



## Doctoral Thesis

# **Characterisation, Calibration, and Design of Visual-Inertial Sensor Systems for Robot Navigation**

**Author(s):**

Nikolic, Janosch

**Publication Date:**

2016

**Permanent Link:**

<https://doi.org/10.3929/ethz-a-010874630> →

**Rights / License:**

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

DISS. ETH NO. 23546

Characterisation, Calibration, and Design of  
**VISUAL-INERTIAL SENSOR SYSTEMS**  
for Robot Navigation

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH  
(Dr. sc. ETH ZURICH)

presented by

JANOSCH NIKOLIC

Dipl. El.-Ing. FH, Hochschule für Technik Rapperswil, Schweiz  
M.Sc. in Signal Processing and Communications,  
University of Edinburgh, UK

born on 4. January 1983

citizen of  
Salenstein TG, Switzerland

accepted on the recommendation of

Prof. Dr. Roland Y. Siegwart  
Prof. Dr. Davide Scaramuzza

2016

Autonomous Systems Lab  
ETH Zürich  
Switzerland

© 2016 Janosch Nikolic. All rights reserved.

# Abstract

Many robotic systems require a three dimensional map of the environment and knowledge of their position, orientation, and velocity within this map. This thesis advocates the use of visual (cameras) and inertial sensors (gyroscopes and accelerometers) for motion estimation and mapping. Its aim is to provide an in-depth understanding of important aspects to consider when combining data from these two sensing modalities. To this end, this thesis presents contributions in the design, characterisation, calibration, and application of visual-inertial sensor systems.

Mechanisms for the tight fusion of visual and inertial data can benefit from certain features in the sensing hardware such as time synchronisation. We therefore present the design of a novel visual-inertial sensor unit with the aim of maximising, at the hardware level,

*Abstract*

the benefits from the fusion of the two sensing modalities. The unit incorporates global shutter cameras, gyroscopes, and accelerometers; all selected specifically to address the requirements of robotic localisation and mapping systems. The rigid spatial alignment and the deterministic synchronisation between the sensors facilitate an optimal exploitation of all sensor measurements, and allows us to draw meaningful conclusions from experiments. Furthermore, the pre-calibrated unit facilitates research on visual-inertial sensor fusion at a low entry barrier and enables us to effortlessly equip robotic platforms with real-time localisation and mapping capabilities.

When combining measurements from different sources, accurate sensor models are crucial. To this end, we present a novel data-driven method which can automatically determine optimal noise model parameters for inertial sensors. The proposed method can identify stochastic model parameters for composite noise processes that span large time scales and magnitudes – as they often appear in micro electromechanical gyroscopes and accelerometers. Results for real data from several sensors of different technologies (micro electromechanical, fibre optic, and ring laser) show that the method performs well assuming that the model possesses sufficient expressive power. Furthermore, we observe that modern consumer chip inertial measurement units can outperform more expensive gyroscopes and accelerometers in terms of noise performance. This raises interesting questions regarding sensor selection and grows expectations for high performance, miniature visual-inertial sensor units in the future.

High sensor noise performance cannot be leveraged unless the translation and orientation between the sensors and the sensor intrinsic calibration parameters are known precisely. In addition, all measurements need to be accurately aligned in time, accounting for delays between the sensors and sensor internal delays. To address this, we present a batch optimal calibration routine which jointly estimates all calibration parameters using every available measurement. The estimator is simple to implement and requires – in contrast to

existing methods – no tuning parameters. The parameter estimates are asymptotically unbiased and attain the Cramér-Rao lower bound up to linearisation errors. We reach a precision of less than 1 mm in extrinsic translation, 1 mrad in orientation, and 10  $\mu$ s in time shift. Detailed results from repeated calibration experiments highlight the scale of the intrinsic errors in commercial gyroscopes and accelerometers and the need to take them into consideration when designing low-cost, precision visual-inertial sensor units.

To conclude, we demonstrate the potential of visual-inertial sensor suites for real-world applications through field tests. We task an unmanned aerial inspection vehicle to navigate in the boiler unit of a thermal power plant relying on motion estimates from an on-board visual-inertial sensor unit – the first mission of its kind. The field tests reveal the motion estimation performance and the robustness that can be achieved under difficult environmental conditions using such a sensor suite and dedicated visual-inertial motion estimation algorithms.

**Keywords:** Visual-inertial sensor unit, calibration, maximum likelihood, noise model parameter identification, gyroscopes, accelerometers, cameras.

# Zusammenfassung

Viele Systeme in der Robotik benötigen eine dreidimensionale Karte ihrer Umgebung und müssen ihre eigene Position, Orientierung und Geschwindigkeit in dieser Karte kennen um zu funktionieren. Diese Dissertation schlägt den Gebrauch von visuellen (Kameras) und inertialen Sensoren (Gyroskope und Beschleunigungsmesser) für die Eigenbewegungsschätzung und Kartenerstellung vor. Das Bestreben unserer Arbeit ist es, ein tief gehendes Verständnis für die wichtigen Aspekte bei der Kombination dieser beiden Messmodalitäten zu schaffen. Zu diesem Zweck präsentieren wir Beiträge in den Bereichen der Entwicklung, Charakterisierung, Eichung und Anwendung von visuell-inertialer Sensorik.

Mechanismen zur engen Verknüpfung von visuellen und inertialen Messdaten können von bestimmten Eigenschaften der Sensorik

profitieren, beispielsweise von einer genauen zeitlichen Synchronisierung. Zu diesem Zweck präsentieren wir den Entwurf einer neuartigen visuell-inertialen Sensoreinheit mit dem Ziel, diese Eigenschaften auf der Hardwareebene zu optimieren. Die Einheit beinhaltet mehrere Kameras mit globalem Verschluss, Kreisel und Beschleunigungssensoren. Alle Sensoren wurden spezifisch ausgewählt um den Ansprüchen von Lokalisierungs- und Kartenerstellungssystemen in der Robotik gerecht zu werden. Die rigide räumliche Anordnung und die deterministische zeitliche Synchronisierung zwischen den Sensoren ermöglicht eine optimale Verwertung aller Messungen und ermöglicht es uns, aussagekräftige Schlüsse aus Experimenten zu ziehen. Eine solche vorkalibrierte Einheit erleichtert die Erforschung der visuell-inertialen Sensorfusion und erlaubt es uns des weiteren, Roboterplattformen einfach mit Fähigkeiten in der Echtzeit-Lokalisierung und Kartenerstellung auszustatten.

Wenn Informationen von verschiedenen Sensoren verbunden werden, sind genaue Sensormodelle unumgänglich. Zu diesem Zweck präsentieren wir eine neuartige, datenbasierte Methode, welche automatisch optimale Rauschmodellparameter für Inertialsensoren ermitteln kann. Die vorgeschlagene Methode kann Parameter für zusammengesetzte stochastische Prozesse, welche grosse Variationen in der Korrelationszeit und Amplitude aufweisen – wie sie oft in mikro-elektromechanischen Gyroskopen und Beschleunigungssensoren auftreten – problemlos identifizieren. Resultate basierend auf Daten verschiedener Sensoren und verschiedener Sensortechnologien (mikro-elektromechanisch, faseroptisch und Ringlaser) zeigen, dass die Methode gut funktioniert solange die Modelle geeignet sind, die Phänomene ausreichend zu beschreiben. Weiter beobachten wir, dass moderne mikro-elektromechanischen Chip-Trägheitsnavigationssensoren teurere und grössere Sensoreinheiten bezüglich des Rauschverhaltens übertreffen können. Dies wirft interessante Fragen zur Auswahl der Sensoren auf und wird voraussichtlich die Entwicklung von kleinsten, präzisen visuell-inertialen Sensoreinheiten in der

## *Zusammenfassung*

Zukunft ermöglichen.

Gute Sensorrauscheigenschaften bringen nur dann einen Vorteil, wenn die relative Position und Orientierung zwischen den Sensoren, wie auch die Sensor intrinsischen Kalibrier-Parameter, präzise bestimmt sind. Zudem müssen die Daten aller Sensoren zeitlich genau abgestimmt sein, wobei die Verzögerung zwischen den Sensoren wie auch sensorinterne Verzögerungen berücksichtigt werden müssen. Zu diesem Zweck präsentieren wir einen blockoptimalen Kalibrieralgorithmus, der, basierend auf sämtlichen zur Verfügung stehenden Messwerten, gleichzeitig alle diese Parameter bestimmt. Der Schätzalgorithmus ist einfach zu implementieren und benötigt – im Gegensatz zu existierenden Methoden – keine Abstimmungsparameter. Die geschätzten Parameter sind asymptotisch unverzerrt und erreichen die untere Cramér-Rao Schranke bis auf Fehler in der Linearisierung. Wir erlangen damit eine Präzision von unter 1 mm in der extrinsischen Translation, 1 mrad in der relativen Orientierung, und 10 µs in der Bestimmung des zeitlichen Versatzes. Detaillierte Resultate aus wiederholten Kalibrierexperimenten zeigen die typische Grösse der internen Fehlerquellen von Gyroskopen und Beschleunigungssensoren und unterstreichen die Wichtigkeit, diese in der Entwicklung von präzisen visuell-inertialen Einheiten zu berücksichtigen.

Abschliessend demonstrieren wir das Potenzial von visuell-inertialen Sensorkombinationen mittels Feldversuchen. Dabei soll ein unbemannter Flugroboter selbstständig in der Brennkammer eines thermischen Kraftwerks navigieren um eine visuelle Vorinspektion der Kammer zu ermöglichen – die erste Mission ihrer Art. Zu diesem Zweck wurde der Inspektionsroboter mit einer visuell-inertialen Sensoreinheit und Algorithmen zur Eigenbewegungsschätzung ausgestattet. Die Feldversuche zeigen die Genauigkeit und die Robustheit auf, welche selbst unter schwierigen Umgebungsbedingungen mit einer solchen Sensoreinheit und dedizierten Algorithmen zur visuell-inertialen Bewegungsschätzung erreicht werden können.

**Schlüsselwörter:** Visuell-inertiale Sensoreinheit, Kalibrierung, Rauschmodellparameterbestimmung, Maximum Likelihood, Gyroskope, Beschleunigungsmesser, Kameras.



# Contents

<b>Abstract</b> . . . . .	i
<b>Zusammenfassung</b> . . . . .	iv
<b>Preface</b> . . . . .	1
<b>1. Introduction</b> . . . . .	5
1.1    Aim and Scope . . . . .	8
1.2    Approach . . . . .	12
1.3    Thesis Outline . . . . .	18
<b>2. Contributions</b> . . . . .	19
2.1    Summary of Publications . . . . .	20
2.2    List of Publications . . . . .	26
2.3    Supervised Student Projects . . . . .	28

*Contents*

<b>3. Future Directions . . . . .</b>	<b>33</b>
3.1 Calibration . . . . .	34
3.2 Visual-Inertial Sensor Unit Design . . . . .	36
<b>Paper I. Maximum Likelihood Identification of Inertial Sensor Noise Model Parameters . . . . .</b>	<b>39</b>
1. Introduction . . . . .	41
2. Stochastic Sensor Models . . . . .	47
3. Automatic Model Identification . . . . .	58
4. Experiments . . . . .	65
5. Conclusion . . . . .	86
Acknowledgements . . . . .	87
<b>Paper II. Non-Parametric Extrinsic and Intrinsic Calibration of Visual-Inertial Sensor Systems . . . . .</b>	<b>89</b>
1. Introduction . . . . .	91
2. Related Work . . . . .	93
3. The Visual-Inertial Sensor System . . . . .	96
4. Calibration . . . . .	102
5. Experiments and Results . . . . .	111
6. Conclusion and Future Work . . . . .	120
Acknowledgements . . . . .	121
<b>Paper III. A Synchronized Visual-Inertial Sensor System with FPGA Pre-Processing for Accurate Real-Time SLAM . . . . .</b>	<b>125</b>
1. Introduction . . . . .	127
2. Related Work . . . . .	129
3. The Visual-Inertial SLAM Sensor . . . . .	130
4. Calibration . . . . .	139
5. Visual-Inertial Motion Estimation . . . . .	140
6. Results . . . . .	142
7. Conclusion and Outlook . . . . .	145
Acknowledgements . . . . .	146

<b>Paper IV. A UAV System for Inspection of Industrial Facilities</b>	147
1. Introduction	149
2. Related Work	151
3. Aerial System Design and Control	153
4. Integrated Sensor Design and State Estimation	157
5. Experiments and Results	160
6. Conclusion	166
Acknowledgements	167
<b>Bibliography</b>	169
<b>About the Author</b>	183



# Preface

This thesis is a cumulative dissertation. At its core are four self-contained publications which are included in this text in their full length. Chapter 1 introduces the topic of visual-inertial sensing. It outlines the scope and the aims of our work and presents the approaches we chose to address these aims. Chapter 2 summarises the contributions of this thesis to engineering in robotics. It discusses the context and the contribution of each publication and contains a list of related publications which resulted from our work. Chapter 3 briefly discusses future work in the field of visual-inertial sensing. The bibliographies of all chapters, including those of the publications, are combined in a single bibliography at the end of this text.

## ACKNOWLEDGEMENTS

I am deeply grateful to my thesis supervisor Prof. Roland Siegwart, head of the Autonomous Systems Lab, ETH Zürich, for his enduring support, encouragement, and trust. With his lab he provided a research and engineering environment which is second to none. It is also with a great deal of pleasure that I acknowledge the contributions of my colleagues Joern Rehder and Michael Burri. I greatly benefited from their scientific advice, knowledge, engineering skills, and hard work.

My thanks go to Dr. Paul Furgale for his guidance, and to Stefan Bertschi, Amir Melzer, Prof. Stefan Leutenegger, and Konrad Rudin, for their help and encouragement. I am grateful to Dr. Christoph Hürzeler and Prof. Konstantinos Alexis for their supervision and their great leadership in the AIRobots project. For the exceptional quality of their work I would like to thank Markus Bühler and Dario Fenner from the ASL workshop. Similarly, I would like to thank Luciana Borsatti, Cornelia Della Casa, and Margot Ziekau for their generous help and support.

I am thankful for the help of Dr. Markus Achtelik, Simon Lynen, Dr. Laurent Kneip, Dr. Stephan Weiss, Thomas Baumgartner, Dr. François Pomerleau, Dr. Mark Höpflinger, Philipp Krüsi, and Fadri Furrer. I would also like to acknowledge the contributions of Pascal Gohl to the success of the CTI project. I wish to thank Dr. Igor Gilitschenski and Dr. Juan Nieto for their expert scientific advice and for revising this manuscript. For their guidance and support I am also indebted to Dr. Samir Bouabdallah, Dr. Gilles Caprari, Prof. Margarita Chli, and Prof. Davide Scaramuzza, former post docs of the Autonomous Systems Lab.

Throughout my dissertation I had the privilege to supervise students at the Bachelor and Master level. They contributed significantly to this work and I would like to express my gratitude to them. Amongst all of them, the work of Rainer Voigt was probably the

most influential, and indeed laid the foundations for many of our contributions.

I would like to thank Dr. Johnny Lee and the members of his team from the Advanced Technology and Projects group at Google for hosting me. I thank Prof. Lorenzo Marconi and his team from the University of Bologna, Italy, for the fruitful collaboration and their hospitality. My thanks also go to everyone involved in the Narcea field trials, without whose support our experiments would not have been possible. The efforts and cooperation with our CTI partners and the consortium of the Armasuisse projects, headed by Dr. Quentin Ladetto, are also acknowledged.

My deepest thanks, however, go to my family, friends, and loved ones.

Zürich, April 2016

Janosch Nikolic

## FINANCIAL SUPPORT

The European Commission's Seventh Framework Programme (FP7) and armasuisse S+T funded the application oriented part of this research under grant agreement no. FP7-248669 (AIRobots) and project no. 8003501880. The Swiss Commission for Technology and Innovation (CTI) funded the development of the visual-inertial sensor unit under grant 13394.1 PFFLE-NM.



# CHAPTER 1

## Introduction

STATE estimation and mapping are core problems in robotics: many robots require a map of the environment and they need to know their state within this map in order to operate. Numerous solutions to this problem were developed over the past century. Most of them relied on external infrastructure such as the global positioning system (GPS) and optical tracking systems. In contrast, Simultaneous Localisation and Mapping (SLAM) more broadly tackles the problem of mapping a previously unknown environment while at the same time estimating the robot's state within this map. SLAM is ideally suited for applications where external infrastructure is not available – for example inside the boiler unit of a thermal power plant.

Many sensing technologies are used to provide measurements to these SLAM algorithms. Arguably the most prominent are range

## *Chapter 1. Introduction*

sensors such as 2D and 3D laser scanners, RGB-D sensors, wheel odometry, and ultrasonic range finders. Over the past decades, cameras gained in popularity due to their small form factor, their low price, and the massive amount of information they can provide for SLAM algorithms. The availability of sufficient computational resources on embedded platforms enabled their use for real-time SLAM and further advanced their deployment.

Cameras provide a projection of the three dimensional world into the camera's two dimensional image plane. Structure can only be inferred from motion or from multiple, rigidly connected cameras. In turn, inferring motion from images requires visual texture in the environment, i.e. salient and distinctive regions in the images that can be tracked across camera frames over time. This makes the use of cameras for SLAM challenging and prone to failures when operated on dynamic robots in real, poorly textured environments. Additional challenges are posed by repetitive structures, motion blur, and poor illumination. Furthermore, processing camera images is computationally demanding which prohibits their use for low latency, high rate applications such as unmanned aerial vehicle (UAV) guidance and control or augmented reality (AR).

To mitigate the problem, camera systems are often augmented with supplementary sensors. Inertial measurement units – gyroscopes and accelerometers – are an ideal addition to cameras. They measure the robot's angular velocity and the specific force, sensing modalities entirely different from images. Gyroscopes and accelerometers are interoceptive sensors, i.e. their measurements are independent of the robot's environment: they function equally well in darkness, dust, and under rapid motion. They can aid the lower layer camera image processing and increase the robustness and accuracy of SLAM systems substantially. In addition, their high bandwidth and sampling rate enable low latency motion estimation at a high rate.

The recent appearance of small form factor, low cost micro electromechanical (MEMS) chip gyroscopes and accelerometers facilitates

the integration of inertial sensors into SLAM systems in countless applications. However, inertial sensors suffer from their own deficiencies. MEMS sensors in particular [47] are prone to noise as well as deterministic errors such as non-ideal scale factors or axes misalignment. If a tight fusion of their measurements with visual cues from cameras is desired, proper modelling and calibration of these sensors is vital.

Due to its potential impact in various fields visual-inertial motion estimation received considerable attention in industry and academia [52, 32, 89]. Nevertheless, certain aspects in the design of visual-inertial navigation solutions still remain poorly explored. For instance, the noise processes which should model inherent sensor noise are often inflated as a heuristic to capture errors due to wrong timing or insufficient calibration – with or without the knowledge of the engineer. Furthermore, inertial sensor intrinsic calibration, in contrast to camera calibration, is rarely addressed rigorously in the context of visual-inertial sensor systems. This is the case despite the fact that insufficiently calibrated inertial sensors invalidate any stochastic sensor model immediately. Another example of an only partially explored problem in visual-inertial navigation is the importance of temporal alignment between the visual and the inertial sensors which is only sparsely explored in existing literature. IMU intrinsic delays are usually disregarded which leads to suboptimal performance.

Furthermore, combined visual-inertial sensor units are, with a few exceptions, not readily available. This hinders research in visual-inertial state estimation and mapping and prohibits the wide-spread use of visual-inertial SLAM systems for robotic applications.

## 1.1 Aim and Scope

The primary aim of this work is to advance the development of visual-inertial sensor systems for robotic applications. To this end, this thesis contributes to key problems in the design, calibration, characterisation, and application of such systems and presents solutions to them. This facilitates an in-depth understanding of these topics and can guide the future design of visual-inertial sensor systems. The following sections outline the aims and the scope of this thesis. They are split into four parts, and each part will later be addressed by one of the publications included in this thesis.

We start by discussing a field test performed in a boiler unit of a thermal power plant in northern Spain. This field test motivated the subsequent development of a visual-inertial sensor unit and the methods to calibrate and characterise it. During these field tests, which are described in Paper IV, we realised the potential that visual-inertial sensor systems have for robot guidance and navigation under difficult conditions. We also became aware of the limitations that our collection of loosely assembled cameras, inertial sensors, and illumination modules had.

We then started to build a visual-inertial sensor unit targeted at robot state estimation and mapping applications. This led to the hardware design we present in Paper III. The unit was a key to facilitate further research in tightly coupled visual-inertial SLAM. The unit was deployed on a wide range of platforms.

The availability of this visual-inertial sensor system immediately raised the question of how to optimally calibrate it. Furgale, Rehder, and Siegwart [25] developed a continuous-time estimator to calibrate the relative transforms between the cameras and the IMU. Later on, we developed a discrete-time method for a full calibration of the sensor unit. The method is presented in Paper II. The calibration relies on sensor noise models in order to efficiently combine the measurements

### 1.1 Aim and Scope

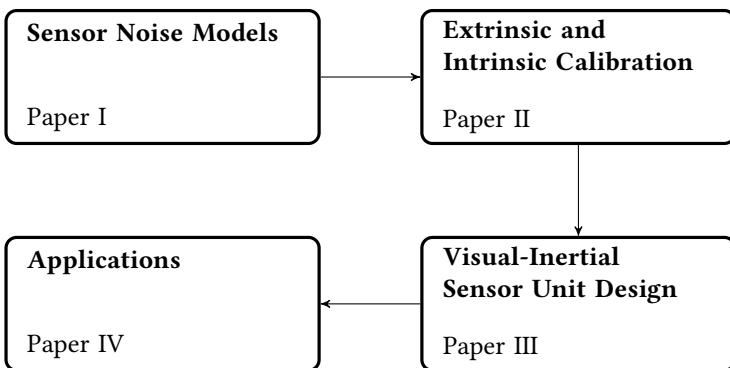


Figure 1.1: The scope of this thesis: Topics in the design, modelling, calibration, and application of visual-inertial sensor systems for robotics.

from all sensors and ultimately compute optimal calibration parameters. Parameter determination for the noise models is a difficult task and raised the question if an algorithm which can automatically determine noise models from data can be designed. Our findings on this problem are presented in Paper I.

In this thesis, the topics are presented in a logical rather than in a chronological order. Fig. 1.1 shows the core topics presented in this thesis and the corresponding papers which address them. The remainder of this section describes the aims in more detail and outlines the scope of each of these topics.

## Inertial sensor noise models

Accurate sensor models are crucial for the design of consistent estimators – in the context of calibration as well as state estimation

## *Chapter 1. Introduction*

and mapping. The sensor error models usually consist of a stochastic (noise) and a deterministic part. In contrast to the often simple noise models associated with cameras, random errors in gyroscopes and accelerometers are usually described with more complex, continuous-time composite noise processes. The parameters of these models carry units such as  $\text{rad}/\text{s}^2$  per  $\sqrt{\text{Hz}}$ , and are often difficult to infer unambiguously from sensor data sheets [4]. This can lead to error and often necessitates manual parameter tuning.

We aim at a method which can automatically infer the parameters of gyroscope and accelerometer noise models from data and thus limit the potential for errors. Furthermore, a rigorous probabilistic approach must be followed to maximise the validity of the model and ultimately to attain optimal calibration or motion estimation performance. We assume that the structure of the noise model itself is given, as it is the case in most applications. Automatic model selection is outside the scope of this thesis.

## **Extrinsic and intrinsic calibration**

The best sensors are of limited use if their spatial and temporal alignment are not inferred precisely or if sensor intrinsic calibration is missing. If the sensor intrinsics are disregarded, overall performance can be far from the theoretically possible, noise-bounded performance. We aim to develop a routine that is capable of providing a full extrinsic and intrinsic calibration. The calibration routine should be conceptually simple, generic, and require a minimal number of design choices and user defined parameters. In order to be widely applicable, the method should not rely on specialised equipment such as rate tables or tracking devices. Most importantly, the method must be subject to stringent probabilistic criteria.

We aim at offline calibration in a controlled environment. In order to be widely applicable the method should be agnostic to the motion that the sensor unit undergoes for calibration. This facilitates different

### *1.1 Aim and Scope*

calibration procedures although the quality of the calibration depends on the motion. In our case, the sensor unit moves in front of a visual calibration target. Calibration based on natural landmarks is outside the scope of this thesis.

## **Standardised visual-inertial sensor unit**

The most important aim of our work is to contribute towards accurate motion estimation and mapping solutions for robotics. In this regard, the goal of the development of a custom visual-inertial sensor unit is twofold: in practice, if a localisation and mapping framework is based on visual and inertial data, it requires a hardware sensor suite that can provide this data. Secondly, visual-inertial estimation frameworks can benefit from certain hardware features such as rigidly mounted sensor subsystems, deterministic temporal alignment, and a full pre-calibration of all subsystems.

To this end, we aim to provide a unified sensor hardware front-end and the corresponding interfaces to facilitate an effortless deployment of visual-inertial SLAM frameworks on any robotic system. This also enables research on visual-inertial state estimation or SLAM at a low entry barrier. Furthermore, we aim to give definite answers to questions related to sensor modelling and calibration. A custom sensor setup gives us full control over the hardware and allows us to draw rigorous conclusions.

## **Applications**

We aim to validate our designs in realistic real-world scenarios to improve our developments and to assess their performance in real applications. This should result in a better hardware design, improved firmware, and more information for a proper sensor selection.

The scope of the application oriented part of this thesis is to further show how visual-inertial sensor systems can be used in demanding

## *Chapter 1. Introduction*

scenarios: payload restricted platforms, enclosed industrial environments, poor visual conditions (illumination and repetitive texture), and in cases where high navigation accuracy is required. The field tests should convince the reader that combined visual-inertial sensors are an ideal sensor suite for localisation and mapping in many demanding use-cases.

## **1.2 Approach**

This section outlines the methods and techniques that are developed in this work to address the aims defined in Section 1.1. For a detailed description of the algorithms, system designs, and applications, the reader is referred to the articles which are included in this thesis (Paper I - IV). Corresponding references to the papers are made throughout this section.

### **Inertial sensor noise model parameter identification**

The development of our automatic inertial sensor noise model identification method is guided by four principles:

1. the method must be automatic,
2. the method should depend on as few design parameters as possible,
3. the method must be optimal with respect to stringent probabilistic criteria,
4. the structure of the stochastic model is defined by the user.

In accordance with these principles we develop an estimator based on the maximum likelihood paradigm.

## 1.2 Approach

To successfully apply standard maximum likelihood estimation techniques to this problem, two simple but crucial elements are required. The straight forward approach would be to design an estimator which develops models that directly explain the noise in the gyroscope and accelerometer measurements. However, this turns out to be sub-optimal, and the reason is the following: the sensor measurement noise characteristics are dominated by effects which are irrelevant for the navigation states we are ultimately interested in for sensor fusion; the orientation, velocity, and position of the sensor unit. Device internal low-pass filters, for example, have a drastic effect on the gyroscope and accelerometer measurement noise but not on the integrated quantities we are interested in (the orientation, velocity, or position). An estimator which tries to explain measurement noise directly will use its expressive power to model these effects.

Weak, correlated bias fluctuations, on the other hand, have almost no effect on measurement noise, but a drastic impact on the integrated processes – yet they would likely not be modelled. We therefore design the estimator such that the effects are captured according to their impact on the integrated process. This results in a simple but crucial modification to the estimation problem as we outline in Paper I. By changing the objective in this fashion, the desired parameters are obtained immediately.

In addition, the integrated measurements are sampled logarithmically in time. In other words, we select logarithmically spaced samples instead of the most common choice of a regular sampling pattern. This comes with two advantages: i) the identified models are forced to be valid across all time-scales. And ii) the computational complexity is manageable for almost arbitrarily long data sequences without any additional pre-processing.

We use synthetic data to quantitatively compare the performance of the algorithm with competing approaches. To verify that the proposed estimator works on real sensor data we experimentally test it on data captured with a variety of different inertial sensors. The full

## *Chapter 1. Introduction*

estimator and detailed results can be found in Paper I.

### **Discrete-time maximum likelihood calibration**

In order to address the calibration problem we regard the moving visual-inertial sensor unit as a continuous-time dynamic system that provides discrete-time measurements. We then treat calibration as a standard state and parameter estimation problem. To this end, we develop an estimator based on the maximum likelihood principle to jointly estimate the state and the calibration parameters. In other words, we determine the state trajectory and the calibration parameters such that the probability of observing the measurements we have observed is maximised.

All time varying quantities, such as the orientation, position, and velocity of the sensor unit, as well as varying sensor biases, are lumped into the system state. The extrinsic and intrinsic calibration quantities are considered to be fixed but unknown and treated as parameters.

The offline calibration procedure developed in this thesis relies on a visual calibration target (a checker board). The visual-inertial sensor unit is required to move in front of the target for calibration. The estimator then determines the full state trajectory of the sensor unit as well as the calibration parameters using all available measurements. Concretely, the value of the state is estimated at every time instance where a camera frame is captured. This is necessary since the camera observations depend on the value of the state (i.e. the pose of the sensor unit) at these time instances. In-between, the state and its covariance are integrated using a first order (Euler) method. This is in contrast to modern parametric methods such as [25] which represent the motion of the unit, as well as sensor biases, with a parametric representation (e.g. splines). We therefore refer to our method as “non-parametric” even though it can be argued that the term is not entirely correct. Paper II presents the method in full detail.

We test the estimator on real sensor data in order to quantify its

## 1.2 Approach

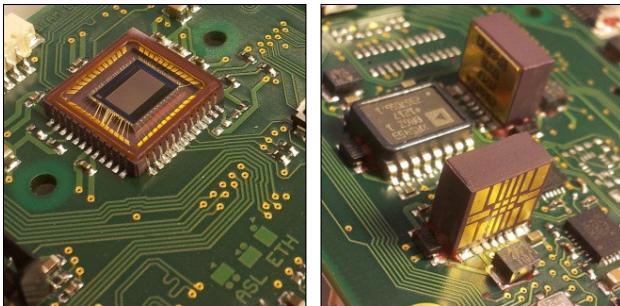


Figure 1.2: Camera (left) and gyroscope (right) sensors on one of the visual-inertial sensor unit prototypes, the “P1”.

performance and assess its applicability for precise extrinsic and intrinsic calibration of visual-inertial sensor units. Parameter estimates from repeated experiments are then compared against ground-truth where available. In addition, we conduct a direct comparison with a state of the art parametric method.

### Development of a visual-inertial sensor unit

We designed a custom visual-inertial sensor unit in order to i) gain full control over the hardware, ii) facilitate research on visual-inertial estimation at a low entry barrier, and iii) enable effortless deployment of these sensors on robots in order to equip them with robust real-time localisation and mapping capabilities.

The sensors we use were selected for properties that make them well suited for visual-inertial localisation and mapping applications. At the core of our module’s design is a Xilinx Zynq “System on Chip” (SoC), a combination of an industry standard dual-core ARM A9 CPU and FPGA fabric. It directly interfaces up to four global shutter, low light sensitive CMOS camera chips, shown in Fig. 1.2.

## *Chapter 1. Introduction*

The comparatively small form factor of the cameras enables the use of widely available 12 mm S-mount lenses. The measurements of all sensors are synchronised relative to the center of the camera exposure times to enforce a more deterministic behaviour when the camera exposure times vary; a feature introduced by Rehder in [64]. All sensors are rigidly connected with each other in order to facilitate factory pre-calibration and a tight combination of visual and inertial cues without the need to track extrinsic calibration parameters.

One factory-calibrated IMU is integrated into the design to provide angular rate and acceleration measurements. The unit is pre-calibrated over a large temperature range and serves as a reference. In addition, commercial low-cost chip IMUs are mounted behind each of the camera chips. They were selected primarily for their superior noise performance in comparison with other sensors of the same class. They allow us to compare and contrast different sensors and to test our calibration mechanisms. Additionally, they enable non-rigid combinations of visual-inertial sensor modules if desired. In order to facilitate an effortless deployment on various platforms and systems, the sensor units are designed for automatic fabrication although still in a prototype stage. The reader is referred to Paper III for detailed information about the sensor unit.

## **Application**

To assess the potential of our visual-inertial sensor unit for real applications we conduct field tests with an early prototype of the unit. A field test gives valuable feedback for design improvements and proves that the system actually works.

The use-case presented here was largely determined through our involvement in the Airobots project [57] and proved to be ideal for our cause. The scenario we address is the visual inspection of power plant boiler systems. Fig. 1.3 shows the site where we conducted the field trials – a thermal power plant in Narcea, northern Spain. The

## 1.2 Approach



Figure 1.3: Top: Narcea thermal power plant, Spain. Bottom: View inside the boiler Unit II of the plant.

field tests were conducted in the plant's boiler unit II. These vast boiler chambers require planned and unplanned maintenance and repair. However, accessing them can be dangerous to humans and outage periods are extremely expensive. We had to undergo several days of safety training just to access the site.

For field testing, the visual-inertial sensor unit is mounted on a prototype unmanned aerial vehicle small enough to enter the boiler unit. The vehicle is fitted with two custom, high-power light-emitting diode (LED) modules that locally illuminate the boiler. The illumi-

## *Chapter 1. Introduction*

nation modules are synchronised with the camera shutters to avoid excessive power consumption and thermal problems. If unsynchronised, i.e. permanently on, power used for illumination would amount to a significant portion of the power required for hovering.

Prior to flight an operator defines a set of trajectories which the vehicle must follow. A filter based visual-inertial motion estimation framework [87] is employed to provide state estimates to the aerial vehicle. Based on these state estimates, the UAV navigates from way-point to way-point and returns to its take-off position.

The performance of the navigation system is evaluated through flight tests and using a Leica “Total Station” TS-15 that tracked a prism mounted on a hand-held version of the visual-inertial sensor unit. A detailed description of the field trials and the evaluation are presented in Paper IV.

### **1.3 Thesis Outline**

The remainder of this thesis is structured as follows: Chapter 2 presents the context and contribution of each paper included in this work. In addition, it provides a list of related publications and a list of graduate and undergraduate projects that contributed to this work. Chapter 3 gives an outlook on further research, engineering, and development which is required in this field from our perspective. Finally, the papers are included as Paper I to Paper IV at the end of this text.

# CHAPTER 2

## Contributions

THIS chapter summarises the contributions of this thesis. Section 2.1 puts the individual publications into context and highlights their role in achieving our objectives. The contribution of each paper is outlined in an overview for the reader. Publications which are not included but are related to this thesis are listed separately in Section 2.2. Many of our results were obtained in collaboration with graduate and undergraduate students. A list of corresponding technical reports that resulted from these projects can be found in Section 2.3 at the end of this chapter.

## 2.1 Summary of Publications

This thesis discusses topics in the design, characterisation, calibration, and application of visual-inertial sensor units. To this end, we include four publications in this text each addressing one of these topics. Rather than chronologically, the publications are included in a logical order. Paper I presents a method that can automatically identify noise model parameters for inertial sensors from data. Paper II then proposes a method for extrinsic and intrinsic calibration of visual-inertial sensor systems that uses these models. Paper III presents the design of a custom visual-inertial sensor unit which we used for our research, experimentation, field testing, and deployment on different robotic platforms. The last publication we include in this thesis, Paper IV, presents field tests we conducted with an early prototype of this visual-inertial sensor unit.

### Paper I

J. Nikolic, P. Furgale, A. Melzer, and R. Siegwart. “Maximum Likelihood Identification of Inertial Sensor Noise Model Parameters”. *IEEE Sensors* (2015).

#### *Context*

Localisation and mapping systems require accurate sensor noise models for motion estimation as well as for calibration. This is particularly important when information from different sensing modalities is combined in a tightly coupled estimator. If the noise models are incorrect, calibration and motion estimation will be inconsistent and the results suboptimal. Parameters for noise models are often not available and difficult to obtain for inertial sensors. For this reason, Paper I develops a method that can automatically obtain such models from measurements.

## 2.1 Summary of Publications

### *Contribution*

The contribution of this work is a method to determine noise model parameters for gyroscopes and accelerometers. In contrast to most existing procedures the method proposed in Paper I is automatic. The novelty of our method is its ability to provide inertial sensor noise model parameters which are optimal with respect to strict probabilistic criteria. The results show that the models obtained with our approach are similar to those which could be obtained with manual, Allan variance based methods by an expert in the field. This proves that our method is applicable in practice.

What makes the method unique is its ability to identify the model parameters in a direct maximum likelihood fashion. This leads to an estimator with good statistical properties. The paper presents a simple but crucial modification of the estimator's objective function that allows us to do that. In addition, the paper introduces a novel logarithmic sampling strategy to ensure that the model is valid across all time scales. Furthermore, it presents simulations showing that the proposed method works well also for extreme composite noise processes such as the ones found in modern MEMS gyroscopes and accelerometers. The simulations show that competing methods fail in these cases. Repeated experiments on different devices from different device classes verify the findings from the simulations and prove that the method works well in practice and on real sensor data.

## Paper II

J. Nikolic, M. Burri, I. Gilitschenski, J. Nieto, and R. Siegwart. “Non-Parametric Extrinsic and Intrinsic Calibration of Visual-Inertial Sensor Systems”. *Accepted for publication in IEEE Sensors* (2016).

## *Chapter 2. Contributions*

### *Context*

A prerequisite for precise visual-inertial motion estimation and mapping is accurate calibration. Unless the spatio-temporal alignment between the sensors is known, the performance of the sensors cannot be leveraged. The same holds true for sensor intrinsic calibration. Paper II addresses this by proposing a batch type estimator for the calibration of camera-IMU systems. The estimator relies on the method developed in Paper I for obtaining optimal inertial sensor noise model parameters. It then jointly estimates the relative translation and orientation between the sensors (i.e. the extrinsics), sensor intrinsic calibration parameters such as scale factors and cross-axes sensitivity, and the temporal alignment between the sensors.

### *Contribution*

Paper II presents an estimator for extrinsic, intrinsic, and temporal calibration of visual-inertial sensor systems. In contrast to existing calibration methods the estimator requires no tuning parameters. This eliminates the need for specifying and tuning parameters related to, for example, the sensor noise models. In contrast to established filtering approaches to calibration, we do not need to specify and tune initial state and parameter covariances or probabilistic models for the time evolution of calibration parameters which we know are fixed.

We treat calibration as a joint state and parameter estimation problem for a continuous-time system with discrete-time measurements. We then apply the maximum likelihood paradigm to estimate the state trajectory and the calibration parameters of the sensor unit in a batch fashion. This allows us to determine the calibration parameters in an almost optimal manner (in the maximum likelihood sense and up to linearisation errors).

Experimental results show that the estimator is well suited for the extrinsic and intrinsic calibration of visual-inertial sensor systems. It reaches a calibration precision of less than 1 mm in extrinsic

## 2.1 Summary of Publications

translation, less than 1 mrad in extrinsic orientation, and about 10 µs in time-shift. The results show that our relatively simple estimator is able to compete with modern parametric methods in terms of both accuracy and precision – without relying on a parametric representation of the state trajectory and the sensor biases. This insight is novel. Prior to our work, to the best of our knowledge, reaching such a high precision in time delay estimation with a non parametric method was not considered to be feasible.

### Paper III

J. Nikolic, J. Rehder, M. Burri, P. Gohl, S. Leutenegger, P. Furgale, and R. Siegwart. “A Synchronized Visual-Inertial Sensor System with FPGA Pre-Processing for Accurate Real-Time SLAM”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2014, pp. 431–437.

#### Context

Visual-inertial motion estimation and mapping solutions rely on sophisticated algorithms. These algorithms, in turn, can benefit from certain hardware features. Temporal alignment between camera and IMU, for example, can be estimated on-line and in software, or fixed in hardware. Some aspects are particularly difficult to solve with commodity, off the shelf hardware components. Precise temporal alignment, for example, cannot be achieved but only emulated in software.

Our field tests in Spain revealed the necessity for a fully synchronised, rigidly combined multi camera-IMU setup. Such setups are not readily available, and those solutions which do exist are closed source, which often makes them difficult to use in research projects. In addition, a standardised hardware setup would facilitate easy re-use of visual-inertial odometry or SLAM frameworks across projects and

## *Chapter 2. Contributions*

platforms. These aspects motivated the development of the visual-inertial sensor unit described in Paper III.

### *Contribution*

Paper III contributes with the design of a prototype visual-inertial sensor unit for motion estimation and mapping applications. The unit supports estimation frameworks which seek a tight combination of visual and inertial cues by providing high quality, synchronised sensor data. The novel hardware module is comprised of up to four CMOS, global shutter, low resolution “navigation cameras” and a combination of several MEMS gyroscopes and accelerometers. The sensors are carefully selected and address the needs of high performance visual-inertial motion estimation frameworks.

All sensors are interfaced through an industry standard ARM-FPGA combination. Their measurements are pre-processed and available at a high rate over a standard interface (Gigabit Ethernet). The unit is an ideal platform for prototyping visual-inertial estimation algorithms.

## **Paper IV**

J. Nikolic, M. Burri, J. Rehder, S. Leutenegger, C. Hürzeler, and R. Siegwart. “A UAV System for Inspection of Industrial Facilities”. In: *Proceedings of the IEEE Aerospace Conference*. 2013, pp. 1–8.

### *Context*

SLAM frameworks that rely on cameras and IMUs can be robust, accurate, and well suited for employment in real world applications. Paper IV highlights this by reporting on field tests conducted inside the boiler unit of a thermal power plant. The field tests were conducted in the framework of the Airobots project [57]. Our task was to assess the possibility to perform visual inspection tasks using an unmanned

## *2.1 Summary of Publications*

aerial inspection vehicle which enters the boiler unit and conducts a series of manoeuvres within close proximity to the boiler surface. If such a system could be deployed to localise broken pipework, for example, it would reduce the situations where humans have to enter these structures at a risk to their personal safety.

Installing external infrastructure such as laser trackers or optical tracking systems for navigation purposes is a difficult, time-consuming, and dangerous task in these environments – and is not an option. A system that carries its navigation solution on-board is therefore required.

Paper IV presents a prototype navigation module which was designed exactly for this purpose. It was these field tests that later led to the development of more advanced visual-inertial sensor systems and calibration routines. The paper presents detailed results and a critical assessment of the field tests.

### *Contribution*

Paper IV presents a micro unmanned aerial system (UAS) capable of navigating in enclosed industrial environments. The vehicle is fitted with a prototype visual-inertial sensor unit, including two global shutter navigation cameras and a custom inertial measurement unit. All sensors are hardware synchronised. In addition, active illumination in the form of high-power LED modules provides on-board illumination capabilities.

The vehicle then relies on a visual-inertial state estimator [87] in conjunction with a model predictive controller [12] to follow a pre-defined trajectory in close proximity with the boiler surface; the first experiment of its kind. The paper demonstrates that such semi autonomous operations of UAVs within this challenging environment are feasible, and show what a high level of navigation and flight performance can be reached.

## 2.2 List of Publications

This section provides a list of relevant papers that were published during this dissertation or are under review for publication. The first four publications are an integral part of this thesis and are included in their full length.

### Publications included in this thesis

- J. Nikolic, P. Furgale, A. Melzer, and R. Siegwart. “Maximum Likelihood Identification of Inertial Sensor Noise Model Parameters”. *IEEE Sensors* (2015).
- J. Nikolic, M. Burri, I. Gilitschenski, J. Nieto, and R. Siegwart. “Non-Parametric Extrinsic and Intrinsic Calibration of Visual-Inertial Sensor Systems”. *Accepted for publication in IEEE Sensors* (2016).
- J. Nikolic, J. Rehder, M. Burri, P. Gohl, S. Leutenegger, P. Furgale, and R. Siegwart. “A Synchronized Visual-Inertial Sensor System with FPGA Pre-Processing for Accurate Real-Time SLAM”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2014, pp. 431–437.
- J. Nikolic, M. Burri, J. Rehder, S. Leutenegger, C. Hürzeler, and R. Siegwart. “A UAV System for Inspection of Industrial Facilities”. In: *Proceedings of the IEEE Aerospace Conference*. 2013, pp. 1–8.

### Related publications

- M. Burri, J. Nikolic, C. Hürzeler, G. Caprari, and R. Siegwart. “Aerial Service Robots for Visual Inspection of Thermal Power Plant Boiler Systems”. In: *Proceedings of the International Con-*

## 2.2 List of Publications

ference on Applied Robotics for the Power Industry (CARPI). IEEE. 2012, pp. 70–75.

- R. Voigt, J. Nikolic, C. Hürzeler, S. Weiss, L. Kneip, and R. Siegwart. “Robust Embedded Egomotion Estimation”. In: *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*. 2011, pp. 2694–2699.
- C. Hürzeler, R. Naldi, V. Lippiello, R. Carloni, J. Nikolic, K. Alexis, L. Marconi, and R. Siegwart. “AIRobots: Innovative Aerial Service Robots for Remote Inspection by Contact”. In: *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*. 2013, pp. 2080–2080.
- M. Burri, J. Nikolic, H. Oleynikova, M. Achtelik, and R. Siegwart. “Maximum Likelihood Parameter Identification for MAVs”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2016.
- J. Rehder, J. Nikolic, T. Hinzmann, and R. Siegwart. “Extending Kalibr: Calibrating the Extrinsics of Multiple IMUs and of Individual Axes”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2016.
- T. Kazik, L. Kneip, J. Nikolic, M. Pollefeys, and R. Siegwart. “Real-Time 6D Stereo Visual Odometry with Non-Overlapping Fields of View”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012, pp. 1529–1536.
- F. Perodin, J. Nikolic, J. Busset, and R. Siegwart. “Design and Calibration of Large Microphone Arrays for Robotic Applications”. In: *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*. 2012, pp. 4596–4601.
- P. Gohl, M. Burri, S. Omari, J. Rehder, J. Nikolic, M. Achtelik, and R. Siegwart. “Towards Autonomous Mine Inspection”. In:

## *Chapter 2. Contributions*

*Proceedings of the International Conference on Applied Robotics for the Power Industry (CARPI).* IEEE. 2014, pp. 1–6.

- M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. Achtelik, and R. Siegwart. “The EuRoC MAV Datasets”. *The International Journal of Robotics Research IJRR* (2015).

## **2.3 Supervised Student Projects**

Many students contributed significantly to this research with their knowledge, engineering skills, and hard work. Below is a list of projects that students conducted in collaboration with the author sorted by the type of project. B.Sc. theses that were combined with a “Studies on Mechatronics” are listed under the bachelor theses.

### **Master Thesis**

*An M.Sc. thesis corresponds to a six months full time research project.*

- [S1] Burri, M. “Trajectory Control for Unmanned Aerial Inspection Vehicles Using Uncertain Measurements,” *ETH M.Sc. Thesis* (2011).
- [S2] Voigt, R. “Design of an Embedded Navigation Module for Aerial Inspection of Industrial Boilers,” *ETH M.Sc. Thesis* (2010).
- [S3] Muehlburger, B. “Real-Time People Tracking with Stereovision,” *ETH M.Sc. Thesis* (2012).
- [S4] Perrodin, F. “Design and Calibration of Large Microphone Arrays for Robotic Applications,” *ETH M.Sc. Thesis* (2011).
- [S5] Sarabadiani, A. “Robust, Efficient, and Accurate Visual-Inertial SLAM,” *ETH M.Sc. Thesis* (2011).

### 2.3 Supervised Student Projects

- [S6] Ammann, S. “Experimental Evaluation of Binocular Visual Pose Estimation Methods for UAV Localisation,” *ETH M.Sc. Thesis* (2010).
- [S7] Kazik, T. “Real-time 6D Visual Odometry with a Non-Overlapping Field of View Stereo Rig,” *ETH M.Sc. Thesis* (2011).

## Master Semester Project

*An M.Sc. semester project is a semester long part time research project.*

- [S8] Ruesch, A. “Dynamics Identification and Validation, and Position Control for a Quadrotor,” *ETH M.Sc. Semester Project* (2010).
- [S9] Laube S., and Gasser, L., “Design of the Electronic Systems for an Autonomous Blimp,” *ETH M.Sc. Semester Project* (2012).
- [S10] Krebs, C. “Generic IMU-Camera Calibration Algorithm,” *ETH M.Sc. Semester Project* (2012).
- [S11] Nisser, M. “Visual-Inertial Compassing,” *ETH M.Sc. Semester Project* (2014).
- [S12] Boosfeld, M. “EKF State Estimator for a Dual-Flight Envelope UAV,” *ETH M.Sc. Semester Project* (2014).
- [S13] Kleber, K. “Visual-Inertial Guidance System for a Robotic Arm,” *ETH M.Sc. Semester Project* (2015).
- [S14] Laube, S. “Multi-Target Pulsed LIDAR for Robotic Applications,” *ETH M.Sc. Semester Project* (2012).
- [S15] Schneeberger, J. “Sound Source Tracking using Acoustic Cameras,” *ETH M.Sc. Semester Project* (2015).
- [S16] Bleiker, T. “Electrical design for the TRALOC snake-bot,” *ETH M.Sc. Semester Project* (2011).

## *Chapter 2. Contributions*

### **Bachelor Thesis**

*A B.Sc. thesis is a semester long part time research project.*

- [S17] Wagner, L. “Inertial Sensor Testing, Characterisation and Calibration,” *ETH B.Sc. Thesis* (2014).
- [S18] Schmid, S. “3D Perception Module for a Ship Inspection Robot,” *ETH B.Sc. Thesis* (2013).
- [S19] Foerderer, N. “Acoustic Alarm Detection for Inspection of Oil and Gas Sites,” *ETH B.Sc. Thesis* (2015).
- [S20] Simpson, R. “State Estimation of a Wall Racing Robot,” *ETH B.Sc. Thesis* (2015).
- [S21] Mueller, T., and Mueller, R. “Design and Integration of a Printing Head for the Focus Project BeachBot,” *ETH B.Sc. Thesis* (2014).
- [S22] Brescianini, D. “Flying Like a Bat – Development of a Smart Obstacle Avoidance Sensor,” *ETH B.Sc. Thesis* (2010).
- [S23] Huessy, J. “State Estimation for a Ballbot,” *ETH B.Sc. Thesis* (2010).
- [S24] Zank, M. “Docking Mechanism for Contact Inspection with UAVs,” *ETH B.Sc. Thesis* (2010).
- [S25] Schaeppi, A., and Buholzer, P. “Attitude Estimation and Control of the Sepios Robot,” *ETH B.Sc. Thesis* (2014).

### **B.Sc. Studies on Mechatronics**

*A Studies on Mechatronics consists of a semester long literature study and review.*

### 2.3 Supervised Student Projects

- [S26] Muehlebach, M. “Camera Auto Exposure Control for VSLAM Applications,” *ETH Studies on Mechatronics* (2010).
- [S27] Renaut, F. “MEMS Inertial Sensor Technology,” *ETH Studies on Mechatronics* (2013).
- [S28] Faessler, M. “Force Sensing Technologies,” *ETH Studies on Mechatronics* (2010).
- [S29] Brescianini, D. “Ultra Wideband Radar for Micro Aerial Vehicles,” *ETH Studies on Mechatronics* (2010).



# CHAPTER 3

## Future Directions

**I**N this dissertation we presented the prototype of a novel visual-inertial sensor unit for robot localisation and mapping. We demonstrated the importance of sensor intrinsic calibration, particularly for device internal delays: if the timing is incorrect, calibration of other intrinsic parameters is futile. We developed an estimator which is able to determine the extrinsic, intrinsic, and timing parameters of visual-inertial sensor units without the need for specialised calibration equipment – arbitrary device motion in front of a visual target sufficed. The estimator, based on stringent probabilistic considerations, produced accurate calibration results and can compete with modern parametric methods. In addition, we presented an algorithm that can reliably determine noise model parameters for gyroscopes and accelerometers from data. This eliminates the need for manual

### *Chapter 3. Future Directions*

procedures that require expert know-how.

The field tests we conducted in a power plant demonstrated the performance that visual-inertial navigation systems can offer in terms of accuracy and robustness. The test results proved that combined visual-inertial sensor systems are well suited for robot navigation and mapping, and showed that such systems have a large potential in future applications.

The contributions of this thesis are a small step towards a robust, accurate, and reliable device for robot guidance and mapping in real, challenging applications. Several questions remain unanswered, and a number of improvements to the existing designs and algorithms are possible. To this end, the remainder of this chapter is structured as follows: future directions in the area of calibration and identification are outlined in Section 3.1. Improvements and new ideas for the design of visual-inertial sensor units are presented in Section 3.2.

## **3.1 Calibration**

Many topics in calibration and sensor modelling with application to visual-inertial SLAM have not been addressed adequately to date. We highlight some of these topics in this section.

### **Continuous self calibration**

Sensor calibration is only useful if the calibration is valid under the conditions encountered during the deployment of the sensor unit. An accurate calibration that was conducted under room temperature, for example, may well be inaccurate under the actual operating temperature. Furthermore, the calibration can suffer if the sensor unit is exposed to mechanical stress, shock, or vibration. In addition, sensor intrinsic calibration parameters may be subject to ageing. Calibration

### *3.1 Calibration*

may therefore change over time even in the absence of external influences. For large multi camera - IMU systems, such as cars and trucks, even stable extrinsic calibration parameters may not be guaranteed.

Some form of self calibration mechanisms are therefore unavoidable. A periodic re-calibration that requires infrastructure, even if it is just a visual calibration target, is likely not an option in many applications. Even infrastructure-less procedures, if they require some form of user interaction, can be impractical. Natural landmark based, automatic calibration algorithms that incorporate temperature related effects will therefore be required in the future.

## **Extended inertial sensor models**

The models we used in this thesis to describe visual-inertial sensor setups assume that the accelerometers have one common origin. The axes may be misaligned, but they are located in one single point. This should be extended to more generic configurations, which in turn requires a mean to estimate angular velocity and angular acceleration, since our non-parametric approach does not immediately give access to these quantities. One option could be to include angular velocities and angular acceleration into the state, for example with an angular acceleration random walk model.

Once the noise performance of consumer MEMS gyroscopes and accelerometers improves, modelling these sensors properly will become more and more important. In this thesis we used simple, generic models without any physical basis. Models from first principle may lead to better results, and this should be considered in future work. Also, lumping all dynamic effects of the inertial sensors into one single time delay parameter is suboptimal.

Furthermore, if the performance of consumer MEMS gyroscopes continues to improve, visual-inertial gyrocompassing might become feasible. This would allow one to align not only relative to the gravity vector, but also to true north – without relying on magnetometers.

## 3.2 Visual-Inertial Sensor Unit Design

The recent availability of vast computational resources on embedded platforms, miniature global shutter camera modules, and the ever increasing performance of MEMS inertial sensors is giving rise to small, affordable visual-inertial sensor platforms with integrated computational resources: “SLAM in a box”, one of the holy grails of robotics, is becoming a reality. A number of issues can be addressed in this context.

### Sensor selection

We elaborated in detail on the importance of precise time-synchronisation and sensor intrinsic modelling. However, the influence of sensor noise characteristics on navigation performance remains largely unexplored in the current literature. As soon as visual cues are incorporated, this question immediately becomes difficult to answer, and more research is needed in this domain to facilitate informed choices when it comes to sensor selection. The same holds true for the parameters of the visual subsystem: the influence of camera resolution, sensitivity, field-of-view, frame rate, and camera arrangement needs further investigation.

### Hardware design

The industry will certainly continue to miniaturise visual-inertial sensor systems, which will include computational resources, and reduce their cost. However, these solutions may or may not be suitable for field robotic applications. The current prototype of our visual-inertial sensor unit also lacks features that prevent its use for real field robotics. A proper housing should protect it from dust, humidity, and shock. An option to synchronise individual visual-inertial sensor units over longer distances would also be useful and eliminate the

### *3.2 Visual-Inertial Sensor Unit Design*

need for inter-module synchronisation in software. For applications that require a very high precision, integrating higher quality inertial sensors could also be considered.

Furthermore, for those use-cases which require active illumination because sufficient ambient light can not be guaranteed, an integrated solution for active illumination would be useful and reduce platform integration efforts. The active illumination should tightly integrate into the system, and the link between motion estimation, exposure control, and illumination should be closed. In addition, since we have full control over the hardware, sophisticated exposure control mechanisms that operate on the frames in an interleaved fashion (i.e. tracking different exposure set points on alternating frames) are conceivable. This could result in a form of “soft high dynamic range (HDR)” mode – possible since, with our setup, all elements throughout the visual-inertial SLAM chain have access and control over these settings.

Several applications not only require a sparse map and the robot pose within this map, but ultimately a high quality, coloured dense map or images (e.g. for inspection). The integration of additional camera chips which are tailored for this, rather than for navigation, would be useful.



# Paper I

## Maximum Likelihood Identification of Inertial Sensor Noise Model Parameters

Janosch Nikolic · Paul Furgale · Amir Melzer · Roland Siegwart

### Abstract

Accurate visual-inertial localization and mapping systems require accurate calibration and good sensor error models. To this end, we present a simple offline method to automatically determine the parameters of inertial sensor noise models. The proposed methodology identifies noise processes across a large range of strength and time-scales, for example weak gyroscope bias fluctuations buried in broadband noise. This is accomplished with a classical maximum likelihood estimator, based on the integrated process (i.e. the angle, velocity, or position), rather than on angular rate or acceleration as is standard in the literature. This trivial modification allows us to capture noise processes according to their effect on the integrated process, irrespective of their contribution to rate or acceleration noise. The cause of the noise is not discussed in this article.

## *Paper I*

The method is tested on different classes of sensors by automatically identifying the parameters of a standard inertial sensor noise model. The results are analysed qualitatively by comparing the model's Allan variance to the Allan variance computed directly from sensor data. A simulation that resembles one of the devices under test facilitates a quantitative analysis of the proposed estimator. Comparison with a competing, state-of-the-art method shows the advantages of the algorithm.

Published in *IEEE Sensors*, 2015.

©2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The final publication is available at <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7239526>, DOI 10.1109/-JSEN.2015.2476668.

# 1. Introduction

**I**NERTIAL sensors are employed in countless applications, ranging from consumer electronics to autonomous vehicles and unmanned aerial systems. Many different types of sensors are used, from low-cost, multi-axis microelectromechanical systems (MEMS) devices with a footprint below  $20\text{ mm}^2$ , to ring laser gyroscopes (RLGs). In order to understand the characteristics of these sensors, both stochastic (“noise”) and deterministic (sensitivity, axes misalignment, etc.) errors have to be considered.

When inertial measurements are combined with data from other sensors, such as precise landmark observations from a camera system

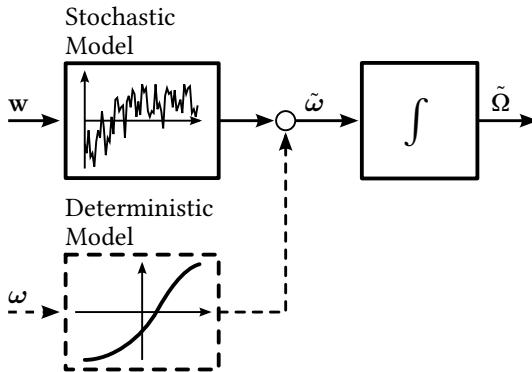


Figure 1.: Inertial sensor model with a deterministic and a random component. Here, the true angular rates  $\omega$  are corrupted with deterministic errors, for example a scale factor that varies with temperature, as well as non-deterministic errors, such as additive broadband noise. This report presents a method to identify noise processes according to their contribution to the angular increments  $\Omega$ .

for example, an accurate model for the different sources of error in the inertial sensor data is vital. Such a model facilitates the optimal design of an estimator, and allows us to verify the proper operation of all software and hardware components.

A good sensor model often incorporates stochastic as well as deterministic components, as illustrated in Fig. 1. The stochastic part of the model includes errors such as broadband noise, or a slowly (randomly) varying bias, and describes them in a *probabilistic* sense. These errors can not be calibrated for, and limit the performance of the device fundamentally.

The instrumental noise figures are usually identified under static mechanical (i.e. non moving) and thermal conditions [4]. The stochastic model then provides an *upper bound* on the sensor performance, and crucial parameters for an estimator (i.e. noise densities).

This work focuses on how to identify such noise models automatically (for low-cost devices), since deriving a good noise model for an inertial sensor can be a difficult task for a non-expert. However, we also emphasise the importance of a good deterministic sensor model, since a stochastic model alone usually leads to over-confidence in the inertial data, especially when using uncalibrated sensors.

## 1.1 Existing Methods

Section 2 reviews the most commonly used stochastic models for inertial sensors. The Allan variance (AV) [3] and the non-parametric power spectral density (PSD), the classical tools to describe noise in inertial sensors, are revisited.

In practice, the noise model parameters are usually determined manually by “fitting” noise processes to estimates of the AV or the PSD. A detailed description of these procedures is provided, for example, in the Appendix of the IEEE standard 952-1997 [38]. These procedures are reliable, particularly those based on the Allan variance. Computing the sample based AV is standardized, and by the nature of the AV, the

## 1. Introduction

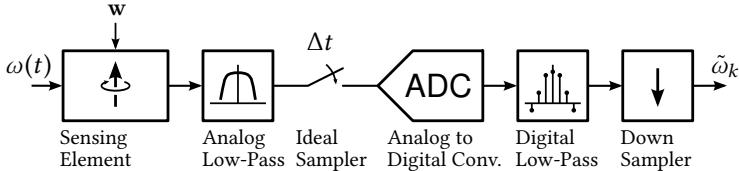


Figure 2.: Conceptual diagram of the data acquisition and pre-processing steps in inertial sensors. The sensing element captures the quantity of interest, here angular rate  $\omega(t)$ . Sensing electronics, often followed by an anti-aliasing mechanism, capture and condition the signal from the sensing element before it is sampled and digitized. The discrete-time rate or acceleration is then filtered with a digital low-pass filter and downsampled, *reducing noise* and introducing *delay*.

fitted noise models will be valid across the desired noise strength- and time-scales. The technical note [92], which is critical about applying classical automatic noise modelling techniques to inertial sensor data, presents convincing arguments for using this manual “AV method”.

However, these manual procedures require expertise to map the sample-based AV to parameters of a model which is useful for sensor fusion algorithms. The “Kalibr” toolbox<sup>1</sup> for spatial and temporal calibration of visual-inertial systems [25], for example, requires the user to specify the parameters of a particular inertial sensor noise model. An automatic method could eliminate the need for a manual identification of these noise model parameters.

Several methods were proposed for automatic identification of

---

<sup>1</sup> *Kalibr* is available at [www.github.com/ethz-asl/kalibr](http://www.github.com/ethz-asl/kalibr) (2016).

inertial sensor noise model parameters in the literature. The approach in [85] address the problem with a method that computes the theoretical covariance of the AV (given a model hypothesis), and then uses an optimal linear estimator to derive the parameters of a model whose AV matches the AV computed from measurement data. The method is intuitive and inherits the desirable properties of the manual AV method. It was developed in [85] for a model that incorporates white noise and a random walk process, and was tested on MEMS gyroscope and accelerometer data. In contrast to this method, the algorithm we propose does not aim to identify model parameters such that the model's AV resembles the sample-based AV of the observed measurements.

The “Generalized Method of Wavelet Moments” (GMWM), proposed in [29], can be viewed as a generalization of this approach. Instead of the Allan variance, it uses a wavelet basis (“wavelet variance”), in which the sensor measurements are decomposed across different time-scales. Model parameters are then identified such that the model's wavelet variance optimally matches the wavelet variance computed from the observed measurements. This state-of-the-art method was specifically developed for the identification of various noise processes in inertial sensors, and applied extensively for this purpose in [78]. An implementation of the method was provided by its developers, and we compare the performance of the method proposed in this article with the GMWM in Section 5.

[60] presents an innovation-based adaptive estimator for INS/GPS applications. A Kalman filter estimates the system state, while a maximum likelihood method uses a block of data to *independently* update the process and measurement covariance matrices. This results in a significant performance increase of their INS, and the method is suitable for on-line applications. In contrast, our method provides a full maximum likelihood estimate of the noise model state and its parameters, and is intended for offline usage.

## 1.2 Automatic Noise Model Parameter Identification

Section 3 develops a conceptually simple method to automatically identify the parameters of an inertial sensor noise model. A small change in the identification problem allows us to successfully apply *classical maximum likelihood* estimation to identify the parameters of the noise model. The method performs well for composite noise processes that span many strength- and time scales, such as weak MEMS gyroscope bias fluctuations buried in broadband noise.

In contrast to existing algorithms, the estimator we present works on the *integrated process* (i.e. the angle or velocity), rather than on the rate or acceleration measurements directly. The integrated process is an entirely different process, but arguably more relevant in sensor fusion applications, where information is often fused at an attitude, velocity, or position level. Noise processes which contribute to rate or acceleration noise only marginally may have a drastic effect on the integrated process, and vice versa. The proposed method exploits this by identifying the noise model parameters such that the resulting noise model is accurate for the integrated process, and captures noise processes even if their contribution to rate or acceleration noise is negligible. The algorithm is therefore called the “integrated maximum likelihood estimator”,  $\text{ML}_i$ . Model selection is not discussed.

## 1.3 Deterministic Errors and Calibration

The deterministic part of the inertial sensor model captures non-random effects such as scale factor errors, axes misalignment, or frequency response. These non-idealities can, to some extent, be calibrated for, which often requires a per-device (factory) calibration across temperature, or on-line calibration mechanisms [55]. Different deterministic models of varying complexity are used in practice, depending on the type of sensor and application requirements. A

simple model is used in Section 5 to briefly quantify deterministic error sources. A dedicated experiment with a visual-inertial sensor setup highlights the uncomfortable fact that, particularly for uncalibrated low-cost devices, a sensor error model which is based on static sensor data alone, captured at constant temperature, does not capture inertial measurement errors well under *realistic operating conditions* (i.e. when the sensor undergoes motion, or the temperature varies).

For accurate visual-inertial localization and mapping systems, precise time-synchronization of cameras and inertial sensors is essential. Both software and hardware mechanisms exist to achieve this. We highlight the phase delay that a sensor internal decimation stage, a common feature in MEMS inertial sensors, can introduce. This is often overlooked in practice, but becomes apparent immediately when considering the entire “data acquisition process”, illustrated in Fig. 2. Sophisticated software-synchronization methods [25, 54] will automatically compensate for this, but hardware-synchronized systems will *not*.

## 1.4 Experiments

Section 5 tests the proposed method on sensor data from different device-classes, and verifies the identified noise models qualitatively using the classical Allan variance and the PSD as a reference. Additional tests show that the method is repeatable, and reliably captures the relevant noise processes. The accuracy of the proposed method is analysed quantitatively, using a simulation that resembles one of the devices under test. The method is also compared to a competing, state-of-the art method (GMWM).

We show the limitations of purely stochastic sensor models, whose parameters were identified from static sensor data alone, in a dedicated experiment with a visual-inertial sensor system.

## 2. Stochastic Sensor Models

The stochastic part of the sensor model captures errors which are considered random, and describes them in a probabilistic sense. Good examples of random errors are wideband “electronic noise”, originating from transducer and amplifier stages, or fluctuations in drive frequency that result in slowly varying, random gyroscope bias errors in MEMS devices [47]. Note that our aim here is not to link the noise model to a physical model of the sensor, but merely to derive generic, simple noise models which are suitable for *sensor fusion applications*, based on measurements.

This section summarizes the most important concepts in the description and analysis of noise processes in inertial sensor measurements, as far as they are relevant for the derivations that follow. The reader is referred to [58] for an excellent in-depth treatment of stochastic processes in general, and to [38] for a detailed collection of random processes that are useful to model noise in inertial sensors.

### 2.1 Spectral Density and Allan Variance

At first, we restrict our attention to zero-mean, stationary Gaussian processes. This excludes, for example, quantization and flicker noise [48], for the benefit of tractability. We denote the random process that describes the sensor noise as  $z$ . Depending on the context,  $z$  refers to the random process, or to a particular realization (a sample path) of it.

The correlation of the process  $z$  is defined as

$$\psi_{zz}(t_1, t_2) = E[z(t_1)z(t_2)] = \psi_{zz}(\tau) \quad (1)$$

where  $E[\dots]$  denotes expectation. Since the process is Gaussian,  $\psi_{zz}$  fully characterizes it, and due to the processes stationarity,  $\psi_{zz}$  is a function of the time difference  $\tau = t_1 - t_2$  only. The variance (power)

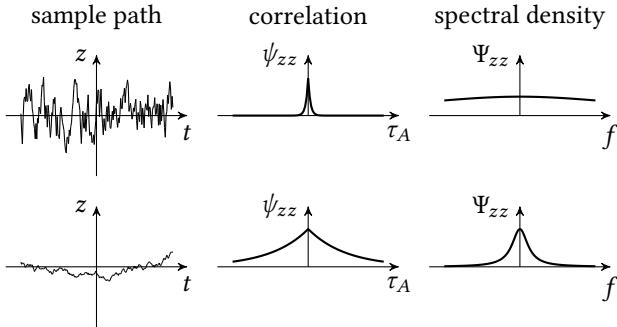


Figure 3.: Correlation and spectral density: The top row illustrates a sample path, the correlation, and the power spectral density of a rapidly fluctuating noise process (e.g. broadband noise in a gyroscope). Such a process has a peaked correlation, and power spread across a large frequency range. The bottom row depicts a slowly varying noise process, such as sensor bias variation, with opposite correlation and spectral characteristics.

of the process is given by  $\psi_{zz}(0)$  [39]. A noise process that fluctuates rapidly, such as wideband “electronic noise”, has a “peaked” correlation (i.e. a short correlation time). A process that varies slowly, for example a gyroscope bias drift, has a “broad” correlation. Figure 3 illustrates this.

### 2.1.1 Power Spectral Density

The power spectral density is a standard tool for the description of stationary stochastic processes, and it is commonly used to characterize noise processes in gyroscopes and accelerometers [4, 38]. The power spectral density is a standard tool for the description of stationary stochastic processes, and it is commonly used to characterize noise processes in gyroscopes and accelerometers [4, 38]. The PSD

## 2. Stochastic Sensor Models

of a process is its correlation, expressed in the frequency domain (Wiener-Khinchin):

$$\Psi_{zz}(f) = \int_{-\infty}^{\infty} \psi_z(\tau) e^{-i2\pi f \tau} d\tau \quad (2a)$$

$$\psi_{zz}(\tau) = \int_{-\infty}^{\infty} \Psi_{zz}(f) e^{i2\pi f \tau} df. \quad (2b)$$

Similar to the Fourier transform of a deterministic signal, it expresses (noise) power per frequency band (therefore the name power spectral *density*, and the unit “per Hz”). It is an intuitive quantity. A rapidly fluctuating noise process with a “peaked” correlation has a broad power spectrum (i.e. the noise power is spread across a large frequency range), whereas the power of a slowly varying process is concentrated around the lower frequencies, see Fig. 3. If the noise model and its parameters are known, the correlation and the PSD can be computed using (1) and (2).

A number of *non-parametric* methods exist to compute the PSD from sampled data. A classical example is Welch’s method [90]. A good example for a parametric method is the identification of an autoregressive (AR) model by solving the Yule-Walker equations, based on sample covariance matrices.

Such classical methods struggle to capture weak processes with long correlation times, when they are buried deeply in broadband noise processes – a typical case in MEMS devices (see [92, 91] and Section 4.3). The Allan variance is a tool that is particularly well suited to handle exactly such cases.

### 2.1.2 Allan Variance

The Allan variance [3], originally developed to characterize the frequency stability of oscillators, addresses these issues and is an important tool in practice. Many manufacturers of inertial sensors include it in the device datasheet (see e.g. [16]), and engineers deduce key

noise parameters from it. It is often simply denoted as  $\sigma_z^2(\tau_A)$ . The AV (or its square-root, the Allan *deviation*) is a time-domain method, intuitive to interpret, and in contrast to the PSD there is a standard procedure to compute it from sampled data (with small variations). The Allan variance is briefly reviewed here, since it is an important tool for characterizing noise in inertial sensors, and since it is used to qualitatively assess our method in Section 5. The reader is referred to [38] and the references therein for a more detailed treatment.

The AV of a random process  $z$ ,  $AV_z(\tau_A)$ , is a function of the integration time  $\tau_A$ . In words, it describes how much the integrated process “varies” as a function of the time-difference  $\tau_A$ . Formally, the AV is computed as follows:  $z$  is integrated over blocks of length  $\tau_A$ . The mean of the variances of successive blocks is then the Allan variance for this particular integration time  $\tau_A$  [3]. This can be written concisely as follows:

$$\bar{z}_k(\tau_A) = \frac{1}{\tau_A} \int_{k\tau_A}^{(k+1)\tau_A} z(t) dt \quad (3a)$$

$$AV_z(\tau_A) = \frac{1}{2} E \left[ (\bar{z}_{k+1}(\tau_A) - \bar{z}_k(\tau_A))^2 \right]. \quad (3b)$$

Heuristically, the AV of a white noise process decreases with increasing integration time  $\tau_A$ , since the means  $\bar{z}_k$  are estimated more accurately – and therefore the variance between successive estimates decreases. Correlated processes result in distinct Allan variances, depending on the properties of the process. A comprehensive list of noise processes and their corresponding AV can be found in [38].

The AV can be determined from discrete samples of  $z$ , i.e. angular rate or acceleration measurements, by computing the sample mean of each segment  $\bar{z}_k(\tau_A)$ , and the sample variance of successive segments – exactly the procedure used in this work. The AV can also be deduced from the spectral density of a process, using the following relationship

## 2. Stochastic Sensor Models

from [38] (Appendix C.1):

$$AV_z(\tau_A) = 4 \int_0^\infty \Psi_{zz}(f) \frac{\sin^4(\pi f \tau_A)}{(\pi f \tau_A)^2} df. \quad (4)$$

Examples are given in the next section.

### 2.2 The Standard Noise Model

We now describe one of the most commonly used inertial sensor noise models – the same model whose parameters we identify in Section 5. The model combines “white noise”, capturing rapidly fluctuating noise processes, with an *exponentially time-correlated* process that captures slowly varying noise processes (biases). In engineering notation, we write:

$$z = x + v \quad (5a)$$

$$\dot{x} = -\frac{1}{\tau_b}x + w \quad (5b)$$

where  $z$  is the modelled noise process, corrupting rate and acceleration measurements.  $v$  represents the white noise component of the model.  $x$  is the slowly varying (bias) process with correlation time  $\tau_b$ , “driven” by the white noise process  $w$ .  $w$  and  $v$  are independent, zero mean, white Gaussian noise processes of strength  $\sigma_b$  and  $\sigma_w$ . Formally:

$$E[w(t_1)w(t_2)] = \sigma_b^2 \delta(t_1 - t_2) \quad (6a)$$

$$E[v(t_1)v(t_2)] = \sigma_w^2 \delta(t_1 - t_2) \quad (6b)$$

where  $\delta(\cdot)$  denotes the Dirac delta function.

The model is popular due to its relatively large expressive power despite the small number of parameters  $\theta = \{\sigma_w, \sigma_b, \tau_b\}$ . It can be incorporated into state estimation frameworks efficiently due to its

Gauss-Markov properties. It is for example used to model gyroscopes and accelerometers in a GPS aided INS [17], or in the context of a sun-sensor aided EKF attitude estimator in [84] (modelling noise in the gyroscopes, although with  $\frac{1}{\tau_b} = 0$ , i.e. a “pure” random walk bias model). [32, 52, 55] employ it in the context of visual-inertial simultaneous localization and mapping (SLAM).

### 2.2.1 Terminology

$\sigma_w$ , the strength of the rate or acceleration white noise process, is often termed “angular/velocity random walk” (ARW/VRW). IEEE standard [38] denotes  $\sigma_w$  as  $N$ , and device datasheets often refer to it simply as (rate or acceleration) noise density (see e.g. [16]). Sometimes, the noise density that corresponds to the single-sided PSD is reported (twice the double-sided density). To be sure to interpret the indicated values correctly, measuring is often unavoidable.

$\sigma_b$ , the strength of the white noise process  $w$  that drives the bias fluctuations, is termed “bias diffusion” in this report. [38] uses  $K$  for this quantity when  $\frac{1}{\tau_b} = 0$ , and  $q_c$  otherwise. This is not equivalent to the “bias (in)stability”, which is used to parametrize flicker noise [37].  $\sigma_b$  is rarely reported in device datasheets. Table 1 summarizes the terminology.

### 2.2.2 PSD and AV of the Standard Noise Model

The PSD of the noise process  $z$  is the superposition of the PSD of the white noise process  $v$  and the time-correlated bias process  $x$ . The PSD of  $v$  is a constant,

$$\Psi_{vv}(f) = \sigma_w^2 \quad (7)$$

and the Allan variance of  $v$  is:

$$AV_v(\tau_A) = \frac{\sigma_w^2}{\tau_A} \quad (8)$$

## 2. Stochastic Sensor Models

Table 1.: Parameters of the standard inertial sensor noise model (5): terminology and units.

Param.	Description	Units
$\sigma_w$	(rate or acceleration) noise density, white noise density, angular or velocity random walk (ARW/VRW)	rad/(s $\sqrt{\text{Hz}}$ ) m/(s <sup>2</sup> $\sqrt{\text{Hz}}$ )
$\sigma_b$	bias diffusion, bias noise, random walk	rad/(s <sup>2</sup> $\sqrt{\text{Hz}}$ ) m/(s <sup>3</sup> $\sqrt{\text{Hz}}$ )
$\tau_b$	bias correlation time	s

The PSD of the exponentially correlated bias process  $x$  (5b) can be found as the Fourier-transform of its autocorrelation (2). To find the autocorrelation, we note down the state-transition of the bias model:

$$\Phi(t + \tau, t) = \exp\left(-\frac{1}{\tau_b}|\tau|\right). \quad (9)$$

We can now formally write

$$\begin{aligned} \psi_{xx}(\tau) &= E[x(t + \tau)x(t)] \\ &= E\left[\left(\Phi(t + \tau, t)x(t) + \int_t^{t+\tau} \Phi(t + \tau, s)w(s)ds\right)x(t)\right] \\ &= \Phi(t + \tau, t)E[x(t)x(t)] \\ &= \underbrace{\exp\left(-\frac{1}{\tau_b}|\tau|\right)}_{\Phi(t + \tau, t)} \underbrace{\frac{\sigma_b^2 \tau_b}{2}}_{\psi_{xx}(0)}. \end{aligned} \quad (10)$$

$\psi_{xx}(0)$  is the expected steady-state variance of the process. Hence,

$$\Psi_{xx}(f) = \frac{\sigma_b^2 \tau_b^2}{(2\pi \tau_b f)^2 + 1}. \quad (11)$$

For long correlation times,  $\lim_{\tau_b \rightarrow \infty} \Psi_{xx}(f) = \frac{\sigma_b^2}{(2\pi f)^2}$ . The process is in this case referred to as “Brownian motion”, a “Wiener process”, or often simply as an (unbounded) random walk.

The AV of  $x$  is ([38], Table B.1):

$$AV_x(\tau_A) = \frac{\sigma_w^2 \tau_b^2}{\tau_A} \times \left[ 1 - \frac{\tau_b}{2\tau_A} \left( 3 - 4 \exp\left(-\frac{\tau_A}{\tau_b}\right) + \exp\left(-\frac{2\tau_A}{\tau_b}\right) \right) \right]. \quad (12)$$

For  $\tau_A \ll \tau_b$ ,  $AV_x$  is similar to the the AV of a random walk, and

$$AV_x(\tau_A) \approx \frac{\sigma_b^2}{3} \tau_A. \quad (13)$$

Figure 4 shows the power spectral density and the Allan deviation for the standard noise model, in log-log scale.

### 2.2.3 Discretization

We briefly discuss the discretization of the standard noise model (5), since it involves a simple, but crucial assumption. We first determine the discrete-time equivalent of the bias process (5b), i.e. a discrete-time model which has exactly the same properties as the continuous-time model at the sampling instances  $k\Delta t$ . In continuous-time, the bias process formally evolves from  $(k-1)\Delta t$  to  $k\Delta t$  as follows:

$$x(k\Delta t) = \Phi(\Delta t)x((k-1)\Delta t) + \int_{(k-1)\Delta t}^{k\Delta t} \Phi(k\Delta t, s)w(s)ds \quad (14)$$

## 2. Stochastic Sensor Models

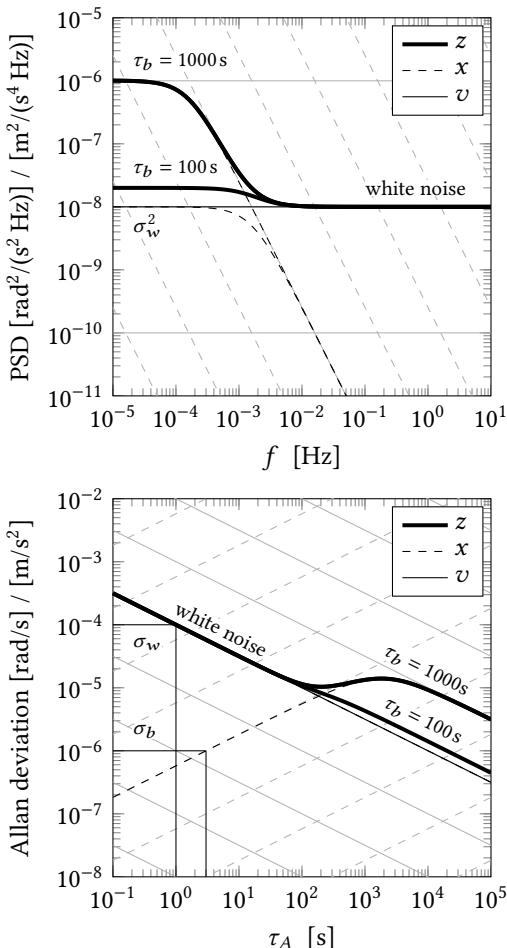


Figure 4.: Power spectral density (top) and Allan deviation (bottom) of the standard noise model (5). Under certain conditions, the model parameters can be read off the sample based PSD and AV plots directly, as explained in Section 2.3.

with  $\Phi(\Delta t) = \Phi(k\Delta t, (k-1)\Delta t)$ . A discrete-time equivalent is then found as:

$$x_k = \exp\left(-\frac{1}{\tau_b}\Delta t\right)x_{k-1} + w_k \quad (15a)$$

$$\text{E}[w_k w_k] = -\frac{\sigma_b^2 \tau_b}{2} \left[ \exp\left(-\frac{2\Delta t}{\tau_b}\right) - 1 \right] = \sigma_{bd}^2 \quad (15b)$$

where  $w_k$  is a zero-mean, discrete-time white Gaussian noise process of strength  $\sigma_{bd}^2$ . If the bias correlation time is large compared to the sampling time ( $\tau_b \gg \Delta t$ ),  $\sigma_{bd}^2 \approx \Delta t \sigma_b^2$ .

Discretization of the formal, infinite variance continuous-time “white noise” component  $v$  of the standard model requires more care. If one argues that  $v$  is ideally band-limited to  $\frac{1}{2\Delta t}$  before sampling, then

$$\text{E}[v_k v_k] = \frac{1}{\Delta t} \sigma_v^2 = \sigma_{wd}^2 \quad (16)$$

and

$$z_k = x_k + v_k \quad (17)$$

where  $v_k$  is discrete-time, zero mean white Gaussian noise of strength  $\sigma_{wd}^2$ . The argument that  $v$  is integrated in-between samples would lead to the same result. If the measurements are simply subsampled without prior filtering, scaling  $\sigma_w^2$  with  $\frac{1}{\Delta t}$  is *not correct*. In other words, what should determine  $\sigma_{wd}^2$  is the *bandwidth* of  $v$  (determined by a correctly configured device-internal decimation stage, see Fig. 2), and not the sampling rate. Algorithm 1 summarizes the discrete-time equivalent implementation of the standard noise model.  $w \leftarrow \mathcal{N}(0, \sigma^2)$  denotes a zero-mean, normally distributed random variable with variance  $\sigma^2$ .

---

**Algorithm 1** Discrete-time equivalent of the std. noise model (5).

---

**init:**

$$\begin{aligned}\sigma_{wd}^2 &\leftarrow \frac{1}{\Delta t} \sigma_w^2 && \triangleright \text{assuming } v \text{ is band-limited to } \frac{1}{2\Delta t} \\ \sigma_{bd}^2 &\leftarrow \Delta t \sigma_b^2 && \triangleright \text{assuming } \tau_b \gg \Delta t \\ \Phi_d &\leftarrow \exp\left(-\frac{1}{\tau_b} \Delta t\right) \\ x_0 &\leftarrow \begin{cases} 0 & \text{if } \frac{1}{\tau_b} = 0 \text{ (by definition)} \\ \mathcal{N}\left(0, \frac{\sigma_b^2 \tau_b}{2}\right) & \text{otherwise} \end{cases}\end{aligned}$$

**for**  $k \leftarrow 1$  **to**  $n$  **do**

$$\begin{aligned}w_k &\leftarrow \mathcal{N}\left(0, \sigma_{bd}^2\right), v_k \leftarrow \mathcal{N}\left(0, \sigma_{wd}^2\right) \\ x_k &\leftarrow \Phi_d x_{k-1} + w_k \\ z_k &\leftarrow x_k + v_k\end{aligned}$$

**end for**

## 2.3 Manual Identification

The parameters of the standard noise model (5) can be directly read off of an Allan deviation plot under two conditions: a.) the noise can be approximated with the standard model, and b.) the noise is dominated by white noise at  $\tau_A = 1$  s. If these conditions are met,  $\sigma_w$  corresponds approximately to the Allan deviation at  $\tau_A = 1$  s (see (8)).  $\sigma_b$  can be read off at  $\tau_A = 3$  s, as is clear from (13), and illustrated in Fig. 4. [21] describes manual procedures to identify other common noise processes, such as flicker noise (bias instability). The advantage of such manual procedures is that one can identify the characteristics of the noise process, and immediately see if the model fits the data well in terms of the Allan deviation.

### 3. Automatic Model Identification

Obtaining the noise model parameters from the Allan variance or the PSD involves manually “fitting” graphs that correspond to known noise processes to the AV or the PSD that was computed from sensor data. This requires expertise. A method which can automatically identify the noise statistics will require less know-how, is less prone to error, and provides model parameters which are optimal with respect to probabilistic criteria.

The concept of *maximum likelihood* (ML) is well suited for problems that involve joint state, model, and noise statistics estimation, and is adopted here. ML estimators are asymptotically unbiased and attain the Cramér-Rao lower bound, even for problems that include the estimation of noise densities (discussed in detail in [68]).

#### 3.1 The Integrated Noise Process

We propose one simple but crucial modification to the parameter identification problem. This modification allows us to apply classical maximum likelihood to identify typical noise processes in inertial sensors. Instead of estimating the parameters such that the model explains noise in the measurements directly (angular rate or acceleration), we identify the parameters such that noise in the *integrated* process is explained. This allows us to identify composite noise processes that span large strength- and time-scales, such as weak bias fluctuations “buried” deeply in broadband noise (a common case in MEMS inertial sensors, see [92] for a critical review of existing automatic methods). While some noise processes contribute to noise power in acceleration or rate only marginally, they can have a drastic effect on the - arguable more relevant - integrated process (i.e. the velocity, position, or angle).

### 3. Automatic Model Identification

## 3.2 Augmented Noise Model

To identify the integrated noise process, the noise model is augmented with an integrator. This increases the dimension of the model by one, but only for identification purposes. The final noise model will have the desired structure again.

We will derive the likelihood function using a continuous-time formulation of the noise model, which allows us to think about the diffusion and the correlation of the bias processes in a natural way; in terms of noise power per unit of frequency and in seconds. Our attention is restricted to linear Gaussian noise models. We formally define the *augmented noise model* as follows:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\boldsymbol{\theta})\mathbf{x}(t) + \mathbf{G}\mathbf{w}(t) \quad (18a)$$

$$\bar{z}(t_i) = \mathbf{H}\mathbf{x}(t_i). \quad (18b)$$

$\bar{z}(t_i)$  denotes the integrated measurement process (sampled at time  $t_i$ ),  $\mathbf{x}$  denotes the state of the model,  $\mathbf{F}$  is the continuous-time “system matrix” (the desired noise model, augmented with an integrator), and  $\mathbf{w}$  is a continuous-time Gaussian “white noise” process of strength  $\mathbf{Q}$ ,  $E[\mathbf{w}(t_i)\mathbf{w}^T(t_j)] = \mathbf{Q}\delta(t_i - t_j)$ .  $\mathbf{H}$  denotes the 1-by- $p$  “measurement matrix”, and  $\mathbf{G}$  maps the driving noise  $\mathbf{w}$  on the state  $\mathbf{x}$ . The parameters of the model are summarized in  $\boldsymbol{\theta}$  and include, in particular, also the noise densities  $\mathbf{Q}$ . Note that the standard noise model (5) can be cast into this form.

## 3.3 Maximum Likelihood Estimator

If the model parameters  $\boldsymbol{\theta}$  were known, estimating the state trajectory  $\mathbf{X}_m = \{\mathbf{x}(t_0), \mathbf{x}(t_1), \dots, \mathbf{x}(t_i)\}$  could be done efficiently due to the independence and the Gaussian assumption on  $\mathbf{w}$ . A Kalman filter [42] would in this case provide a sub-optimal solution for  $\mathbf{X}_m$ , and a smoother would give the optimal solution, given all measurements.

Jointly estimating the states  $\mathbf{X}_m$  and the parameters  $\boldsymbol{\theta}$  is more complex, but a known problem as well. [27] and [76] apply the method of Expectation Maximization (EM) to parameter estimation in general linear systems. This involves a forward-backward Kalman smoothing step to obtain estimates for  $\mathbf{X}_m$  and its covariances, followed by a simple update of the model parameters. The two steps are then iterated.

However, this method is not applicable here, since we usually require the model to have a specific form (e.g.  $\mathbf{F}$  has a specific structure, or  $\mathbf{Q}$  is diagonal, etc.). We therefore resort to the classical “innovations likelihood” formulation as used by, for example, [30].

Due to the Markov and independence properties of the model, we can write the likelihood function in innovations form as

$$p(\bar{\mathbf{Z}}_m | \hat{\boldsymbol{\theta}}) = \prod_{i=0}^m p(\bar{z}(t_i) | \bar{\mathbf{Z}}_{t_{i-1}}, \hat{\boldsymbol{\theta}}) \quad (19)$$

where  $\bar{z}(t_i)$  denotes the value of the integrated noise process at time  $t_i$ , and

$$\bar{\mathbf{Z}}_{t_i} = \{\bar{z}(t_0), \bar{z}(t_1), \dots, \bar{z}(t_i)\}.$$

As we will see later,  $\bar{z}$  need not be uniformly sampled in time, hence  $\Delta t_i = t_i - t_{i-1}$  is not necessarily constant. For the ML<sub>i</sub> log method, the measurement sequence is divided into  $m$  logarithmically increasing time-steps such that the entire sequence of  $n$  measurements is spanned.

Evaluation of (19) requires an expression for the conditional expectation of the mean and the covariance of each  $\bar{z}(t_i)$ . Given the model parameters  $\boldsymbol{\theta}$ , they can be computed exactly. From time  $t_{i-1}$  to  $t_i$ ,  $\bar{z}(t)$  formally evolves as follows:

$$\begin{aligned} \bar{z}(t_i) &= \mathbf{H} \times \\ &\left[ \Phi(t_i, t_{i-1}) (\hat{\mathbf{x}}(t_{i-1}) - \Delta \mathbf{x}(t_{i-1})) + \int_{t_{i-1}}^{t_i} \Phi(t_i, s) \mathbf{w}(s) ds \right] \end{aligned} \quad (20)$$

### 3. Automatic Model Identification

where

$$\Phi(t_i, t_{i-1}) = \exp(\mathbf{F}\Delta t_i) \quad (21)$$

is the state transition matrix, and  $\Delta\mathbf{x}(t_{i-1}) = \mathbf{E}[\mathbf{x}(t_{i-1})|\bar{\mathbf{Z}}_{t_{i-1}}] - \mathbf{x}(t_{i-1})$  denotes the difference between the true state  $\mathbf{x}$  and our estimate  $\hat{\mathbf{x}}$  at time  $t_{i-1}$ .

Since the expectation of the stochastic integral in (20) is zero, the expected value of  $\bar{z}(t_i)$  is

$$\hat{z}_i = \mathbf{E}[\bar{z}(t_i)|\bar{\mathbf{Z}}_{t_{i-1}}, \hat{\theta}] = \mathbf{H}\Phi(\Delta t_i)\hat{\mathbf{x}}(t_{i-1}). \quad (22)$$

For the covariance, we find

$$\begin{aligned} \mathbf{E}[\Delta\bar{z}(t_i)] &= \mathbf{H}\Phi(\Delta t_i)\mathbf{E}[\Delta\mathbf{x}(t_{i-1})\Delta\mathbf{x}^T(t_{i-1})]\Phi^T(\Delta t_i)\mathbf{H}^T + \\ &\quad \mathbf{H}\int_{t_{i-1}}^{t_i} \Phi(t_i, s)\mathbf{E}[\mathbf{w}(s)\mathbf{w}^T(s)]\Phi^T(t_i, s)ds\mathbf{H}^T \\ &= \mathbf{H}(\mathbf{P}_{t_{i-1}|t_{i-1}} + \mathbf{Q}_d(\Delta t_i))\mathbf{H}^T \\ &= B(t_i) \end{aligned} \quad (23)$$

where

$$\Delta\bar{z}(t_i) = \bar{z}(t_i) - \hat{z}(t_i) \quad (24)$$

is the “prediction error”, and  $\mathbf{P}_{t_{i-1}|t_{i-1}} = \mathbf{E}[\Delta\mathbf{x}(t_{i-1})\Delta\mathbf{x}^T(t_{i-1})]$  denotes the covariance of the state at time  $t_{i-1}$ , given all  $\bar{\mathbf{Z}}_{t_{i-1}}$ .  $\mathbf{Q}_d(\Delta t_i)$ , the uncertainty in  $\mathbf{x}(t_i)$  introduced through  $\mathbf{w}$  since  $t_{i-1}$ , is written as

$$\mathbf{Q}_d(\Delta t_i) = \int_{t_{i-1}}^{t_i} \Phi(t_i, s)\mathbf{E}[\mathbf{w}(s)\mathbf{w}^T(s)]\Phi^T(t_i, s)ds. \quad (25)$$

Due to the properties of our model (linearity) and the assumptions on  $\mathbf{w}$ , all  $\Delta\mathbf{x}(t_i)$  and  $\Delta\bar{z}_i$  are normally distributed. The negative logarithm of the integrated process (19), the negative log-likelihood function, can therefore be written as

$$L(\hat{\theta}) = \frac{1}{2} \sum_{i=0}^m \left( \log B(t_i) + \Delta\bar{z}^2(t_i)B(t_i)^{-1} \right) + c \quad (26)$$

where  $c$  captures terms that do not depend on  $\hat{\theta}$ .

The  $\hat{\theta}$  which minimizes (26) is then the maximum likelihood estimate of  $\theta$ . Minimization is performed iteratively, where each iteration requires a filtering step using the current estimate  $\hat{\theta}$ . Given  $\hat{\theta}$ , all terms in (26) can be computed recursively using the Kalman recursions. Additional methods are proposed in [30].

In classical state and parameter estimation problems,  $\log B(t_i)$  would be considered constant, and minimization of the log-likelihood would not depend on it. This is not the case here. Indeed,  $\log B(t_i)$  appears naturally, and penalizes “*model complexity*”.

### 3.4 Implementation

To obtain  $\bar{z}$ ,  $Z_n$  (the  $n$  raw, i.e. not integrated rate or acceleration measurements) are integrated numerically:

$$\bar{z}(k\Delta t) = \bar{z}((k-1)\Delta t) + \Delta t z_{k-1} \quad (27)$$

where  $z_{k-1}$  is the  $k$ -th raw measurement sample and  $\bar{z}(0) = 0$ . Integration is performed only once, but at the base sampling rate  $\Delta t$ .

The “sampling pattern”, the points in time  $t_i$  where  $\bar{z}$  is evaluated, is a design choice. We chose a logarithmic pattern ( $\Delta t_i$  increases logarithmically), but one could also argue for a uniform pattern where  $\Delta t_i$  is constant. A logarithmic pattern ensures that the noise characteristics of the sensor are well modelled at any time-scale, and that the computational complexity is reasonable for long experiments. Note that only the complexity of the numerical integration of  $Z_n$  depends on the number of raw measurements  $n$ . One evaluation of the likelihood  $\bar{L}$  has complexity linear in  $m$ , the number of samples in the sampling pattern, and  $m \ll n$  for a logarithmic sampling pattern. For all our experiments, we chose  $m = 1000$ . Algorithm 2 summarizes the proposed method.

When computing  $\Phi(\Delta t_i, \hat{\theta})$  and  $Q_d(\Delta t_i, \hat{\theta})$ , it is important to consider that the strengths and correlation times of noise processes in

### 3. Automatic Model Identification

typical MEMS inertial sensors span several orders of magnitude. The broadband noise processes have short correlation times  $\ll 1\text{ s}$ , while bias fluctuations often have correlation times larger than 100 s. The first-order approximation,  $\Phi \approx \mathbf{I} + \mathbf{F}\Delta t_i$ , is therefore not applicable here, since the integration time  $\Delta t_i$  can be long compared to the (broadband) dynamics in  $\mathbf{F}$ . For the same reason, the common approximation  $\mathbf{Q}_d \approx \Delta t_i \mathbf{G} \mathbf{Q} \mathbf{G}^T$  is not valid. If the order of the model is small ( $\leq 2$ ), or the model has a special structure,  $\Phi$  and  $\mathbf{Q}_d$  can be computed analytically. Otherwise,  $\hat{\Phi}(\Delta t_i, \hat{\theta})$  can be approximated in many ways [61], and we use a scaling and squaring method to compute it.

---

**Algorithm 2**  $\text{ML}_i$  algorithm for automatic inertial sensor noise model parameter identification. Corresponding equations for implementation are given in brackets.

---

**input:** raw gyroscope or accelerometer measurements in rad/s or m/s<sup>2</sup> (unfiltered, at sampling rate) →  $Z_n$

**procedure**  $\text{ML}_i(\text{model}, \hat{\theta}_{\text{init}})$

- integrate  $Z_n$  numerically →  $\bar{z}(k\Delta t)$  ▷ (27)
- select sampling pattern (e.g. “log”) →  $\Delta t_i$  ▷ 3.4
- augment noise model with an integrator →  $F$
- minimize** log-likelihood  $\bar{L}$  numerically
- for**  $i \leftarrow 0$  **to**  $m$  **do**
  - compute state transition  $\Phi(\hat{\theta}, \Delta t_i)$  ▷ (21), (28)
  - compute  $Q_d(\hat{\theta}, \Delta t_i)$  ▷ (28)
  - compute prediction error  $\Delta \bar{z}(t_i)$  ▷ (24)
  - compute prediction error cov.  $B(t_i)$  ▷ (23)
- end for**
- compute log-likelihood  $\bar{L}$  ▷ (26)

**end**

verify model (PSD, AV) ▷ 2.1

**return**  $\hat{\theta}$

**end procedure**

---

To compute  $Q_d$ , the variance induced into the model from  $t_{i-1}$  to  $t_i$  through the integral in (25), we make use of the following relationship from [86]:

$$\exp\left(\begin{bmatrix} -F & GQG^T \\ 0 & F^T \end{bmatrix} \Delta t_i\right) = \begin{bmatrix} B & \Phi(\Delta t_i)^{-1} Q_d \\ 0 & \Phi(\Delta t_i)^T \end{bmatrix}. \quad (28)$$

$Q_d$  can then be extracted from the r.h.s. of (28). Important: The expression on the l.h.s. of (28) needs to be *balanced* before exponentiation (for long integration times  $\Delta t_i$ ).

## 4. Experiments

We now demonstrate that the proposed method accurately identifies noise models for different inertial sensors. Section 4.1 describes the experimental setup. In Section 4.2, noise model parameters are automatically derived from real sensor data. The identified models are analysed qualitatively, using the classical Allan variance as a reference. The performance of the proposed method is then quantified in Section 4.3, using a simulation that resembles the behaviour of one of the devices under test. The method is then compared to a competing, state-of-the-art method for noise process identification in inertial sensors. Section 4.4 assesses the validity of the models under real operating conditions.

### 4.1 Experimental Setup

To acquire sensor data for the experiments, the devices under test (DUTs) were attached to a rigid, non-moving object. Some of the experiments were conducted while the DUTs were mounted on a rate table, with all table axes and the temperature control equipment switched off entirely, to prevent spurious vibrations. Other sources of vibration, such as the laboratory power supply that was used to power the DUTs, were separated from the setup to prevent vibrations.

Since even a small temperature change can cause a large variation in the sensor bias (up to  $100^{\circ}/(hK)$  for low-cost MEMS devices), the experiments were conducted in a room with minimal variation in ambient temperature, although not temperature-controlled. Sensor data was only used for evaluation after a device warm-up time of  $\geq 1\text{ h}$ . Temperature was monitored with three temperature sensors (type: MCP9804, accuracy:  $\pm 0.25^{\circ}\text{C}$ , resolution:  $0.0625^{\circ}\text{C}$ ), and recorded at a rate of 1 Hz. Sensor internal temperature, if available, was recorded along with the ambient temperature. Figure 5 shows the devices used for the experiments.

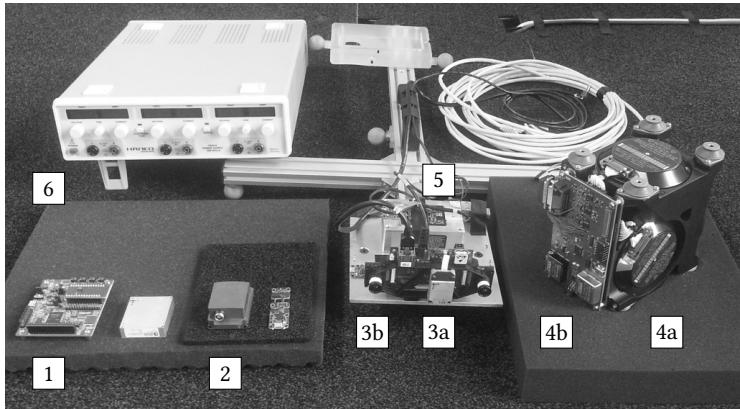


Figure 5.: Devices and laboratory equipment used during the experiments.  
1. BMX055 MEMS IMU evaluation kit, 2. MTI100 MEMS IMU, 3. visual-inertial sensor unit (3a. ADIS16448, 3b. MPU9150), 4. inertial sensor assembly (4a. GG1320AN RLG, 4b. readout electronics), 5. DSP3000 FOG, 6. laboratory power supply used to power the devices under test.

All sensors we tested have a digital interface, and a maximum sampling rate of  $\geq 100$  Hz. For a precise characterization of slower noise processes, the experiment durations were always  $> 10$  h. Raw data was recorded without any device external decimation (i.e. at sampling rate). Device-internal settings, such as range and bandwidth, were stored along with the raw data. IEEE standard 1554-2005 [70] provides additional useful recommendations on how to design experiments involving inertial sensors in general.

## 4.2 Models Derived from Real Sensor Data

We now apply the  $ML_i$  method, proposed in Section 3, to identify noise models for real gyroscopes. We show that the automatically derived

#### 4. Experiments

models are realistic by comparing their Allan variance and PSD – the classical tools in the analysis of inertial sensor noise characteristics – with the AV and non-parametric PSD computed directly from sensor data.

We employ the standard noise model (5), with the addition of a fixed, but unknown “*turn-on*” bias. This is important since in many low-cost MEMS gyroscopes, the turn-on bias (in the order of 1°/s) is much larger than the random bias variations that occur during operation. An alternative would be to remove the sample mean from the raw data prior to processing, as it is done in [29].

For our gyroscope at rest, we therefore derive the following model from the standard noise model:

$$\tilde{\omega}(t) = b(t) + v(t) + b_{\text{on}} \quad (29a)$$

$$\dot{b}(t) = -\frac{1}{\tau_b}b(t) + w(t) \quad (29b)$$

where  $\tilde{\omega}$  denotes the measured angular rate,  $b$  is the sensor bias, and  $b_{\text{on}}$  models the unknown turn-on gyroscope bias.  $v$  and  $w$  are white Gaussian noise process of strength  $\sigma_v^2$  and  $\sigma_b^2$ , respectively, modelling broadband noise and driving bias fluctuations. Except for the turn-on bias  $b_{\text{on}}$ , the model is identical with the standard model. Figure 6 presents the gyroscope model as a block diagram.

This model can be cast in the *standard integrated form* (18) as follows:

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -\frac{1}{\tau_b} & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} \mathbf{w}(t) \quad (30a)$$

$$\bar{z}(t) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \mathbf{x}(t) \quad (30b)$$

where  $\bar{z}$  corresponds to the integrated rate measurement (i.e. the angle  $\tilde{\Omega}$ ),  $\mathbf{x} = [b \ b_{\text{on}} \ \tilde{\Omega}]^T$ , and  $\mathbf{w} = [v \ w]^T$ . The model parameters are  $\theta = \{\sigma_v, \sigma_b, -\frac{1}{\tau_b}\}$ .  $b_{\text{on}}$ , the turn-on bias, is a nuisance parameter.

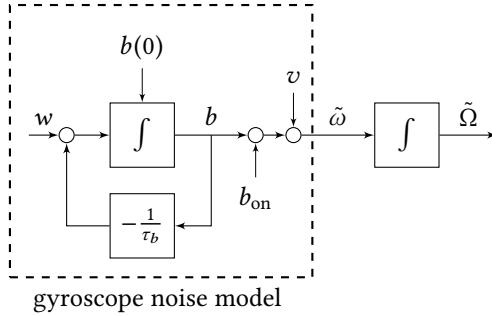


Figure 6.: Block diagram of the standard gyroscope sensor noise model for a single sensor axis (29), derived from the standard model (5).  $\tilde{\omega}$  denotes the measured angular rate,  $b$  the slowly fluctuating gyroscope bias,  $b_{\text{on}}$  a fixed but unknown turn-on bias, and  $w$  and  $v$  are white Gaussian noise processes.

#### 4.2.1 Long Term Experiment

We now apply the proposed integrated maximum likelihood estimator ( $\text{ML}_i$ ) to  $> 100$  h of data captured with the  $z$ -gyroscope of a BMX055 MEMS inertial measurement unit (IMU) from Bosch [8]. At a sampling rate of 400 Hz, this amounts to approximately  $n = 150$  M samples. The number of samples at which the integrated process is evaluated,  $m$ , is set to 1 000. The model parameters are coarsely *initialized* to

$$\hat{\sigma}_b = 0 \quad (31a)$$

$$-\frac{1}{\hat{\tau}_b} = 0 \quad (31b)$$

$$\hat{\sigma}_w = \left[ \frac{1}{n} \sum_{k=0}^{n-1} (z_k - \mu_z)^2 \Delta t \right]^{\frac{1}{2}} \quad (31c)$$

#### 4. Experiments

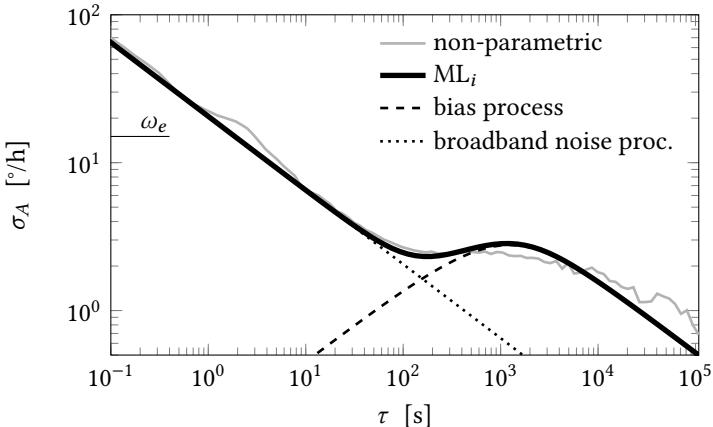


Figure 7.: Allan deviation of the BMX055 z-gyroscope. Grey: non-parametric (sample) Allan deviation. Black (solid): Allan deviation corresponding to the automatically identified noise model (29). It consists of a broadband noise component (dotted), and an exponentially time-correlated bias process (dashed). The earth's rotational rate  $\omega_e$  is shown for reference.

where  $\mu_z$  is the sample mean of  $Z_n$ . In other words, the broadband noise strength is initialized with the continuous-time equivalent of the sample standard deviation of the measurements, and the initial strength of the bias process is set to zero. The same initialization procedure was used for all experiments.

Figure 7 shows the results. The non-parametric standard Allan deviation, computed directly from  $Z_n$ , is shown in grey. The Allan deviation corresponding to the identified model was computed analytically, and is shown in black (solid). It consists of the white noise process  $v$  (dotted), and the exponentially time-correlated bias process  $b$  (dashed).

Qualitatively, the model captures the noise characteristics at any time-scale  $> 0.1$  s. The model does not have enough expressive power to capture the non-white process visible at  $\tau \approx 2$  s, in addition to the slow bias variations at  $\tau > 1000$  s. Identification of the integrated noise process ensures that the processes which are most relevant for *angle noise power* are captured, even if they cause less rate noise power.

Note that the bias diffusion parameter  $\sigma_b$  is exceptionally small for a low-cost MEMS gyroscope (no other sensor from this class showed a better bias stability in our tests). This means that the bias fluctuations are “buried” particularly deeply in the broadband noise floor in this device. We found that the  $x$  and  $y$  gyroscopes of the BMX055 (not shown) perform worse than the  $z$  axis (in terms of noise), due to a different, out-of-plane implementation. Table 2 summarizes the results.

Table 2.:  $ML_i$  estimator results for the long term experiment with the BMX055  $z$ -gyroscope.

Parameter	Value	Unit
Experiment		
Device	BMX055 $z$ -gyro	–
Sampling Rate	400	Hz
Device-Internal LPF	47	Hz
Experiment Duration	120	h
Model Parameters ( $ML_i$ )		
White Noise Strength	$\sigma_w$	$^{\circ}/(h \sqrt{Hz})$
Bias Diffusion	$\sigma_b$	$rad/(s^2 \sqrt{Hz})$
Bias Correlation Time	$\tau_b$	s

Remark: Given the model parameters, we can compute the rate

## 4. Experiments

Table 3.: Summary of repeated experiments with the BMX055  $z$ -gyroscope, processed with  $\text{ML}_i$ . The statistics are computed from ten 12h datasets.

Model Parameter		Mean	Std. Dev.	Units
White Noise	$\sigma_w$	21.66	0.66	$^{\circ}/(h \sqrt{\text{Hz}})$
Bias Diffusion	$\sigma_b$	$1.26 \times 10^{-6}$	$0.15 \times 10^{-6}$	$\text{rad}/(s^2 \sqrt{\text{Hz}})$
Bias Corr. Time	$\tau_b$	530.51	73.90	s

noise power generated by the band-limited white noise and the bias noise process. Assuming a bandwidth of 47Hz, broadband noise accounts for  $2 \cdot 47\text{Hz} \cdot (9.97 \times 10^{-5} \text{ rad}/(s \sqrt{\text{Hz}}))^2 = 9.35 \times 10^{-7} \text{ rad}^2/\text{s}^2$ , or  $0.0554^{\circ}/\text{s}$  RMS. The bias fluctuations account for only  $\frac{\sigma_b^2 \tau_b}{2} = 4.76 \times 10^{-10} \text{ rad}^2/\text{s}^2$ , or  $0.0012^{\circ}/\text{s}$  RMS rate noise. This corresponds to a broadband to bias rate noise power ratio of  $\approx 2000 : 1$ , or 33dB.

### 4.2.2 Repeatability

To confirm that the proposed method is repeatable, we applied it to ten datasets from the same device. The duration of each dataset is 12h.

Figure 8 shows the results. The sample based Allan deviations are shown in grey, and those corresponding to the identified models are shown in black (solid). Qualitatively, the resulting models have similar characteristics across the entire time-scale. The flickering tendency at long integration times can not be captured with the model. The estimates of the broadband noise strength ( $\sigma_w$ ) are more repeatable than the estimates of the parameters related to the slow bias fluctuations ( $\sigma_b$  and  $\tau_b$ ).

$\sigma_b$  and  $\tau_b$  are inherently more difficult to estimate, since the likelihood function has a “flatter optimum” with respect to these param-

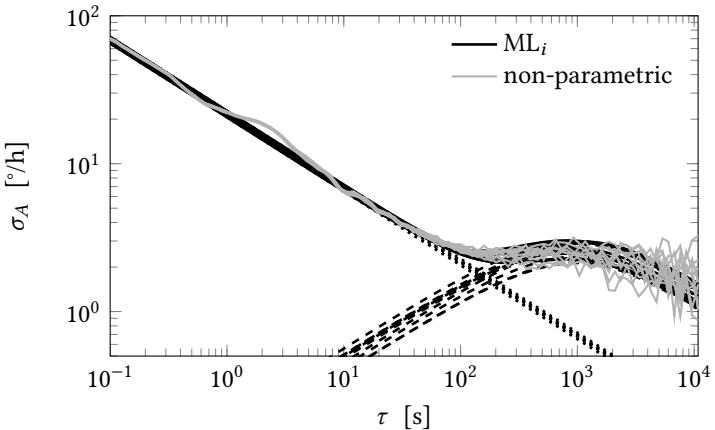


Figure 8.: Repeatability: Allan deviations of ten 12h datasets captured with the z-gyroscope of the BMX055 IMU. Grey: non-parametric estimates of the Allan deviations. Black: Allan deviations corresponding to the models that were automatically identified using  $ML_i$ .

eters, even for long observations. In other words, a larger range of parameters explains the long-term behaviour of the observed sample path well. The best point estimate is therefore more sensitive to the specific realization of the noise process that was observed. The sample-based Allan deviation also suffers from this, as Fig. 8 shows. Table 1 summarizes these results. Note that the statistics of the parameter estimates that are indicated are computed using estimation results from different datasets.  $ML_i$  itself does not provide estimates of parameter uncertainties.

## 4. Experiments

### 4.2.3 Testing with Different Devices

The proposed estimator was tested on different gyroscopes, again using the standard noise model (29). Figure 9 shows the Allan deviation for one of the axes of each of the devices under test. For multi-axes devices, the axis which is perpendicular to the device footprint (typically  $z$ ) is shown. This is important since for single chip MEMS devices, in contrast to e.g. the ADIS16448 or the MTI100, the performance between in-plane and out-of-plane axes often varies. Table 4 lists the identified parameters (for all axes).

Qualitatively, in terms of the Allan deviation,  $ML_i$  captures the noise processes well, considering the expressive power of the standard model. With one exception: the noise governing the GG1320AN dithered ring laser gyroscope [26] can not be adequately modelled with the standard noise model (since it is dominated by white *angle* noise at the time-scale of interest). This shows that the method can only be used for a fair comparison of inertial sensor noise performance if the assumed noise model structure is applicable.

### 4.2.4 Noise Processes near the Sampling Frequency

Many MEMS inertial sensors have device-internal, digital filtering and sub-sampling stages that *decimate* the measurements before readout, see Fig. 2. The decimation stage is often configurable, and allows the engineer to select a combination of bandwidth and output sampling rate. This has two important implications: a.) a significant *delay* can be introduced, depending on the selected bandwidth and filter architecture. An extreme bandwidth setting of 12 Hz, for example, can introduce a delay  $> 30$  ms. If the inertial data is combined with e.g. precise 2D landmark observations from a camera system, this needs to be accounted for. b.) The decimation stage introduces additional dynamics in the noise processes (close to the sampling frequency). An automatic noise model parameter estimation mechanism that works on rate or acceleration directly will aim to model these dynamics –

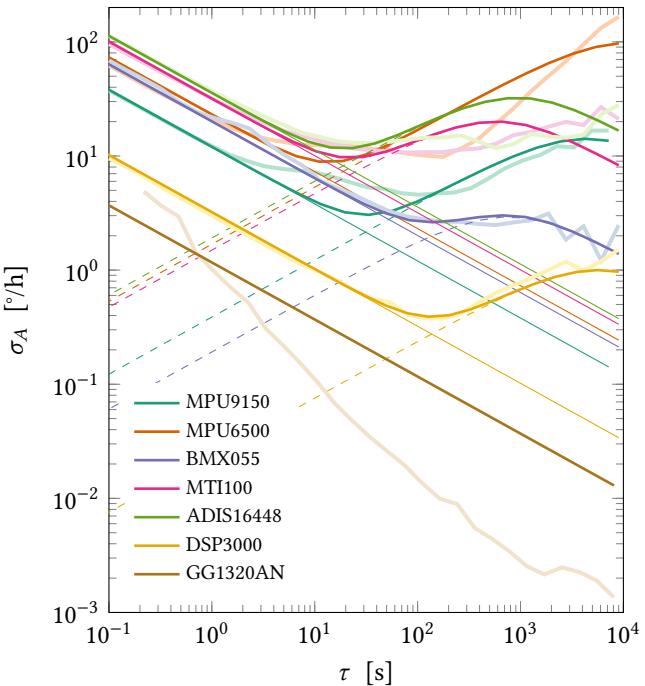


Figure 9.: Allan deviations of different gyroscopes under test (for multi-axes devices, only the gyroscope  $z$ -axis is shown). The sample based Allan deviations are shown in pastel, with an overlay of the Allan deviations that correspond to the identified standard noise models ( $ML_i$ ).

since it is designed to explain rate or acceleration noise power. As a result, the method may quickly fail to capture weak bias fluctuations, even though they contribute more significantly to noise in the integrated process. In contrast,  $ML_i$  copes well with such cases, since it

#### 4. Experiments

Table 4.: Parameters of the standard noise model (5), identified with the proposed  $\text{ML}_i$  estimator, for different devices.

<b>Device</b>	<b>Axis</b>	$\sigma_w$ rad/(s $\sqrt{\text{Hz}}$ )	$\sigma_b$ rad/(s $^2$ $\sqrt{\text{Hz}}$ )	$\tau_b$ s
<b>MPU9150</b>	$x$ -gyro	$7.28 \times 10^{-5}$	$1.25 \times 10^{-5}$	$4.9 \times 10^3$
	$y$ -gyro	$7.47 \times 10^{-5}$	$8.73 \times 10^{-6}$	$6.6 \times 10^3$
	$z$ -gyro	$5.83 \times 10^{-5}$	$3.25 \times 10^{-6}$	$2.3 \times 10^3$
<b>MPU6500</b>	$x$ -gyro	$1.42 \times 10^{-4}$	$5.66 \times 10^{-6}$	$1.7 \times 10^4$
	$y$ -gyro	$1.08 \times 10^{-4}$	$1.09 \times 10^{-5}$	$6.0 \times 10^2$
	$z$ -gyro	$1.12 \times 10^{-4}$	$1.42 \times 10^{-5}$	$5.9 \times 10^3$
<b>BMX055</b>	$x$ -gyro	$1.24 \times 10^{-4}$	$1.65 \times 10^{-6}$	$5.4 \times 10^2$
	$y$ -gyro	$1.99 \times 10^{-4}$	$2.41 \times 10^{-6}$	$2.4 \times 10^3$
	$z$ -gyro	$9.97 \times 10^{-5}$	$1.20 \times 10^{-6}$	$6.6 \times 10^2$
<b>MTI100</b>	$x$ -gyro	$1.49 \times 10^{-4}$	$2.78 \times 10^{-5}$	$1.5 \times 10^2$
	$y$ -gyro	$2.16 \times 10^{-4}$	$1.06 \times 10^{-5}$	$2.1 \times 10^3$
	$z$ -gyro	$1.54 \times 10^{-4}$	$1.26 \times 10^{-5}$	$3.1 \times 10^2$
<b>DSP3000</b>	–	$1.56 \times 10^{-5}$	$2.01 \times 10^{-7}$	$3.1 \times 10^3$
<b>ADIS16448</b>	$x$ -gyro	$1.75 \times 10^{-4}$	$1.55 \times 10^{-5}$	$4.0 \times 10^2$
	$y$ -gyro	$1.64 \times 10^{-4}$	$1.27 \times 10^{-5}$	$2.9 \times 10^2$
	$z$ -gyro	$1.73 \times 10^{-4}$	$1.61 \times 10^{-5}$	$5.0 \times 10^2$

tries to explain the *integrated process* itself.

Figure 10 shows the PSD of the BMX055  $z$ -gyroscope measurements close to the sampling frequency. A non-parametric estimate of the PSD, computed using Welch’s method [90], is shown in grey for

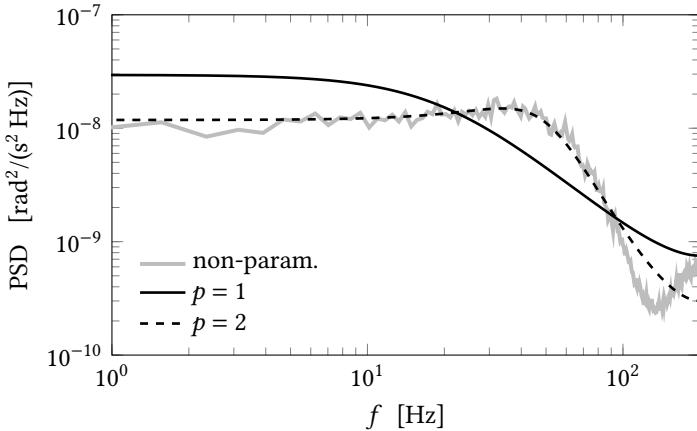


Figure 10.: Power spectral density (PSD) of the BMX055  $z$ -gyroscope near the sampling frequency (double-sided density). The non-parametric PSD, computed using Welch's method, is shown in grey. The PSD of the automatically identified models of order  $p = 1$  and  $p = 2$  are shown in black.

frequencies from 1Hz to 200Hz (half the sampling frequency). The effect of a decimation stage, with a cut-off frequency setting of 47Hz, is clearly visible; it limits the bandwidth of the broadband noise. If required, these processes can be identified using any of the classical methods. Once the model parameters are identified, the exact PSD of the process generated by the model can be computed. The PSD corresponding to the identified model of order one and two are shown in black.

Since the dynamics of these processes are usually faster than the rate at which information from an aiding sensor, e.g. a camera, is incorporated, they are usually not part of an inertial sensor model.

### 4.3 Simulations

This section uses simulated data to analyse the proposed estimator quantitatively. The simulation resembles, to some extent, the noise characteristics of one of the devices analysed in Section 4.2: the z-gyroscope of the BMX055 MEMS IMU. The standard model (29) is used, with parameters drawn according to the values reported in Table 1. In contrast to the experiments in the previous sections, we now have access to the true noise model parameters  $\theta$ .

#### 4.3.1 Comparison with Direct Methods

Figure 11 shows the statistics of the bias diffusion strength  $\sigma_b$  (top) and the bias correlation time  $\hat{\tau}_b$  (bottom) estimates, for different estimators. The bias diffusion strength  $\sigma_b$  was varied in order to emulate increasing levels of bias fluctuation, while  $\sigma_w$  and  $\tau_b$  were held constant. The lowest  $\sigma_b$  of  $1.15 \times 10^{-6}$  rad/(s<sup>2</sup>  $\sqrt{\text{Hz}}$ ) corresponds approximately to the bias diffusion of the BMX055 gyroscope (see Table 1) – a small value for this class of sensors and this level of broadband noise.

For high levels of bias diffusion, all methods perform well, including classical maximum likelihood on uniformly sampled rate noise (ML uniform). Once the bias fluctuations approach realistic values, identification of the integrated (angle) noise process (ML<sub>i</sub>) has an advantage. The logarithmic sampling strategy of the ML<sub>i</sub> log estimator, favouring models which represent the noise characteristics at any time-scale, leads to accurate estimates of the bias diffusion strength and the bias correlation time at all levels of bias diffusion. The characteristics of the broadband noise processes were estimated accurately in all configurations (not shown).

#### 4.3.2 Comparison with other Methods

We now compare the proposed estimator with the state-of-the-art “Generalized Method of Wavelet Moments” (GMWM), developed

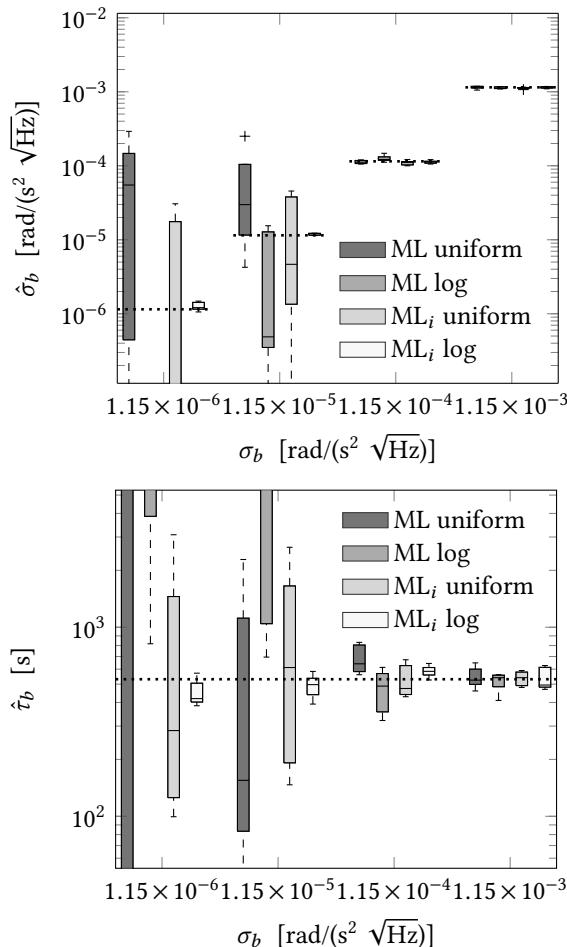


Figure 11.: Estimated sensor bias diffusion ( $\hat{\sigma}_b$ ) and correlation time ( $\hat{\tau}_b$ ) for classical (ML) and integrated ( $\text{ML}_i$ ) maximum likelihood methods, using uniform (uniform) and logarithmic (log) sampling schemes, for varying levels of bias diffusion.

#### 4. Experiments

specifically for the identification of such noise processes [29], and applied extensively to inertial sensor noise modelling [78]. GMWM provides models which are accurate across all time-scales by decomposing the measurements in a wavelet basis. It can also be used to identify more complex noise processes, such as quantization noise and bias instability (flickering), something our method is not capable of.

GMWM requires the selection of a specific wavelet basis, and the choice has an influence on the resulting parameter estimates. Our method, on the other hand, requires the selection of a sampling pattern. The pattern also has an influence on the parameter estimates, since it will ultimately determine the time-scales at which the model is forced to represent the observed (integrated) noise process well.

Figure 12 shows a comparison of the two methods, again using a simulation of the standard model (29) with parameters derived from Table 1. Both methods were employed on ten simulated datasets with a length of 12h each. The resulting estimates are shown in grey ( $ML_i$ ) and white (GMWM).<sup>2</sup>

The broadband noise strength  $\sigma_w$  is estimated accurately by both methods. The bias diffusion  $\sigma_b$  is generally estimated less accurately, and the estimates provided by the GMWM appear slightly biased. The bias correlation time  $\tau_b$ , a difficult parameter to estimate for this noise process (see Section 4.3), is estimated to within approximately  $\pm 120$  s of  $\tau_b = 530$  s.

#### 4.4 Device Characteristics under Normal Operating Conditions

Two important questions remain unanswered:

1. Do the noise models still capture the sensor errors well enough

---

<sup>2</sup>An implementation of GMWM [29] was provided by the “Geodetic Engineering Laboratory” (Ecole Polytechnique Fédérale de Lausanne).

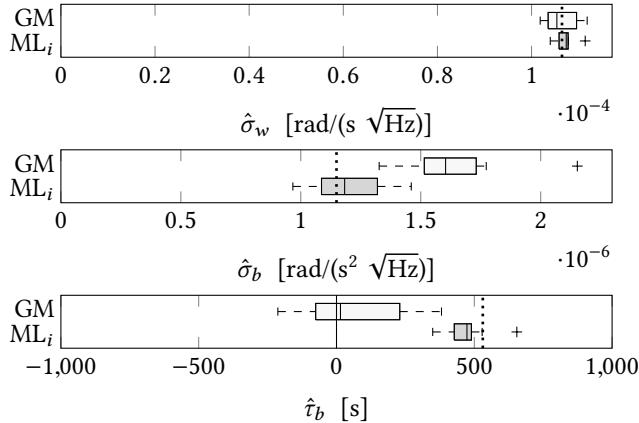


Figure 12.: Parameter estimates from two different methods: the generalized method of wavelet moments (GM) [29, 78], and integrated maximum likelihood ( $ML_i$ ). The standard model parameters were estimated for 10 synthetic datasets, each 12h long. Ground truth is indicated with a dotted line.

under normal operating conditions, i.e. when the sensor is in motion?

2. Can we actually achieve the performance predicted by the noise model, with and without device calibration?

To address these questions, a series of experiments were conducted, with the aim to give a qualitative insight into the different sources of error in low-cost inertial sensors.

#### 4.4.1 Devices under Motion

We first conduct an experiment where the sensors undergo *arbitrary motion*. The experiment relies on our visual-inertial sensor unit [64], shown in Fig. 4. The unit incorporates several MEMS IMUs, synchro-

#### 4. Experiments

nized and rigidly attached to two global-shutter CMOS camera sensors. We compare the integrated state of the unit (using gyroscopes and accelerometers) with pose estimates from the camera system. This then allows us to qualitatively evaluate whether the noise models are valid under these conditions, or not.

During the experiment, the visual-inertial sensor unit followed a smooth trajectory (hand-held) in front of a visual calibration target. Rate and acceleration were in the range of  $\pm 180^\circ/\text{s}$  and  $\pm 2\text{g}$ . No shocks or vibrations were present, and device temperature was constant to within  $< 1.5^\circ\text{C}$ . The IMUs were sampled at 200Hz, the cameras at 20Hz. The cameras were calibrated intrinsically prior to the experiments, using Kalibr. Target observations with the camera system then provided a noisy, potentially biased through inaccuracies in camera in- and extrinsics, but *error-bounded* ground-truth for the orientation and the position of the sensor unit with respect to the target, at a rate of 20Hz.

Figure 1 shows a schematic overview of the experimental setup. To make a precise statement about the accuracy of the inertial sensor data under motion, the following parameters have to be estimated:

$\mathbf{p}_B(t), \mathbf{q}_B(t)$	Pose and velocity of the sensor units (body) frame $B$ , expressed in an inertial frame of reference $I$ . $B$ coincides with the input reference axes of the IMU.
${}^B\mathbf{p}_C, \mathbf{q}_{CB}$	Camera $C$ to IMU transformation.
$\mathbf{q}_{TI}$	The orientation of the calibration target $T$ with respect to the gravity vector $\mathbf{g}$ (two degrees of freedom).
$t_{BC}$	A fixed time-difference between the cameras and the IMUs (IMU internal delays).
$\mathbf{C}_I$	IMU intrinsic calibration parameters (scale factors, axes misalignment).
$\mathbf{b}_g(t), \mathbf{b}_a(t)$	Gyroscope and accelerometer biases.



Figure 13.: Visual-inertial sensor unit [64] used for assessing inertial sensor performance under motion, in front of a visual calibration target. An MPU9150 MEMS IMU is located behind each of the cameras, and used for testing.

All of these states and parameters were estimated using a classical, discrete-time maximum likelihood (batch) estimator. The estimator incorporates all measurements (i.e. every sample from the gyroscope and the accelerometer, and all undistorted 2D camera observations of corners in the visual target). The states were estimated at each time-instance at which a camera frame was captured. The parameters, such as the camera to IMU extrinsics or the IMU calibration parameters, were assumed fixed but unknown, and estimated at the same time. Up to linearisation, and under the assumption that the sensor models that were used are correct, this estimator is asymptotically optimal. The details of the estimator are not discussed in this report.

Figure 15 shows the difference between the gyroscope integrated attitude, and the attitude which was estimated using only observations from the cameras. Following the “full batch” estimation, outlined above, the state of the system was fixed at  $t_0$  to the estimate from the full batch estimator (at  $t_0$ ). The initial covariances of the sensor biases were obtained from the respective elements of the inverse information matrix of the estimator [41]. The state was then integrated using only inertial measurements (from the gyroscopes and accelerometers). The orientation of the cameras with respect to the target (computed using the cameras only) was transformed using the estimated parameters of the camera-IMU extrinsics, time delay, and orientation of the target

#### 4. Experiments

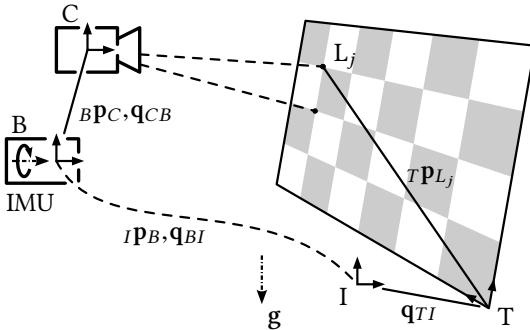


Figure 14.: Schematic overview of the experiment involving the visual-inertial sensor unit and the calibration target. Once the setup is accurately calibrated, errors in the (moving) inertial sensors can be quantified.

with respect to gravity. Shown in Fig. 15 is the square root of the variance of the rotation (magnitude) between the two frames.

Also shown is the uncertainty as predicted by the automatically identified IMU noise model (dashed). The variance of the model was computed numerically (to first order). The predicted uncertainty is composed of a.) an angular random walk component due to the white noise part of the model, b.) the integrated bias processes, and c.) uncertainty in the state (biases) of the model at  $t_0$ , the beginning of integration.

The square root of the variance of the integrated attitude for the *uncalibrated* device (averaged over five experiments) is larger than predicted by the model. This indicates that the pure noise model is too optimistic, and does not account for errors that are introduced once the uncalibrated device is in motion. This is not surprising, considering the scale of the deterministic errors that are present in a sensor of this class. Gyroscope axes misalignment angles, for example, were estimated to be in the order of  $0.5^\circ$ . In other words, even a small rate

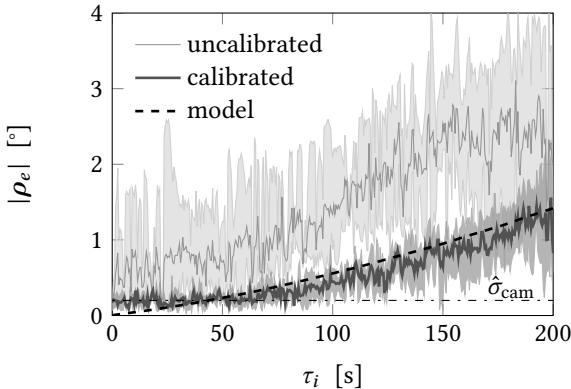


Figure 15.: Attitude error (RMS) for five experiments with the visual-inertial sensor unit in motion. Black (dashed): predicted square root of the integrated attitude error, computed using the automatically identified gyroscope noise models. Grey: differences between the integrated attitude (calibrated and uncalibrated) and the attitude estimated with the camera system.  $\hat{\sigma}_{\text{cam}}$ : approximate uncertainty in the transformed camera attitude estimates (determined from this figure).

on one axis may cause an almost instantaneous apparent “bias” on an other axis which, considering only the static noise model, would be very unlikely to occur. Clearly, the errors introduced through deterministic errors depend on the motion.

Also shown are the results from the same experiment, but incorporating IMU intrinsic calibration terms in the integrator (scale factors and misalignment). This leads to a better agreement between predicted (model) variance and measured variance. Indeed, the performance of the gyroscopes are in the order of the performance predicted by the static noise model.

## 4. Experiments

### 4.4.2 Temperature

The results presented above suggest that IMU intrinsic calibration reduces the deterministic errors in the inertial data. Indeed, a low-cost MEMS IMU IC *outperformed* a factory-calibrated IMU that consists of several single-axis (MEMS) gyroscopes and accelerometers – simply due to *better noise performance*. It is crucial to note, however, that the experiments were conducted at constant temperature (and no temperature-cycles in-between experiments), and with smooth motion (no vibrations). In practice temperature may vary rapidly, not only due to changes in ambient temperature, but also due to device internal temperature changes.

Temperature plays an important role in MEMS inertial sensors, and its influence can quickly invalidate the noise model. We show this with the example sensor sensitivity changes as a function of temperature. To this end, we conducted experiments on a three-axis rate table with a temperature chamber [1]. Temperature was ramped from room temperature to approximately 65 °C, while 42-element motion profiles (based on the classical 18-element sequence proposed in [15]) were executed repeatedly. The total experiment duration is below 1 h per three axis device (either angular rate or acceleration, since the motion profile is not the same). Extrinsic (IMU input reference axes to rate table) and IMU intrinsic calibration parameters were jointly estimated.

Figure 16 shows the change in the scale factor (the sensitivity) of one of the MEMS gyroscopes we tested [8]. This particular device showed a change in the scale factor with temperature of up to 150 ppmK<sup>-1</sup>. In the same sensor, a gyroscope bias temperature sensitivity of up to 40 °/(h K) was registered (not shown). For typical operating conditions, such temperature induced bias changes can easily outweigh the random bias variations captured with the noise model. This is one of the reasons why low-cost inertial sensor noise model parameters are often considered “tuning parameters”. It is also

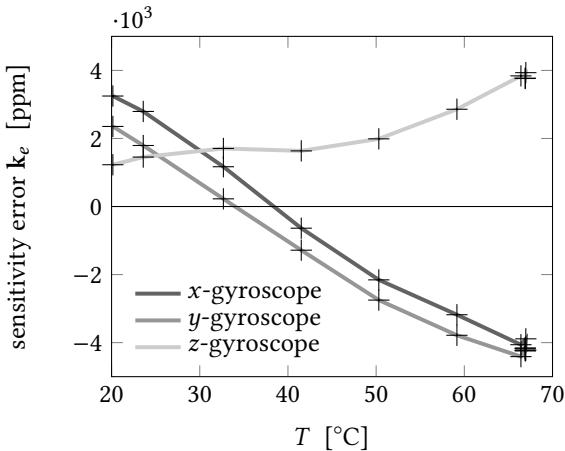


Figure 16.: Estimated sensitivity error of the BMX055 gyroscopes as a function of temperature.

apparent that device calibration, if performed at constant temperature, is *only valid for this temperature range*, and may even increase error if applied at a different operating temperature. [55] addresses this (in a visual-inertial SLAM context), by tracking IMU intrinsic model parameters (scale factors, misalignment, and g-sensitivity) *on-line*.

## 5. Conclusion

This report presented a method to automatically identify the parameters of inertial sensor noise models. By identifying noise processes according to their contribution to the integrated process, processes that span many time- and strength scales were easily captured with

### *Acknowledgements*

a classical maximum likelihood estimator. The resulting models are accurate at an attitude, velocity, or position level, and are therefore appropriate for usage in sensor fusion applications.

A qualitative analysis of the method on different devices showed that it performs well on real sensor data. An experiment involving a visual-inertial sensor unit demonstrated the usefulness of the method for *practical applications*. And a quantitative analysis, based on simulated data, highlighted the capability of the method to accurately characterize challenging composite noise processes.

We also concluded that the derived error models are too optimistic for uncalibrated devices under real operating conditions. A useful extension of this work would therefore be to automatically tune noise model parameters *under these conditions*. In a visual-inertial estimation context, for example, a joint optimization of deterministic and stochastic model parameters is possible – in principle.

## **Acknowledgements**

The authors wish to express their gratitude to Joern Rehder, Michael Burri, Pascal Gohl, and Stefan Leutenegger, amongst others, for their commitment and their tireless engineering efforts that lead to the visual-inertial sensor unit which facilitated our experiments. We thank Gabriel Agamennoni and Simon Lynen (ETH Zürich) for their scientific contributions. We thank Joel Hesch, Renzo de Nardi, and Johnny Lee (Google) for supporting us with their expertise and equipment, and the Geodetic Engineering Laboratory at EPFL for providing a well documented implementation of the GMWM.



# Paper II

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

Janosch Nikolic · Michael Burri · Igor Gilitschenski ·  
Juan Nieto · Roland Siegwart

### Abstract

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

*Paper II*

in extrinsic translation, 1 mrad in orientation, and 10  $\mu$ s in time shift – within a calibration window of 20 s.

Accepted for publication in *IEEE Sensors*, 2016.

©2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

## 1. Introduction

**V**ISUAL-INERTIAL sensor units are a popular choice for localisation and mapping systems since the combination of these complementary sensing modalities leads to increased robustness and accuracy. However, calibration is required to optimally leverage the measurements of all sensors. Exact knowledge of the transform between the sensors is vital [52]. Most precise systems also align the sensor data temporally to compensate for camera exposure time and inertial measurement unit (IMU) intrinsic delays [25].

While most visual-inertial localisation and mapping frameworks employ an intrinsic calibration for the camera, intrinsic calibration for the inertial sensors is rarely considered in this context (with some notable exceptions, e.g. [54]). To mitigate the problem, factory calibrated IMUs are used [64], or the inertial sensor noise model parameters are tuned to account for unmodelled effects in the gyroscopes and accelerometers [89].

Many modern localisation and mapping systems employ low-cost, micro electromechanical (MEMS) chip gyroscopes and accelerometers. Due to the technologies used, these can exhibit significant scale factor (sensitivity) and misalignment errors (in the order of 1% and 1°, respectively), see [47]. Once the sensor unit undergoes motion, these terms cause correlated errors which quickly render a purely stochastic sensor model invalid. This leads to errors in the extrinsic calibration and to suboptimal performance.

Our contribution is a probabilistic framework for the joint estimation of sensor extrinsic and IMU intrinsic calibration parameters. Including IMU intrinsic parameters into the entire calibration procedure improves the quality of the calibration, and allows us to achieve an overall performance which is limited largely by the sensor's noise performances alone. The novel method we present requires no tuning parameters and is optimal with respect to stringent probabilistic

criteria. The specification of probabilistic models for those calibration parameters which are considered constant is not necessary.

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

Quantitative results from repeated experiments provide an insight into achievable calibration accuracy, and show the typical range of deterministic errors in consumer MEMS gyroscopes and accelerome-

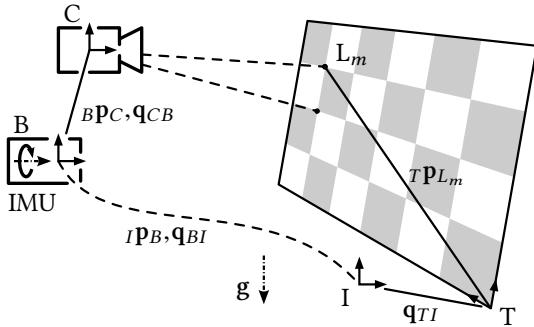


Figure 1.: Visual-inertial sensor system in front of a visual calibration target. The camera  $C$  and the IMU are rigidly attached to the sensor unit's body frame  $B$ , which coincides with the IMU's input reference axes. Gyroscopes and accelerometers measure the unit's angular rate and acceleration,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

## 2. Related Work

ters. The results also show that the precision and the accuracy of the extrinsic calibration can be improved significantly by incorporating IMU intrinsic calibration terms into the optimisation.

We compare our method directly to more complex, parametric approaches [24]. Parametric approaches use temporal basis functions, for example splines, to represent the motion of the sensor unit and other time varying states such as fluctuating sensor biases. In general, these parametric approaches are regarded as being superior particularly for time delay estimation. However, our experiments reveal that the accuracy of both approaches is comparable. To the best of our knowledge, no such comparison exists in the literature to date.

The remainder of this paper is structured as follows: Related approaches with a focus on works that address IMU intrinsic calibration are summarised in Section 2. The sensor models we employ for the calibration of visual-inertial sensor units are described in Section 3. The maximum likelihood state and parameter estimator we developed for calibration is presented in Section 4, and the experimental setup and results from repeated calibration experiments are presented in Section 5.

## 2. Related Work

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

or off-line methods v) calibration infrastructure: natural landmarks, visual calibration target, and other equipment.

Early works by Alves et al. [5] and Lobo et al. [56] rely on dedicated hardware (a pendulum and a rate table, respectively) to determine the relative pose (position and orientation) between a camera and an IMU, and IMU intrinsic parameters. The different calibration parameters are estimated independently in both approaches, which is suboptimal. In contrast, the method we present requires only a visual calibration target (a checker board, for example), which simplifies the calibration process.

Mirzaei et al. [59] present a calibration method based on an extended Kalman filter (EKF), using a visual calibration target and hand-held motion. The calibration parameters (extrinsics) are incorporated into the system state and estimated along with the pose, velocity, and sensor biases. An extension to a full batch solution is also proposed but does not include sensor intrinsic parameters. Kelly et al. [46] present a similar approach using an unscented Kalman filter. In addition, they propose an extension by using only natural landmarks, without the need for a calibration target. Zachariah et al. [93] use a sequential linear filter based on a sigma point Kalman filter and additionally track IMU intrinsic calibration parameters. Li et al. [55] recently proposed an EKF based method to estimate a multitude of calibration parameters of a rolling-shutter camera-IMU setup. Li includes IMU and camera intrinsics, as well as time delays into the system state. This filter based method works on-line and requires no calibration infrastructure.

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

## 2. Related Work

considered “fixed but unknown” and does not require the specification of a probabilistic model for these parameters.

In contrast to these recursive methods, Hol et al. [33] treat calibration as a gray-box system identification problem. A full state and state covariance of the system are propagated using the IMU, and a linearised cost function with visual residuals (from checker board observations), weighted according to their predicted covariance, is minimised with respect to the camera-IMU extrinsic calibration parameters. This approach is similar to our method, but it does not include IMU intrinsic parameters.

Furgale et al. [25] propose a full maximum likelihood estimator to determine sensor extrinsics and a time delay between camera and IMU. The method uses a parametric representation of the unit’s motion and its sensor biases (B-splines). This enables sophisticated calibration features and direct computation of inertial measurement residuals. We will refer to this method as *Kalibr*, since a corresponding toolbox was released under this name<sup>1</sup>. Krebs [50] extends this work by including IMU intrinsic calibration parameters and is similar to our work in terms of sensor modelling. Rehder [71] pushes further to leverage the parametric spline representation of the sensor unit’s trajectory to determine the position of individual accelerometers. In contrast to [71], we cannot estimate the position of individual accelerometers. We do not estimate angular accelerations which renders our method unsuitable for the estimation of such parameters. However, extensions to our method to address this are feasible.

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

---

<sup>1</sup>*Kalibr* is available at [www.github.com/ethz-asl/kalibr](http://www.github.com/ethz-asl/kalibr) (April 2016).

target points are detected and inertial measurements are integrated in-between.

### 3. The Visual-Inertial Sensor System

A typical visual-inertial sensor system is comprised of one or more cameras  $C_i$  which are rigidly connected to a strapdown IMU. We define a moving body coordinate frame  $B$  on the sensor unit such that it coincides with the IMU's input reference axes (IRA). Fig. 1 illustrates the sensor system in front of the calibration target  $T$ . The corresponding notation is briefly introduced in Section 3.1. As the sensor unit moves, the camera observes points from the calibration target through a projection onto its camera image plane, where the observations are corrupted by noise. The corresponding camera measurement model is summarised in Section 3.2. At the same time, the noisy gyroscopes and accelerometers measure the sensor unit's angular rate and acceleration. The corresponding gyroscope and accelerometer sensor models are described in Section 3.3.

#### 3.1 Notation

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

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

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

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

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

### 3. The Visual-Inertial Sensor System

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

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

where  $[\cdot \times]$  denotes the skew-symmetric operator [84]:

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

## 3.2 Camera Measurement Model

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

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

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

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

$$= \pi(C_{CB}(C_{IT}\mathbf{p}_{L_m} - I\mathbf{p}_B) - B\mathbf{p}_C) + \mathbf{v} \quad (5c)$$

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

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

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

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

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

### 3.3 Inertial Sensor Model

The gyroscope and accelerometer sensor models have a stochastic and a deterministic component. The stochastic (noise) model we use are described in Section 3.3.1, and the deterministic models in Section 3.3.2.

#### 3.3.1 Stochastic Model

A variety of stochastic processes are used for modelling inertial sensor noise [38]. We employ a simple model that is widely used in the context of visual-inertial sensing (for example in [53, 84, 32]): a combination of a rapidly fluctuating (white) noise process and a slowly varying, correlated noise process (a bias). A typical source for the wideband noise component in MEMS inertial sensors is electronic noise from transducer and amplifier stages, and fluctuations in drive frequency are a source for bias variation [47].

We denote the gyroscope and accelerometer noise processes as  $\mathbf{n}_g$  and  $\mathbf{n}_a$ . To keep the notation simple we will not distinguish between

### 3. The Visual-Inertial Sensor System

random processes and a particular realisation (a sample path) of the process and write

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

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

$\mathbf{w}_g$  and  $\mathbf{w}_a$  denote continuous-time, white Gaussian noise processes of strength  $\sigma_g$  and  $\sigma_a$ ,

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

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

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

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

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

where  $\mathbf{w}_{bg}$  and  $\mathbf{w}_{ba}$  are white noise processes of strength  $\sigma_{bg}$  and  $\sigma_{ba}$ , the bias “diffusions”.

During calibration we will set the correlation times to  $\tau_g = \tau_a = \infty$ . We do this for two reasons: i) low-cost MEMS gyroscopes and accelerometers exhibit turn-on biases which remain constant during the operation of the instrument (at constant temperature) but are significant (in the order of a few degrees per second and up to  $0.5 \text{ m/s}^2$ ). Setting the correlation times to infinity prevents us from having to estimate the turn-on biases since no a-priori mean is required for an unbounded (i.e.  $\tau = \infty$ ) random walk. ii) an exponentially time-correlated process is rarely a good approximation for true bias variation which often exhibits flickering characteristics. In addition, bias variation is mostly driven by temperature changes which is itself

typically not exponentially time-correlated. Nevertheless, exponentially correlated bias processes and turn-on biases, if desired, would fit naturally into the concept presented here.

A viable procedure to obtain the noise strengths (the “sigmas”) manually using the Allan variance is given in [38]. We obtain these noise model parameters automatically using the maximum likelihood method outlined in [65]. This reduces the potential for error since it requires no user input. An alternative method is presented in [78].

### 3.3.2 Deterministic Model

Various, often identical, deterministic inertial sensor models exist in the literature (see, for example, [50, 15, 72, 83, 77]). A more generic accelerometer model is proposed in [54], where each accelerometer has an arbitrary orientation and scale on which the specific force is projected.

For tractability, we choose a simple model based on the model of [77]. It incorporates scale errors, axes misalignment (cross axes sensitivity), and gyroscope g-sensitivity:

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

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

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

$$\begin{aligned} \mathbf{K}_g &= \begin{bmatrix} k_{gx} & 0 & 0 \\ 0 & k_{gy} & 0 \\ 0 & 0 & k_{gz} \end{bmatrix} \mathbf{M}_g = \begin{bmatrix} 1 & 0 & 0 \\ \gamma_z & 1 & 0 \\ -\gamma_y & \gamma_x & 1 \end{bmatrix} \\ \mathbf{K}_a &= \begin{bmatrix} k_{ax} & 0 & 0 \\ 0 & k_{ay} & 0 \\ 0 & 0 & k_{az} \end{bmatrix} \mathbf{M}_a = \begin{bmatrix} 1 & -\alpha_{yz} & \alpha_{zy} \\ \alpha_{xz} & 1 & -\alpha_{zx} \\ -\alpha_{xy} & \alpha_{yx} & 1 \end{bmatrix}. \end{aligned}$$

### 3. The Visual-Inertial Sensor System

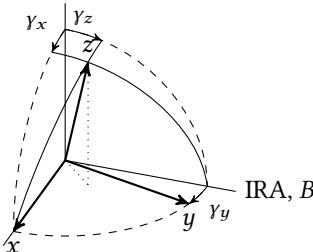


Figure 2.: Misalignment (or cross-axis sensitivity) of the gyroscope triad. The  $x$ -axis of the gyroscope is aligned with its corresponding input reference axis (IRA). The (approximate) small angles  $\gamma$  are contained in the misalignment matrix  $\mathbf{M}_g$ . Note that the input axes  $x$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

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

When data from accurate sources is fused, precise temporal alignment is essential. The visual-inertial sensor unit we used performs a shutter-centric alignment between camera frames and inertial measurements in hardware. IMU intrinsic delays, however, have to be determined from measurements. The main contributor to commercial MEMS gyroscope and accelerometer delays are device internal (digital) low-pass filters. We summarise their effect in one single, constant delay  $\Delta t_{BC}$ . This is an approximation, since the true delay is frequency dependent. Depending on the bandwidth of these filters (often user configurable), typical delays range from below 1 ms to 30 ms or more. This is significant and needs to be taken into account when designing precision visual-inertial estimation algorithms.

## 4. Calibration

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

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

The states and the parameters which are estimated are summarised in Section 4.1. The equations that link the inertial sensor measurements and the motion of the sensor unit are stated in Section 4.2, and an outline of the state and parameter estimator is given in Section 4.3.

### 4.1 System State and Parameters

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

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

$$\mathbf{x} = \begin{bmatrix} \mathbf{q}_{BI}^T & {}_I\mathbf{p}_B^T & {}_I\mathbf{v}_B^T & \mathbf{b}_g^T & \mathbf{b}_a^T \end{bmatrix}^T. \quad (11)$$

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

#### 4. Calibration

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

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

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

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

## 4.2 Equations of Motion

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

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

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

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

where

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

We can substitute the true body angular rates  ${}_B\omega$  and accelerations  ${}_B\mathbf{a}$  with the noisy, erroneous gyroscope and accelerometer measurements  $\tilde{\omega}$  and  $\tilde{\mathbf{a}}$  using the sensor models (10a) and (10b). In summary, we can then write

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

with  $\mathbf{u} = [\tilde{\omega}^T \quad \tilde{\mathbf{a}}^T]^T$ , and  $\mathbf{w}$  collecting all the driving white noise processes

$$\mathbf{w} = [\mathbf{w}_g^T \quad \mathbf{w}_{bg}^T \quad \mathbf{w}_a^T \quad \mathbf{w}_{ba}^T]^T.$$

### 4.3 State and Parameter Estimation

We use the maximum likelihood principle to estimate the state trajectory and the calibration parameters of the sensor unit. In other words we select the state and the parameters such that the probability of observing the measurements that have occurred is maximal. This leads to estimates which are asymptotically unbiased and attain the Cramér-Rao lower bound [58].

Consequently, we aim to solve the following optimisation problem:

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

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

$$\begin{aligned} \hat{\mathbf{X}} &= \{\hat{\mathbf{x}}(t_0), \hat{\mathbf{x}}(t_1), \dots, \hat{\mathbf{x}}(t_K)\} \\ \mathbf{Z} &= \{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_K\} \\ \mathbf{U} &= \{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_L\}. \end{aligned}$$

#### 4. Calibration

Table 1.: The states and the parameters which are estimated during calibration.

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

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

Since the process noise terms  $\mathbf{w} = [\mathbf{w}_g^T \quad \mathbf{w}_{bg}^T \quad \mathbf{w}_a^T \quad \mathbf{w}_{ba}^T]^T$  are white  $\mathbf{x}(t)$  is Markov (strictly, this is only true later once we linearised around a nominal state trajectory and nominal parameters, since  $\mathbf{w}$  enters the system multiplicatively). Furthermore, since the camera measurement noise process  $\mathbf{v}$  is independent, the camera measure-

ments at time instance  $k$  only depend on the value of  $\mathbf{x}$  at that time  $t_k$ . We can therefore factor the joint pdf in (14) as follows:

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

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

#### 4.3.1 Nominal Trajectory and Parameters

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

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

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

where the  $\boxplus$  operator is simply component wise addition, except for the quaternions where quaternion multiplication is used.

#### 4.3.2 Minimal State and Parameter Representation

The attitude of the sensor system  $B$  and two calibration parameters, the camera extrinsic orientation and the target orientation, are represented with unit quaternions; a non-minimal representation in  $\mathbb{S}^3$ .

#### 4. Calibration

We therefore employ minimal coordinates  $\delta\chi_k \in \mathbb{R}^{15}$  and  $\delta\eta \in \mathbb{R}^{34}$  and mappings  $\Psi_x$  and  $\Psi_\theta$  to transform from minimal coordinates to tangent space [52]:

$$\delta\mathbf{x}_k = \exp(\Psi_x(\delta\chi_k)) \quad (17a)$$

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

The error state dynamics can then be written as:

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

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

##### 4.3.3 ML Calibration as a Least Squares Problem

Instead of maximising the joint pdf (15) directly we minimise its negative logarithm, the “negative log-likelihood function”. Since the system is linearised and the different noise sources are modelled as independent Gaussians, minimising the log-likelihood becomes a least squares problem:

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

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

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

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

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

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

For the state or “inertial” residuals we write

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

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

The derivative of the state transition with respect to the calibration parameters,  $\mathbf{J}_{k-1}$ , would be zero if only the extrinsics or camera

#### 4. Calibration

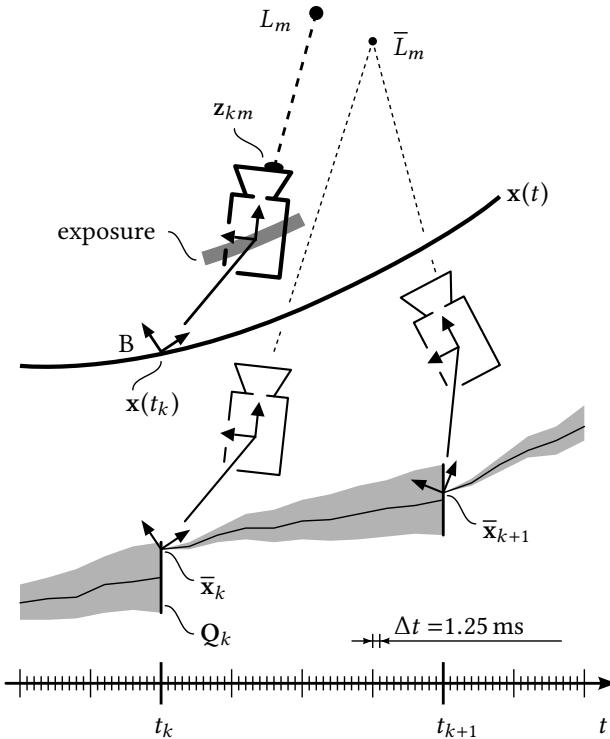


Figure 3.: Illustration of the maximum likelihood state and parameter estimator for calibration. At every time instance  $t_k$  a camera frame is captured, providing  $M$  target point observations  $\mathbf{z}_{km}$ . In-between frames the state is integrated starting from a nominal state trajectory  $\bar{\mathbf{x}}_k$  and nominal parameters  $\theta$ , resulting in the visual and inertial error terms in (19).

intrinsics were estimated. If IMU intrinsic parameters are estimated  $J_{k-1}$  is non-zero. We compute the Jacobians with respect to  $\delta\eta$  at each IMU sample time and post multiply them with the remaining state

transition matrices up to time  $t_k$  to obtain  $\mathbf{J}_{k-1}$ . The derivatives with respect to the time delay parameter  $\Delta t_{BC}$  is computed in exactly the same manner.

Finally, we normalise the weighted linear least squares problem (19) by pre-multiplication with the matrix square root of the inverse of the weighting matrices  $\mathbf{R}$  and  $\mathbf{Q}_k$  [19]. For the residuals, we write

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

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

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

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

with

$$\mathbf{A}' = \begin{bmatrix} \bar{\Phi}'_0 & -\mathbf{I}'_1 & & \mathbf{J}'_0 \\ & \ddots & & \vdots \\ & & \bar{\Phi}'_{K-1} & -\mathbf{I}'_K & \mathbf{J}'_{K-1} \\ \mathbf{H}'_{\boldsymbol{\chi}_0} & \mathbf{H}'_{\boldsymbol{\chi}_1} & & \mathbf{H}'_{\boldsymbol{\eta}_0} \\ & & \ddots & \mathbf{H}'_{\boldsymbol{\eta}_1} \\ & & & \mathbf{H}'_{\boldsymbol{\chi}_K} & \mathbf{H}'_{\boldsymbol{\eta}_K} \end{bmatrix} \quad (25)$$

## 5. Experiments and Results

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

### 4.3.4 Algorithm

To find the maximum likelihood solution for the full non-linear problem we successively linearise, compute the Jacobians, and solve (25) to obtain the least squares solution for  $\delta\chi$  and  $\delta\eta$ . Since (25) is large but very sparse we use a sparse QR solver [18]. The state trajectory and calibration parameter estimates are then updated and the system re-linearised. This procedure is then iterated for a fixed number of iterations (see Section 5).

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

## 5. Experiments and Results

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

the *precision* of the estimator. Biases introduced through modelling errors are in these cases not assessed. In addition, the results were compared quantitatively with a competing, state of the art parametric calibration method (an extended version of Kalibr, denoted here as *EKalibr* [71]). We refer to our method as discrete maximum likelihood (DML) here.

Section 5.1 describes the experimental setup that was used for the experiments. Section 5.2 reports the inertial sensor noise model parameters that were used in the estimator and how they were determined. Section 5.3 reports calibration results for the camera-IMU extrinsics, Section 5.4 analyses the temporal alignment, and Section 5.5 the IMU intrinsic parameter estimates.

## 5.1 Experimental Setup

We now describe the setup that was used to conduct the experiments. It is important to highlight that temperature variation has an important effect on most inertial sensors [91]. However, this is outside the scope of this report; calibration is performed at a constant temperature only.

### 5.1.1 Hardware Setup

We used a visual-inertial sensor unit that provided hardware synchronised measurements from two cameras and three IMUs. Fig. 4 depicts the unit. A factory calibrated ADIS16448 MEMS IMU from Analog Devices [16] and the two consumer grade MPU9150 chip MEMS IMUs from Invensense [63] provided gyroscope and accelerometer measurements at a rate of 800 Hz each. The cameras consisted of MT9V034 global shutter CMOS chips from Aptina in conjunction with BM2820 S-mount lenses with a diagonal field of view of 122°. The cameras were intrinsically calibrated prior to data collection using the camera intrinsic calibration functionality of Kalibr. Images were captured at a rate of 20 frames per second per camera. The cameras were pre-

## 5. Experiments and Results

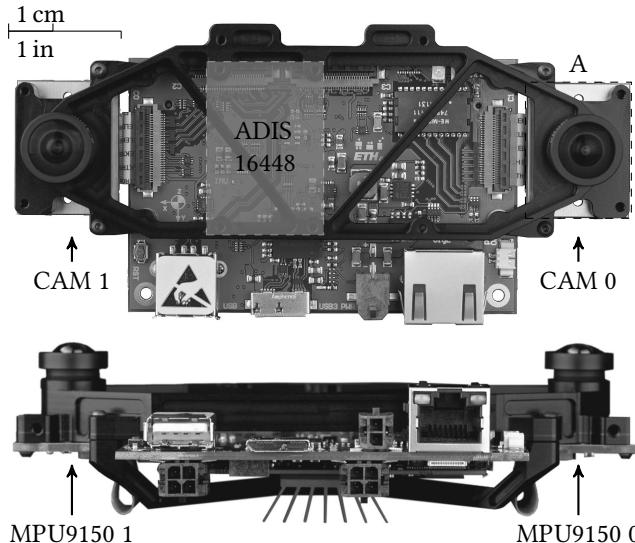


Figure 4.: Visual-inertial sensor unit [64] used for assessing the accuracy and precision of the proposed calibration method. An MPU9150 MEMS IMU is located behind each of the cameras, and a factory calibrated ADIS16448 IMU was attached to the unit. Photo: François Pomerleau.

triggered in order to temporally align all sensor data to the center of the camera exposure time – a core feature of the sensor unit’s FPGA firmware developed by Rehder [64].

### 5.1.2 Data Collection

The sensor unit was moved in front of a checkerboard with  $6 \times 7$  corners and a corner spacing of 60 mm for dataset collection. A total of 50 datasets with a length 20 s each were captured in sequence and stored for post-processing.

### 5.1.3 Estimator

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

### 5.1.4 Kalibr

Kalibr was run in its extended version [71]. It was configured to use the same measurements and the exact same noise parameter settings that were used in the estimator presented here.

## 5.2 Noise Model Parameters

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

It is often difficult to obtain these parameters unambiguously from data sheets particularly for low-cost devices. We therefore used the method presented in [65] to automatically determine the IMU noise model parameters for all axes of each sensor. The models were identified from sensor data captured when the sensor unit was at rest at constant room temperature. An alternative procedure to obtain these parameters using the Allan variance is presented in [38]. This procedure can lead to similar results but it is a manual method.

We picked the noise model parameters that corresponded to the worst performing sensor axis per gyroscope and accelerometer triad and then used these parameters for all axes. This is not strictly necessary and selecting individual noise model parameters instead would not increase the computational complexity. This is the case since  $Q_k$

## 5. Experiments and Results

Table 2.: Identified gyroscope and accelerometer noise model parameters for the ADIS16448 and the MPU9150 MEMS IMUs as used for the experiments.

	<b>ADIS</b>	<b>MPU9150</b>	<b>unit</b>
<b>Gyroscopes</b>			
White noise str.	$\sigma_g$	$1.86 \times 10^{-4}$	$8.94 \times 10^{-5}$ rad/(s $\sqrt{\text{Hz}}$ )
Bias diffusion	$\sigma_{bg}$	$2.66 \times 10^{-5}$	$1.08 \times 10^{-5}$ rad/(s <sup>2</sup> $\sqrt{\text{Hz}}$ )
Bias corr. time	$\tau_b$	$\infty$	s
<b>Accelerometers</b>			
White noise str.	$\sigma_a$	$1.86 \times 10^{-3}$	$2.24 \times 10^{-3}$ m/(s <sup>2</sup> $\sqrt{\text{Hz}}$ )
Bias diffusion	$\sigma_{ba}$	$4.33 \times 10^{-4}$	$7.53 \times 10^{-5}$ m/(s <sup>3</sup> $\sqrt{\text{Hz}}$ )
Bias corr. time	$\tau_b$	$\infty$	s

in (23) cannot be pre-computed in general. Fig. 5 shows the sample Allan deviation (grey) for the gyroscopes and accelerometers of the MPU9150 (MPU 0). The synthetic Allan deviations corresponding to the model parameters used by the estimator are shown in black (solid). The model's expressive power is not large enough to capture the long-term bias fluctuations accurately but captures the noise behaviour well for short correlation times. Table 2 lists the identified noise model parameters used for the gyroscopes and accelerometers by the estimator. Note that the MPU9150, a  $4 \times 4 \times 1$  mm consumer IMU, compares favourably with the ADIS16448 in terms of noise performance.

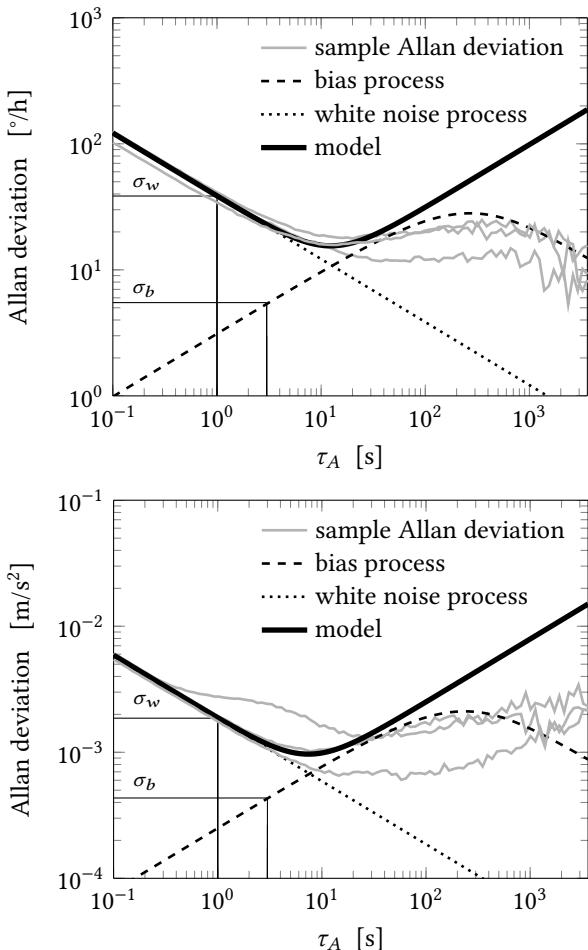


Figure 5.: Allan deviation of the gyroscopes (top) and accelerometers (bottom) of the MPU9150 MEMS IMU. Conventional sample Allan deviations [3] for each sensor axis are shown in grey and Allan deviations corresponding to the noise models used for the experiments are shown in black.

### 5.3 Extrinsics

The translation and the rotation between camera and IMU are key parameters in visual-inertial sensor systems. The estimates for the corresponding calibration parameters are shown in Table 3. The column “intr. off” refers to the estimates when sensor intrinsic estimation is disabled (but not the delay, which is always estimated), and “intr. on” to when they were enabled.

The results for the MPU 0 and MPU 1 show that even for short calibration sequences in the order of 20 s, a precision of  $< 1 \text{ mm}$  on all axes in translation and  $< 1 \text{ mrad}$  in rotation can be achieved if IMU intrinsics are estimated (see column “intr. on” in Table 3). The precision for the ADIS is lower which is most likely due to i) its poorer noise performance and ii) the fact that the ADIS relies on separate accelerometers which results in “size effects” [35] (which we do not model).

Fig. 6 depicts the printed circuit board (PCB) of the visual-inertial sensor units camera-IMU module (section “A” in Fig. 4) with the MPU9150 MEMS IMU and the Aptina MT9V034 CMOS camera chip (on top layer, not visible). Extrinsic translation estimates in  $x$  and  $y$  from all 50 calibration experiments, expressed in the camera reference frame, are indicated for calibration with ( $\Delta$ ) and without ( $\nabla$ ) IMU intrinsics. These results indicate that the extrinsics are estimated more accurately when the IMU intrinsics are taken into consideration, removing a significant bias and improving the precision of the estimates significantly. We assume that the accelerometers of the MPU9150 are not located in the center of the sensor package which could explain the offset of 1.87 mm (in mean) along the  $y$  axis.

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

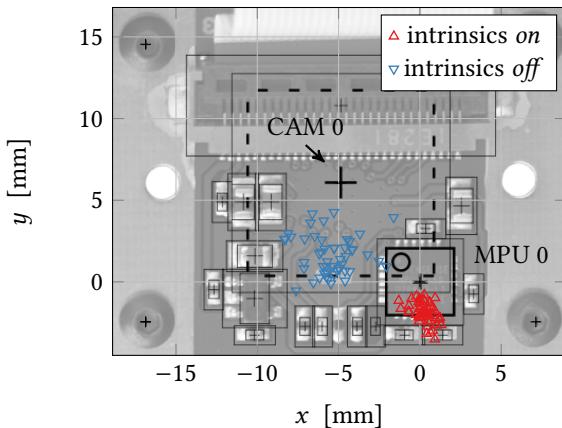


Figure 6.: Bottom view of the camera-IMU printed circuit board of the visual-inertial sensor unit (section “A” in Fig. 4). Translation extrinsics estimates, expressed in the camera reference frame, are indicated for calibration with ( $\Delta$ ) and without ( $\nabla$ ) IMU intrinsics.

parameters.

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

## 5.4 Time Delay

Table 4 summarises the time delay estimates we obtained using the method presented here (DML) with and without IMU intrinsics. It also shows the results obtained with EKalibr. Outliers from the EKalibr estimates were removed prior to computing the standard deviations

## 5. Experiments and Results

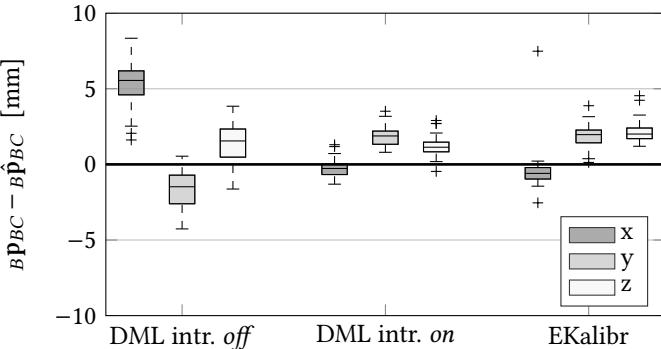


Figure 7.: Camera-IMU extrinsic translation estimation error statistics using ground truth from CAD data to assess the accuracy of the calibration. The error statistics correspond to the translation between CAM 0 and MPU9150 0 and refer to the center of the MPUs sensor package.

across all calibration experiments.

The precision of the time delay estimates increases when IMU intrinsic parameters are incorporated into the calibration problem (“intrinsic *on*”). It is in this case in the order of 10 $\mu$ s which corresponds to less than one-hundredth of the IMU sampling time of 1.25 ms. This is surprising given i) the different nature of the estimators, and ii) that different effects are lumped into the time delay parameter. It also highlights the importance of a precise temporal alignment between the camera and the gyroscopes. This is particularly true for systems that exhibit fast dynamics or are subject to vibrations.

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

## 5.5 IMU Intrinsicics

The estimates of the IMU intrinsic calibration parameters are shown in Table 5. The misalignment and scale factor errors of the MPU’s gyroscopes are in the order of up to 1 % and 1.5°, respectively, which is significant. The misalignment and scale factor errors of the MPU’s accelerometers appear to be slightly lower. The estimation results for the g-sensitivities of the gyroscopes (not shown) are less conclusive.

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

## 6. Conclusion and Future Work

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

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

## *Acknowledgements*

# **Acknowledgements**

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

Table 3.: Results for the extrinsic calibration parameter estimates for the MPU9150 0, MPU9150 1, and the ADIS16448, with respect to CAM 0. The standard deviations ( $\pm 1\sigma$ ) are computed over 50 calibration datasets and give an indication of the precision of estimator, but not its accuracy.

<b>device</b>	<b>parameter</b>		<b>intr. off</b>	<b>intr. on</b>
<b>MPU 0</b>	$B\mathbf{p}_C$ [mm]	$x$	$0.60 \pm 1.46$	$-5.15 \pm 0.58$
		$y$	$4.41 \pm 1.19$	$7.92 \pm 0.63$
		$z$	$-19.33 \pm 1.25$	$-19.54 \pm 0.65$
	$\beta_{CB}$ [mrad]	$\beta x$	$4.80 \pm 1.26$	$6.78 \pm 0.61$
		$\beta y$	$-0.06 \pm 1.16$	$5.34 \pm 0.54$
		$\beta z$	$5.52 \pm 0.68$	$3.86 \pm 0.51$
<b>MPU 1</b>	$B\mathbf{p}_C$ [mm]	$x$	$0.63 \pm 1.49$	$-3.73 \pm 0.55$
		$y$	$118.25 \pm 1.06$	$113.04 \pm 0.60$
		$z$	$-26.78 \pm 0.71$	$-18.60 \pm 0.62$
	$\beta_{CB}$ [mrad]	$\beta x$	$-1.45 \pm 1.02$	$-7.65 \pm 0.52$
		$\beta y$	$0.51 \pm 1.37$	$5.23 \pm 0.43$
		$\beta z$	$6.70 \pm 0.37$	$7.43 \pm 0.53$
<b>ADIS</b>	$B\mathbf{p}_C$ [mm]	$x$	$34.02 \pm 1.84$	$34.58 \pm 1.91$
		$y$	$-11.95 \pm 1.35$	$-9.05 \pm 2.86$
		$z$	$0.95 \pm 0.92$	$1.69 \pm 3.52$
	$\gamma_{CB}$ [mrad]	$\gamma_x$	$0.95 \pm 1.16$	$1.42 \pm 2.45$
		$\gamma_y$	$2.98 \pm 1.27$	$4.42 \pm 1.86$
		$\gamma_z$	$7.37 \pm 0.53$	$5.77 \pm 1.62$

### *Acknowledgements*

Table 4.: Mean and standard deviation of the time delay estimates  $\Delta\hat{t}_{BC}$  of the MPUs with respect to CAM 0, obtained by the method proposed here (DML) and *EKalibr*.

<b>Method</b>	<b>MPU9150 0</b>	<b>MPU9150 1</b>
<b>DML intrinsics off</b>	$3.0226 \text{ ms} \pm 24.0 \mu\text{s}$	$2.9886 \text{ ms} \pm 15.2 \mu\text{s}$
<b>DML intrinsics on</b>	$3.0306 \text{ ms} \pm 9.6 \mu\text{s}$	$3.0486 \text{ ms} \pm 8.4 \mu\text{s}$
<b>EKalibr</b>	$3.0305 \text{ ms} \pm 15.5 \mu\text{s}$	$3.0321 \text{ ms} \pm 8.3 \mu\text{s}$

Table 5.: IMU intrinsic calibration parameter estimates (DML intrinsics *on*) for the MPU9150 0, the MPU9150 1, and the ADIS16448. Indicated are the mean and the standard deviation over 50 calibration experiments using the proposed method.

<b>Param.</b>		<b>ADIS</b>	<b>MPU0</b>	<b>MPU1</b>
<b>Gyroscopes</b>				
$\Delta k_g$	$k_x$	$0.68 \pm 2.71$	$2.18 \pm 0.59$	$-7.45 \pm 0.56$
[‰]	$k_y$	$-0.54 \pm 2.95$	$\underline{-9.77 \pm 0.51}$	$-2.11 \pm 0.59$
	$k_z$	$2.28 \pm 1.16$	$2.33 \pm 0.25$	$-3.00 \pm 0.24$
$M_g$	$\gamma_x$	$7.83 \pm 3.19$	$1.36 \pm 0.42$	$-1.78 \pm 0.44$
[mrad]	$\gamma_y$	$0.73 \pm 2.11$	$8.81 \pm 0.59$	$9.54 \pm 0.47$
	$\gamma_z$	$1.28 \pm 3.01$	$-16.30 \pm 0.47$	$\underline{24.64 \pm 0.42}$
<b>Accelerometers</b>				
$\Delta k_a$	$k_x$	$21.60 \pm 6.46$	$-1.39 \pm 1.51$	$-5.93 \pm 1.93$
[‰]	$k_y$	$-8.45 \pm 13.72$	$6.19 \pm 1.03$	$1.39 \pm 0.89$
	$k_z$	$-3.41 \pm 7.03$	$5.74 \pm 1.58$	$6.37 \pm 1.80$
$M_a$	$\alpha_{xz}$	$3.12 \pm 6.42$	$0.55 \pm 2.39$	$-0.31 \pm 2.41$
['']	$\alpha_{xy}$	$1.51 \pm 4.47$	$-0.33 \pm 2.17$	$5.73 \pm 2.61$
	$\alpha_{yx}$	$40.20 \pm 16.73$	$-2.38 \pm 1.38$	$-2.02 \pm 1.25$
	$\alpha_{yz}$	$-7.92 \pm 15.19$	$-7.29 \pm 1.16$	$5.07 \pm 1.28$
	$\alpha_{zy}$	$-0.79 \pm 7.65$	$-5.31 \pm 1.01$	$-4.39 \pm 1.37$
	$\alpha_{zx}$	$-10.44 \pm 7.51$	$9.57 \pm 1.39$	$-15.63 \pm 1.03$

# Paper III

## A Synchronized Visual-Inertial Sensor System with FPGA Pre-Processing for Accurate Real-Time SLAM

Janosch Nikolic · Joern Rehder · Micahel Burri ·  
Pascal Gohl · Stefan Leutenegger · Paul Furgale ·  
Roland Siegwart<sup>1</sup>

### Abstract

Robust, accurate pose estimation and mapping at real-time in six dimensions is a primary need of mobile robots, in particular flying Micro Aerial Vehicles (MAVs), which still perform their impressive maneuvers mostly in controlled environments. This work presents a visual-inertial sensor unit aimed at effortless deployment on robots in order to equip them with robust real-time Simultaneous Localization and Mapping (SLAM) capabilities, and to facilitate research on this important topic at a low entry barrier.

---

<sup>1</sup>Joern Rehder and Janosch Nikolic contributed equally to this work.

### *Paper III*

Up to four cameras are interfaced through a modern ARM-FPGA system, along with an Inertial Measurement Unit (IMU) providing high-quality rate gyro and accelerometer measurements, calibrated and hardware-synchronized with the images. This facilitates a tight fusion of visual and inertial cues that leads to a level of robustness and accuracy which is difficult to achieve with purely visual SLAM systems. In addition to raw data, the sensor head provides FPGA-pre-processed data such as visual keypoints, reducing the computational complexity of SLAM algorithms significantly and enabling employment on resource-constrained platforms.

Sensor selection, hardware and firmware design, as well as intrinsic and extrinsic calibration are addressed in this work. Results from a tightly coupled reference visual-inertial motion estimation framework demonstrate the capabilities of the presented system.

Published in the proceedings of the *IEEE International Conference on Robotics & Automation (ICRA)*, 2014.

©2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The final publication is available at <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6906892>, DOI 10.1109/ICRA.2014.6906892.

## 1. Introduction

MANY mobile robots require on-board localization and mapping capabilities in order to operate truly autonomously. Control, path planning, and decision making rely on a timely and accurate map of the robots surroundings and on an estimate of the state of the system within this map.

Accordingly, Simultaneous Localization and Mapping (SLAM) has been an active topic of research for decades [20]. Tremendous advances led to successful employments of SLAM systems on all sorts of platforms operating in diverse environments. Different interoceptive and exteroceptive sensors such as 2D and 3D laser scanners, wheel odometry, cameras, inertial sensors, ultrasonic range finders, and radar, amongst others, provide the necessary data.

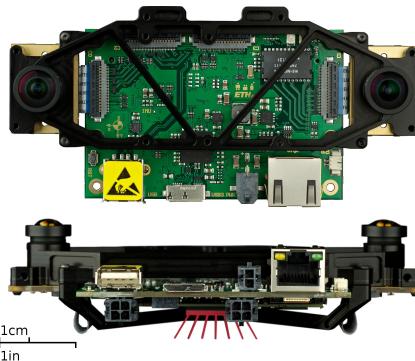


Figure 1.: The *SLAM Sensor* unit in a fronto-parallel “stereo” configuration (front- and side-view). The sensor interfaces up to four cameras and incorporates a time-synchronized and calibrated inertial measurement system. Access to high quality raw- and pre-processed data is provided through simple interfaces.

### *Paper III*

Yet it is often a challenge to equip a platform with a reliable and accurate real-time SLAM or state estimation system that fulfills payload, power, and cost constraints. A “plug-and-play” SLAM solution that achieves all requirements and runs robustly under the given conditions is seldom readily available, and thus significant engineering efforts often have to be undertaken.

Visual SLAM systems that rely on cameras have received particular attention from the robotics and computer vision communities. A vast amount of data from low-cost, lightweight cameras enables incredibly powerful SLAM or structure-from-motion (SfM) systems that perform accurate, large-scale localization and (dense) mapping in real-time [49, 80]. However, SLAM algorithms that rely only on visual cues are often difficult to employ in practice. Dynamic motion, a lack of visible texture, and the need for precise structure and motion estimates under such conditions often renders purely visual SLAM inapplicable.

Augmenting visual SLAM systems with inertial sensors tackles exactly these issues. MEMS Inertial Measurement Units (IMUs) provide valuable measurements of angular velocity and linear acceleration. In *tight combination* with visual cues, this can lead to more robust and accurate motion estimation systems that are able to operate in *less controlled, sparsely textured, and poorly illuminated* scenes while undergoing dynamic motion. However, this requires all sensors to be well calibrated, rigidly connected, and precisely time-synchronized.

This work makes a step towards a *general-purpose SLAM system* by providing these capabilities. The sensor head evolved through the development of several prototypes and was tested in many applications, for instance in a coal fired power plant [67] or on a car [53]. Fig. 1 shows our final hardware iteration.

The remainder of this article is organized as follows: in Section 3, we outline the design concept, FPGA-pre-processing (see Section 3.3), and the calibration of such a visual-inertial sensor unit (see Section 4). We provide an overview of our reference tightly coupled visual-

## *2. Related Work*

inertial motion estimation framework in Section 5, which we use in Section 6 to illustrate the capabilities of the sensor system.

## **2. Related Work**

There exist different FPGA vision systems particularly geared to robotics. The GIMME platform [2] is similar in scope to this hardware in that it computes visual interest points on an FPGA and transmits those to a host system in order to bring visual pose estimation to platforms with computational and power constraints. However, it is a purely visual sensor setup and hence does not require elaborated synchronization or calibration between different types of sensors.

Another system that employs inertial sensors has been developed by the DLR [74]. In this system, a general purpose computer and an FPGA are closely interleaved in order to enable ego-motion estimation and depth computation on a handheld device. In contrast to our setup, cameras and inertial sensors are not as tightly integrated into the system, and images appear to be timestamped at the start of sensor exposure, resulting in a varying, exposure dependent offset to IMU measurements. Furthermore, its weight might prohibit application on very payload-constrained platforms.

As heterogeneous sensor systems for motion estimation and localization become increasingly popular, spatial calibration has attracted some attention and resulted in a variety of frameworks [22, 44, 59]. More recently, the importance of accurate synchronization of the sensors became apparent and was addressed in [45, 22, 25]. While this work makes use of the calibration presented in [25] to determine the transformation between cameras and IMU and to determine fixed delays present when polling inertial data, its approach to the problem is exactly antithetic: rather than connecting a set of stand-alone sensors to a general purpose computer and calibrating for potentially

time-variant time-offsets afterwards, we pursued a tight integration of all hardware components with a central unit capable of concurrent triggering and polling of all sensors.

### **3. The Visual-Inertial SLAM Sensor**

This section outlines important design concepts and “lessons learned” throughout the development of three successive prototypes that led to the sensor presented here.

Subsection 3.1 provides a conceptual overview of the sensor. Subsection 3.2 describes a synchronization method that guarantees ideal alignment of all sensors in time. Subsection 3.3 describes the FPGA implementation of image processing operations such as keypoint detection to reduce CPU-load of successive SLAM software.

#### **3.1 Sensor Design Concept**

At the core of the SLAM sensor, we employ a modern XILINX Zynq System-on-Chip (SoC), a device that combines FPGA resources with a dual ARM Cortex-A9 on a *single chip*. Hardware programmability allows a direct, lowest-level interface to the CMOS imagers and inertial sensors, enabling precise synchronization and a reliable data acquisition process.

At the same time, the chip offers a powerful, industry standard CPU running Linux. This facilitates simple development and efficient execution of processes that are time-consuming to implement on an FPGA (e.g. host-communication or even a simple SLAM framework). In contrast to previous prototypes which featured a XILINX Spartan-6 FPGA - Intel ATOM combination, this also offers a better integration and a higher bandwidth between logic and CPU. Fig. 2 gives an overview of the hardware architecture, and Table 1 summarizes the

### 3. The Visual-Inertial SLAM Sensor

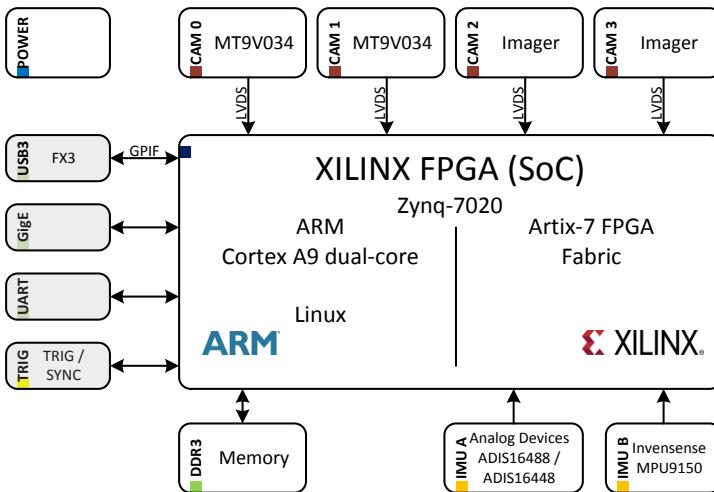


Figure 2.: Block-diagram of the SLAM sensor hardware architecture. Camera chips and inertial sensors interface the ARM-FPGA system-on-chip directly. Standard interfaces provide fast access to the data provided by the module.

technical specifications of the sensor unit.

#### 3.1.1 Visual Subsystem

The SLAM sensor offers *four* camera extension ports. The ability to rely on several cameras is crucial for many real-world applications. Even when wide Field-of-View (FoV) optics are used, a single camera may still point into a direction where keypoint tracking is difficult (lack of texture, temporary obstruction of the FoV, bad illumination). With the option to use four cameras simultaneously, configurations such as the combination of a “fronto-parallel” stereo pair and two fish-eye modules are quickly realized.

Basic Characteristics	Value	Unit
Mass (for diff. configurations)		
1 cam+MPU	60	g
2 cams+mount+ADIS16448	130	g
4 cams+mount+ADIS16488	185	g
Embedded Processing	XILINX Zynq 7020	
Processor	2xARM Cortex A9	
FPGA	ARTIX-7	
Interfaces	GigE, USB2/3	
<b>Camera System</b>		
Sensors	Aptina MT9V034	
Shutter type	global	
Opt. resolution	752×480	pixel
Max. frame rate	60	fps
<b>Inertial System</b>		
Rate Gyroscope		
Measurement Range	$\pm 1000$	$^{\circ}/\text{s}$
Noise Density	0.007	$^{\circ}/\text{s Hz}^{-1/2}$
Accelerometer		
Measurement Range	$\pm 177$	$\text{ms}^{-2}$
Noise Density	$0.66 \cdot 10^{-3}$	$\text{ms}^{-2} \text{ Hz}^{-1/2}$
Max. sampling rate	2.4	kHz

Table 1.: Overview of the SLAM sensors technical specifications. High quality sensors that perform well in low-light scenarios and when undergoing dynamic motion are integrated in the sensor unit. The module's relatively low weight facilitates employment on payload-constrained platforms.

In the current configuration, camera chips were selected according to their *low-light sensitivity* and *global shutter* functionality. Aptina's MT9V034 CMOS sensors offer good performance and a direct interface

### 3. The Visual-Inertial SLAM Sensor

to the FPGA through LVDS ports. By default, Lensagon lenses of the type BM2820 (122° diagonal FoV) or BM2420 (132° FoV) are used.

In addition, a FLIR Tau 2 *thermal imager* can be connected, which then occupies one of the camera ports. Similar to the camera modules, it directly interfaces with the Zynq providing time-synchronized digital (14 bit dynamic range) thermal images to the host.

#### 3.1.2 Inertial Subsystem

The current prototype allows two options with regard to the IMU subsystem. By default, each camera module is fitted with a low-cost MEMS IMU offering a triple axis gyroscope, accelerometer, and magnetometer in a single package. The MPU-9150 was selected due to its high range in both angular rates and acceleration. Chip internal filtering and processing are switched off, and only raw data is used.

In addition, a factory-calibrated MEMS IMU system from the ADIS family of Analog Devices can be connected. The ADIS16448 and ADIS16488 are equipped with higher quality gyroscopes and accelerometers, and they are factory-calibrated over a large scale and temperature range. Depending on the application, one can trade-off sensor weight versus accuracy of the inertial subsystem.

## 3.2 Sensor Synchronization and Data Acquisition

We configure the image sensors for external triggering. At the same time, the inertial sensors are polled for data acquisition. As stated earlier, accurate synchronization of different sensors was the driving motivation for a tight integration in hardware. It is an established fact in photogrammetry, that images should be timestamped by their mid-exposure time, and in previous work [25], it could be shown that neglecting image exposure time in timestamping data has an observable effect, which suggests that it could adversely affect image-based state estimation. We made the design choice to not correct for

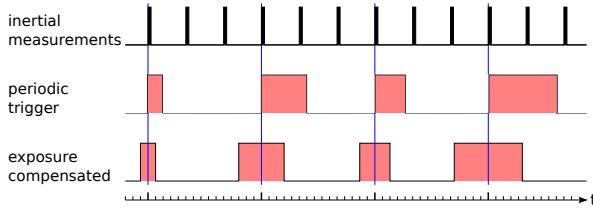


Figure 3.: This timing diagram shows strictly periodic polling of an IMU as well as two schemes of camera synchronization, where high levels mark exposure times. Triggering the camera at the instance an inertial measurement is retrieved is a common approach to synchronization. However, the exposure is asymmetrical with respect to the inertial measurement. By taking varying exposure into account and shifting each triggering instance accordingly, significantly improved synchronization can be achieved, as demonstrated in Fig. 6.

the exposure time in timestamping images, but to account for the exposure time when triggering the sensors. This way, the middle of the exposure times will still be equally spaced despite varying lighting conditions, which exhibits certain advantages when representing states in a time-discrete manner. Fig. 3 illustrates the synchronization scheme in comparison with periodically triggering, where varying lighting conditions result in exposure midpoints that are not equally spaced.

Note that also the inertial measurements may exhibit a delay. This delay is in general fixed and can be a combination of communication, filter and logic delays. Section 4 will detail on estimating this delay, which is compensated for in the same way the exposure delay is addressed, by moving the moment when a polling request is initiated with respect to the point in time when the measurement is timestamped. As part of the results section, Fig. 6 reproduces an experiment from [25]. The results demonstrate that the delays can be

### 3. The Visual-Inertial SLAM Sensor

accounted for in the sensor data acquisition, thereby improving the synchronization between sensors significantly.

### 3.3 FPGA Accelerated Image Processing

As depicted in Fig. 7, the detection of interest points consumes a significant share of the processing time in the state estimation pipeline. At the same time, many interest point detectors operate on a rather confined neighborhood of pixels and can be implemented exclusively using fixed-point arithmetic, which renders them well suited for an implementation as dedicated logic blocks inside an FPGA. For this project, a fixed-point version of the Harris corner detector [31] as well as the FAST corner detector [73] have been implemented. While the resources of the FPGA used in the setup are not sufficient to integrate them both at the same time, it is possible to load the FPGA with different configurations depending on the requirements of the experiment. Note that the quantities reported in Table 2 have been acquired for an earlier prototype based on the Xilinx XC6SLX45T.

*Harris Corner Detection:* The Harris corner detector is based on an approximation of the auto-correlation function for small image patches. With  $I_x$  denoting the derivative in x-direction of the image intensity at pixel  $x + u, y + v$ , and  $w(u, v)$  denoting a weighted averaging function, the approximated local auto-correlation is calculated as [31]

$$\mathbf{A}(x, y) = \sum_u \sum_v w(u, v) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}. \quad (1)$$

With  $\mathbf{A}$  and a weighting factor  $k$ , the corner response function  $r$  is calculated as

$$r = |\mathbf{A}| - k \operatorname{tr}(\mathbf{A})^2. \quad (2)$$

Larger positive values of this function correspond to corner regions,

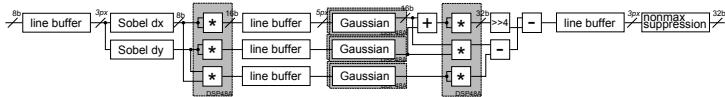


Figure 4.: Block diagram illustrating the implementation of the Harris corner detector. For improved readability, only the width of the topmost path is shown, which also applies to any other path in the same column. Note that bit widths are shown in oblique font, while line buffer widths in terms of numbers of pixels of the respective input bit width are displayed in italics. Unless marked otherwise, bit widths propagate through blocks. For operations that potentially lead to an overflow, a saturation operation is performed. Each gray box illustrates the location of a single DSP slice within the processing pipeline.

while negative results indicate edges. Flat regions trigger a small response. Examining this function reveals pixel differencing operations, cascaded multiplications as well as local averaging. Fig. 4 depicts the FPGA implementation of the corner score function. Derivatives of image intensities are computed by means of Sobel filters, while local averaging is performed by convolution with a Gaussian kernel. As in [2], weighting the Trace of the matrix in the cost function has been realized by a bit shift operation. Individual blocks like Sobel and Gaussian filters as well as the multipliers in the pipeline operate at higher frequency than the pipeline itself—25 MHz and 125 Mhz respectively—allowing for the re-utilization of resources. Furthermore, by making use of the separability properties of Sobel and Gaussian filters, resource utilization can further be reduced. The resulting resource utilization is shown in Table 2. The maximum clock rate is limited and thus imposes upper bounds on the degree to which resources can be shared. However, the pixel rate of the sensors used in this sensor setup allows for a excessive re-utilization of resources, resulting in a core that can be conveniently duplicated for four cameras without exceeding the area of the FPGA.

### 3. The Visual-Inertial SLAM Sensor

*FAST Corner Detection:* The FAST corner detector is a heuristically motivated approach to interest point detection, which compares intensities of image points on a circle around the point in question. It identifies a pixel as an interest point based on the number of pixels in a segment that is either coherently lighter or darker than the central element. In [73], different scores for nonmaximum suppression are proposed. Taking the mere number of coherent intensity comparisons can be efficiently implemented, but results in a rather coarsely quantized score. On the other hand, considering the sum of absolute differences (SAD) of this segment with the center pixel yields finer granularity in the score at the expense of occupying a larger area on-chip. In this project, both scores have been implemented with the resource utilization displayed in Table 2. Fig. 5 illustrates a detail of the implementation as a block diagram, which depicts the path testing for lighter pixels, which is duplicated for the test for darker pixels. The central and surrounding pixel, grouped in sets of four consecutive elements, feed into the block. The design heavily employs identical blocks, which are only shown in a number sufficient to convey the underlying interconnection principles. As for the Harris implementation, individual components of the detector are clocked at a higher rate than the overall pipeline, resulting in a reduction in resource utilization. To this end, the comparison with the central pixel is executed in four clock cycles, decreasing the number of comparators that operate on image data. Counting of coherent segment lengths is done for each potential starting point of the segment in parallel. The appropriate signal connecting the counting units with the registers holding the intensity comparisons are represented by a routing network block in the schematic. Per clock cycle, each segment length counter evaluates four comparisons. To this end, the counter block depicted in Fig. 5 determines the position of the first zero in the 4 bit segment, and accumulates these. Once the coherency of a segment is interrupted, further accumulations are blocked. In order to determine the maximum coherent segment length from the parallel counter

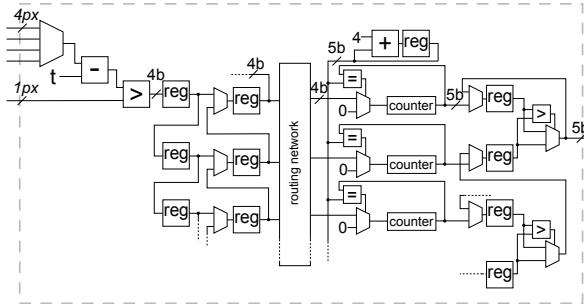


Figure 5.: Logic diagram of a detail of the fast implementation. By reusing blocks, the area footprint of the core can be reduced significantly.

Table 2.: Resource utilization of the implemented interest point detectors for a WVGA image on a Xilinx Spartan 6 architecture. The number in brackets indicates the device utilization for a Xilinx XC6SLX45T.

	<b>RAMB16B</b>	<b>DSP48A</b>	<b>Slices</b>
<b>Harris</b>	17 (14%)	8 (13%)	774 (11%)
<b>FAST</b>	5 (4%)	0 (0%)	1,124 (16%)
<b>FAST+SAD</b>	5 (4%)	0 (0%)	1,913 (28%)

units, a recursive comparator structure has been implemented. The comparison for darker is implemented accordingly and results from the two paths which are fused using an additional comparator stage. The figure does not depict the extraction of the central pixel and the surrounding circle that precedes the block shown, as well as the non-maximum suppression succeeding the block. Note that Fig. 5 depicts the case where the mere segment length is employed.

## 4. Calibration

In order to achieve accurate motion estimates, the sensor setup needs to be calibrated. As a factory calibrated IMU is employed in the setup, the remaining quantities that need to be estimated are

- the camera intrinsics,
- the extrinsics of the stereo setup,
- the transformation between the cameras and the IMU,
- and the fixed time delay between camera and IMU measurements.

The camera intrinsics and stereo extrinsics are determined from a set of stills of a checkerboard using the well-established camera calibration toolbox by Bouguet<sup>2</sup>. The toolbox is based on the pinhole camera model and employs the radial-tangential distortion model established by Brown [10].

The transformations between the cameras and the IMU as well as the time delay is estimated using a unified framework, presented in our previous work [25] and specifically developed to estimate filter and communication delays to deliver optimal synchronization of sensors. The framework is based on the idea of parameterizing time-variant quantities as B-splines—introduced in detail in [24]—and solving for these as well as a set of time-invariant calibration parameters in a batch optimal fashion. Apart from requiring fewer parameters when fusing measurements of significantly different rates such as images and inertial data, this approach allows for an accurate estimation of the fixed time delay between camera and IMU. Like other frameworks [59, 44], the calibration procedure requires waving the setup in front of a checkerboard, while exciting all rotational

---

<sup>2</sup> Available at [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/)

degrees of freedom sufficiently in order to render the displacement of camera and IMU well observable. We also experimented with incorporating the calibration for the stereo extrinsics directly into the unified calibration framework, but observed degraded performance when used in visual-inertial SLAM, an explanation to which may be that the setups between calibration and SLAM vary (mostly as far as scene depth is concerned).

The calibration process describes the position and orientation of the IMU with respect to the world frame in continuous-time, which also includes a continuous-time representation of respective derivatives (velocity, acceleration, and angular velocity). Furthermore, both accelerometer and gyroscope biases—modeled as random walks—are represented as continuous-time quantities. The calibration may then be formulated as a batch optimization that combines reprojection error  $\mathbf{e}_y$  of checkerboard corners with errors on the acceleration  $\mathbf{e}_\alpha$  and  $\mathbf{e}_\omega$ , as well as terms concerning the compliance of the biases with the random walk processes ( $\mathbf{e}_{b_\alpha}$  and  $\mathbf{e}_{b_\omega}$ ).

## 5. Visual-Inertial Motion Estimation

Since the sensor was designed to perform real-time visual-inertial motion estimation, we applied our framework [53] to an outdoor dataset. In short, the method is inspired by recent advances in purely vision-based SLAM that solve a sparse non-linear least-squares problem. Such approaches optimize the reprojection error of a fairly large number of landmarks as observed by various camera frames. Our method tightly integrates inertial measurements into the cost function  $J$  by combining reprojection error  $\mathbf{e}_r$  with an IMU error term  $\mathbf{e}_s$  obtained from propagation using standard IMU kinematics in-between

## 5. Visual-Inertial Motion Estimation

successive image frames:

$$J(\mathbf{x}) := \sum_{i=1}^I \sum_{k=1}^K \sum_{j \in \mathcal{J}(i,k)} \mathbf{e}_r^{i,j,k T} \mathbf{W}_r^{i,j,k} \mathbf{e}_r^{i,j,k} + \sum_{k=1}^{K-1} \mathbf{e}_s^k T \mathbf{W}_s^k \mathbf{e}_s^k. \quad (3)$$

$\mathbf{x}$  denotes the variables to be estimated, composed of the states at all camera snapshot time steps  $k$ , as well as all the 3D positions of the landmarks. Note that the states cover not only 6D poses, but also the velocity as well as biases of both gyroscope and accelerometer sensors. The velocity is needed for state propagation in-between time steps, and the slowly varying biases are tracked in order to remove them from gyro and accelerometer readings.  $i$  stands for the camera index of the sensor assembly, and  $j$  for the landmark index. Landmarks visible in the  $i^{\text{th}}$  camera are summarized in the set  $\mathcal{J}(i,k)$ . Furthermore,  $\mathbf{W}_r^{i,j,k}$  denotes the information matrix of reprojection errors related to detection uncertainty in the image plane. Finally,  $\mathbf{W}_s^k$  represents the information of the  $k^{\text{th}}$  IMU error, as obtained from the IMU sensor noise models, provided by the manufacturer (see Table 1). We furthermore include the extrinsic calibration of the cameras in the optimization.

This fully probabilistically motivated batch optimization problem over all cameras and IMU measurements quickly grows intractable. We therefore bound the optimization window by applying the concept of marginalization. This allows us to keep a fixed number of keyframes that are arbitrarily spaced in time and that are still related to each other with (linearized) IMU error terms. Consequently, drift during stand-still is avoided, and nevertheless we are able to track dynamic motions.

## 6. Results

We now highlight a few results which we obtained with the visual-inertial sensor unit. Experimental results from sensor synchronisation are shown in Section 6.1, timing results are presented in Section 6.2, and an evaluation of the visual-inertial motion estimation framework is presented in Section 6.3.

### 6.1 Sensor Synchronization

The box plot in Fig. 6 depicts the effect of exposure-compensated sensor synchronization in comparison to a synchronization scheme, where the camera trigger is temporally aligned with polling the IMU. For each synchronization paradigm, we collected about ten datasets for three fixed exposure times by dynamically moving the sensor setup in front of a checkerboard. The algorithm outlined in Section 4 was used to estimate the time-offset between the measurements. The figure clearly shows the exposure dependency of the inter-sensor delay for the synchronization where the camera trigger events are equally spaced in time. In addition, a fixed offset becomes apparent, which can be estimated when extrapolating the graph for zero exposure time. As detailed on in Section 3.2, the sensor setup compensates for the exposure as well as for the fixed time-offset, resulting in an average inter-sensor delay of only about  $7\ \mu\text{s}$ .

### 6.2 Timing

Figure 7 shows profiling results for the visual-inertial SLAM system. Timings were generated on our flying platform equipped with a Core2Duo host computer. The sensor assembly was operated in a two-camera configuration, with both cameras running at 20 Hz, and with an IMU rate of 200 Hz.

The most expensive operation in this configuration is keypoint

## 6. Results

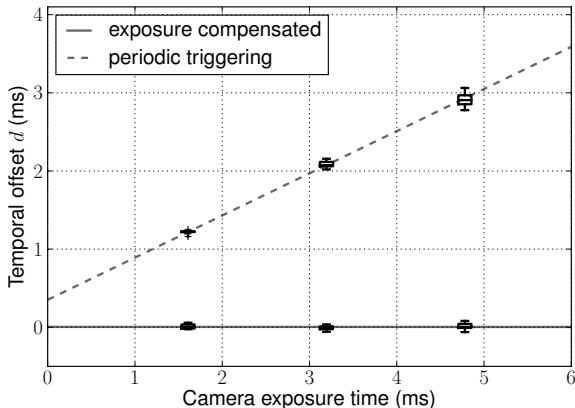


Figure 6.: Results for compensating relative delays of camera and IMU. The dotted line marks the estimated time offset between camera and IMU for a synchronization scheme, where the camera is triggered periodically and the timestamp represents the trigger time. This paradigm clearly results in an exposure dependent delay. Note that there also exists a fixed time-offset, which is induced by filter and communication delays in the IMU and can be estimated by extrapolating for zero exposure time. Our setup compensates for both types of delay, resulting in an almost perfect synchronization with an average estimated delay of only about 7 $\mu$ s.

detection using an SSE-accelerated CPU implementation of Harris corners, followed by optimization in the visual-inertial SLAM backend algorithm. With an optimization window of more than five keyframes, the optimization is not able to finish in time and starts dropping frames. Using the FPGA for corner detection resolves this issue.

The computational complexity of the detection further grows when camera resolution or frame rate is increased, or when more cameras are integrated. Outsourcing detection to the FPGA thus significantly reduces CPU load. The remaining parts of the visual-

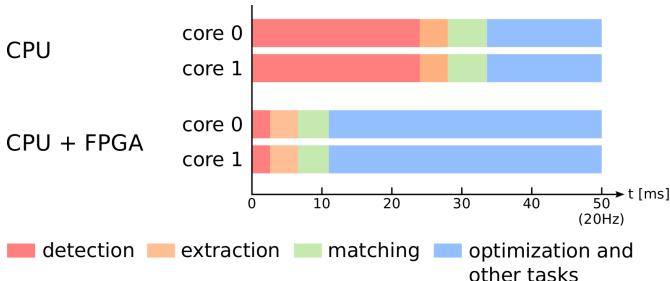


Figure 7.: Profiling for visual-inertial SLAM with and without FPGA accelerated keypoint detection on a Core2Duo. Detection complexity is directly related to camera resolution and consumes a significant amount of time. Outsourcing this operation to the FPGA frees up resources and thus enables processing on resource-constrained platforms, larger optimization windows, or other tasks.

inertial motion estimation algorithm are then largely independent of the system's hardware configuration.

### 6.3 Visual-Inertial Motion Estimation Evaluation

We recorded a dataset walking around the ETH main building. The sequence contains changing illumination, varying depth, and dynamic objects such as people and cars. The length of the trajectory was 700 m. Two video streams were captured at 20 Hz and the IMU at 200 Hz. Processing was performed with the algorithm outlined in Section 5.

Fig. 8 shows the trajectory and structure reconstruction manually overlaid onto an orthophoto. The position error at the end of the trajectory amounts to 5 m laterally and 1 m vertically, thus about 0.7 % of the distance traveled. Note that no loop-closure constraint was applied when reaching the point of origin.

## 7. Conclusion and Outlook

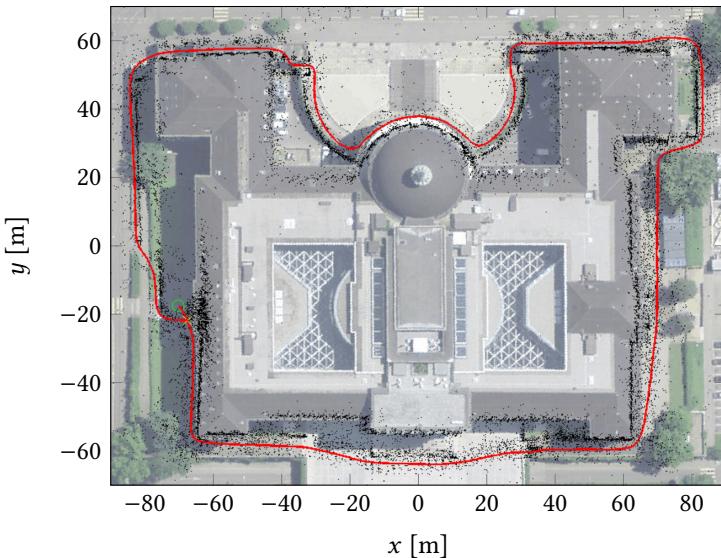


Figure 8.: Reconstructed trajectory (red) and estimated landmarks (black) for a hand-held sequence with the SLAM system in stereo configuration. A distance of 700m around the ETH main building was covered, and drift accumulates to approximately 5m laterally and 1m vertically.

## 7. Conclusion and Outlook

This work presented the design of a time-synchronized, calibrated sensor head which is targeted at mobile robotic applications in need of accurate, robust, real-time pose estimation and mapping in uncontrolled environments. Hardware synchronization includes compensation for variable shutter opening, resulting in provably virtually zero time offset between images and IMU measurements. Low-level image processing tasks such as keypoint detection were implemented

### *Paper III*

in programmable hardware in order to speed up processing and free CPU resources. The measurements taken by the presented sensor head were finally fed to a tightly-coupled real-time visual-inertial SLAM framework, the output of which demonstrated the capabilities of the sensor head.

The modular design is ready for integration of higher resolution imagers. Our future activities will on the one hand focus on integration on different platforms ranging from fixed-wing unmanned aircraft to legged robots. On the other hand, we plan to port a light-weight visual-inertial SLAM solution onto the ARM of the sensorhead, in order to obtain a true “SLAM in a box” module.

## Acknowledgements

The research leading to these results has received funding from *armasuisse Science and Technology*, project No. 050-23, research contract No. 8003501880. This project also received funding from the *Swiss Commission for Technology and Innovation (CTI)*, project No. 13394.1 PFFLE-NM (Visual-Inertial 3D Navigation and Mapping Sensor), and from the European Commission’s Seventh Framework Program under grant agreement nr. 285417 (ICARUS), nr. 600958 (SHERPA), and nr. 231143 (ECHORD/TUAV). The authors would also like to thank Markus Bühler, Dario Fenner and Fabio Diem for mechanical design and fabrication, and Simon Lynen for support in driver development.

# Paper IV

## A UAV System for Inspection of Industrial Facilities

Janosch Nikolic · Michael Burri · Joern Rehder ·  
Stefan Leutenegger · Christoph Huerzeler ·  
Roland Siegwart

### Abstract

This work presents a small-scale Unmanned Aerial System (UAS) capable of performing inspection tasks in enclosed industrial environments. Vehicles with such capabilities have the potential to reduce human involvement in hazardous tasks and can minimize facility outage periods. The results presented generalize to UAS exploration tasks in almost any GPS-denied indoor environment. The contribution of this work is two-fold. First, results from autonomous flights inside an industrial boiler of a power plant are presented. A lightweight, vision-aided inertial navigation system provides reliable state estimates under difficult environmental conditions typical for such sites. It relies solely on measurements from an on-board MEMS inertial measurement unit and a pair of cameras arranged in a classical stereo configuration. A model-predictive controller allows for efficient trajectory following and enables flight in close proximity to the boiler surface. As a second contribution, we highlight

## *Paper IV*

ongoing developments by displaying state estimation and structure recovery results acquired with an integrated visual/inertial sensor that will be employed on future aerial service robotic platforms. A tight integration in hardware facilitates spatial and temporal calibration of the different sensors and thus enables more accurate and robust ego-motion estimates. Comparison with ground truth obtained from a laser tracker shows that such a sensor can provide motion estimates with drift rates of only few centimeters over the period of a typical flight.

Published in the proceedings of the *IEEE Aerospace Conference*, 2013.

©2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The final publication is available at <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6496959>, DOI 10.1109/-AERO.2013.6496959.

## 1. Introduction

INDUSTRIAL sites often contain areas and facilities that are difficult to access or hazardous to humans. The boiler unit of a coal-fired thermal power plant constitutes one example of that environment type. Fig. 2 shows such a unit, and visual inspection thereof will serve as an application example throughout this work. Whenever it is necessary to inspect such areas, tremendous precautions have to be

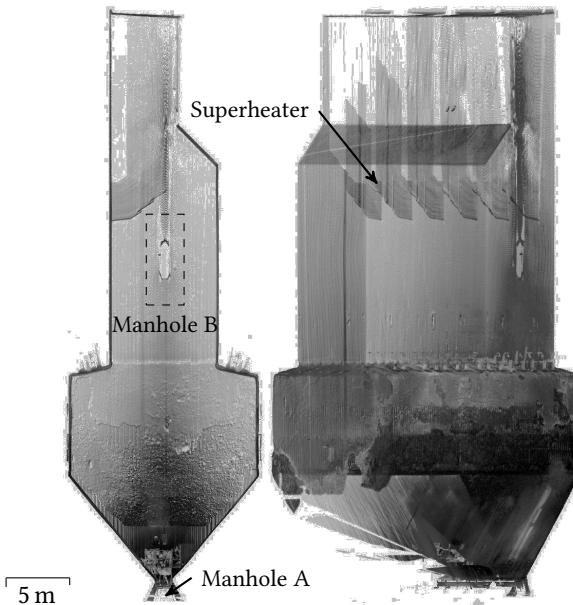


Figure 1.: Laser scan of power plant boiler unit II (side view). With a height of almost 50m, this is a medium-sized unit. The flight tests were conducted in the lower part of the boiler.

taken in order to minimize the risks to human workers. Furthermore, scaffolding often needs to be installed in order to grant access to sections that are otherwise hard to reach. These measures result in significant downtime of the plant and are thus associated with vast economic costs.

Autonomous aerial systems have the potential to reduce human involvement in a variety of industrial inspection tasks and can in turn reduce outage periods. They neither require extensive safety measures nor supporting structures to access elevated or confined areas. In contrast to existing and very successfully employed robotic platforms such as e.g. magnetic crawlers [82], flying inspection vehicles can access virtually any area with ease and are not restricted to ferromagnetic surfaces.

However, the employment of flying systems poses a number of challenges such as very limited payload capabilities, and hence limited computational resources. One of the main difficulties is reliable state or vehicle ego-motion estimation in absence of GPS signals, high-quality inertial measurement units (IMUs), and the like.

Section 3 describes the platform, its visual/inertial state estimation and the autopilot: a model predictive control scheme enables the vehicle to follow pre-programmed trajectories in close vicinity of the structure efficiently. By arranging these trajectories around an area of interest, a human operator can automatically acquire sets of images for visual inspection. In this respect, the autonomy of our system facilitates efficient inspection of large structures and enables remote inspection in areas, where occlusions or impaired radio reception had rendered it infeasible before.

Section 4 of this work presents a custom-designed visual-inertial sensor-head. The size of our aerial vehicle imposes tight constraints on weight and energy consumption onto the state perception module. At the same time, a high degree of robustness and accuracy is required to enable autonomous operation in poorly and repetitively textured areas. To address these requirements, we are in the course of

## *2. Related Work*

developing an integrated visual/inertial sensor that allows for precise hardware synchronization of multiple different sensors and on-board estimation of motion and structure. Tight integration of cues from a pair of cameras arranged in a forward-looking stereo configuration and a MEMS inertial measurement unit on a hardware level allows for accurate spatial and temporal calibration of the sensor unit and ultimately enables the required level of robustness and accuracy.

Section 5 presents results with respect to the described hardware and algorithms: the system was deployed in a shut-down boiler, in order to assess the on-board state estimation and control capabilities in relation to recorded ground truth motion. Furthermore the integrated sensor performance including a tightly coupled visual-inertial odometry approach is evaluated again with respect to ground truth structure and pose measurements.

Despite the fact that our flight and sensor tests were conducted inside an industrial boiler, we believe that the significance of our results extend to a much broader class of scenarios. With our system, we successfully demonstrate that path following, the predominant prerequisite for all indoor navigation tasks, is possible in almost any GPS-denied, unstructured environment providing a minimum of visible texture.

## **2. Related Work**

Autonomous navigation of aerial vehicles in GPS-denied environments has recently gained attention in the robotics community. Using a laser range sensor, Shen et al. demonstrated indoor navigation of a quadrotor [75]. However, the requirements, imposed onto the system by a localization based on scan matching along with altitude measured by a deflected portion of the scan, are not met in most industrial environments. Approaches using an RGB-D sensor instead of a laser

scanner [34] do not share the same constraints on the structure of the environment, but are only applicable to indoor operations. While this would suffice to conduct the experiment presented in this work, it would fail in the case of an outdoor inspection task. Weiss et al. [88] demonstrated visual/inertial enabled outdoor flights of a small aerial vehicle. In contrast to their approach based on a monocular camera, we employ a stereo setup; we believe that the additional information introduced by measurements from synchronized cameras displaced by a known transformation adds necessary robustness.

Most similar projects use either PID or LQR controllers to keep the UAV at a desired position. The main advantage of a PID controller ([34] and [23]) is the simple design. Furthermore, those controllers do not suffer from offsets in presence of disturbances. However, the control performance achievable with such a basic control schemes is often not satisfying. The second approach is motivated by the theory of optimal control and is used in [75], [7] and [6]. An optimal feedback strategy is calculated that minimizes a quadratic cost function, using a linear model of the quadrotor. The following two requirements are necessary for the solution to be optimal: actuator saturation must not occur and the reference position has to be constant. Both preconditions are typically not fulfilled; in order to overcome these requirements, a model predictive controller is proposed.

There exist several FPGA based stereo vision platforms such as [2], [40]. Unlike these systems, the sensor presented in this work also incorporates an inertial measurement unit (IMU), as its application domain is more geared towards motion estimation rather than dense 3D reconstruction. We believe that what sets our integrated system apart from solutions comprised of individual sensors is its precise calibration for inter-sensor spatial transformations and time delays. Commonly, the transformation between a camera and inertial sensors is estimated recursively [59][46], while time delays are determined in a separate process [45]. In contrast, we estimate these quantities in a unified, batch optimal framework, achieving sub-millisecond

### *3. Aerial System Design and Control*

accuracy in the synchronization of the different sensors [25].

In the computer vision and robotics literature, the Visual Odometry (VO), or visual Simultaneous Localization and Mapping (SLAM) problem is well studied: it is formulated as estimation of both structure, typically in the form of sparse landmarks, and camera poses observing those as two-dimensional projections. Recently, it has been shown that sparse nonlinear optimization over a selection of key-frames is computationally more efficient and provides more accurate results than filtering approaches [79]. Since the optimization problem becomes inherently intractable over time when constantly adding observations, various approaches have been suggested to keep the algorithms limited in complexity and real-time capable: a common procedure, employed e.g. in the monocular real-time SLAM algorithm PTAM [49], is keeping the optimized frames bounded, yielding some form of a sliding-window optimization. PTAM has been adapted [88] to operate on-board MAVs and providing a six-dimensional pose input to an Extended Kalman Filter (EKF) for loosely-coupled IMU-vision integration. In terms of tightly coupling such vision algorithms with inertial sensing, we base our approach on [81], where a nonlinear batch optimization is suggested that jointly optimizes over reprojection error of landmarks as well as deviation of the motion from the prediction based on inertial measurements.

## **3. Aerial System Design and Control**

This section describes the platform that was used during the experiments in boiler unit II. First, the aerial vehicle prototype is described, and modelling, identification and control are presented. The next section then provides an overview over the on-board vision-aided inertial navigation subsystem that provides real-time motion estimates during flight and allows the controller to stabilize the vehicle.



Figure 2: Prototype aerial vehicle used for experimentation. The sensorhead consists of a pair of cameras arranged in a classical, front-looking stereo configuration. An on-board active illumination module provides sufficient illumination to operate inside the boiler. The IMU is mounted on the back side (not visible).

### 3.1 Modelling and Control

The base of the aerial platform is the AscTec Hummingbird attitude controller and brushless motor controllers. In other words, roll, pitch and yaw velocity are controlled by the base platform, relying on the attitude controllers integrated IMU. A custom airframe protects the propellers during voluntary or involuntary contact with the environment. The platform is lightweight enough in order to survive most crashes, easy to use and generally safe for humans to operate.

The commands to the attitude controller are the roll and pitch angles  $\phi^*$  and  $\theta^*$ , the total thrust  $T^*$  and the yaw rate  $\dot{\psi}^*$ . To find a relation between those commands and the response of the quadrotor a simple point mass model as described in [7] is used. Fig. 3 provides

### 3. Aerial System Design and Control

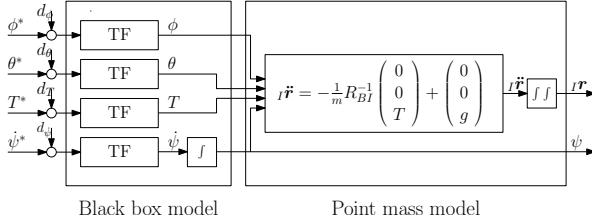


Figure 3.: Complete model of the quadrotor using a simple point mass and four transfer functions representing the behaviour of the on-board attitude controller to the commands  $(\phi^*, \theta^*, \dot{\psi}^*, T^*)$ .

an overview of this approach. The unknown response of the attitude controller to reference commands was identified with a blackbox modelling approach (prediction error method, PEM). Models of first order followed by time delays were found to be accurate enough for the task at hand. The interested reader is referred to [12] for additional details regarding modelling and identification of the vehicle.

Different approaches for trajectory following were evaluated. A model predictive control strategy is conceptually simple and offers a number of benefits in practice, and was therefore selected. The key idea is to take the current state  $x$ , use the model to predict a fixed number  $N$  of future states and find an optimal sequence of commands that minimizes the quadratic cost function

$$J_N^*(x) := \min \frac{1}{2} x_N^T P x_N + \frac{1}{2} \sum_{k=0}^{N-1} x_k^T Q x_k + u_k^T R u_k \quad (1)$$

s.t.  $x_{k+1} = Ax_k + Bu_k, \quad \forall k = 0, \dots, N-1,$   
 $u_k \in \mathbb{U}, \quad \forall k = 0, \dots, N-1,$   
 $x_0 = x.$

Only the first command  $u_0$  is then applied to the vehicle and the problem is reiterated in the next time step.

The cost function can be extended to include any desired trajectory, which allows the controller to react to reference changes in advance, reducing the necessary control actions, but still following the trajectory in an optimal sense. This leads to less aggressive flight maneuvers, which reduces motion blur in the camera images, an important aspect for the visual navigation subsystem. In addition, the cost function can be adapted to punish fast attitude changes to further smoothen the behaviour of the quadrotor. The weighting matrices  $Q = Q^T \geq 0$  and  $R = R^T > 0$  are therefore chosen to track the position and avoid fast changes in attitude. A terminal penalty matrix  $P$  is necessary to guarantee stability.

### 3.2 Vision-Aided Inertial Navigation System

The following subsection provides an overview of the system that was developed for motion estimation of the aerial vehicle. The vehicle state consists of the following variables

$$x = [p_I \quad q_{IB} \quad v_I \quad b_a \quad b_\omega]^T \quad (2)$$

where  $p_I$  denotes the position of the sensorhead,  $q_{IB}$  the attitude quaternion,  $v_I$  the velocity and  $b_a$ ,  $b_\omega$  the accelerometer and gyroscope biases, respectively. Fusion of visual and inertial cues is, strictly speaking, implemented in a loosely coupled fashion following the stochastic cloning approach outlined in [62]. Only an overview of the processing pipeline is provided here, and the interested reader is referred to [87] for further details.

In the first stage, a FAST [73] keypoint detector in conjunction with an adaptive thresholding scheme provides well distributed image features that are then tracked in-between camera frames using a hamming matcher based on BRIEF [14] descriptors. A motion hypothesis is provided by the inertial measurements and then refined in a bundle adjustment step over the current and the most recent image

#### *4. Integrated Sensor Design and State Estimation*

pairs. This refined estimate is then used in an update step following a stochastic cloning approach. Gyroscope and accelerometer biases, modelled as random walks, are estimated within the same framework.

Having an IMU in the loop which is time-synchronized with the camera shutters allows the estimation of confidence intervals, in the image plane, of where tracked features must re-appear in subsequent camera frames. Hence, once inertial sensor biases are tracked accurately, one can guide feature matching and gain robustness when operating in poorly or repetitively textured scenes. Also, we are able to continuously estimate the vehicle state even when visual tracking fails completely, but only for a short period of time.

## **4. Integrated Sensor Design and State Estimation**

The following section presents the prototype design of a *SLAM in a Box* system, a general-purpose module able to provide synchronized and calibrated raw-data. Alternatively, it directly provides reliable motion estimates capabilities to different types of robotic platforms, and in a large variety of environments. The hope is that such a module will be helpful for engineers that simply need a sensor that can provide pose estimates in GPS-denied environments, and to researchers in need of high-quality raw data alike.

### **4.1 Sensor Design**

The core of the visual/inertial sensor is a Xilinx Spartan 6 FPGA that implements interfaces to up to four Aptina MT9V034 global shutter CMOS image sensors and an Analog Devices ADIS16488 factory-calibrated MEMS IMU. Additionally, a set of low- and medium cost MEMS gyroscopes and accelerometers are mounted for applications

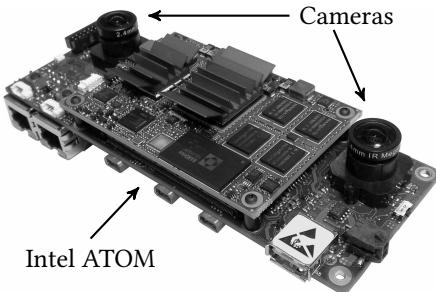


Figure 4.: The integrated visual-inertial SLAM sensor. Only the two cameras mounted in a fronto-parallel stereo setup were used in this study. Like the aerial vehicle, the sensor is equipped with a camera exposure synchronous LED flash (not illustrated) to enable operation in the boiler. The FPGA is on the bottom side of the PCB, in the center-top is the commercially available Intel ATOM module.

where weight or cost are crucial. The FPGA connects through a Gigabit Ethernet interface to a host computer. By routing all sensor streams through the FPGA, it is possible to accurately timestamp all measurements with the same clock source. Fig. 4 shows the sensor hardware.

In order to fuse measurements from a set of sensors contributing towards a motion estimate, the spatial transformation between the sensors has to be known accurately. In addition—and despite our efforts to synchronize the sensors in hardware—there still exists a deterministic time delay between the cameras and inertial sensors, caused by communication delays as well as filter delays in the inertial sensor that have not been accounted for in the time-stamping. Hence, a calibration for these fixed delays is required to achieve optimal estimation results.

In order to calibrate for these inter-sensor transformations and

#### *4. Integrated Sensor Design and State Estimation*

time delays, we employed a unified framework based on continuous-time batch estimation as proposed by Furgale et al. [24] (see [25] for details).

The sensor unit can then either provide synchronized, intrinsically and extrinsically calibrated visual and inertial measurements, a map of features such as e.g. Harris scores, computed in the FPGA, or directly motion estimates computed on the CPU.

## **4.2 Tightly Coupled Visual-Inertial Odometry**

We applied a loosely-coupled approach for visual-inertial state estimation [87] running on-board the aerial vehicle: this algorithm serves as the basis for the model-predictive controller described above. As a complementary method, we also experimented with tight integration of inertial sensing and visual odometry. Keypoint observations and associations with map landmarks are created using the BRISK detector [51] and binary descriptor extractor. Inertial error terms are formulated using accelerometer and gyro readings integrated between camera frames using standard IMU kinematics following largely the method described in [81]. Knowledge on the statistical properties of the inertial sensors on the one hand and keypoint detection accuracy on the other hand allow for statistically near-optimal fusion. Inspired by the recent success of key-frame based visual odometry algorithms we applied some generalization of the aforementioned batch fusion to handle a constant size sliding window of key-frames. We select key-frames by analyzing the field-of-view overlap. Consequently, no key-frames are inserted in the optimization when keeping the pose, which results in almost non-existent drift in this relevant case.

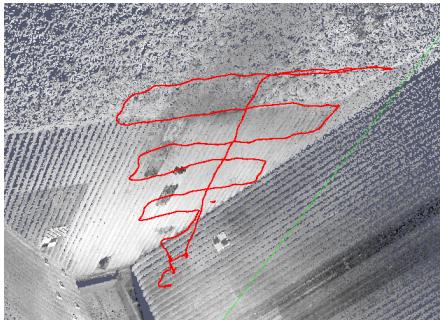


Figure 5.: Actual flight path executed by the aerial vehicle, tasked to follow a pre-programmed trajectory with a forward velocity of 0.25 m/s.

## 5. Experiments and Results

All experiments presented in this section were conducted in a boiler unit of a thermal power plant in northern Spain. This particular unit was temporarily shut down, and could be safely accessed through a man-hole at the bottom. Fig. 2 shows a side-view of the boiler.

First, results on the flight experiments are discussed, and results obtained with the integrated sensor introduced in Section 4 are presented subsequently.

### 5.1 Flight Experiments

The results presented here aim at demonstrating that the methods presented in Section 3 work in this particular environment. They give an indication of the performance that can be achieved using only on-board sensors, and no external tracking devices or infrastructure.

The task of the aerial vehicle was to follow a pre-defined trajectory at a distance of about one meter to the boiler surface. For the first run, the desired forward velocity was set to 0.25m/s, and then increased

## 5. Experiments and Results



Figure 6.: Aerial vehicle during a flight test. At a distance of approximately one meter to the boiler wall, the MAV passes an area where the surface is insulated with an non-ferromagnetic, slightly eroded heat shield. The scene is illuminated by high-power LEDs carried on-board, synchronized with the camera shutter to save power and avoid over-heating.

to 0.8m/s for the second run, to test the dynamic capabilities of the vehicle and the visual-inertial motion estimation system to the limit. Fig. 5 shows an overlay of the path the vehicle followed during a typical test run. Fig. 6 shows the vehicle during the flight.

The MPC loop was closed at a frequency of 50Hz, with a prediction horizon of  $N = 30$  steps. Image processing was performed off-board, and the vehicle was provided with a data link over a tether cable from the ground station.

Fig. 7 indicates that the vehicle was able to follow the desired trajectory accurately for both demanded forward velocities. It is important to note that the reference is compared to the pose estimates provided by the on-board sensors, and *not* against an externally measured ground-truth. This is justified by the fact that these estimates are accurate, as demonstrated in the following section, but strictly speaking, not correct.

Compared to alternative control strategies evaluated, the control actions of the MPC controller are smooth, which in turn supports the

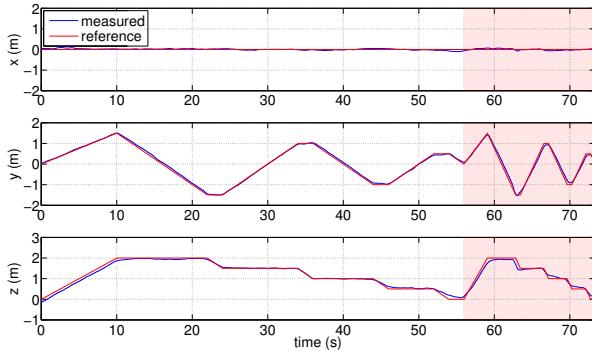


Figure 7.: Aerial vehicle following a predefined trajectory with a velocity reference of 0.25 m/s. During the second run (highlighted), the forward velocity was increased to 0.8m/s.

vision-based motion estimation module. Also, there is virtually no overshoot, an important requirement when flying in close proximity with the environment.

Table 1 lists the RMS tracking errors for the two different forward velocities. Keeping a constant distance to the wall in direction  $x$  was not an issue for the controller. However, especially in the vertical ( $z$ ), external disturbances caused by the tether cable negatively affect the tracking error.

## 5.2 Sensor Evaluation

In order to evaluate the performance of the tightly integrated visual/inertial sensor and motion estimation framework (Section 4), the sensor unit was mounted on an aluminum beam and had to undergo movements that resemble the motion experienced during a typical flight.

For quantitative evaluations, a Leica Total Station TS15 was used

## 5. Experiments and Results

Table 1.: Measured position error (RMS) in world coordinates, following the trajectory shown in Fig. 7.

<b>Axis</b>	<b>RMS error</b>	<b>RMS error</b>
	$v = 25 \text{ cm/s}$	$v = 80 \text{ cm/s}$
$x$	$3.68 \times 10^{-2} \text{ m}$	$3.88 \times 10^{-2} \text{ m}$
$y$	$5.25 \times 10^{-2} \text{ m}$	$10.55 \times 10^{-2} \text{ m}$
$z$	$7.32 \times 10^{-2} \text{ m}$	$11.30 \times 10^{-2} \text{ m}$

to track a reflective prism mounted on the sensor, with sub-millimeter precision. In addition, a visual calibration target (a checkerboard) was attached onto the boiler wall, and visited at the beginning of every dataset. The purpose of this target was two-fold. First, it was used to align the coordinate frame of the Total Station with the somewhat arbitrarily chosen coordinate frame of the motion estimation framework. The alignment was realized by measuring a sufficient amount of corners using the laser tracker and thus determining the rigid-body transformation between the coordinate frame of the tracker and the pattern. Subsequently, about 200 observations of the calibration pattern were used to determine camera poses in the coordinate frame of the tracker by means of a perspective N-point (PnP) algorithm using the implementation of [9]. Associating these poses to the motion estimation by timestamps, the tracker frame was then aligned with the estimation frame in a least square fashion over all 200 pose estimates. The other purpose of visiting the visual target as part of every dataset was to determine the unknown time offset between laser tracker and sensor timestamps. Again, this was implemented as a least square optimization over the position estimates provided by the PnP algorithm and the laser tracker.

Although the Leica Total Station provided measurements at sub-millimeter accuracy, the precision of the quantitative evaluations is limited by the uncertainty in the calibrated chain of transformations from the inertial measurement unit to the reflective prism, as well as jitter in recording of the laser tracker measurements, which amounts to a degradation in accuracy to about a few millimeters.

In addition to accurate position ground truth, full three dimensional structure scans of the boiler were recorded using a static Faro laser scanner, although comparisons to our structure reconstruction will remain qualitative in nature here.

Fig. 8 depicts results of our visual-inertial motion estimation framework for an experiment of approximately 8 minutes length. The camera frame rate had been set to 20 Hz and the IMU returned readings at a rate of 200 Hz. In the figure, the estimated sensor path is marked as a solid blue line, while the dashed black line shows ground truth obtained from laser tracker readings. The red curve shows a section of about 200 frames during which the calibration target was visible in the video frames and that was used for path alignment and time offset calibration. In this experiment, the time offset was determined to be about 0.2 seconds, which would significantly distort the ground truth comparison when not accounted for. The results show that the estimation framework is in general capable of recovering the motion. However, it also highlights that the estimate diverges from ground truth over the course of the experiment due to accumulation of errors. Fig. 10 depicts this divergence quantitatively. The motion estimation exhibits comparatively low drift, which results in a displacement from ground truth of about 10 cm over the dataset.

Fig. 9 shows a qualitative comparison of the structure as recovered visually and by a laser scanner. For the laser scan only every thousandth measurement is displayed. The visual reconstruction is significantly more sparse and scarcely corrupted by outliers. Nevertheless, the reconstruction of the lateral walls of the boiler aligns accurately with the laser scan.

## 5. Experiments and Results

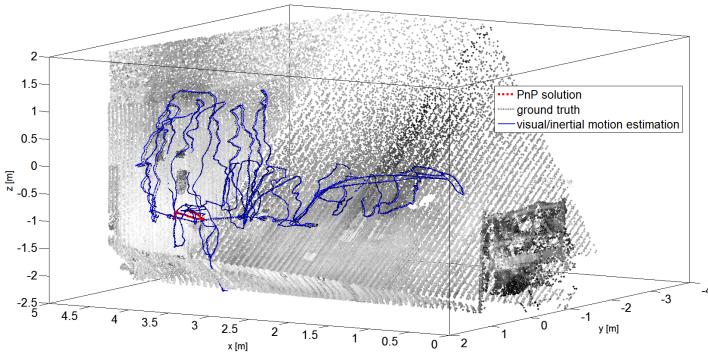


Figure 8.: Comparison of ground truth positions and estimated sensor path. The red curve marks the PnP pose estimation used to align the coordinate frames of the motion estimation and the laser tracker. The motion estimation depicted in blue resembles the ground truth shown in black well. Nevertheless, drift becomes apparent over the course of the dataset.

Fig. 11 illustrates a comparison between a video frame from the dataset and a synthetic view rendered from a laser scan of the structure and the estimated pose of the camera at the moment the video frame was recorded. The rendering is based on camera intrinsics and distortion parameters identical to the ones calibrated for the setup used in conducting the experiments. Although purely qualitatively, this result highlights the orientation component of pose estimation, a quantity that is not accounted for by position ground truth. Camera image and synthetic view exhibit great resemblance, suggesting that not only the camera position but also its orientation has been recovered accurately.

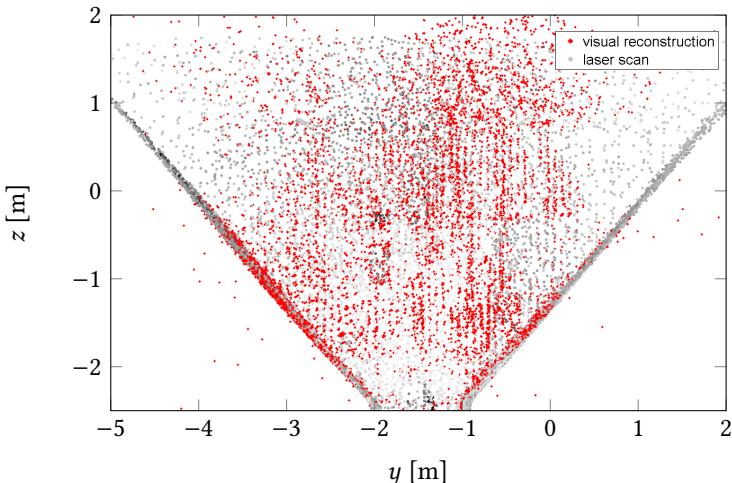


Figure 9.: Overlay of a sparse, visual reconstruction of a section of the boiler onto a laser scan. The visually reconstructed point cloud aligns accurately with the laser scan, which becomes particularly apparent for the lateral walls of the boiler.

## 6. Conclusion

This work presented an unmanned aerial system with application to industrial inspection tasks. To the best of our knowledge, the autonomous flight experiments conducted inside an industrial boiler are unprecedented. As such, the results convey an idea of how autonomous aerial systems could facilitate visual inspection in enclosed industrial environments. By commanding scanning trajectories, we were able to take sets of overlapping, fronto-parallel photographs of the boiler walls at close range, enabling detection of larger structural damages. Quantitative results for our visual-inertial state estimation

### Acknowledgements

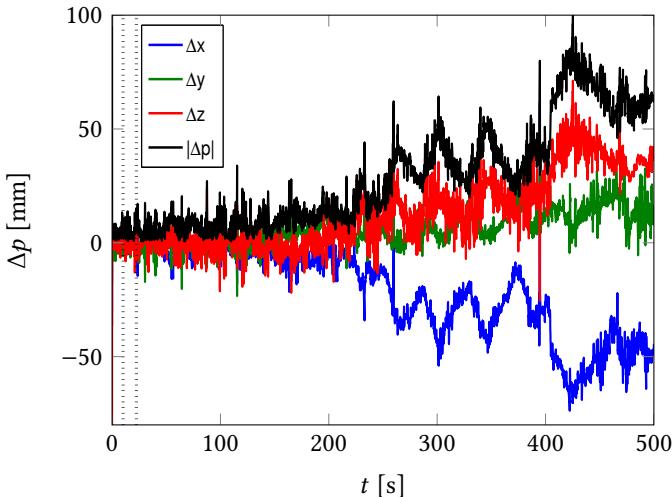


Figure 10.: Visualization of the position error of the visual/inertial motion estimation framework. The vertical, dotted lines frame a section of roughly 10 seconds, which was used for aligning the sensor path with the tracker measurements. The results exhibit a low level of drift, amounting to less than 10 cm over a dataset of roughly 9000 frames spanning 500 s.

framework show a small drift in position in the order of around 10 cm over more than 8 minutes. This suggests that following large trajectories is feasible even without external navigation aids.

## Acknowledgements

The authors would like to thank the following people for their support: Vera De-Vries, Wolfgang Mesters, Elsa Recio, and Eleni Tsigkogianni

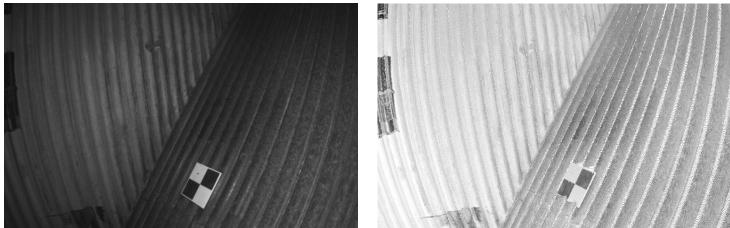


Figure 11.: Comparison of a camera frame from the dataset and a synthetic view using the estimated camera pose and reflectance measurements from a structure scan of the boiler. The rendering applies the camera intrinsics and distortion parameters calibrated for the camera used in the experiments. The figures exhibit great resemblance, suggesting that not only the position of the camera but also its orientation has been estimated to sufficient accuracy.

from Alstom, and Javier Vega Arias from GNF. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2010-2013) under grant agreement no. 248669 (AIRobots). This project also received funding from the Swiss Commission for Technology and Innovation (CTI), project no. 13394.1 PFFLE-NM (Visual-Inertial 3D Navigation and Mapping Sensor).

# Bibliography

- [1] *3-Axis Test System with Thermal Shroud TES-4V33-TG*. Rev. April 2013. Motion Dynamic. 2013.
- [2] C. Ahlberg, J. Lidholm, F. Ekstrand, G. Spampinato, M. Ekstrom, and L. Asplund. “Gimme-A General Image Multiview Manipulation Engine”. In: *Proceedings of the International Conference on Reconfigurable Computing and FPGAs (ReConFig)*. IEEE. 2011, pp. 129–134.
- [3] D. Allan. “Statistics of Atomic Frequency Standards”. *Proceedings of the IEEE* vol. 54 (1966), pp. 221–230.

## BIBLIOGRAPHY

- [4] R. O. Allen and D. H. Chang. *Performance Testing of the Systron Donner Quartz Gyro QRS11-100-420 (EM 343-1297)*. Tech. rep. Jet Propulsion Laboratory, 1993.
- [5] J. Alves, J. Lobo, and J. Dias. “Camera-Inertial Sensor Modeling and Alignment for Visual Navigation”. In: *Proceedings of the International Conference on Advanced Robotics*. 2003, pp. 1693–1698.
- [6] A. Bachrach, R. He, and N. Roy. “Autonomous Flight in Unknown Indoor Environments”. *International Journal of Micro Air Vehicles* vol. 1.4 (2009), pp. 217–228.
- [7] M. Blösch, S. Weiss, D. Scaramuzza, and R. Siegwart. “Vision Based MAV Navigation in Unknown and Unstructured Environments”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2010, pp. 21 –28.
- [8] *BMX055 Data Sheet: Small, Versatile 9-Axis Sensor Module*. Rev. 1.1. Bosch. 2014.
- [9] J.-Y. Bouguet. “Camera Calibration Toolbox for Matlab (2008)”. URL [http://www.vision.caltech.edu/bouguetj/calib\\_doc](http://www.vision.caltech.edu/bouguetj/calib_doc) (2008).
- [10] D. C. Brown. “Close-Range Camera Calibration”. *Photogrammetric engineering* vol. 37.8 (1971), pp. 855–866.
- [11] M. Burri, J. Nikolic, H. Oleynikova, M. Achtelik, and R. Siegwart. “Maximum Likelihood Parameter Identification for MAVs”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2016.
- [12] M. Burri, J. Nikolic, C. Hürzeler, G. Caprari, and R. Siegwart. “Aerial Service Robots for Visual Inspection of Thermal Power Plant Boiler Systems”. In: *Proceedings of the International Con-*

## BIBLIOGRAPHY

- ference on Applied Robotics for the Power Industry (CARPI). IEEE. 2012, pp. 70–75.
- [13] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. Achtelik, and R. Siegwart. “The EuRoC MAV Datasets”. *The International Journal of Robotics Research IJRR* (2015).
  - [14] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. “BRIEF: Binary Robust Independent Elementary Features”. In: *Proceedings of the 11th European Conference on Computer Vision. ECCV’10*. Springer-Verlag, 2010, pp. 778–792.
  - [15] A. Chatfield. *Fundamentals of High Accuracy Inertial Navigation*. American Institute of Aeronautics and Astronautics, Inc., 1997.
  - [16] *Compact, Precision Ten Degrees of Freedom Inertial Sensor ADIS16448*. Rev. C. Analog Devices. 2014.
  - [17] J. Crassidis. “Sigma-Point Kalman Filtering for Integrated GPS and Inertial Navigation”. *IEEE Transactions on Aerospace and Electronic Systems* vol. 42.2 (2006), pp. 750–756.
  - [18] T. A. Davis. *Direct Methods for Sparse Linear Systems*. SIAM, Philadelphia, 2006.
  - [19] F. Dellaert and M. Kaess. “Square Root SAM: Simultaneous Localization and Mapping via Square Root Information Smoothing”. *The International Journal of Robotics Research* vol. 25.12 (2006), pp. 1181–1203.
  - [20] H. Durrant-Whyte and T. Bailey. “Simultaneous Localization and Mapping: Part I”. *IEEE Robotics and Automation Magazine* vol. 13.2 (2006), pp. 99–110.

## BIBLIOGRAPHY

- [21] N. El-Sheimy, H. Hou, and X. Niu. “Analysis and Modeling of Inertial Sensors Using Allan Variance”. *IEEE Transactions on Instrumentation and Measurement* vol. 57.1 (2008), pp. 140–149.
- [22] M. Fleps, E. Mair, O. Ruepp, M. Suppa, and D. Burschka. “Optimization Based IMU Camera Calibration”. In: *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2011, pp. 3297–3304.
- [23] F. Fraundorfer, L. Heng, D. Honegger, G. H. Lee, L. Meier, P. Tanskanen, and M. Pollefeys. “Vision-Based Autonomous Mapping and Exploration using a Quadrotor MAV”. In: *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2012, pp. 4557–4564.
- [24] P. Furgale, T. Barfoot, and G. Sibley. “Continuous-Time Batch Estimation Using Temporal Basis Functions”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2012, pp. 2088–2095.
- [25] P. Furgale, J. Rehder, and R. Siegwart. “Unified Temporal and Spatial Calibration for Multi-Sensor Systems”. In: *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*. 2013.
- [26] *GG1320AN RLG Data Sheet: Digital Ring Laser Gyroscope*. Honeywell. 2010.
- [27] Z. Ghahramani and G. E. Hinton. *Parameter Estimation for Linear Dynamical Systems*. Tech. rep. University of Toronto, 1996.
- [28] P. Gohl, M. Burri, S. Omari, J. Rehder, J. Nikolic, M. Achtelik, and R Siegwart. “Towards Autonomous Mine Inspection”. In:

## BIBLIOGRAPHY

- Proceedings of the International Conference on Applied Robotics for the Power Industry (CARPI).* IEEE. 2014, pp. 1–6.
- [29] S. Guerrier, J. Skaloud, Y. Stebler, and M.-P. Victoria-Feser. “Wavelet-Variance-Based Estimation for Composite Stochastic Processes”. *Journal of the American Statistical Association* vol. 108.503 (2013), pp. 1021–1030.
  - [30] N. Gupta and R. Mehra. “Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculations”. *IEEE Transactions on Automatic Control* vol. 19.6 (1974), pp. 774–783.
  - [31] C. Harris and M. Stephens. “A Combined Corner and Edge Detector.” In: *Proceedings of the Alvey Vision Conference*. Vol. 15. Manchester, UK. 1988, p. 50.
  - [32] J. Hesch, D. Kottas, S. Bowman, and S. Roumeliotis. “Consistency Analysis and Improvement of Vision-aided Inertial Navigation”. *IEEE Transactions on Robotics* vol. 30.1 (2014), pp. 158–176.
  - [33] J. D. Hol, T. B. Schn, and F. Gustafsson. “Modeling and Calibration of Inertial and Vision Sensors”. *The International Journal of Robotics Research* 29.2 (2010), pp. 231–244.
  - [34] A. S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox, and N. Roy. “Visual Odometry and Mapping for Autonomous Flight Using an RGB-D Camera”. In: *Proceedings of the International Symposium of Robotics Research (ISRR)*. 2011.
  - [35] J. C. Hung, J. S. Hunter, W. W. Stripling, and H. V. White. *Size Effect on Navigation Using a Strapdown IMU*. Tech. rep. Army Missile Command Redstone Arsenal and Technology Lab, 1979.

## BIBLIOGRAPHY

- [36] C. Hürzeler, R. Naldi, V. Lippiello, R. Carloni, J. Nikolic, K. Alexis, L. Marconi, and R. Siegwart. “AIRobots: Innovative Aerial Service Robots for Remote Inspection by Contact”. In: *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*. 2013, pp. 2080–2080.
- [37] *IEEE Standard for Inertial Sensor Terminology*. IEEE Std 528-2001. IEEE Aerospace and Electronic Systems Society. 2001.
- [38] *IEEE Standard Specification Format Guide and Test Procedure for Single-Axis Interferometric Fiber Optic Gyros*. IEEE Std 952-1997. IEEE Aerospace and Electronic Systems Society. 1998.
- [39] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, Inc, 1970.
- [40] S. Jin, J. Cho, X. Dai Pham, K. Lee, S. Park, M. Kim, and J. Jeon. “FPGA Design and Implementation of a Real-Time Stereo Vision System”. *IEEE Transactions on Circuits and Systems for Video Technology* vol. 20.1 (2010), pp. 15–26.
- [41] M. Kaess and F. Dellaert. “Covariance Recovery from a Square Root Information Matrix for Data Association”. *Journal of Robotics and Autonomous Systems (RAS)* vol. 57 (2009), pp. 1198–1210.
- [42] R. E. Kalman. “A New Approach to Linear Filtering and Prediction Problems”. *Transactions of the ASME Journal of Basic Engineering* no. 82 (1960), pp. 35–45.
- [43] T. Kazik, L. Kneip, J. Nikolic, M. Pollefeys, and R. Siegwart. “Real-Time 6D Stereo Visual Odometry with Non-Overlapping Fields of View”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012, pp. 1529–1536.

## BIBLIOGRAPHY

- [44] J. Kelly and G. Sukhatme. “Fast Relative Pose Calibration for Visual and Inertial Sensors”. In: *Experimental Robotics*. Springer. 2009, pp. 515–524.
- [45] J. Kelly and G. S. Sukhatme. “A General Framework for Temporal Calibration of Multiple Proprioceptive and Exteroceptive Sensors”. In: *Proceedings of the 12th International Symposium on Experimental Robotics*. 2010.
- [46] J. Kelly and G. S. Sukhatme. “Visual-Inertial Sensor Fusion: Localization, Mapping and Sensor-to-Sensor Self-calibration”. *The International Journal of Robotics Research* vol. 30.1 (2011), pp. 56–79.
- [47] V. Kempe. *Inertial MEMS: Principles and Practice*. Cambridge University Press, 2011.
- [48] M. Keshner. “1/f Noise”. *Proceedings of the IEEE* vol. 70.3 (1982), pp. 212–218.
- [49] G. Klein and D. Murray. “Parallel Tracking and Mapping for Small AR Workspaces”. In: *Proceedings of the Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*. 2007.
- [50] C. Krebs. *Generic IMU-Camera Calibration Algorithm*. Tech. rep. Autonomous Systems Lab (ASL), ETH Zurich, 2012.
- [51] S. Leutenegger, M. Chli, and R. Siegwart. “BRISK: Binary Robust Invariant Scalable Keypoints”. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2011, pp. 2548–2555.
- [52] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale. “Keyframe-Based Visual-Inertial Odometry using Nonlinear

## BIBLIOGRAPHY

- Optimization”. *The International Journal of Robotics Research* vol. 34.3 (2015), pp. 314–334.
- [53] S. Leutenegger, P. Furgale, V. Rabaud, M. Chli, K. Konolige, and R. Siegwart. “Keyframe-Based Visual-Inertial SLAM Using Nonlinear Optimization”. In: *Proceedings of Robotics Science and Systems (RSS)*. 2013.
  - [54] M. Li and A. I. Mourikis. “Online Temporal Calibration for Camera-IMU Systems: Theory and Algorithms”. *The International Journal of Robotics Research* (2014).
  - [55] M. Li, H. Yu, X. Zheng, and A. Mourikis. “High-Fidelity Sensor Modeling and Self-Calibration in Vision-Aided Inertial Navigation”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2014, pp. 409–416.
  - [56] J. Lobo and J. Dias. “Relative Pose Calibration Between Visual and Inertial Sensors”. *The International Journal of Robotics Research* vol. 26.6 (2007), pp. 561–575.
  - [57] L. Marconi, F. Basile, G. Caprari, R. Carloni, P. Chiacchio, C. Hürzeler, V. Lippiello, R. Naldi, J. Nikolic, B. Siciliano, S. Stramigioli, and E. Zwicker. “Aerial Service Robotics: The AIRobots Perspective”. In: *Proceedings of the International Conference on Applied Robotics for the Power Industry (CARPI)*. 2012, pp. 64–69.
  - [58] P. S. Maybeck. *Stochastic Models, Estimation, and Control*. Academic Press, 1979.
  - [59] F. Mirzaei and S. Roumeliotis. “A Kalman Filter-Based Algorithm for IMU-Camera Calibration: Observability Analysis and Performance Evaluation”. *IEEE Transactions on Robotics* vol. 24.5 (2008), pp. 1143–1156.

## BIBLIOGRAPHY

- [60] A. H. Mohamed and K. P. Schwarz. “Adaptive Kalman Filtering for INS/GPS”. *Journal of Geodesy* vol. 73.4 (1999), pp. 193–203.
- [61] C. Moler and C. V. Loan. “Nineteen Dubious Ways to Compute the Exponential of a Matrix”. *SIAM Review* vol. 20 (1978), pp. 801–836.
- [62] A. I. Mourikis and S. I. Roumeliotis. “On the Treatment of Relative-Pose Measurements for Mobile Robot Localization”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2006, pp. 15–19.
- [63] *MPU9150 MEMS IMU Product Specification*. Rev. 4.3. Invensense. 2013.
- [64] J. Nikolic, J. Rehder, M. Burri, P. Gohl, S. Leutenegger, P. Furgale, and R. Siegwart. “A Synchronized Visual-Inertial Sensor System with FPGA Pre-Processing for Accurate Real-Time SLAM”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2014, pp. 431–437.
- [65] J. Nikolic, P. Furgale, A. Melzer, and R. Siegwart. “Maximum Likelihood Identification of Inertial Sensor Noise Model Parameters”. *IEEE Sensors* (2015).
- [66] J. Nikolic, M. Burri, I. Gilitschenski, J. Nieto, and R. Siegwart. “Non-Parametric Extrinsic and Intrinsic Calibration of Visual-Inertial Sensor Systems”. *Accepted for publication in IEEE Sensors* (2016).
- [67] J. Nikolic, M. Burri, J. Rehder, S. Leutenegger, C. Hürzeler, and R. Siegwart. “A UAV System for Inspection of Industrial Facilities”. In: *Proceedings of the IEEE Aerospace Conference*. 2013, pp. 1–8.

## BIBLIOGRAPHY

- [68] J. Paul D. Abramson. “Simultaneous Estimation of the State and Noise Statistics in Linear Dynamical Systems”. PhD thesis. Massachusetts Institute of Technology, 1968.
- [69] F. Perrodin, J. Nikolic, J. Busset, and R. Siegwart. “Design and Calibration of Large Microphone Arrays for Robotic Applications”. In: *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*. 2012, pp. 4596–4601.
- [70] *Recommended Practice for Inertial Sensor Test Equipment, Instrumentation, Data Acquisition, and Analysis*. IEEE Std 1554-2005. IEEE Aerospace and Electronic Systems Society. 2013.
- [71] J. Rehder, J. Nikolic, T. Hinzmann, and R. Siegwart. “Extending Kalibr: Calibrating the Extrinsic of Multiple IMUs and of Individual Axes”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2016.
- [72] R. M. Rogers. *Applied Mathematics in Integrated Navigation Systems*. American Institute of Aeronautics and Astronautics, Inc., 2007.
- [73] E. Rosten and T. Drummond. “Machine Learning for High-Speed Corner Detection”. In: *Proceedings of the 9th European Conference on Computer Vision*. Springer-Verlag. 2006, pp. 430–443.
- [74] K. Schmid and H. Hirschmüller. “Stereo Vision and IMU Based Real-Time Ego-Motion and Depth Image Computation on a Handheld Device”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2013.
- [75] S. Shen, N. Michael, and V. Kumar. “Autonomous Multi-Floor Indoor Navigation with a Computationally Constrained MAV”.

## BIBLIOGRAPHY

- In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2011, pp. 20–25.
- [76] R. H. Shumway and D. S. Stoffer. “An Approach to Time Series Smoothing and Forecasting using the EM Algorithm”. *Journal of Time Series Analysis* vol. 3.4 (1982), pp. 253–264.
  - [77] I. Skog and P. Hndel. “Calibration of a MEMS Inertial Measurement Unit”. In: *Proceedings of the XVII IMEKO World Congress*. 2006.
  - [78] Y. Stebler, S. Guerrier, J. Skaloud, and M.-P. Victoria-Feser. “Generalized Method of Wavelet Moments for Inertial Navigation Filter Design”. *IEEE Transactions on Aerospace and Electronic Systems* vol. 50.2 (2014), pp. 2269–2283.
  - [79] H. Strasdat, J. Montiel, and A. Davison. “Real-Time Monocular SLAM: Why Filter?” In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2010, pp. 2657 –2664.
  - [80] H. Strasdat, A. Davison, J. M. M. Montiel, and K. Konolige. “Double Window Optimisation for Constant Time Visual SLAM”. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2011, pp. 2352–2359.
  - [81] D. Strelof and S. Singh. “Motion Estimation from Image and Inertial Measurements”. *The International Journal of Robotics Research* vol. 23.12 (2004), pp. 1157–1195.
  - [82] E. Stumm, A. Breitenmoser, F. Pomerleau, C. Pradalier, and R. Siegwart. “Tensor Voting Based Navigation for Robotic Inspection of 3D Surfaces Using Lidar Point Clouds”. *The International Journal of Robotics Research* (2012).

## BIBLIOGRAPHY

- [83] D. H. Titterton and J. L. Weston. *Strapdown Inertial Navigation Technology*. The Institution of Electrical Engineers and the American Institute of Aeronautics, 2004.
- [84] N. Trawny and S. I. Roumeliotis. *Indirect Kalman Filter for 3D Attitude Estimation*. Tech. rep. MARS LAB, University of Minnesota, 2005.
- [85] R. Vaccaro and A. Zaki. “Statistical Modeling of Rate Gyros”. *IEEE Transactions on Instrumentation and Measurement* vol. 61.3 (2012), pp. 673–684.
- [86] C. Van Loan. “Computing Integrals Involving the Matrix Exponential”. *IEEE Transactions on Automatic Control* vol. 23.3 (1978), pp. 395–404.
- [87] R. Voigt, J. Nikolic, C. Hürzeler, S. Weiss, L. Kneip, and R. Siegwart. “Robust Embedded Egomotion Estimation”. In: *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*. 2011, pp. 2694–2699.
- [88] S. Weiss and R. Siegwart. “Real-Time Metric State Estimation for Modular Vision-Inertial Systems”. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2011, pp. 4531 –4537.
- [89] S. Weiss, D. Scaramuzza, and R. Siegwart. “Monocular-SLAM Based Navigation for Autonomous Micro Helicopters in GPS-denied Environments”. *Journal of Field Robotics* vol. 28.6 (2011), pp. 854–874.
- [90] P. D. Welch. “The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging over Short, Modified Periodograms”. *IEEE Transactions on Audio and Electroacoustics* vol. 15.2 (1967), pp. 70–73.

## BIBLIOGRAPHY

- [91] Y. Yuksel, N. El-Sheimy, and A. Noureldin. “Error Modeling and Characterization of Environmental Effects for Low Cost Inertial MEMS Units”. *Position Location and Navigation Symposium PLANS* (2010).
- [92] Y. Yuksel and H. Kaygisiz. *Stochastic Errors of Low Cost MEMS Inertial Units*. Tech. rep. www.instk.org, 2012.
- [93] D. Zachariah and M. Jansson. “Joint Calibration of an Inertial Measurement Unit and Coordinate Transformation Parameters using a Monocular Camera”. In: *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. 2010, pp. 1–7.



# About the Author

Janosch Nikolic was born on January 4<sup>th</sup>, 1983. He completed a four year apprenticeship in electronics with Zellweger-Ascom in Uster ZH and Hombrechtikon ZH, Switzerland, and received the diploma in electrical engineering (Dipl. El.-Ing. FH) from the HSR, Rapperswil SG, Switzerland. He wrote his diploma thesis in collaboration with Steven Brossi at Starkey Hearing Technologies, MN, USA. He holds an M.Sc. degree in signal processing and communications from the University of Edinburgh, Edinburgh, UK.



During his studies, Janosch completed an internship with Schlumberger in Fahud, Oman, and a research exchange with Google's Advanced Technologies and Projects Group, CA, USA. Janosch Nikolic's research interest include the design, characterisation, and calibration of visual-inertial sensor systems with application to localisation and mapping, and robot guidance. He is an engineer with a passion for electronics, sensing, and estimation applied to challenging real-world problems.

