow{\mathcal{F}}_A$ and the gyroscope frame $\underrightarrow{\mathcal{F}}_\omega$ | 6.2.7.2 |
| $\mathbf{A}_{\omega}$ | Influence of linear acceleration on gyroscope measurements (“g-sensitivity”) | 6.2.7.2 |
| ${}_W\mathbf{g}$ | Direction of gravity expressed in $\underrightarrow{\mathcal{F}}_W$ | 6.2.7.1 |

$$\begin{aligned}
\Theta, \mathbf{x}(t) = \underset{\Theta, \mathbf{x}(t)}{\operatorname{argmin}} \Big( \quad & e_{h_C}(\mathbf{x}(t), \Theta)^T \mathbf{R}_C^{-1} e_{h_C}(\mathbf{x}(t), \Theta) \\
& + e_{h_\alpha}(\mathbf{x}(t), \Theta)^T \mathbf{R}_\alpha^{-1} e_{h_\alpha}(\mathbf{x}(t), \Theta) \\
& + \int \mathbf{e}_{f_\alpha}(\tau)^T \mathbf{Q}_\alpha^{-1} \mathbf{e}_{f_\alpha}(\tau) \mathrm{d}\tau \\
& + e_{h_\omega}(\mathbf{x}(t), \Theta)^T \mathbf{R}_\omega^{-1} e_{h_\omega}(\mathbf{x}(t), \Theta) \\
& + \int \mathbf{e}_{f_\omega}(\tau)^T \mathbf{Q}_\omega^{-1} \mathbf{e}_{f_\omega}(\tau) \mathrm{d}\tau \Big),
\end{aligned} \tag{6.2}$$

where subscripts $C$, $\alpha$, and $\omega$ identify contributions as originating from the camera, accelerometer, and gyroscope model respectively.

We solve (6.2) iteratively for $\mathbf{x}(t)$ and $\Theta$ using the Levenberg-Marquardt (LM) algorithm (Nocedal and Wright, 2006).

### 6.2.4 State Parametrization

Our implementation extends the open-source toolbox *kalibr* (Furgale et al., 2015b), 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 (Furgale et al., 2015a).

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

$$\Phi(t) := \begin{bmatrix} \phi_1(t) & \dots & \phi_B(t) \end{bmatrix}, \quad \mathbf{x}(t) := \Phi(t)\mathbf{c}, \tag{6.3}$$

where each $\phi_b(t)$ is a $D \times 1$ B-spline and $\Phi(t)$ 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 $6 \times 1$ spline with 3 degrees of freedom for relative translation and 3 degrees of freedom for relative orientation:

$$
\begin{aligned}
{}_W\mathbf{t}_{WA}(t) &:= \Phi_t(t)\mathbf{c}_t & (6.4)\\
\varphi(t) &:= \Phi_\varphi(t)\mathbf{c}_\varphi\,. & (6.5)
\end{aligned}
$$

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

$$\mathbf{C}_{AW}(t) := \mathcal{C}\big(\varphi(t)\big)^T, \tag{6.6}$$

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

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

$$\mathbf{b}(t) := \Phi_b(t)\mathbf{c}_b \tag{6.7}$$

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

### 6.2.5 Baseline Camera Measurement Model

The baseline approach (Furgale, Rehder, and Siegwart, 2013) uses the projection of known three-dimensional (3D) points ${}_W\mathbf{p}_m$ corresponding to corner $m \in [1, \ldots, M]$ in the visual calibration target to model camera measurements.

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

$$\mathbf{h}_C(\mathbf{x}(t), \Theta) := \begin{bmatrix} \mathbf{h}_C^1(\mathbf{x}(t), \Theta) \\ \vdots \\ \mathbf{h}_C^K(\mathbf{x}(t), \Theta) \end{bmatrix}, \tag{6.8}$$

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

$$\mathbf{h}_C^k(\mathbf{x}(t), \Theta) := \begin{bmatrix} \pi(\mathbf{T}_{CW}(t_k + d_C)_W\mathbf{p}_1) \\ \vdots \\ \pi(\mathbf{T}_{CW}(t_k + d_C)_W\mathbf{p}_M) \end{bmatrix}. \tag{6.9}$$

Here, $\pi(\cdot)$ 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 (Furgale, Rehder, and Siegwart, 2013).

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.

### 6.2.6 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

$$\mathbf{h}_C^k\left(\mathbf{x}(t),\Theta\right) := \begin{bmatrix} \mathrm{B}\left({}_I\mathbf{p}_1^k,\mathbf{x}(t),\Theta\right) \\ \vdots \\ \mathrm{B}\left({}_I\mathbf{p}_M^k,\mathbf{x}(t),\Theta\right) \end{bmatrix}, \tag{6.10}$$

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 6.2.6.3. 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 ${}_I\mathbf{p}$:

$$\mathrm{L}({}_W\mathbf{p}) \mapsto \mathrm{B}({}_I\mathbf{p}) \tag{6.11}$$

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 (Debevec and Malik, 2008).

#### 6.2.6.1 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 ${}_Wz = 0$, 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

$${}_I\mathbf{p} = \mathbf{H}_{IW}^{-1}\begin{bmatrix} {}_Wx \\ {}_Wy \\ 1 \end{bmatrix} \tag{6.12}$$

with homography

$$\mathbf{H}_{IW} = \mathbf{K}_{IC}\begin{bmatrix} {}^1\mathbf{R}_{CW} & {}^2\mathbf{R}_{CW} & {}_C\mathbf{t}_{CW} \end{bmatrix}, \tag{6.13}$$

where $\mathbf{K}_{IC}$ denotes the camera matrix, $\mathbf{R}_{CW}$ and ${}_{C}\mathbf{t}_{CW}$ are defined according to (6.1) and superscripts denote individual columns of the rotation matrix $\mathbf{R}_{CW}$.

Here, we will assume that $\pi(\cdot)$ describes a pinhole camera model and that distortion has been compensated for. Other projection models are equally feasible, but require an adaptation of (6.13). For the pinhole model, camera intrinsics can be represented by the camera matrix $\mathbf{K}_{IC}$:

$$\mathbf{K}_{IC} := \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{6.14}$$

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

$$\mathbf{H}_{IW}(t) = \mathbf{K}_{IC} \begin{bmatrix} \left(\mathbf{R}_{CA}{}^{1}\mathbf{R}_{AW}(t+d_C)\right)^T \\ \left(\mathbf{R}_{CA}{}^{2}\mathbf{R}_{AW}(t+d_C)\right)^T \\ \left({}_{C}\mathbf{t}_{CA} + \mathbf{R}_{CA}{}_{A}\mathbf{t}_{AW}(t+d_C)\right)^T \end{bmatrix}^T \tag{6.15}$$

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

#### 6.2.6.2 Radiometric mapping

The radiometric part of the model is given by

$$\mathrm{L} \xmapsto{\mathrm{S}(\cdot)} \mathrm{E} \xmapsto{\int\!\int_A \mathrm{d}A\mathrm{d}t} \mathrm{X} \xmapsto{\mathrm{R}(\cdot)} \mathrm{B} \tag{6.16}$$

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 (6.16) individually in the following.

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

$$\mathrm{L}({}_{W}\mathbf{p}) = \rho({}_{W}\mathbf{p})\alpha({}_{W}\mathbf{p}) \tag{6.17}$$

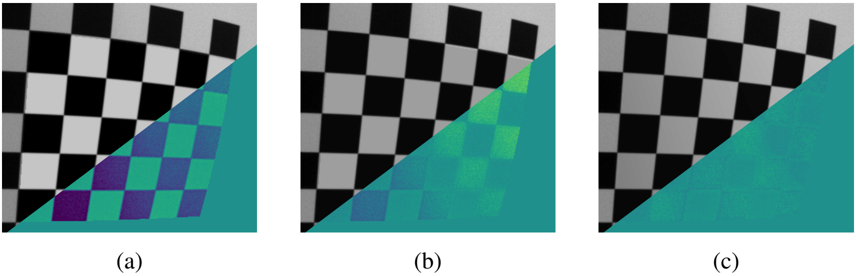


Figure 6.2: The rendered target superimposed onto the corresponding camera image. The triangular insets show the color-coded difference image for clarity. Mismatches between image and model are most visible in two top rows. Fig. 6.2a shows the rendering prior to optimization. The intensity of the model clearly does not match the camera image. Fig. 6.2b displays the post optimization result using an even illumination model as proposed in Rehder et al. (2017). The result exhibits subtle inconsistencies. Fig. 6.2c was generated using the polynomial illumination model (6.19). The image lacks any obvious visual seams, suggesting that a 2$^{\text{nd}}$ degree polynomial is sufficient to capture the lighting environment present in our datasets. Best viewed in color.

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

$$\rho({}_W\mathbf{p}) := \begin{cases} \rho_w & \text{if } (\lfloor \frac{{}_Wx}{\Delta x} \rfloor + \lfloor \frac{{}_Wy}{\Delta y} \rfloor) \bmod 2 = 0 \\ \rho_b & \text{else} \end{cases} \tag{6.18}$$

where the operator $\lfloor \cdot \rfloor$ denotes a floor operation and $\Delta x$ and $\Delta y$ the extent of individual checkerboard tiles in $x$ and $y$ direction. The values $\rho_w$ and $\rho_b$ 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 $\rho_w$ to be 1, while $\rho_b$ is estimated.

Target illumination is modelled as a 2$^{\text{nd}}$ degree polynomial:

$$\begin{aligned} \alpha({}_W\mathbf{p}) :=& a_1 + a_2 {}_Wx + a_3 {}_Wy + a_4 {}_Wx^2 + \\ & a_5 {}_Wy^2 + a_6 {}_Wx {}_Wy, \end{aligned} \tag{6.19}$$

where $\mathbf{a} := [a_1, \ldots, a_6]$ 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. 6.2 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 (Kim and Pollefeys, 2008). 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 (Debevec and Malik, 2008).

We further fold the integration over the finite extent of an individual pixel on the image sensor as well as the effect of imperfectly focussing optics into this step. Fig. 6.3a provides the rationale behind 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 (Heide et al., 2013).

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.

$$\mathrm{E}^*({}_I\mathbf{p},t) := \frac{1}{J^2}\sum_{i=1}^{J}\sum_{j=1}^{J}\mathrm{W}_{ij}\mathrm{E}\left(\mathbf{H}_{IW}^{-1}(t)\left({}_I\mathbf{p}+\mu\begin{bmatrix}\frac{J}{i}-\frac{1}{2}\\ \frac{J}{j}-\frac{1}{2}\end{bmatrix}\right)\right) \tag{6.20}$$

$$\mathrm{X}({}_I\mathbf{p}) = \sum_{n=1}^{N}\mathrm{E}^*\left({}_I\mathbf{p}, t_0+\frac{n}{N-1}t_e\right)\frac{1}{N}t_e \tag{6.21}$$

Here, $N$ denotes the number of images rendered to emulate motion blur. Equation (6.20) 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 $\mu$ denotes an up-scaling factor. The weights $\mathbf{W} := [W_{11}, W_{12}, \dots, W_{JJ}]$ 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 $\hat{\Theta}$ 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 $\hat{\mathbf{x}}$:

$$\hat{\Theta}, \hat{\mathbf{x}}, \hat{\mathbf{W}} = \underset{\hat{\Theta}, \hat{\mathbf{x}}, \hat{\mathbf{W}}}{\operatorname{argmin}} \sum_{k=1}^{K} \left( \mathbf{h}_C^k \left( \hat{\Theta}, \hat{\mathbf{x}}, \hat{\mathbf{W}} \right) - \tilde{\mathbf{m}}_C^k \right)^2 \tag{6.22}$$

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, $\hat{\Theta}$ is initialized according to Section 6.2.8 and $\hat{\mathbf{x}}$ from target corners using the Perspective-n-Point algorithm (Lepetit, Moreno-Noguer, and Fua, 2009), while all weights $\mathbf{W}$ are initialized to 1. Fig. 6.3b depicts the kernel estimated from 50 images for the optics used in the experiments in Section 6.3.

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 (6.22) to determine a suitable combination of $\mu$ 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.

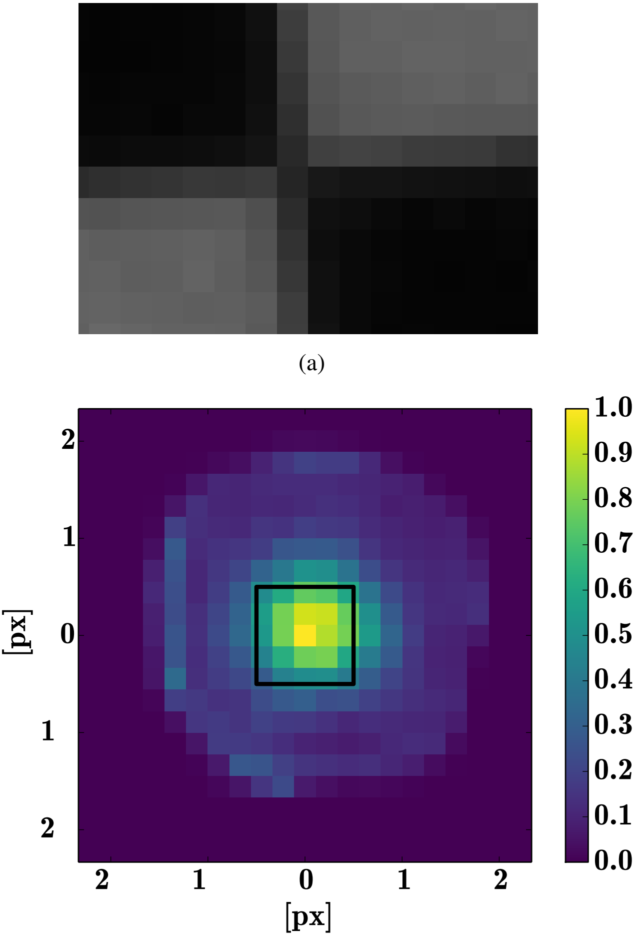
Most sensors are designed for this mapping to be linear, and we disabled all digital processing of the signal in the sensor that would have altered the response curve. Accordingly, the camera response curve is modelled as linear as

$$\mathrm{B}({}_I\mathbf{p}) := s\mathrm{X}({}_I\mathbf{p}) + o, \tag{6.23}$$

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 (6.19) accordingly. Hence, $s$ is assumed to be 1 and its estimation is omitted.

#### 6.2.6.3 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



(a)

(b)

Figure 6.3: Imperfect focussing has a noticeable impact on image forming. Fig. 6.3a 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. 6.3b, and subsequently down-sampling the result.

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 $\tau$ where direct error terms are only evaluated for $\mathrm{C}({}_{I}\mathbf{p}) = 1$:

$$\mathrm{C}({}_{I}\mathbf{p}) := \begin{cases} 1 & \text{if } \left|\nabla \tilde{\mathrm{B}}_{{}_{I}\mathbf{p}}\right| > \tau \\ 0 & \text{else} \end{cases} \tag{6.24}$$

Despite this reduction, the resulting set of error terms will still yield a vastly overconstrained 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 (Furgale, Rehder, and Siegwart, 2013).

Fixating the number of error terms is accomplished through QR decompositions (Golub and Loan, 1996).

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

$$\mathbf{JP} = \mathbf{QR} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix} \tag{6.25}$$

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

$$\hat{\mathbf{J}}^k := \mathbf{R}_1 \mathbf{P}^T \tag{6.26}$$

and the corresponding intensity error is given by

$$\hat{\mathbf{e}}_{h_C}^k := \mathbf{Q}_1^T \mathbf{e}_{h_C}^k. \tag{6.27}$$

### 6.2.7 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 (6.2) as

$$\mathbf{h}_\alpha(\mathbf{x}(t),\Theta) := \begin{bmatrix} \alpha\left(\mathbf{x}(t_1),\Theta\right) \\ \vdots \\ \alpha\left(\mathbf{x}(t_K),\Theta\right) \end{bmatrix} \tag{6.28}$$

and

$$\mathbf{h}_\omega(\mathbf{x}(t),\Theta) := \begin{bmatrix} \varpi\left(\mathbf{x}(t_1),\Theta\right) \\ \vdots \\ \varpi\left(\mathbf{x}(t_K),\Theta\right) \end{bmatrix}, \tag{6.29}$$

with $\alpha(\cdot)$ and $\varpi(\cdot)$ as defined in (6.37) and (6.40) 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 $\tilde{\alpha}_k$ and $\tilde{\omega}_k$ 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.

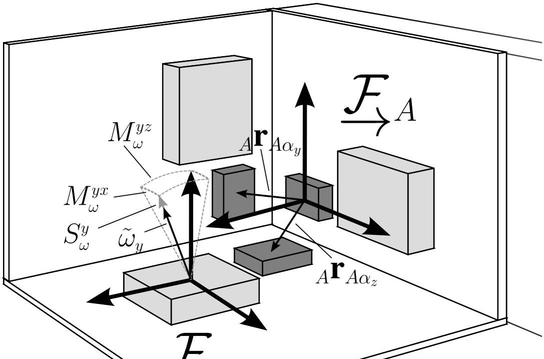

$\underrightarrow{\mathcal{F}}_A$
$M_\omega^{yz}$
${}_A\mathbf{r}_{A\alpha_y}$
$M_\omega^{yx}$
$S_\omega^y$
$\tilde{\omega}_y$
${}_A\mathbf{r}_{A\alpha_z}$
$\underrightarrow{\mathcal{F}}_\omega$

Figure 6.4: 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 $M_\omega^{yx}$ and $M_\omega^{yz}$, 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 IC despite their different mechanical design.

Fig. 6.4 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 6.1.

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 6.2.4.

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

$$ {}_W\ddot{\mathbf{t}}_{WA}(t) = \ddot{\Phi}_t(t)\mathbf{c}_t \tag{6.30} $$

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

With $\mathbf{C}_{AW}(t)$ defined according to (6.6), angular velocity and angular acceleration as perceived in $\underrightarrow{\mathcal{F}}_A$ are computed as

$$ {}_A\omega_{WA}(t) = \mathbf{C}_{AW}(t)\ {}_W\omega_{WA}(t) \tag{6.31} $$

$$ {}_A\dot{\omega}_{WA}(t) = \mathbf{C}_{AW}(t)\ {}_W\dot{\omega}_{WA}(t) \tag{6.32} $$

with

$$ {}_W\omega_{WA}(t) = \mathbf{S}\big(\varphi(t)\big)\dot{\varphi}(t) = \mathbf{S}\big(\Phi(t)\mathbf{c}_\varphi\big)\dot{\Phi}(t)\mathbf{c}_\varphi \tag{6.33} $$

$$ {}_W\dot{\omega}_{WA}(t) = \mathbf{S}\big(\varphi(t)\big)\ddot{\varphi}(t) = \mathbf{S}\big(\Phi(t)\mathbf{c}_\varphi\big)\ddot{\Phi}(t)\mathbf{c}_\varphi \tag{6.34} $$

where $\mathbf{S}(\cdot)$ is the matrix relating parameter rates to angular velocities and accelerations (Hughes, 1986).

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

$$\begin{aligned} {}_A\mathbf{a}_{WA}(t) = \mathbf{C}_{AW}(t)({}_W\ddot{\mathbf{t}}_{WA}(t) - {}_W\mathbf{g}) \\ + \operatorname{diag}(\lfloor\dot{\omega}(t)\rfloor_\times \mathbf{R}_\alpha + \lfloor\omega(t)\rfloor_\times^2 \mathbf{R}_\alpha), \end{aligned} \tag{6.35}$$

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

$$\mathbf{R}_\alpha := \begin{bmatrix} {}_A\mathbf{r}_{A\alpha_x} & {}_A\mathbf{r}_{A\alpha_y} & {}_A\mathbf{r}_{A\alpha_z} \end{bmatrix}. \tag{6.36}$$

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

$$\alpha(t) := \mathbf{S}_\alpha \mathbf{M}_{\alpha A} \mathbf{a}_{WA}(t) + \mathbf{b}_\alpha(t) \tag{6.37}$$

where $\mathbf{S}_\alpha$ is a diagonal matrix comprising scaling effects and $\mathbf{M}_\alpha$ is a lower uni-triangular 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 (Furgale, Rehder, and Siegwart, 2013):

$$\dot{\mathbf{b}}_\alpha(t) = \mathbf{w}_\alpha(t) \tag{6.38}$$

with

$$\mathbf{w}_\alpha(t) \sim \mathcal{GP}\left(\mathbf{0}, \sigma_{b_\alpha}^2 \mathbf{I}\delta(t - t')\right) \tag{6.39}$$

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

#### 6.2.7.2 Gyroscope model

Gyroscope measurements are modelled as

$$\begin{aligned} \varpi(t) := \quad & \mathbf{S}_\omega \mathbf{M}_\omega \mathbf{C}_{\omega A A}\omega_{WA}(t) \\ & + \mathbf{A}_\omega \mathbf{C}_{\omega A A}\mathbf{a}_{WA}(t) \\ & + \mathbf{b}_\omega(t) \end{aligned} \tag{6.40}$$

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 (6.37). 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 (6.38) as

$$\dot{\mathbf{b}}_\omega(t) = \mathbf{w}_\omega(t) \tag{6.41}$$

with

$$\mathbf{w}_\omega(t) \sim \mathcal{GP}\left(\mathbf{0}, \sigma_{b_\omega}^2 \mathbf{I}\delta(t-t')\right) \tag{6.42}$$

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

### 6.2.8 Initialization

Sufficiently faithful initial estimates for the parameters $\Theta$ and the state $\mathbf{x}(t)$ are required in order for (6.2) 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 $\rho_b$ and intensity offset $o$ are initially set to 0. Exposure time $t_e$ is initialized to 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 (Lepetit, Moreno-Noguer, and Fua, 2009). Subsequently, an orientation curve $\hat{\varphi}(t)$, 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:

$$\hat{\varpi}(t) := \mathbf{C}_{\omega A}\mathbf{C}_{CAC}^T \hat{\omega}_{WC}(t) + \hat{\mathbf{b}}_\omega \tag{6.43}$$

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

$$\mathbf{C}_{CA}, \hat{\mathbf{b}}_\omega = \underset{C_{CA},\hat{\mathbf{b}}_\omega}{\operatorname{argmin}} \sum_{k=1}^{K} \left(\hat{\boldsymbol{\varpi}}(t_k) - \tilde{\omega}_k\right)^2 . \tag{6.44}$$

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

Following the approach proposed by Mair et al. (2011), 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 ${}_W\hat{\omega}_{WC_k}$ are sampled from the spline $\hat{\varphi}(t)$ 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:

$$d_{xcorr} = \underset{d}{\operatorname{argmax}} \sum_{k=1}^{K} \left|{}_W\hat{\omega}_{WC(k+d)}\right| \left|\tilde{\omega}_k\right| \tag{6.45}$$

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 $\hat{\varphi}(t)$ at the time instants of the accelerometer readings, the initial value is computed as

$${}_W\bar{\mathbf{a}} = \frac{1}{K}\sum_{k=1}^{K} \hat{\mathbf{C}}_{CW}^{-1}(t_k + d_C)\mathbf{C}_{CA}\tilde{\mathbf{a}}_k \tag{6.46}$$

$${}_W\mathbf{g} = g_0 \frac{{}_W\bar{\mathbf{a}}}{|{}_W\bar{\mathbf{a}}|}, \tag{6.47}$$

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 $\mathbf{T}_{CA}^{-1}$.

# 6.3 Results

## 6.3.1 Experimental setup and dataset collection

All data were recorded with the visual/inertial sensor (Nikolic et al., 2014b) shown in Fig. 6.5 featuring an Analog Devices ADIS16448 IMU, an InvenSense MPU9150

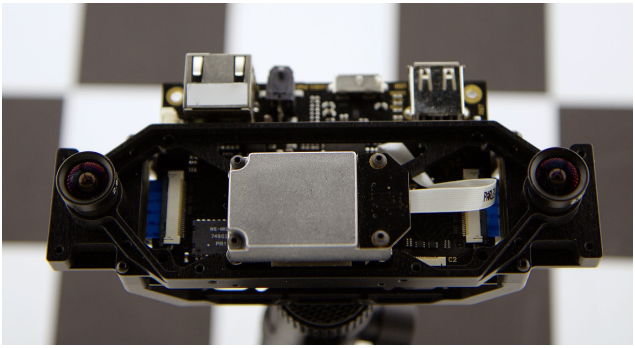


Figure 6.5: The experimental setup comprises an Analog Devices ADIS16448 IMU, an InvenSense MPU9150 IMU, and two Aptina MT9V034 global shutter image sensors of which only one was used.

IMU, and two Aptina Wide Video Graphics Array (WVGA) MT9V034 global shutter image sensors of which only one was used. The ADIS16448 is a factory-calibrated Microelectromechanical Systems (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 Field Programmable Gate Array (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 6.2 lists all variables together with the values used to generate the results.

Table 6.3 compiles the noise model parameters. The noise model parameters for accelerometers and gyroscopes were determined using the approach proposed by Nikolic et al. (2016a). The strength $\sigma_{\mathrm{B}}$ 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.

Table 6.2: Variables used to parametrize the algorithm

| Variable | Description | Value | Section |
|---|---|---|---|
| $O$ | Order of the B-spline | 6 | 6.2.4 |
| $N_{\mathbf{x}}$ | Knots per second supporting the pose spline | 150 | 6.2.4 |
| $N_{\mathbf{b}}$ | Knots per second supporting the bias splines | 50 | 6.2.4 |
| $\tau$ | Threshold on the gradient in the image | 7 | 6.2.6.3 |
| $N$ | Number of images used to emulate motion blur | 5 | 6.2.6.2 |
| $J$ | Size of the weighting window **W** | 17 | 6.2.6.2 |
| $\mu$ | Up-scaling factor for rendering the irradiance image | 3.5 | 6.2.6.2 |

Table 6.3: Noise model parameters

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

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.

### 6.3.2 Appropriate IMU modelling is key to high calibration precision.

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

Using the baseline approach (6.9) 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.

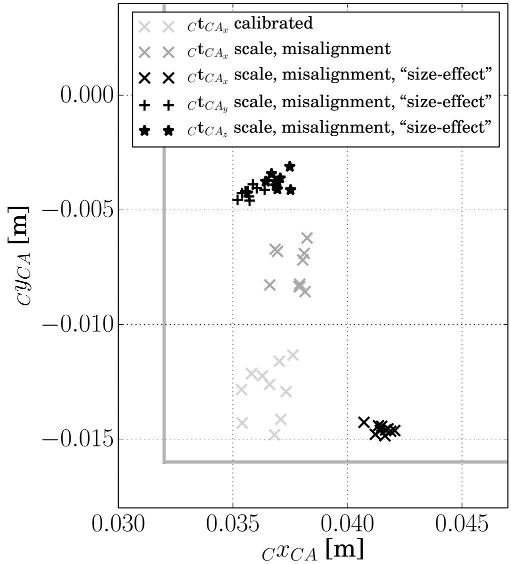
For both IMUs, three models of increasing complexity were considered. These models were

- 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 6.2.7.

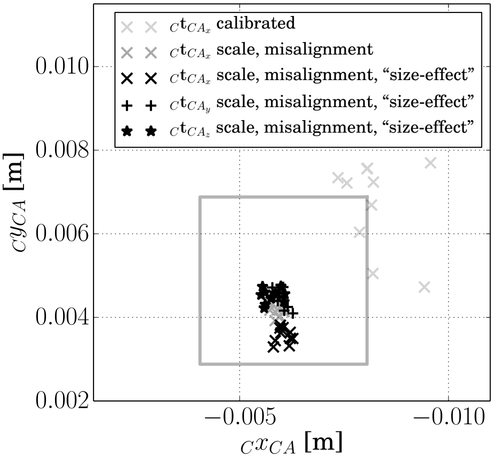
Table 6.4: Calibration results for the baseline camera error (6.9) and IMU models of different fidelity

| **IMU Model** | $\sigma_{{}_C\mathbf{t}_{CA}}$ [mm] | $\sigma_F$ [°] | $\sigma_{d_C}$ [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. 6.6 depicts the ${}_Cx_{CA}$ and ${}_Cy_{CA}$ of the camera/IMU displacement ${}_C\mathbf{t}_{CA}$ as well as the position of individual accelerometer axes where estimated. Table 6.4 displays the same results numerically.



(a) ADIS16448



(b) MPU9150

Figure 6.6: 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.

Fig. 6.6a 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 6.4 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 respect to the Fréchet expectation (Pennec, 1999) denoted as $\sigma_F$—and the temporal offset $d_C$ alike. Fig. 6.6a 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. 6.6a. 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. 6.6b 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 (Krebs, 2012; Nikolic et al., 2016b; Rehder et al., 2016) 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 Nikolic et al. (2016b) 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.

### 6.3.3 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. 6.7 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 6.2.8, resulting in the absence of motion blur in the rendered view shown in Fig. 6.7a. Following calibration, the exposure time is accurately estimated. Consequently, the rendered target closely resembles the image as apparent in Fig. 6.7b.

We use the estimated exposure time $t_e$ as a proxy here to shed light on how accurately motion blur—and consequently the motion of the camera during exposure—can be recovered. Fig. 6.8 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 $0.964 \pm 0.007$, $2.260 \pm 0.013$ and $3.21 \pm 0.016$ ms respectively.

These results are fundamentally different from our previous work (Furgale, Rehder, and Siegwart, 2013), 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 Nikolic et al. (2014b) which renders $t_e$ unobservable for our previous method.

### 6.3.4 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 6.3.2. It further exclusively focuses on the MPU9150 and the

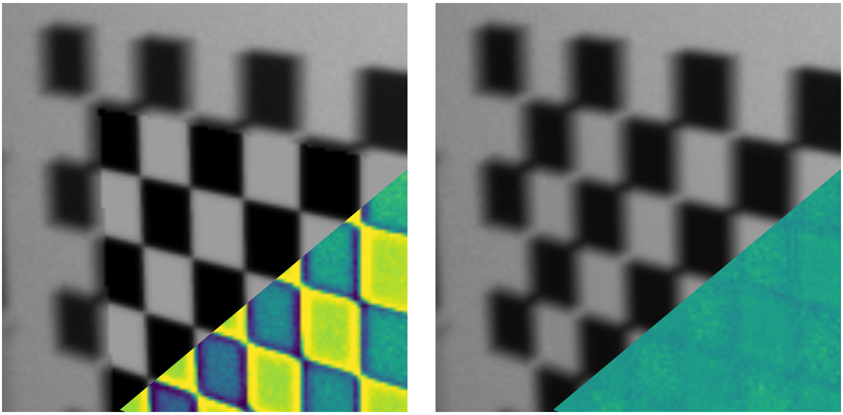

(a) (b)

Figure 6.7: 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. 6.7a depicts the rendered target superimposed onto an image prior to calibration. Fig. 6.7b 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.

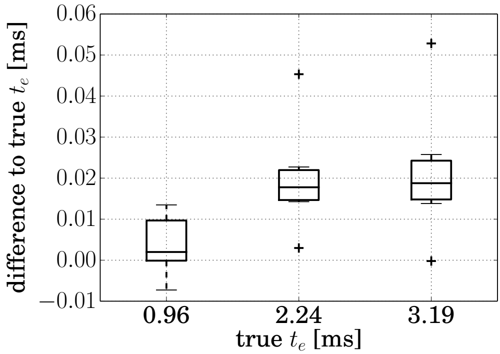


Figure 6.8: Estimation offset versus true exposure times for 3 datasets with different camera settings. The mean estimate is accurate to about $\frac{2}{100}$ ms, suggesting that the camera trajectory during exposure can be equally faithfully recovered.

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 6.5 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 6.5: Calibration results for the direct camera error (6.10) and IMU models of different fidelity

| **IMU Model** | $\sigma_{C\mathbf{t}_{CA}}$ [mm] | $\sigma_F$ [°] | $\sigma_{d_C}$ [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 6.3.2, 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 (6.2).

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 6.1.

## 6.4 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 (2005). 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 (Furgale, Rehder, and Siegwart, 2013; Nikolic et al., 2016b; Rehder et al., 2016; Yang and Shen, 2016) 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 °/s as compared to about 150 °/s in (Rehder et al., 2016) and only about 55 °/s in Furgale, Rehder, and Siegwart (2013).

# 7

# Conclusion and Future Directions

Each of the previous chapters contains a conclusion that provides a particular perspective on the significance of the work within the domain of calibration. Ultimately however, an adequate calibration is merely a prerequisite for accurate and robust state estimation, and the previously introduced methods are solely tools for generating such a calibration. Rather than echoing insights already contained in the previous chapters, this section will examine previous findings for significance that extends beyond the immediate domain of calibration and into those of state estimation and system design.

The findings in this work suggest that synchronization should be informed by a deeper understanding of the inner working of the sensors comprised in a robotic system. Far too often, system designer mistake a signal merely temporally *related* to the measurement instant as an indication of its exact occurrence. This work provides a number of examples of such misunderstandings: Many applications understand cameras as providing an instantaneous irradiance snapshot on assertion of a trigger signal which completely neglects the image exposure process. Similarly, the time of arrival at the host system is commonly assigned to Inertial Measurement Unit (IMU) measurements, as system designers are often unaware of the existence of analog and digital filters which delay the measurement. And while many publications model scanning Laser Range Finders (LRFs) as sampling distances consecutively in time, little work focuses on properly synchronizing these devices to other sensors in the same system. Consequently, the findings of this work serve as a reminder to treat all aspects of a robotic system with the same level of diligence. Negligence at an early stage in the perception pipeline will adversely impact all following efforts and in turn consume some of the improvements from any algorithmic advancements in the state estimation step.

This work proposes a plane-based, probabilistic an LRF measurement model which accounts for the *beam direction*. In contrast, established approaches commonly formulate the measurement error in the direction of the *normal of the plane*. While this established model is computationally less expensive to evaluate, it marks a simplification that induces a deterministic error which grows with the incidence angle. Obviously, this simplification is one of many, and it is difficult to predict its effect on the accuracy of the overall state estimation system. Nonetheless, it is not unlikely that LRF based state estimation approaches would benefit from the presented model—particularly in applications where accuracy is paramount and which can tolerate the additional computational effort required to evaluate it. Furthermore, the presented model is *extensible* as demonstrated by the consistently estimated range bias which allows for further improvements in accuracy in state estimation as the understanding of deterministic errors in LRF measurements advances.

An increasing number of robotic systems rely on visual/inertial state estimation as an input to planning and controls. It has been common wisdom in the robotics community that best state estimation results can be achieved with large, factory-calibrated IMUs composed of single axis, high-quality inertial sensors. Many of the arguments for such a device are valid, but this work adds another nuance to the discussion: Results show that *size* is an important factor as well and that in terms of calibration precision, an inexpensive consumer-grade IMU can outperform a navigation-grade device. Unfortunately, the significance of these findings is limited by imperfections in synchronization that affected the two devices under test in different ways. Nonetheless, it challenges the idea that the specific force can be perceived in a single spot and, by extension, that integrating larger, higher-quality sensor packages always pays off. Applications that neglect the displacement of individual accelerometer axes ultimately face the issue of a perceived shift in the position of the reference frame that depends on the motion and that can be in the order of the distances between the accelerometer axes. The implications of this shift are particularly severe for applications that demand millimeter accurate location estimates.

Modelling camera measurements in terms of image intensities enables an intuitive formulation of rolling shutter calibration with consistent treatment of uncertainties. However, modelling intensities with a fidelity that provides additional value for camera/IMU calibration is a highly involved process which touches on a variety of different aspects ranging from illumination over optics to sensor response curves. Consequently, this work indirectly sheds light on the value in abstraction from intensities: The Point Spread Function (PSF) is difficult to model as it generally depends on the position in image space and on the distance to the object. Yet, it is an essential building block in modelling optics. Sub-pixel accurate interest point detection and polynomial distortion models appear to capture the cumulative effect of the PSF well and hence alleviate approaches from explicit modelling it.

This work could be taken forward in a number of directions.

The understanding of individual sensor models could be advanced further, and in the course of this work, a number of improvements to the sensor models were considered: For inertial sensors, these improvements included individual time delays for accelerometers and gyroscopes to account for different filter characteristics (InvenSense, 2011), non-linearities modelled as polynomials (Looney, 2010), and gyroscope scale factors that vary with linear accelerations (Park et al., 2015). The direct model for camera measurements was augmented to further consider vignetting as well as non-linear sensor response curves. PSF dependence on the position in image space was approximated by estimating multiple kernels on a fixed spatial grid. Similarly, the impact of chromatic aberrations on tracking was investigated, and mitigations using color sensors and channel-wise modelling of the PSF were explored (Schulz, 2016). The LRF model equation was extended by terms reflecting a range dependent offset as a polynomial rather than a constant bias, and modelling gyroscopic effects, a factor hypothesized in Bosse, Zlot, and Flick (2012), was considered (Holtmann, 2016).

Many of these experiments remained inconclusive which may in parts be attributed to the shortcomings in sensor synchronization highlighted in Chapter 6. Consequently, further investigations would likely require improvements in the hardware setup—an effort that appeared to be prohibitively time-consuming within the time frame of this work. However, such improvements could potentially yield novel insights into deterministic sources of measurement errors for these sensors.

Chapter 6 raises the intriguing question of whether motion blur comprises information about the sensor trajectory during image exposure that could be leveraged to improve camera/IMU calibration. The answer to this question does not seem obvious, and designing an experiment that would unambiguously settle it is likewise not trivial. In order to isolate influencing factors, identical trajectories would have to be performed for different exposure times, a task that is more appropriately addressed by a robotic arm than by a human operator. A positive outcome of this test would spawn the follow-up question of an optimal exposure time.

Finally, while offline calibration is a valuable tool to advance the understanding of sensor models, its suitability as part of the deployment process of any robotic application heavily depends on the volatility of the calibration parameters. Sensor parameters may change on short time scales, for example due to temperature changes, or over longer periods of time through mechanical wear and component ageing. It seems plausible that such change affects virtually all types of sensors. Accordingly, offline calibration can only be one facet in successfully deploying robots: A truly robust process would likely employ offline calibration only as a first step to estimate an initial set of parameters that govern its sensor models. It would then identify a subset of these parameters which change over time or through external influences and optimality find a lower-dimensional solution space that sufficiently captures possible changes. Consequently, the system would track the volatile subset of parameters in this lower-dimensional space online, either continuously as part of the state or repeatedly at a frequency dictated by the time constants of the expected changes.

# List of Publications

The following list contains all accepted publications.

## Journals

Rehder, J., Siegwart, R., and Furgale, P. (2016). „A General Approach to Spatiotemporal Calibration in Multisensor Systems“. *IEEE Transactions on Robotics* 32.2, pp. 383–398

Rehder, J. and Siegwart, R. (2017). „Camera/IMU Calibration Revisited“. *IEEE Sensors Journal* 17.11, pp. 3257–3268

## Conferences

Nikolic, J., Burri, M., Rehder, J., Leutenegger, S., Huerzeler, C., and Siegwart, R. (2013). „A UAV system for inspection of industrial facilities“. In: *Aerospace Conference, 2013 IEEE*. IEEE, pp. 1–8

Furgale, P., Rehder, J., and Siegwart, R. (2013). „Unified Temporal and Spatial Calibration for Multi-Sensor Systems“. In: *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*

Nikolic, J., Rehder, J., Burri, M., Gohl, P., Leutenegger, S., Furgale, P. T., and Siegwart, R. (2014a). „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)*, pp. 431–437

Rehder, J., Beardsley, P., Siegwart, R., and Furgale, P. (2014). „Spatio-Temporal Laser to Visual/Inertial Calibration with Applications to Hand-Held, Large Scale Scanning“. In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Chicago, IL, USA, pp. 459–465

Jud, D., Mora, J. A., Rehder, J., Siegwart, R., and Beardsley, P. (2016). „Customized Sensing for Robot Swarms“. In: *Experimental Robotics: The 14th International Symposium on Experimental Robotics*. Ed. by M. A. Hsieh, O. Khatib, and V. Kumar. Cham: Springer International Publishing, pp. 523–534

Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M. W., and Siegwart, R. (2016). „The EuRoC micro aerial vehicle datasets“. *The International Journal of Robotics Research* 35.10, pp. 1157–1163

Rehder, J., Nikolic, J., Schneider, T., Hinzmann, T., and Siegwart, R. (2016). „Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes“. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 4304–4311

Rehder, J., Nikolic, J., Schneider, T., and Siegwart, R. (2017). „A Direct Formulation for Camera Calibration“. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 6479–6486

# Bibliography

Alismail, H., Baker, L. D., and Browning, B. (2012). „Automatic Calibration of a Range Sensor and Camera System“. In: *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*. IEEE, pp. 286–292.

Alves, J., Lobo, J., and Dias, J. (2003). „Camera-inertial sensor modelling and alignment for visual navigation“. *Machine Intelligence and Robotic Control* 5.3, pp. 103–112.

Anderson, S., MacTavish, K., and Barfoot, T. D. (2015). „Relative continuous-time SLAM“. *The International Journal of Robotics Research*, p. 0278364915589642.

Bartels, R. H., Beatty, J. C., and Barsky, B. A. (1987). *An Introduction to Splines for use in Computer Graphics and Geometric Modeling*. Morgan Kaufmann Publishers Inc.

Blake, A. and Zisserman, A. (1987). „Localising discontinuities using weak continuity constraints“. *Pattern recognition letters* 6.1, pp. 51–59.

Bok, Y., Jeong, Y., Choi, D.-G., and Kweon, I. S. (2011). „Capturing village-level heritages with a hand-held camera-laser fusion sensor“. *International Journal of Computer Vision* 94.1, pp. 36–53.

Bok, Y., Choi, D.-G., Vasseur, P., and Kweon, I. S. (2014). „Extrinsic Calibration of Non-overlapping Camera-Laser System using Structured Environment“. In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Bosse, M., Zlot, R., and Flick, P. (2012). „Zebedee: Design of a Spring-Mounted 3-D Range Sensor with Application to Mobile Mapping“. *Transactions on Robotics* 28, pp. 1104–1119.

Bouguet, J.-Y. (2004). *Camera Calibration Toolbox for Matlab*. URL: `http://www.vision.caltech.edu/bouguetj/calib_doc/`.

Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M. W., and Siegwart, R. (2016). „The EuRoC micro aerial vehicle datasets“. *The International Journal of Robotics Research* 35.10, pp. 1157–1163.

Chenga, P., Andersona, M., Heb, S., and Zakhor, A. (2014). „Texture Mapping 3D Models of Indoor Environments with Noisy Camera Poses“. In: *SPIE Electronic Imaging Conference 9020, Computational Imaging XII*.

Comport, A. I., Meilland, M., and Rives, P. (2011). „An asymmetric real-time dense visual localisation and mapping system“. In: *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, pp. 700–703.

de Boor, C. (2001). *A practical guide to splines*. Springer Verlag.

Debevec, P. E. and Malik, J. (2008). „Recovering high dynamic range radiance maps from photographs“. In: *ACM SIGGRAPH 2008 classes*. ACM, p. 31.

Demski, P., Mikulski, M., and Koteras, R. (2013). „Characterization of Hokuyo UTM-30LX laser range finder for an autonomous mobile robot“. In: *Advanced Technologies for Intelligent Systems of National Border Security*. Springer, pp. 143–153.

Eling, C, Wieland, M, Hess, C, Klingbeil, L, and Kuhlmann, H (2015). „Development and evaluation of a UAV based mapping system for remote sensing and surveying applications“. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40.1, p. 233.

Engel, J., Sturm, J., and Cremers, D. (2013). „Semi-dense visual odometry for a monocular camera“. In: *Proceedings of the IEEE international conference on computer vision*, pp. 1449–1456.

Fischler, M. A. and Bolles, R. C. (1981). „Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography“. *Communications of the ACM* 24.6, pp. 381–395.

*Flea3 USB 3.0 Digital Camera Technical Reference* (2016). Version 7.2. Point Grey.

Fleps, M., Mair, E., Ruepp, O., Suppa, M., and Burschka, D. (2011). „Optimization based IMU camera calibration“. In: *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*. IEEE, pp. 3297–3304.

Forster, C., Pizzoli, M., and Scaramuzza, D. (2014). „SVO: Fast semi-direct monocular visual odometry“. In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 15–22.

Furgale, P. T., Barfoot, T. D., and Sibley, G (2012). „Continuous-Time Batch Estimation Using Temporal Basis Functions“. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2088–2095.

Furgale, P., Rehder, J., and Siegwart, R. (2013). „Unified Temporal and Spatial Calibration for Multi-Sensor Systems“. In: *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Furgale, P., Tong, C. H., Barfoot, T. D., and Sibley, G. (2015a). „Continuous-time batch trajectory estimation using temporal basis functions“. *The International Journal of Robotics Research*, p. 0278364915585860.

Furgale, P., Maye, J., Rehder, J., and Schneider, T. (2015b). *kalibr — A unified Camera/IMU Calibration Toolbox*. URL: `https://github.com/ethz-asl/kalibr/`.

Geiger, A., Moosmann, F., Car, O., and Schuster, B. (2012). „Automatic camera and range sensor calibration using a single shot“. In: *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, pp. 3936–3943.

Golub, G. H. and Loan, C. F. van (1996). *Matrix Computations*. The John Hopkins University Press.

Grießbach, D., Baumbach, D., Börner, A., Buder, M., Ernst, I., Funk, E., Wohlfeil, J., and Zuev, S. (2012). „IPS–A system for real-time navigation and 3D-modeling“. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39, B5.

Grossberg, M. D. and Nayar, S. K. (2004). „Modeling the space of camera response functions“. *IEEE transactions on pattern analysis and machine intelligence* 26.10, pp. 1272–1282.

Hartley, R. and Zisserman, A. (2000). *Multiple view geometry in computer vision*. Vol. 2. Cambridge Univ Press.

Heide, F., Rouf, M., Hullin, M. B., Labitzke, B., Heidrich, W., and Kolb, A. (2013). „High-quality computational imaging through simple lenses“. *ACM Transactions on Graphics (TOG)* 32.5, p. 149.

Holtmann, K. (2016). *Characterization of Deterministic Errors in Laser Range Measurements*. Tech. rep. ETH Zurich.

Hughes, C., Denny, P., Glavin, M., and Jones, E. (2010). „Equidistant fish-eye calibration and rectification by vanishing point extraction". *IEEE transactions on pattern analysis and machine intelligence* 32.12, pp. 2289–2296.

Hughes, P. C. (1986). *Spacecraft Attitude Dynamics*. John Wiley & Sons.

Hung, J., Hunter, J., Stripling, W., and White, H. (1979). *Size Effect on Navigation using a Strapdown IMU*. Tech. rep. U.S. Army Missile Research, Development Command, Guidance, and Control Directorate Technology Laboratory.

Hwangbo, M., Kim, J.-S., and Kanade, T. (2013). „IMU self-calibration using factorization". *Robotics, IEEE Transactions on* 29.2, pp. 493–507.

IEEE Aerospace and Electronic Systems Society. Gyro and Accelerometer Panel and Institute of Electrical and Electronics Engineers (1998). *IEEE Standard Specification Format Guide and Test Procedure for Single-axis Interferometric Fiber Optic Gyros*. IEEE (std.) IEEE.

– (1999). *IEEE Standard Specification Format Guide and Test Procedure for Linear, Single-axis, Nongyroscopic Accelerometers*. IEEE (std.) IEEE.

InvenSense (2011). *MPU-9150 Register Map and Descriptions-Document Number: RM-MPU-9150A-00*. Tech. rep. InvenSense.

James, M. R. and Quinton, J. N. (2014). „Ultra-rapid topographic surveying for complex environments: the hand-held mobile laser scanner (HMLS)". *Earth Surface Processes and Landforms* 39.1, pp. 138–142.

Jones, E. S. and Soatto, S. (2011). „Visual-inertial navigation, mapping and localization: A scalable real-time causal approach". *The International Journal of Robotics Research* 30.4, pp. 407–430.

Joshi, N., Szeliski, R., and Kriegman, D. J. (2008). „PSF estimation using sharp edge prediction". In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, pp. 1–8.

Jud, D., Mora, J. A., Rehder, J., Siegwart, R., and Beardsley, P. (2016). „Customized Sensing for Robot Swarms". In: *Experimental Robotics: The 14th International Symposium on Experimental Robotics*. Ed. by M. A. Hsieh, O. Khatib, and V. Kumar. Springer International Publishing, pp. 523–534.

Kannala, J. and Brandt, S. S. (2006). „A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses“. *IEEE transactions on pattern analysis and machine intelligence* 28.8, pp. 1335–1340.

Kelly, J. and Sukhatme, G. (2009). „Fast relative pose calibration for visual and inertial sensors“. In: *Experimental Robotics*. Springer, pp. 515–524.

Kelly, J., Roy, N., and Sukhatme, G. S. (2014). „Determining the Time Delay Between Inertial and Visual Sensor Measurements“. *Robotics, IEEE Transactions on* 30.6, pp. 1514–1523.

Kelly, J. and Sukhatme, G. S. (2011). „Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration“. *The International Journal of Robotics Research* 30.1, pp. 56–79.

Kim, J. H., Cadena, C., and Reid, I. (2016). „Direct semi-dense SLAM for rolling shutter cameras“. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1308–1315.

Kim, S. J. and Pollefeys, M. (2008). „Robust radiometric calibration and vignetting correction“. *IEEE transactions on pattern analysis and machine intelligence* 30.4, pp. 562–576.

Klein, G. S. and Drummond, T. (2005). „A Single-frame Visual Gyroscope.“ In: *BMVC*.

Krebs, C. (2012). *Generic IMU-Camera Calibration Algorithm: Influence of IMU-axis on each other*. Tech. rep. Autonomous Systems Lab, ETH Zurich.

Leica Geosystems (2014). *Leica ScanStation P20*. URL: `http://www.leica-geosystems.com/en/Leica-ScanStation-P20\_101869.htm`.

Lepetit, V., Moreno-Noguer, F., and Fua, P. (2009). „Epnp: An accurate o (n) solution to the pnp problem“. *International journal of computer vision* 81.2, pp. 155–166.

Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., and Furgale, P. (2015). „Keyframe-based visual–inertial odometry using nonlinear optimization“. *The International Journal of Robotics Research* 34.3, pp. 314–334.

Li, G., Liu, Y., Dong, L., Cai, X., and Zhou, D. (2007). „An algorithm for extrinsic parameters calibration of a camera and a laser range finder using line features“. In: *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*. IEEE, pp. 3854–3859.

Li, M. and Mourikis, A. I. (2013). „3-D Motion Estimation and Online Temporal Calibration for Camera-IMU Systems“. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 5689–5696.

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

Li, M. and Mourikis, A. I. (2014). „Online temporal calibration for camera–IMU systems: Theory and algorithms“. *The International Journal of Robotics Research* 33.7, pp. 947–964.

Lobo, J. and Dias, J. (2007). „Relative pose calibration between visual and inertial sensors“. *The International Journal of Robotics Research* 26.6, pp. 561–575.

Looney, M. (2010). „A simple calibration for MEMS gyroscopes“. *EDN (Electrical Design News)* 55.9, p. 21.

Ma, J., Bajracharya, M., Susca, S., Matthies, L., and Malchano, M. (2015). „Real-time pose estimation of a dynamic quadruped in GPS-denied environments for 24-hour operation“. *The International Journal of Robotics Research*, p. 0278364915587333.

Mair, E., Fleps, M., Suppa, M., and Burschka, D. (2011). „Spatio-temporal initialization for IMU to camera registration“. In: *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on*. IEEE, pp. 557–564.

Marquardt, D. W. (1963). „An algorithm for least-squares estimation of nonlinear parameters“. *Journal of the Society for Industrial & Applied Mathematics* 11.2, pp. 431–441.

Maune, D., Photogrammetry, A. S. for, and Sensing, R. (2007). *Digital Elevation Model Technologies and Applications: The DEM Users Manual*. American Society for Photogrammetry and Remote Sensing.

Maye, J., Furgale, P., and Siegwart, R. (2013). „Self-supervised Calibration for Robotic Systems“. In: *IEEE Intelligent Vehicles Symposium (IV)*, pp. 473–480.

Mei, C. and Rives, P. (2006). „Calibration between a central catadioptric camera and a laser range finder for robotic applications“. In: *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*. IEEE, pp. 532–537.

– (2007). „Single view point omnidirectional camera calibration from planar grids“. In: *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, pp. 3945–3950.

Meilland, M. and Comport, A. I. (2013). „Super-resolution 3D tracking and mapping“. In: *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, pp. 5717–5723.

Meilland, M., Drummond, T., and Comport, A. I. (2013). „A unified rolling shutter and motion blur model for 3d visual registration“. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2016–2023.

Mirzaei, F. M. and Roumeliotis, S. I. (2007). *IMU-camera calibration: Bundle adjustment implementation*. Tech. rep. Department of Computer Science and Engineering, University of Minnesota.

Mirzaei, F. M., Kottas, D. G., and Roumeliotis, S. I. (2012). „3D LIDAR–camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization“. *The International Journal of Robotics Research* 31.4, pp. 452–467.

Mirzaei, F. and Roumeliotis, S. (2008). „A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation“. *Robotics, IEEE Transactions on* 24.5, pp. 1143–1156.

Moghadam, P., Bosse, M., and Zlot, R. (2013). „Line-based extrinsic calibration of range and image sensors“. In: *The 2013 IEEE International Conference on Robotics and Automation*. Vol. 2, p. 4.

Moon, S. B., Skelly, P., and Towsley, D. (1999). „Estimation and removal of clock skew from network delay measurements“. In: *INFOCOM’99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE.* Vol. 1. IEEE, pp. 227–234.

Mourikis, A. I. and Roumeliotis, S. I. (2007). „A multi-state constraint Kalman filter for vision-aided inertial navigation“. In: *Robotics and Automation (ICRA), 2007 IEEE International Conference on*. IEEE, pp. 3565–3572.

*Murata announces world’s first surface mount MEMS angular acceleration sensor.* `http://www.murata.com/en-us/about/newsroom/news/product/sensor/2015/0805`. Accessed 15. February 2016.

Nikolic, J., Rehder, J., Burri, M., Gohl, P., Leutenegger, S., Furgale, P. T., and Siegwart, R. (2014a). „A synchronized visual-inertial sensor system with FPGA preprocessing for accurate real-time SLAM“. In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 431–437.

Nikolic, J., Burri, M., Rehder, J., Leutenegger, S., Huerzeler, C., and Siegwart, R. (2013). „A UAV system for inspection of industrial facilities“. In: *Aerospace Conference, 2013 IEEE*. IEEE, pp. 1–8.

Nikolic, J., Rehder, J., Burri, M., Gohl, P., Leutenegger, S., Furgale, P. T., and Siegwart, R. (2014b). „A Synchronized Visual-Inertial Sensor System with FPGA Pre-Processing for Accurate Real-Time SLAM“. In: *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE.

Nikolic, J., Furgale, P., Melzer, A., and Siegwart, R. (2016a). „Maximum likelihood identification of inertial sensor noise model parameters“. *IEEE Sensors Journal* 16.1, pp. 163–176.

Nikolic, J., Burri, M., Gilitschenski, I., Nieto, J., and Siegwart, R. (2016b). „Non-Parametric Extrinsic and Intrinsic Calibration of Visual-Inertial Sensor Systems“. *IEEE Sensors Journal* 16.13, pp. 5433–5443.

Nikon (2014). *ModelMaker MMDx digital laser scanner for portable 3D inspection and reverse engineering*. URL: `http://www.nikonmetrology.com/en\_EU/Products/Laser-Scanning/Handheld-scanning/ModelMaker-MMDx/`.

Nilsson, J.-O., Skog, I., and Handel, P. (2014). „Aligning the Forces—Eliminating the Misalignments in IMU Arrays“. *Instrumentation and Measurement, IEEE Transactions on* 63.10, pp. 2498–2500.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. 2nd. Springer.

Núñez, P., Drews Jr, P., Rocha, R., and Dias, J. (2009). „Data Fusion Calibration for a 3D Laser Range Finder and a Camera using Inertial Data.“ In: *ECMR*, pp. 31–36.

Oth, L., Furgale, P., Kneip, L., and Siegwart, R. (2013). „Rolling shutter camera calibration“. In: *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, pp. 1360–1367.

Ovaska, S. and Valiviita, S. (1998). „Angular acceleration measurement: a review“. *Instrumentation and Measurement, IEEE Transactions on* 47.5, pp. 1211–1217.

Pandey, G., McBride, J. R., Savarese, S., and Eustice, R. M. (2012). „Automatic targetless extrinsic calibration of a 3d lidar and camera by maximizing mutual information“. In: *Proceedings of the AAAI National Conference on Artificial Intelligence*, pp. 2053–2059.

Park, B. S., Han, K., Lee, S., and Yu, M. (2015). „Analysis of compensation for a g-sensitivity scale-factor error for a MEMS vibratory gyroscope“. *Journal of Micromechanics and Microengineering* 25.11, p. 115006.

Patron-Perez, A., Lovegrove, S., and Sibley, G. (2015). „A Spline-Based Trajectory Representation for Sensor Fusion and Rolling Shutter Cameras“. *International Journal of Computer Vision*, pp. 1–12.

Pennec, X. (1999). „Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements.“ In: *NSIP*, pp. 194–198.

Pittelkau, M. E. (2005). „Calibration and attitude determination with redundant inertial measurement units“. *Journal of Guidance, Control, and Dynamics* 28.4, pp. 743–752.

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., and Ng, A. Y. (2009). „ROS: an open-source Robot Operating System“. In: *ICRA workshop on open source software*. Vol. 3. 3.2.

Rehder, J. and Siegwart, R. (2017). „Camera/IMU Calibration Revisited“. *IEEE Sensors Journal* 17.11, pp. 3257–3268.

Rehder, J., Siegwart, R., and Furgale, P. (2016). „A General Approach to Spatiotemporal Calibration in Multisensor Systems“. *IEEE Transactions on Robotics* 32.2, pp. 383–398.

Rehder, J., Beardsley, P., Siegwart, R., and Furgale, P. (2014). „Spatio-Temporal Laser to Visual/Inertial Calibration with Applications to Hand-Held, Large Scale Scanning“. In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 459–465.

Rehder, J., Nikolic, J., Schneider, T., Hinzmann, T., and Siegwart, R. (2016). „Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes“. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 4304–4311.

Rehder, J., Nikolic, J., Schneider, T., and Siegwart, R. (2017). „A Direct Formulation for Camera Calibration“. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 6479–6486.

Scaramuzza, D. (2006). *OCamCalib: Omnidirectional Camera Calibration Toolbox for Matlab*. URL: `https://sites.google.com/site/scarabotix/ocamcalib-toolbox`.

Scaramuzza, D., Harati, A., and Siegwart, R. (2007). „Extrinsic self calibration of a camera and a 3d laser range finder from natural scenes“. In: *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*. IEEE, pp. 4164–4169.

Scaramuzza, D., Martinelli, A., and Siegwart, R. (2006). „A flexible technique for accurate omnidirectional camera calibration and structure from motion“. In: *Fourth IEEE International Conference on Computer Vision Systems (ICVS'06)*. IEEE, pp. 45–45.

Schmid, K. and Hirschmüller, H. (2013). „Stereo vision and IMU based real-time egomotion and depth image computation on a handheld device“. In: *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, pp. 4671–4678.

Schulz, Y. (2016). *Direct Radiometric Calibration of Multispectral Cameras*. Tech. rep. ETH Zurich.

Sheehan, M., Harrison, A., and Newman, P. (2010). „Automatic self-calibration of a full field-of-view 3D n-laser scanner“. In: *Proceedings of the International Symposium on Experimental Robotics*, pp. 1–14.

Shen, S., Michael, N., and Kumar, V. (2015). „Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft MAVs“. In: *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, pp. 5303–5310.

So, E. W. Y. and Menegatti, E. (2012). „A unified approach to extrinsic calibration between a camera and a laser rangefinder using point-plane constraints“. In: *1st Workshop on Perception for Mobile Robots Autonomy*.

Tungadi, F., Kleeman, L., et al. (2008). „Time synchronisation and calibration of odometry and range sensors for high-speed mobile robot mapping“. In: *Proc. Australasian Conf. Robotics and Automation (ACRA), J. Kim and R. Mahony, Eds., Canberra, Australia*.

Tykkälä, T., Audras, C., and Comport, A. I. (2011). „Direct iterative closest point for real-time visual odometry“. In: *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, pp. 2050–2056.

Unnikrishnan, R. and Hebert, M. (2005). *Fast Extrinsic Calibration of a Laser Rangefinder to a Camera*. Tech. rep. CMU-RI-TR-05-09. Carnegie Mellon University.

Vasconcelos, F., Barreto, J. P., and Nunes, U. (2012). „A Minimal Solution for the Extrinsic Calibration of a Camera and a Laser-Rangefinder“. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.11, pp. 2097–2107.

Wahba, G. (1990). *Spline models for observational data*. Vol. 59. Siam.

Weiss, S., Achtelik, M. W., Lynen, S., Chli, M., and Siegwart, R. (2012). „Real-time onboard visual-inertial state estimation and self-calibration of mavs in unknown environments“. In: *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, pp. 957–964.

Xsens. *MTi User Manual*. English. Version MTi 10-series and MTi 100-series, Document MT0605P, Revision I. Xsens. December 20, 2016.

Yang, Z. and Shen, S. (2016). „Monocular Visual-Inertial State Estimation With Online Initialization and Camera-IMU Extrinsic Calibration“. *IEEE Transactions on Automation Science and Engineering* PP.99, pp. 1–13.

Zachariah, D. and Jansson, M. (2010). „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, pp. 1–7.

Zhang, L., Liu, Z., and Honghui Xia, C (2002). „Clock synchronization algorithms for network measurements“. In: *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*. Vol. 1. IEEE, pp. 160–169.

Zhang, Q. and Pless, R. (2004). „Extrinsic calibration of a camera and laser range finder (improves camera calibration)“. In: *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*. Vol. 3. IEEE, pp. 2301–2306.

Zhang, Z. (2000). „A flexible new technique for camera calibration“. *IEEE Transactions on pattern analysis and machine intelligence* 22.11, pp. 1330–1334.

# Curriculum Vitae

## Joern Christian Rehder

born June 14th, 1984 in Hamburg, Germany

| | |
|---|---|
| 2012 – 2017 | *ETH, Zurich, Switzerland*<br>Doctoral studies in robotics with the Autonomous System Lab supervised by Prof. Roland Siegwart |
| 2011 | *Carnegie Mellon University, Pittsburgh, Pennsylvania*<br>Research visit (10 months) with the Field Robotics Center supervised by Prof. Sanjiv Singh |
| 2010 | *Volkswagen AG, Wolfsburg, Germany*<br>Internship (6 months) with the research division |
| 2007 – 2008 | *University of California, Berkeley, California*<br>Study abroad (one academic year) funded by a scholarship granted by the German Academic Exchange Service (DAAD) |
| 2004 – 2012 | *TU Hamburg-Harburg, Hamburg, Germany*<br>Dipl. Ing. in Electrical Engineering (with distinction) |
| 2003 – 2004 | *“Haus am Schüberg”, Ammersbek, Germany*<br>Civil service |
| 1998 – 2003 | *Gymnasium Stormarnschule, Ahrensburg, Germany*<br>Abitur |