Airborne Navigation by Fusing Inertial and Camera Data

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Dipl.-Inform. Daniel Bender
aus
Bad Neuenahr-Ahrweiler

Bonn, 2018

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Daniel Cremers
2. Gutachter: Prof. Dr. Reinhard Klein

Tag der Promotion: 28.02.2018
Erscheinungsjahr: 2018

Acknowledgment

First of all, I would like to express my sincere gratitude to my supervisor Professor Dr. Daniel Cremers. I am very grateful for his guidance during the past years. Writing this thesis during my time at the Fraunhofer FKIE was only possible through the continuous support of Professor Dr. Wolfgang Koch, the head of the SDF department. I would also like to thank Prof. Dr. Reinhard Klein for being my second supervisor and Prof. Dr. Rainer Manthey as well as Jun.-Prof. Dr.-Ing. Ribana Roscher for their contributions to the examination board of the University of Bonn.

I especially thank my former college Dr. Marek Schikora. Without the countless hours of his time on discussions and his motivating words, completing this thesis would not have been possible. Many thanks go to my colleagues, especially to all the members of my research group, for the fruitful discussions, continuous support and unforgettable time we spent together. There are too many people to name them all and the chance of forgetting someone would be far too great.

Moreover, I thank the members of the CVPR group of the Technical University of Munich headed by Professor Dr. Daniel Cremers for the pleasant time during my stays. In particular, I wanna say thanks to Dr. Jürgen Sturm and Dr. Jakob Engel for the discussions which yielded many valuable impulses to my work.

Finally, I would like to apologize to my family, friends and everyone else affected by me for being too busy for everything else in life. Thank you so much for always supporting me.

Contents

Acronyms

6-DOF six-degrees-of-freedom

CPU central processing unit

DGPS differential global positioning system

ECEF earth-centered, earth-fixed

EKF extended Kalman filter

ENU east, north, up

FOG fiber optic gyroscope

g^{2}o general graph optimization

GCP ground control point

GNSS global navigation satellite system

GPS global positioning system

GPU graphics processing unit

IMU inertial measurement unit

INS inertial navigation system

LM Levenberg-Marquardt

LSD-SLAM large-scale direct monocular SLAM

MC Monte Carlo

MEMS microelectromechanical systems

NED north, east, down

PTAM parallel tracking and mapping

RANSAC random sample consensus

RMS root-mean-square

RMSE root-mean-square error

RTK real time kinematic

SBAS satellite based augmentation system

SFM structure from motion

SIFT scale invariant feature transformation

SLAM simultaneous localization and mapping

UAS unmanned aircraft system

UTM universal transverse mercator

WGS84 world geodetic system 1984

Notation

The following gives an overview of the notation used within this thesis. In general the type of variables is indicated by the applied capitalization and font style. Non-capital italic variables like x are used to represent scalar values. A vector is indicated by a non-capital bold variable, for example $\mathbf{x}$. Its i-th entry can be accessed by $\mathbf{x}(i)$. Furthermore, a matrix is noted by a capital bold letter, for instance $\mathbf{X}$. In a consistent manner, functions denote with the introduced style their image space. Examples for this are $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ but $\mathbf{f} : \mathbb{R} \rightarrow \mathbb{R}^2$.

Sets and transformation groups

$\mathbb{R}$	set of all real numbers
$\mathbb{R}_+$	set of all nonnegative real numbers
$\mathbb{R}^n$	n-dimensional real linear space
$SO(3)$	real special orthogonal group of rotations around the origin on $\mathbb{R}^3$; if $\mathbf{R} \in SO(3)$, then $\mathbf{R}^\top \mathbf{R} = \mathbf{I}$ and $\det(\mathbf{R}) = 1$
$so(3)$	tangent space (or Lie algebra) of the special orthogonal group $SO(3)$
$SE(3)$	real special Euclidean group of rigid transformations on $\mathbb{R}^3$; elements are pairs $(\mathbf{t}, \mathbf{R})$ with translation $\mathbf{t} \in \mathbb{R}^3$ and rotation $\mathbf{R} \in SO(3)$
$se(3)$	tangent space (or Lie algebra) of the real special Euclidean group $SE(3)$
$Sim(3)$	group of similarity transformations on $\mathbb{R}^3$; elements are triples $(\mathbf{t}, \mathbf{R}, s)$ with translation $\mathbf{t} \in \mathbb{R}^3$, rotation $\mathbf{R} \in SO(3)$ and scaling $s \in \mathbb{R}$

Operators

$\hat{\mathbf{x}}$	skew symmetric matrix of vector $\mathbf{x} \in \mathbb{R}^n$
$\breve{\mathbf{x}}$	initial guess of variable $\mathbf{x}$
$\circ$	pose composition operator

$\lvert x \rvert$	absolute value of scalar x
$\lvert\lvert \mathbf{x} \rvert\rvert_2$	Euclidean norm stating length of the vector $\mathbf{x} = [x_1, x_2, ..., x_n]$ by $\sqrt{x_1^2 + ... + x_n^2}$
$\forall$	for all
$\mathbf{x} \cdot \mathbf{y}$	scalar product (or dot product) of the vectors $\mathbf{x} = [x_1, x_2, ..., x_n]$ and $\mathbf{y} = [y_1, y_2, ..., y_n]$ defined by $\sum_{i=0}^{n} x_i y_i$
$\mathbf{x} \times \mathbf{y}$	cross product of the two linearly independent vectors $\mathbf{x}$ and $\mathbf{y}$ defining a vector which is perpendicular to both input vectors
$\exp(\mathbf{X})$	exponential of quadratic matrix $\mathbf{X}$ given by the power series $\sum_{k=0}^{\infty} \mathbf{X}^k/k!$
$\log(\mathbf{X})$	logarithm of quadratic matrix $\mathbf{X}$ defined as the inverse function of the matrix exponential

Important symbols

α	angle describing the maximal visual odometry drift
$\mathbf{b}$	rigid body displacement describing the mapping of a 3D point $\mathbf{p}$ from its initial to its final configuration according to a rigid body motion $\mathbf{B}$
$\mathbf{B}$	rigid body motion (or rigid-body transformation) describing how the coordinates of a moving object change over time
C, C_i	camera coordinate system, at time t_i
d	function describing the inverse depth of all pixels in the image plane as $\mathrm{d} : \mathbf{\Omega} \rightarrow \mathbb{R}_+$
E	set of graph edges connecting vertices V in a graph G
g	function describing the mapping of a grayscale image from its image plane to the intensity values as $\mathrm{g} : \mathbf{\Omega} \rightarrow \mathbb{R}$
G	2D graph containing a set of vertices V connected by edges E
$\mathbf{I}$	identity matrix
I, I_i	INS coordinate system, at time t_i
$\mathbf{k}$	set of intrinsic camera calibration parameters, describing the focal length, the principal point and the image distortion
λ	geographical longitude, an angle which states for each coordinate on earth the east or west distance to the Prime Meridian
$\mathbf{\Omega}$	image plane $\subseteq \mathbb{R}^2$ containing all pixels $\mathbf{x}$
$\mathbf{p}$	3D point $\in \mathbb{R}^3$
$\boldsymbol{\pi}$	image projection, which performs the mapping from a 3D point $\mathbf{p}$ to a pixel $\mathbf{x}$

Φ	Euclidean distance between two vectors
Φ_r	angle difference between two quaternions
ϕ	geographical latitude, an angle which states for each coordinate on earth the north or south distance to the equator
φ	roll angle rotating around the y-axis
ψ	yaw angle rotating around the z-axis
r_d	travel distance ration stating the mean distance from Monte Carlo runs in relation to the optimal distance
$\mathbf{r}$	ray with origin in the current start vertex $\mathbf{v}_s$ and intersecting the goal vertex $\mathbf{v}_g$
$\mathbf{r}_l, \mathbf{r}_r$	ray with origin in the current start vertex $\mathbf{v}_s$ and direction rotated by the angle of the maximal visual odometry drift α to the left / right of $\mathbf{r}$
$\mathbf{R}$	rotation matrix
σ_x	standard deviation of the scalar random variable x
$\mathbf{\Sigma_x}$	covariance matrix whose element in the i, j position is the covariance between the i^{th} and j^{th} elements of a random vector $\mathbf{x}$
$\mathbf{t}$	translation vector $\in \mathbb{R}^3$
θ	pitch angle rotating around the x-axis
v	function describing the depth variance of all pixels in the image plane as $\mathrm{v} : \mathbf{\Omega} \rightarrow \mathbb{R}_+$
$\mathbf{v}_i$	coordinates of a vertex $\in \mathbb{R}^2$
$\mathbf{v}_g$	coordinates of the goal vertex $\in \mathbb{R}^2$
$\mathbf{v}_s$	coordinates of the start vertex $\in \mathbb{R}^2$
V	set of vertices of a graph G
V	video coordinate system
W	world coordinate system
$\boldsymbol{\omega}$	axis-angle representation $\in \mathbb{R}^3$ of a rotation whereby the rotation angle in radians is encoded as $\|\|\boldsymbol{\omega}\|\|_2$
$\mathbf{x}$	pixel $\in \mathbb{R}^2$ describing a position in the image with its x- and y-coordinate
$\boldsymbol{\xi}$	twist coordinates $\in \mathbb{R}^6$; elements are pairs $(\mathbf{t}, \boldsymbol{\omega})^\top$ with translation $\mathbf{t} \in \mathbb{R}^3$ and rotation $\boldsymbol{\omega} \in \mathbb{R}^3$

Abstract

Unmanned aircraft systems (UASs) are often used as measuring system. Therefore, precise knowledge of their position and orientation are required.

This thesis provides research in the conception and realization of a system which combines GPS-assisted inertial navigation systems with the advances in the area of camera-based navigation. It is presented how these complementary approaches can be used in a joint framework. In contrast to widely used concepts utilizing only one of the two approaches, a more robust overall system is realized.

The presented algorithms are based on the mathematical concepts of rigid body motions. After derivation of the underlying equations, the methods are evaluated in numerical studies and simulations. Based on the results, real-world systems are used to collect data, which is evaluated and discussed.

Two approaches for the system calibration, which describes the offsets between the coordinate systems of the sensors, are proposed. The first approach integrates the parameters of the system calibration in the classical bundle adjustment. The optimization is presented very descriptive in a graph based formulation. Required is a high precision INS and data from a measurement flight. In contrast to classical methods, a flexible flight course can be used and no cost intensive ground control points are required. The second approach enables the calibration of inertial navigation systems with a low positional accuracy. Line observations are used to optimize the rotational part of the offsets. Knowledge of the offsets between the coordinate systems of the sensors allows transforming measurements bidirectional. This is the basis for a fusion concept combining measurements from the inertial navigation system with an approach for the visual navigation. As a result, more robust estimations of the own position and orientation are achieved. Moreover, the map created from the camera images is georeferenced. It is shown how this map can be used to navigate an unmanned aerial system back to its starting position in the case of a disturbed or failed GPS reception. The high precision of the map allows the navigation through previously unexplored area by taking into consideration the maximal drift for the camera-only navigation.

The evaluated concept provides insight into the possibility of the robust navigation of unmanned aerial systems with complimentary sensors. The constantly increasing computing power allows the evaluation of big amounts of data and the development of new concept to fuse the information. Future navigation systems will use the data of all available sensors to achieve the best navigation solution at any time.

Zusammenfassung

Unbemannte Flugsysteme (eng.: unmanned aircraft systems, UAS) werden vermehrt als Messsysteme eingesetzt. Dabei sind zur effizienten Nutzung Kenntnisse über ihre Position und Orientierung erforderlich.

Die vorliegende Arbeit erforscht die Konzeption und Realisierung eines Systems, welches GPS-gestützte inertiale Navigationssysteme mit den Fortschritten im Bereich der kamerabasierten Navigation verknüpft. Es wird vorgestellt, wie diese zwei komplementären Ansätze in einem gemeinsamen Framework vereint werden können. Im Vergleich zu weit verbreiteten Konzepten, welche nur einen der beiden Ansätze nutzen, wird ein robusteres Gesamtsystem realisiert.

Die präsentierten Algorithmen basieren auf dem mathematischen Konzept der Festkörpertransformationen. Nach erfolgter Herleitung, werden die Verfahren zunächst in numerischen Studien und Simulationen untersucht. Aufbauend auf den erzielten Ergebnissen werden Daten mit realen Systemen aufgezeichnet, ausgewertet und diskutiert.

Zur Systemkalibrierung, welche die Offsets zwischen den Koordinatensystemen der zugrunde liegenden Sensoren bestimmt, werden zwei Verfahren vorgestellt. Der erste Ansatz integriert die Parameter der Systemkalibrierung in den klassischen Bündelausgleich. Durch die Nutzung eines auf Graphen basierenden Verfahrens wird die Optimierung sehr anschaulich dargestellt. Der Ansatz erfordert die Nutzung eines präzisen inertialen Navigationssystems und verwendet die Daten eines Messfluges. Dabei ermöglicht die vorgestellte Methode eine flexible Flugplanung ohne die Nutzung kostenintensiver Bodenkontrollpunkte. Das zweite Verfahren ermöglicht die Systemkalibrierung auch für inertiale Navigationssysteme mit großen Fehlern in der Positionsbestimmung. Um die Optimierung des Rotationsanteils der Offsets ohne Nutzung der Positionsmessungen zu realisieren nutzt die Methode Linien als Beobachtungen. Die Kenntnis der Offsets zwischen den Koordinatensystemen der Sensoren ermöglicht die Transformation von Messungen in das jeweils andere Koordinatensystem. Basierend darauf wird ein Konzept vorgestellt, welches die Fusion von Messungen des inertialen Navigationssystems und eines Verfahrens zur visuellen Navigation realisiert. Dies führt zu robusteren Ergebnissen in der Bestimmung von Position und Orientierung des UAS. Darüber hinaus ist, im Gegensatz zu rein kamerabasierten Verfahren, die erstellte Karte des überflogenen Gebietes georeferenziert. Es wird gezeigt, wie diese Karte genutzt werden kann um ein unbemanntes Flugsystem, bei fehlendem oder gestörtem GPS Empfang, rein kamerabasiert zum Startpunkt seiner Mission zurückzukehren zu lassen. Dabei ermöglicht die hohe Genauigkeit der Karte die Nutzung von Abkürzungen durch zuvor nicht besuchte Gebiete unter Beachtung des maximalen Drifts.

Das evaluierte Konzept gibt einen Einblick in die Möglichkeiten zur robusten Navigation von unbemannten Flugsystemen mit komplementären Sensoren. Die weiterhin steigende Rechenkraft ermöglicht neben der Auswertung großer Datenmengen auch die Entwicklung neuer Konzepte zur Datenfusion. Zukünftige Navigationssysteme, werden die Daten aller vorhandenen Sensoren nutzen um jederzeit die bestmögliche Navigationslösung zu berechnen.

CHAPTER I

Introduction

"The real voyage of discovery consists not in seeking new landscapes, but in having new eyes"

— Marcel Proust

Sensor data fusion is a sub area of applied computer science. It is based on the principles from nature. All living creatures fuse information from their sense organs to interact with the world and achieve their goals. This concept is emulated by human beings, which created technology to sense the environment. Most sensors, like cameras or microphones, are based on the model of sense organs from living beings. Others create completely new concepts not found in nature. An example is the positioning based on the travel time from satellite signals by a global navigation satellite system (GNSS) like the global positioning system (GPS). All sensors have in common, that they only perceive partial aspects of the world. Therefore it is an essential task to fuse the information from multiple sensors to observe the situation as precise as possible [Koc14].

In the aerial context, multiple sensors are mounted on flying platforms like airplanes, helicopters or drones. The latter are a result from the technological development in the last centuries, which allows to control and steer aerial platforms without a pilot on board. They were initially developed for dangerous areas or missions by the military forces, but are nowadays also used in civil applications. Analog to the manned platforms multiple concepts exists, which can roughly be divided in rotary and fixed wing systems (Figure 1.1). Platforms using rotary wings are traditionally based on the concept of classical helicopters with one main and one tail rotor. The latter is required to balance out the torque of the main rotor. Nowadays, small drones often use multiple small rotors. By one half of the rotors rotating in the opposite direction of

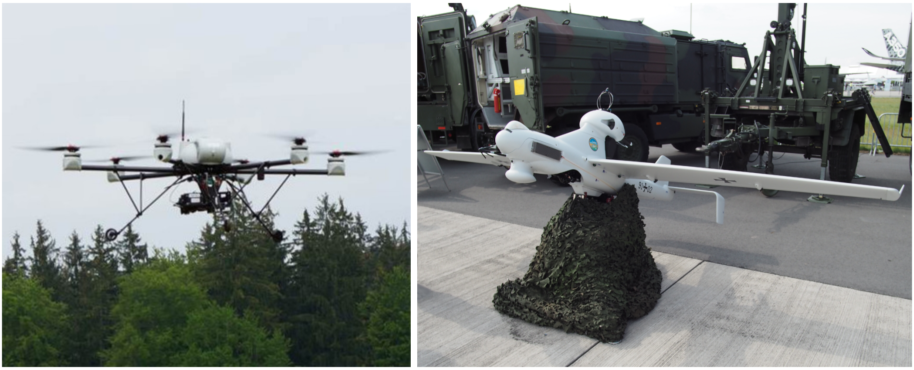


Figure 1.1: Left: Example of a rotary wing drone. The hexacopter of the type AR-200 is produced by AirRobot® GmbH & Co. KG. **Right:** Fixed wing drone of the type LUNA from EMT Ingenieurgesellschaft. The UAS is currently used by the German Army.

the others, the torque is balanced out. All rotary wings platforms have the advantage of vertical take-off and landing, which is possible at nearly every location, and the possibility to hover. In contrast to these systems, fixed wing drones are based on the concept of classical airplanes. They need a runway for starting and landing and are not able to hover. Their main advantage is the great efficiency by traveling long distances. Nowadays, also first platforms combining both concepts exist. By tilting the rotors or even the whole wings after a vertical take-off an efficiency close to the classical fixed wing platforms is achievable.

In the last years, an evolution of the general naming for these systems occurred. Starting from the military phrase drone, the name unmanned aerial vehicle (UAV) was introduced. Both terms are nowadays obsolete, but still widely used on the internet and in articles. According to the International Civil Aviation Organization (ICAO) the currently preferred term to describe the platform is unmanned aircraft (UA). This name includes remotely piloted and autonomous systems as well as a mixture of both approaches. To state that there is a ground control station and not only the aerial vehicle, unmanned aircraft system (UAS) is enforced, which is the phrase used in the following. An Equivalent term is remotely piloted aircraft system (RPAS), which brings the ground control station even more into focus.

During the last decade research and development in the area created platforms in nearly ever size. As a consequence of this diversity and the associated opportunities, UAS are nowadays used, or at least intended, for tasks in nearly all industrial sectors. The transfer of the pilot from the aerial platform to the ground allows smaller systems leading in many cases to cheaper production and operation costs. The downsizing also results in hard payload restrictions and for most systems only small and light sensors can be considered. This degrades the capability of a single sensor and brings into focus the fusion of the collected information of all available sensors to generate the best result for the application.

1.1 Motivation

Human beings have five traditionally recognized senses: sight, hearing, smell, taste and touch. According to the definition of modern physiology there exist a lot more. One of them is balance, which is sensed by the vestibular system in the inner ears. This organ provides the angular momentums and linear accelerations of the head and gives us the perception of spatial orientation and allows us to coordinate movements to keep our body in balance. Another non-traditional sense is kinesthesia. It provides information about the body movement by sensing the relative positions of neighbored body parts [PG09]. The brain integrates its information with balance and sight in a virtual sense of movement (Figure 1.2). Some animals have developed other senses to improve their capabilities of localize themselves and their movement. For example,

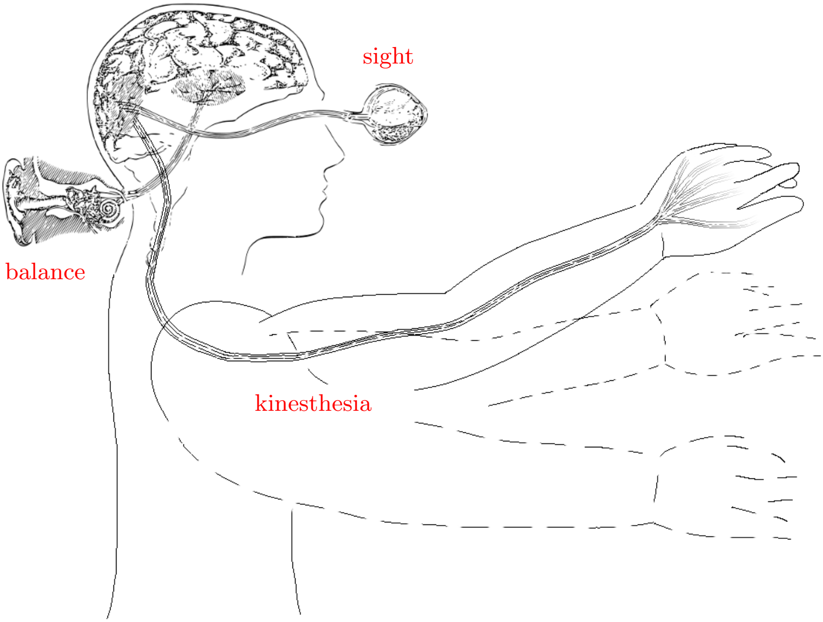


Figure 1.2: To determine their own movement, human beings combine sight with their sense of balance provided by the inner ear and kinesthesia, perceiving the position of body parts.

bats use echolocation to sense their environment by emitting ultrasound signals and listening for the echo reflected by nearby objects. Some birds sense the magnetic field, which allows them to perceive their current location and direction [CLD07].

Analog to the virtual sense of movement used by human beings and animals this thesis contributes to the development of this sense for robots. Especially for mobile platforms, the knowledge of the own pose is of particular importance. It ensures to travel along the desired path and allows to locate all other measurements in a consistent map. The self-localization of a platform is described as six-degrees-of-freedom (6-DOF) pose estimation including position and orientation. Its not only of interest for aerial platforms considered within this thesis, but also for ground based or maritime platforms and constitutes a fundamental task for many applications in the area of robotics, augmented reality or tracking. Various solutions have been proposed, but the challenge to provide a low-cost, real-time and robust solution, remains unchanged.

1.2 Why using a camera and an inertial navigation system?

Aerial photogrammetry is widely used for orthophoto and topographic map creation and originated over 100 years ago. The generation of these models require knowledge of the exterior orientations of the collected aerial images. These poses are determinable by the usage of a bundle adjustment software [TMHF00]. Prerequisites are a sufficient number of images from different viewpoints and a high image overlaps between them. In the aerial photogrammetry, this is achieved by collecting images from at least one downward looking camera while performing flight strips with a high image side overlap. The bundle adjustment defines the problem of simultaneously determining the observed scene geometry, the camera poses and the camera calibration parameters. Its name refers to the bundles of light rays between the observed 3D structure and the optical camera centers at the time of image exposure, which are adjusted to optimally satisfy the corresponding pixel positions. By using only one camera, the determined camera motion is not absolute but states a relative path at an arbitrary scale. The scaling to real-world coordinates is usually done by measuring the GPS positions of multiple ground control points (GCPs) in the outer regions of the observed area. The observations of the GCPs have to be identified in the images to determine the unknown scale factor. This makes clear that the bundle adjustment is a time consuming post processing step, which has to be performed after the measurement flight is completed.

In contrast to aerial map creation, time-critical surveillance and rescue tasks have a high demand for a more flexible flight planning and the direct determination of object localizations [SBKC10, BSK11]. One widely used approach is the online processing of images to estimate the current camera pose. This is known as visual odometry, a key component for the navigation with camera-based systems. For a long period, these systems were working only with a small subset of feature points to achieve real-time capabilities [KM07, KM09]. Nowadays, the increased computing power leads to semi-dense and dense approaches, which achieve a superior pose estimation by processing all image regions with gradients [ESC12, ESC14] or even the whole image [NIH$^+$11].

Despite the great development in visual odometry, the usage of an inertial navigation system (INS) leads to the most accurate real-time pose information in outdoor scenarios. Nowadays, measuring or surveillance flights are usually performed by a manned aircraft equipped with a digital large-format camera and a high precision GPS-corrected INS. This equipment enables direct georeferencing of the captured images and therefore provides very flexible flight planning without the necessity of GCPs. The utilization of the measured positions and attitudes for the exterior camera orientations eliminates the need for the time consuming post processing in form of a bundle adjustment or computational expensive visual odometry. The main drawback is that, in addition to the expensive devices, the execution of flights with manned aircrafts are very cost-intensive and consequently uneconomic for small areas.

There is ongoing research regarding the photogrammetric usage of small UAS with payload capabilities of a few kilograms. This leads to a technological challenge for the equipped sensors. They can not be freely selected anymore but have to fulfill the hard constraints of weight and size determined by the UAS. For a wide range of applications, and in particular for the aerial photogrammetry, these platforms are in general equipped with a miniaturized GPS-corrected INS and a small-format camera [Eis08]. As a result, the sensors will most likely produce less than optimal results compared to their unconstrained counterparts.

Nevertheless, clever processing, especially with methods combing the output from these complementary sensors, leads to accuracies suitable for most tasks. To use the pose information of the INS as outer orientation for the images captured by the camera, the devices have to be time synchronized and rigidly mounted. Using the measurements from an INS without considering the mounting offsets will introduce biases, which can highly degrade the overall system performance. Within the so called INS-camera calibration the static coordinate system transformation between the sensors is determined. More precisely the translational offsets (lever-arm) and the angle misalignments (boresight) are estimated. This is known as system calibration or more precisely INS-camera calibration.

A general framework for combining a camera and an INS is presented in Figure 1.3. The 3D world is reconstructed out of images captured by the camera, while the INS estimates the motion parameters. Latter can to some extend also be extracted from the camera images. By fusing the information of both systems the 3D reconstruction of the world is simplified and can be achieved at a lower computational cost. The

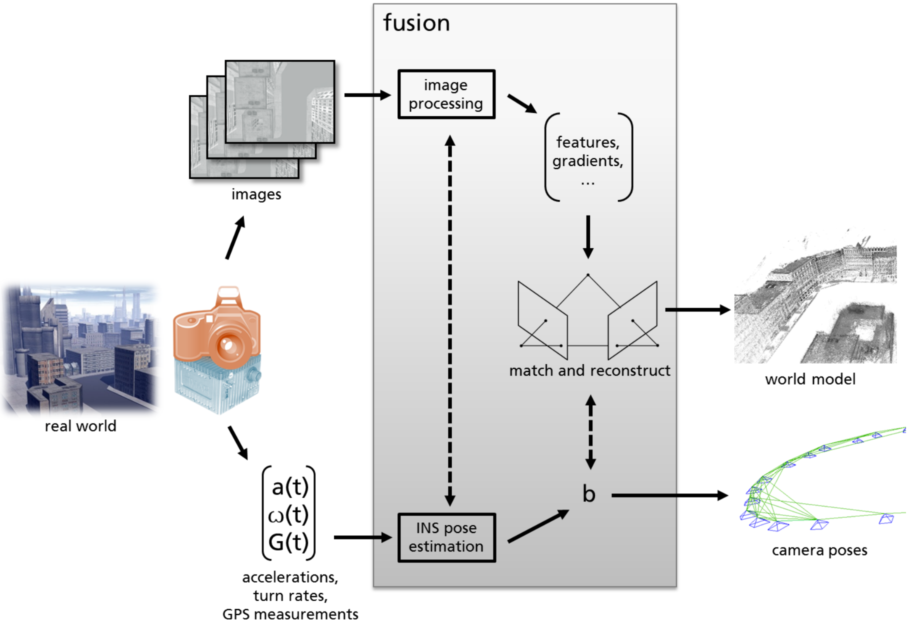


Figure 1.3: General framework combining a camera and an INS. The fusion between the sensor may occur at various levels while processing the measurements.

fusion between these sensors can roughly be divided into two concepts: loosely and tightly coupled systems. Latter feed the measurements of the inertial sensors and information extracted by the image processing in one single optimized filter. Realizations of this estimation problem were solved with an extended Kalman filter [QCZ01] or an unscented Kalman filter [ACVT04]. For loosely coupled approaches, which are considered in the following of this thesis, the INS and the image processing run independently, but use results from the other to improve their own performance. For example pose information from the INS can be used to predict the motion of features and visual information may be used to restrict the integration drift of the inertial sensors [CLD07].

To sum it up, the sensors have useful complementaries, which can compensate each others disadvantages. This is especially important if the data quality of the individual sensors is suboptimal on the basis of hardware miniaturization or cost savings. Clever processing of the sensed data leads to new applications in areas like augmented reality [KD03] or robotics [BCK16] and is expected to yield further benefits in the near future.

1.3 Focus

This thesis develops a multi-sensor system for accurate pose estimations with a sensor suite consisting of an INS and a camera with regard to small airborne platforms. The usage of an INS is the standard approach for the navigation since decades. Nevertheless, new developments in the field of camera technology and the constantly increasing computing power leads to improvements in the area of visual navigation. A combination of both approaches has the potential to be a low-cost and robust solution for the real-time pose estimation.

As part of this thesis two approaches to perform the system calibration between an INS and a camera are presented. Knowing the offsets between the underlying coordinate systems of the two sensors allows an easy transformation between them and makes various improvements possible. The integration of the INS pose estimations into a simultaneous localization and mapping (SLAM) approach will reduce ambiguities and increase the robustness by a smaller search space for the visual pose estimations. This results also in a reduction of the required computing power and even allows the operation in regions with poor textures. Further, the geolocations of the images are determined by measurements of the integrated GPS receiver. As a result the 3D map generated from these images by a SLAM approach are georeferenced. This allows the precise camera only navigation in previously observed regions and even the usage of shortcuts through unobserved areas. Combining the complementary sensors makes the overall system more robust and adds a redundancy to the navigation process. In the case of a sensor failure, the remaining one can be used to perform a controlled termination of the current mission.

It is also an essential point of this thesis that any methodological innovation is examined closely in numerical studies and verified by real-world experiments. The latter are realized by evaluating sensor data collected in measurement flights with aerial platforms.

1.4 Outline

After the introduction given in the current chapter, this thesis is organized as follows:

Chapter 2 provides the basic concepts used in the following chapters. Besides a definition of the used coordinate systems and how to convert coordinates between them, the mathematical concept of rigid body motions is introduced. The latter are used to describe the movement off the used sensors and also the mounting offsets between them. The mapping principle of a camera is described. By introducing the concepts of an INS, also a description of the operating principle of the contained inertial measurement unit (IMU) and the GPS is provided.

In Chapter 3 two approaches for the calibration of a sensor suite of a rigidly mounted camera and an INS are presented. The first allows the integration of the calibration parameters in the classical bundle adjustment without the requirement of expensive GCP observed in the calibration flight. The second approach enables the calibration of systems using a low-cost INS with a low positional accuracy.

Chapter 4 describes the proposed approach of developing a framework, which fuses INS measurements with poses extracted from a visual navigation system. It is shown how to integrate the INS measurements into the large-scale direct monocular SLAM (LSD-SLAM) framework. Integrating precise position updates from differential corrected GPS-signals leads to accurate maps of the observed area.

These maps are used in Chapter 5 to navigate an UAS in a previously visited area without a GPS. In the presented approach, an UAS returns to its starting position by only using its camera and the previously generated output of the SLAM procedure. Thereby, the path planing procedure considers also shortcuts through previously not explored areas.

Finally, this thesis is closed in Chapter 6 by giving a summary of the achieved contributions and obtained results. Possible extensions and a vision of future navigation systems are presented.

CHAPTER II

Fundamentals

"By failing to prepare, you are preparing to fail."

— Benjamin Franklin

The position as well as the orientation of a moving platform is of interest in nearly all of its applications. This results in a 6-DOF localization problem. The proposed solution fuses measurements from a sensor suite containing a camera and an INS. In this chapter the principles and methods which form the basis for the rest of this thesis are introduced. The used coordinate systems and the concept of rigid body motions are defined. Further, a short introduction of the working principles of the used sensors and the benefits of using them in a combined approach is given.

2.1 Coordinate systems

The different sensors take measurements in their associated coordinate systems and the fusion of these measurements requires a global coordinate system. An overview of the different types of these coordinate systems and how to convert between them is given in this section.

The self-localization dealt with in this thesis takes place in the three-dimensional space that we live in and thus the considered coordinate systems are of the same dimensionality. Two generic types of coordinates systems are used. Cartesian coordinate systems describe points in space by three distance values, most frequently named x, y and z, on three orthogonal axes. In contrast to that, spherical coordinate systems describe a point by its radial distance to the origin and two angles, describing a unique direction starting at the origin.

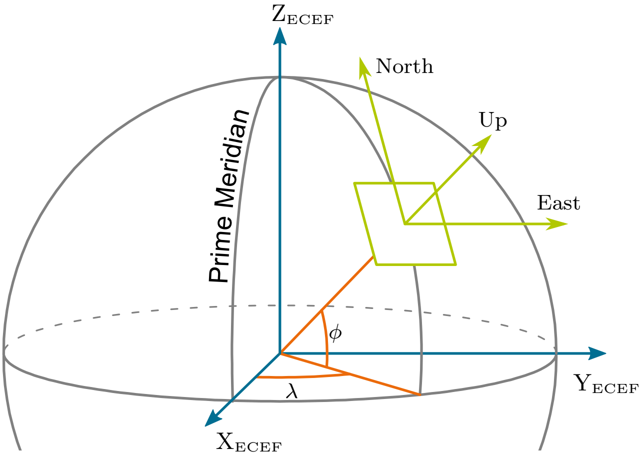


Figure 2.1: The sensor frames as well as the global world frame are ENU coordinate systems (green). They are local Cartesian coordinate systems with the origin tangential to the earth ellipsoid. Further the global ECEF coordinate system (blue) is used to convert between ENU-coordinates and GPS measurements. The latter are taken in latitude ϕ and longitude λ as polar coordinates and an altitude above the ellipsoid (orange).

The known representation of GPS coordinates as latitude, longitude and altitude is a description of a point in space with a spherical coordinate system (Figure 2.1). Thereby, the latitude value describes the angular distance to the equator with values from 0° to 90° up and from 0° to −90° down of the equator. In contrast, there is no natural reference for the second angle known as longitude. The prime median at a longitude of zero is by convention passing Greenwich, a district of London. Further, longitude values are defined in the range between 0° to 360° growing in east direction. The altitude value is not stated as distance from the center of earth, but above an ellipsoid fitted to the earth surface. The world geodetic system 1984 (WGS84) defines a reference ellipsoid, which is used for GPS measurements and also constitutes the current standard for most applications. Nevertheless, it must be noted, that some GPS devices do not state the altitude related to the reference ellipsoid but to the geoid. The latter describes the shape of the earth surface within the earth gravity field. In a good approximation, the geoid is represented by the mean sea level of the world's oceans. In contrast to the mathematically idealized reference ellipsoid, the surface of the geoid is irregular. The absolute difference of up to 100 m between them is known as undulation.

All other coordinate systems used in this thesis are Cartesian. The first to be introduced is the earth-centered, earth-fixed (ECEF) coordinate system. It is used as the basis of most satellite systems, because in contrast to the WGS84 no ellipsoid has to be chosen. Its origin is defined to be in the center of the earth. The positive z-axis points through the North Pole and the xy-plane contains the equator. Thereby the x-axis is defined to intersect the prime median and the y-axis to span a right-handed coordinate system (Figure 2.1). As a consequence, the ECEF coordinate system rotates with the earth.

Describing positions close to the earth surface with ECEF coordinates results in large numbers, which inhibit an intuitive interpretation. As an alternative, local ground reference frames are used. The origin of these coordinate systems is usually given as a point in WGS84 coordinates. The up-down direction given by the z-axis is defined to be orthogonal to the tangent plane of the reference ellipsoid. There are two primary used definitions differing in the axes orientation. For the north, east, down (NED) convention the x-axis points north, y-axis east and the z-axis downwards. This coordinate system is widely used in aviation. Nevertheless, in this thesis only the second convention named east, north, up (ENU) is considered. In this system the x-axis points east, the y-axis north and the z-axis up (Figure 2.1). Both conventions describe right-handed coordinate systems and a transformation between coordinates in these frames is straight forward. The conversion between WGS84 and ENU, with ECEF coordinates as an intermediate step, is well described by Drake [Dra02].

An often used alternative to describe global coordinates is the universal transverse mercator (UTM) system. A map projection, which uses a two-dimensional Cartesian coordinate system to describe a location on earth. Therefore it divides the earth surface in sixty zones, each with a width of six degrees longitude and its favorable transverse Mercator projection. Each of these zones can map a region from north to south with low distortion. The division in latitude bands is not part of the UTM system, but often used to define grid zones. A point on earth is defined by its grid zone and position. The latter is a coordinate pair stating the easting and northing in the zone.

2.2 Rigid body motion

The mathematical concept of Lie groups, which represent smooth topological groups, is used to describe transformations. Examples are the three-dimensional special orthogonal group $SO(3)$ for all 3D rotations and the special Euclidean group $SE(3)$ representing rigid body motions in the 3D space. The algebraic structures $so(3)$ and $se(3)$, named Lie algebras, are vector spaces which allow a minimal representation of the corresponding Lie groups. Some more insight of the associated theory is given in the following definition of rigid body motions. A comprehensive introduction is given in the book of Ma [MSKS03, Hal15].

In general, a rigid body motion $\mathbf{B} \in SE(3)$ describes how the points of a rigid object change over time. Instead of considering the continuous path of the movement, the

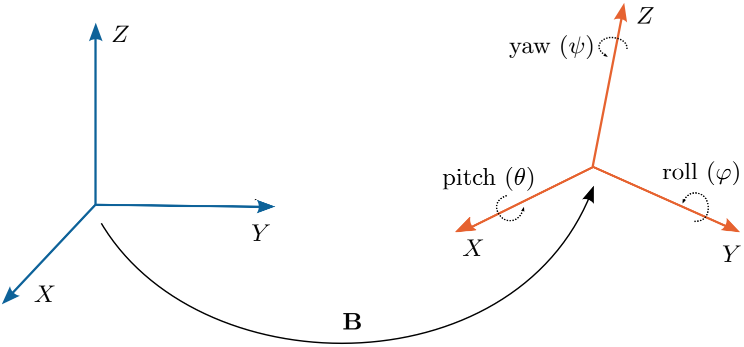


Figure 2.2: A rigid body motion **B** describes the movement of a rigid object, here depicted with its coordinate system. Instead of the whole movement, this figure visualizes the mapping from the initial to the final configuration.

mapping between the initial and the final configuration of the rigid body motion is considered. This mapping **B** can be described by a rotation matrix $\mathbf{R} \in SO(3)$ and a translation vector $\mathbf{t} \in \mathbb{R}^3$ (Figure 2.2). Consequently, the rigid body displacement **b** of a 3D point $\mathbf{p} \in \mathbb{R}^3$ can be performed by:

$$\mathbf{b} : SE(3) \times \mathbb{R}^3 \to \mathbb{R}^3 \,, \tag{2.1}$$

$$\mathbf{b}(\mathbf{B}, \mathbf{p}) = \mathbf{R}\mathbf{p} + \mathbf{t} \,. \tag{2.2}$$

The representation of the rotational part of **B** in form of the overdetermined rotation matrix **R** is not suitable for the optimizations performed in this thesis. Therefore, in the following chapters two different minimal representation, namely Euler angles and skew-symmetric matrices are used. The utilized form of the first are Z-X-Y Euler angles $(\psi, \theta, \varphi) \in \mathbb{R}^3$, called yaw, pitch and roll. These describe the rotational part of the movement as a mapping from $\mathbb{R}^3$ to $SO(3)$ by consecutive rotations around the principal axes in the order Z, X and finally Y:

$$\mathbf{R}(\psi, \theta, \varphi) = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\varphi & 0 & \sin\varphi \\ 0 & 1 & 0 \\ -\sin\varphi & 0 & \cos\varphi \end{bmatrix} . \tag{2.3}$$

The Z-X-Y rotation order leads to singularities at $\theta = \pm{}^{\pi}/_{2}$. This corresponds to a pitch angle of ±90°, which will never be achieved with the utilized aerial platforms and therefore does not constitute a problem. Otherwise, the rotational part of the rigid body motion can be reformulated with a different rotation order of the Euler angles or skew-symmetric matrices.

The latter are part of twist coordinates, which can be described by a vector $\boldsymbol{\xi} = (\mathbf{t}, \boldsymbol{\omega})^\top \in \mathbb{R}^6$. Thereby, $\mathbf{t} \in \mathbb{R}^3$ describes the translational and the skew-symmetric matrix $\hat{\boldsymbol{\omega}} \in so(3)$ the rotational part of the full motion. The rotation angle in radians is encoded as $||\boldsymbol{\omega}||_2$. An element $\hat{\boldsymbol{\xi}} \in se(3)$ can be written in its homogeneous representation as:

$$\hat{\boldsymbol{\xi}} = \begin{pmatrix} \hat{\boldsymbol{\omega}} & \mathbf{t} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\omega_3 & \omega_2 & t_1 \\ \omega_3 & 0 & -\omega_1 & t_2 \\ -\omega_2 & \omega_1 & 0 & t_3 \\ 0 & 0 & 0 & 0 \end{pmatrix} . \tag{2.4}$$

Given $\boldsymbol{\xi}$, a rigid body motion is described by the matrix exponential, which is defined as the always converging power series:

$$\exp : se(3) \to SE(3) \ , \tag{2.5}$$

$$\exp(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} \frac{\hat{\boldsymbol{\xi}}^k}{k!} \ . \tag{2.6}$$

The inverse of this exponential map is defined as the matrix logarithm and therefore fulfills the following relation:

$$\exp(\log(\mathbf{B})) = \mathbf{B} \ . \tag{2.7}$$

This leads to an alternative formulation of equation (2.2) using twist coordinates to describe the displacement of a point $\mathbf{p} \in \mathbb{R}^3$ according to the rigid body motion $\mathbf{B}$ as follows:

$$\mathbf{b} : SE(3) \times \mathbb{R}^3 \to \mathbb{R}^3 \ , \tag{2.8}$$

$$\mathbf{b}(\mathbf{B}, \mathbf{p}) = \exp(\boldsymbol{\xi})\mathbf{p} \ . \tag{2.9}$$

The concatenation of two rigid body motions is a simple multiplication in the $SE(3)$ space and allows defining the pose concatenation operator:

$$\circ : SE(3) \times SE(3) \to SE(3) \ , \tag{2.10}$$

$$\mathbf{B}_{kj} \circ \mathbf{B}_{ji} = \mathbf{B}_{kj}\,\mathbf{B}_{ji} = \mathbf{B}_{ki} \ . \tag{2.11}$$

This relationship can easily be utilized for the twist coordinate representation by applying the exponential mapping stated in equation (2.6).

2.3 Camera

A digital camera is a device which collects light on an electric sensor area to form an image of the observed area. The operating principle is based on the classical pinhole model, whereby the object, projection center and sensor element, known as pixel, are positioned on a line (Figure 2.3). Nowadays, cameras are equipped with lenses, which allow the collection of more light in the same amount of time compared to the

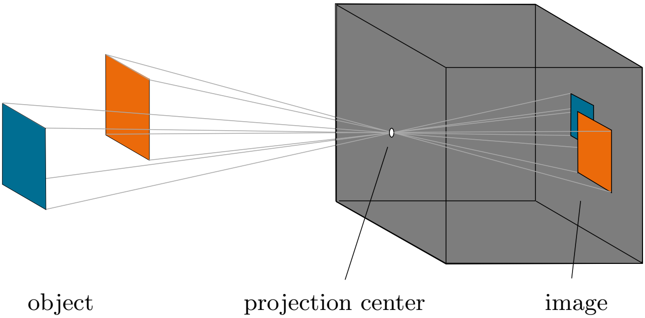


Figure 2.3: The image generation off all cameras is based on the pinhole camera. An inverted image of the scene is captured on the sensor area by collecting light passed through the pinhole for a fixed amount of time.

classical pinhole camera. Nevertheless, the process of image generation remains the same. Opening the camera shutter for a very short period forms an image by collecting light emitted or reflected from the observed objects. The different amount of light collected for the individual pixels characterizes the created image. A mathematical description of the image is given by the function g describing the mapping from the image plan $\boldsymbol{\Omega} \subset \mathbb{R}^2$ to real numbers:

$$\mathrm{g} : \boldsymbol{\Omega} \rightarrow \mathbb{R} \,. \tag{2.12}$$

In practice this relationship is often discretized by a two dimensional array of pixels mapped to 8-bit values in the range of $[0, 255]$. This corresponds to a grayscale image. To create a color image, the concept of the Bayer pattern [Bay76] is applied to nearly all image sensors. Color filters in the basic colors red, green and blue are attached in front of the pixels in an alternating pattern. Based on the brightness perception of the human eye, half of these filters are green, one quarter red and one quarter blue. Thus, each pixel only collects information for one of the three basic colors. To create an image with three complete color channels, the missing values are interpolated.

Various settings of the camera influence the image creation process. The exposure time, specifying the time the shutter is open and light is collected, is a critical value. Depending on the brightness of the scenario the appropriate values for images that are neither to dark nor to bright differ. It has also to be considered, that high exposure values in combination with egomotion of the camera or moving objects in the scene result in motion blur. In low light scenarios the sensor gain can be increased, which allows the collection of more light in the same amount of time. As a downside, the signal-to-noise ratio gets worse. Modern industrial cameras (Figure 2.4) automatically adapt these settings in ranges defined by the user beforehand to achieve a target brightness of the image. In contrast to the classical pinhole model, the usage of lenses

Figure 2.4: Industrial camera Prosilica GX3300 from Allied Vision Technologies GmbH with an attached wide-angle lens from Carl Zeiss AG. The camera allows grabbing the images in real-time via a standard Ethernet port while the image exposure can be externally controlled via a signal connected to a trigger jack.

leads to images which are not sharp for all mapped distances. The focus distance has to be set. In the context of aerial images this is in nearly all cases done by setting the focus to infinity.

To use images captured by a camera for photogrammetric measurements, the projection from a 3D point to a pixel on the sensor area is modeled. The process is described by the intrinsic camera parameters which converts the measurements process of cameras using lenses to the pinhole camera projection.

The set of intrinsic camera calibration parameters is defined by:

$$\mathbf{k} = \{f_x, f_y, o_x, o_y, k_1, k_2, k_3, p_1, p_2\} \ , \tag{2.13}$$

whereby (f_x, f_y) describe the focal length and (o_x, o_y) the principal point of the camera. Furthermore, the radial distortion with the parameters (k_1, k_2, k_3) and the tangential distortion by (p_1, p_2) are expressed. The projection $\boldsymbol{\pi}$ performs the mapping from a transformed 3D point $\mathbf{p} = (x, y, z)^\top$ to pixel coordinates as:

$$\boldsymbol{\pi} : \mathbb{R}^9 \times \mathbb{R}^3 \rightarrow \mathbb{R}^2 \ , \tag{2.14}$$

$$\boldsymbol{\pi}(\mathbf{k}, \mathbf{p}) = \left(\frac{df_x x}{z} - o_x + t_x, \frac{df_y y}{z} - o_y + t_y \right)^\top \ , \tag{2.15}$$

with the radial distortion factor d and the tangential distortion offsets t_x and t_y being defined as follows [Bro66]:

$$d = 1 + k_1 \left(\frac{x^2 + y^2}{z^2} \right) + k_2 \left(\frac{x^2 + y^2}{z^2} \right)^2 + k_3 \left(\frac{x^2 + y^2}{z^2} \right)^3 \ , \tag{2.16}$$

$$t_x = 2p_1 \frac{xy}{z} + p_2 \left(\frac{3x^2 + y^2}{z^2} \right) \ , \tag{2.17}$$

$$t_y = p_1 \left(\frac{x^2 + 3y^2}{z^2} \right) + 2p_2 \frac{xy}{z} \ . \tag{2.18}$$

The described camera projection assumes that the position of the 3D point is described in the camera coordinate system. Knowledge of the extrinsic camera parameters, describing the camera pose with its position and orientation in a global coordinate frame, allows the transformation of a 3D point to the camera coordinate system. This can be realized by applying the rigid body displacement of the camera pose as described in the previous section. Altogether, the mapping from a 3D point $\mathbf{p}$ in the global coordinate system to a pixel position in the image created by a camera at pose $\mathbf{B}$ is given by $\boldsymbol{\pi}(\mathbf{k}, \mathbf{b}(\mathbf{B}, \mathbf{p}))$ applying the previously introduced equations (2.9) and (2.15).

2.4 Inertial navigation system

Inertial measurement system are since many years used for navigational purposes of air, sea and ground vehicles. By measuring the acceleration, information about the speed and position are generated by integration. The combination of accelerometers with gyroscopes, measuring the turn rates, in one device is designated as IMU. Summing up measurements of the internal sensors over a period leads to a drift in the determined position and orientation, which is in most outdoor applications compensated by independent GPS measurements. An INS combines these two subsystems to produce accurate pose information while moving in the 3D space. The state-of-the-art integration techniques for those systems are realized by the application of Kalman filtering [GWA07]. The six-dimensional pose estimation is in general stated with a position in global latitude, longitude and altitude coordinates and the rotational part related to the ENU frame with origin in the estimated position. Thus, the INS states its transformation from the ENU frame to its body coordinate system (Figure 2.5).

Commercially available INS devices are very different from one another. They can roughly be divided based on the gyroscope technology of the IMU, which has a direct relation to the accuracy, size and cost of the whole system. Currently there are two technologies mainly used. Namely, the more classical fiber optic gyroscope (FOG), which are very accurate, and the uprising microelectromechanical systems (MEMS) gyroscopes. The latter are a lot smaller and are available in various accuracy levels. A rise of research activities in the area of MEMS technology results in improved error characteristics, a better sensitivity and advanced fusion methods on embedded hardware. Starting by the entry-level category, which are very cheap and tiny. They provide an output of the device orientation and thanks to MEMS based GPS chipsets also in the position. This device class is widely used in todays smartphones and sensor boards. An example for latter is the Razor IMU from SparkFun Electronics, which integrates three accelerometers, three gyroscopes and three magnetometers with an onboard processor and a serial interface. Such sensor boards are for example used to stabilize quadcopters. However, they are not considered in the sensor concepts

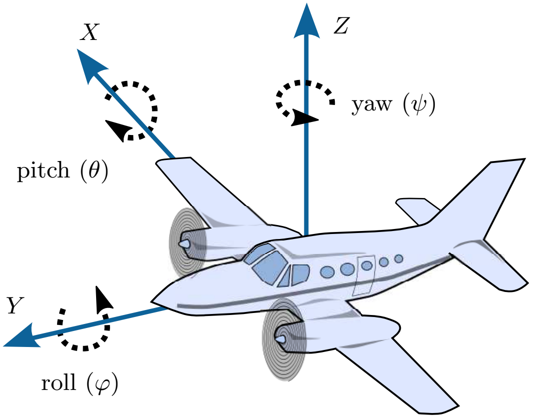


Figure 2.5: Body coordinate system of an INS integrated into an airplane. If the coordinate system of the platform is realized with a different orientation, the axes are usually switched internally to match the depicted configuration.

Table 2.1: General information and stated accuracies of three selected INSs

	Razor IMU (SparkFun Electronics)	Ellipse-D (SBG Systems)	iTraceRT-F200-E (iMar GmbH)
size	2.8×4.1 cm	$8.7 \times 6.7 \times 3.15$ cm	$14.8 \times 14.8 \times 10.4$ cm
weight	6 g	180 g	2400 g
gyroscopes	MEMS	MEMS	FOG
connection	serial interface	USB	USB
specialty	no GPS	dual GPS, RTK	RTK
position	-	1.2 m	1.8 m
		2 cm + 2 ppm (RTK)	2 cm + 2 ppm (RTK)
pitch/roll	n/a	0.1°	0.01°
yaw	n/a	<0.2° (dual GPS)	0.025°
velocity	n/a	0.03 m/s	0.01 m/s

evaluated in this thesis. At the other end of the MEMS spectrum, there are highly optimized INS, like the in this thesis used Ellipse-D from SBG Systems [SBG15]. This device is smaller, lighter and cheaper but still not as accurate as the also in the real-world evaluations utilized iMar iTraceRT-F200-E based on FOG technology [iMA13]. The specifications of the devices are stated in Table 2.1 and an approximation of the size ratio between the devices is visualized in Figure 2.6.

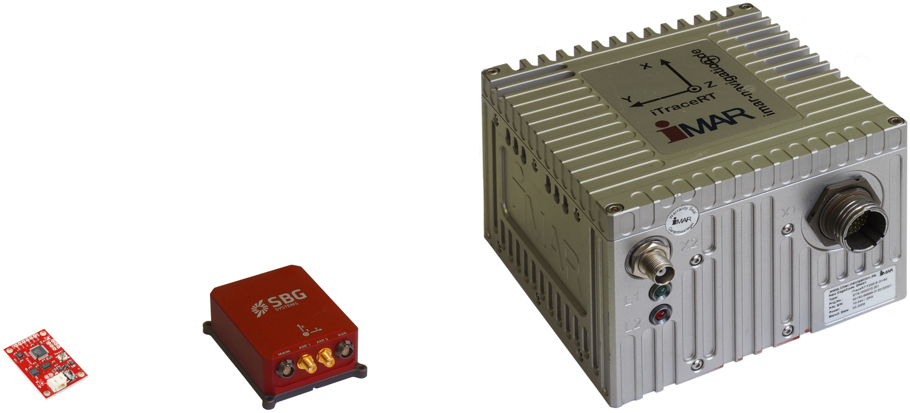


Figure 2.6: Approximated visualization of the size ratio between three selected INSs. **Left:** Razor IMU from SparkFun Electronics **Middle:** Ellipse-D from SBG Systems **Right:** iTraceRT-F200-E from iMAR Navigation GmbH

2.4.1 Inertial measurement unit

A typical IMU consists of three accelerometers, three gyroscopes, a clock and a processing unit. The accelerometers are arranged orthogonally to each other, producing output in 3D space. The measurement principle of an IMU is based on the second law of motion from Newton, which states that the movement of a body is determinable, if the forces acting on the body are known. For measurements on earth this means that the force of gravity and the earth rotation rate have to be compensated.

Classical systems decouple the accelerometers from the angular movements of the platform by mounting rotating gyroscopes in gimbals. As a result, the accelerometers are held in their fixed inertial frame, which allows an easy compensation of the external forces, producing acceleration measurements of the device. At the same time, the turn rates are given by extracting the angles of the gimbals through potentiometer pickoffs. These systems are based on a complex mechanical setup.

The progress in electronic components leads to the development of strapdown IMUs. In contrast to the constant inertial frame of the classical systems, all sensors are rigidly mounted and move with the platform. The navigation computer uses the angular measurements of the gyros to compensate the force of gravity and the earth rotation, which now effect accelerometer measurements depending on the orientation of the device.

The underlying concept of summing up measurements, and thus also small errors, over time leads to drift in the absolute position and orientation. A commonly used sensor type to improve the accuracy are vector magnetometers, which measure the earth's magnetic field in one direction. Analog to the accelerometers and gyroscopes, the typical usage are three orthogonal mounted magnetometers. Their measurements of the local magnetic field direction and strength are combined to determine the north direction. Magnetometers are affected by local disturbances caused by close magnetic objects. Nevertheless, they can be used to improve the accuracy of the calculated orientations, while the overall system remains self contained.

Another common approach is the integration of absolute pose estimations. This concept uses the high sampling rate of the inertial sensors to fill the gap between the drift free updates from an absolute positioning system, like GPS [Woo07].

2.4.2 Global positioning system

The NAVSTAR GPS, an acronym for navigation system with time and ranging global positioning system, was developed on behalf of the United States Department of Defence for military applications and later opened for worldwide civilian use. It allows the precise determination of location, velocity and time at each position on earth.

The principle of GPS is based on the measurement of signals constantly broadcasted by at least 24 satellites orbiting the earth. They are positioned in six different orbits in a manner, that at every time and any position on earth at least four satellites are in direct line-of-sight. If an object like a mountain or a building blocks one of the paths, the radio transmission from that satellite will not arrive at the receiver. Each signal contains information about the satellite status, ephemeris data describing the location of the satellite and a high-precision time stamp. The latter is generated from a synchronized atomic clock installed in the satellite.

A GPS receiver determines the traveling time of received signals by comparing the timestamps encoded in the satellite messages with its time at the moment of reception. Since the signals are traveling at a known speed, the distances to the satellites can be calculated. This information combined with the transmitted locations of the satellites allows the receiver to determine its own position by trilateration. Required are measurements from at least four satellites, because in addition of solving for the three positional components, the clock error of the receiver has to be corrected.

In addition to the determination of the position, receivers are also capable of deriving their current velocity from the GPS signals by using the Doppler shift. This effect describes how a radio signal transmitted or received by a moving object is compressed or stretched and allows the computation of the receiver's velocity up to an accuracy of a few centimeters per second. An improved performance can be achieved by processing differences in consecutive carrier phase measurements.

Some errors, especially signal delays by changing ionospheric conditions and clock errors, can be significant reduced by using differential measurement methods. The usage of differential global positioning system (DGPS) requires at least one additional reference receiver, placed at known coordinates. This receiver determines its own position according to the methods explained before. The difference between these values and its know position allows the generation of correction data, which is usually transmitted wireless to a second, possible moving, receiver. This so called rover uses the correction data to calculate more accurate estimations of its own position, velocity and time. One special realization of the DGPS principle is a technique called real time kinematic (RTK). Instead of the content of the signal, the carrier wave is used for the distance estimation between the satellite and the rover. The higher frequency of the carrier wave compared to the modulated signal leads to a major improvement in the achievable accuracy. Instead of errors of a few meters, accuracies in the centimeter range are possible. The difficulty is the alignment of the carrier phase where all cycles are similar to each other. This ambiguity is resolved to some extent by statistical methods in a differential setup. In practice, the reference receiver transmits its observed phase of the carrier to the rover, which compares this measurement with its own to very precisely determine its relative position.

In the context of GPS there are further aspects like the military code or the ground control segment, which are not required to understand the work presented in this thesis. A deeper insight is given in the book of Kaplan and Hegarthy [KH05]. They also discuss alternative GNSS. Namely, the Russian GLONASS, the European Galileo, the Chinese BeiDou and the Japanese QZSS program. The last three are currently not in full service and more satellite launches are planned for these systems in the coming years. Without limiting generality, this thesis will always refer to GPS, although the concepts are applicable to any GNSS.

CHAPTER III

System calibration

"We have an unknown distance yet to run,
an unknown river to explore."

— John Wesley Powell

To achieve the best possible performance of a multisensor system, a calibration of all existing parameters which have a direct influence on the collected data is required. This includes parameters describing the measurement process of the individual sensors and in addition parameters representing the relation between the sensors. The INS and the camera are rigidly mounted on an aerial platform, which allows reusing the results of the calibration process for following measurements flights. This enables to fully exploit the best possible conditions to calibrate the whole system, while not depending on calibration constraints during the data collection.

Own publication on this subject: The work presented in this chapter has been partly published in [BSSC13], [BSSC14] © 2014 IEEE and [BCK16] © 2016 IEEE.

3.1 Introduction

The investigations for the INS-camera calibration performed in the last decades were targeted at manned aircrafts equipped with high-precision INSs and aerial cameras at high altitude. In contrast, the goal in this chapter is the direct georeferencing of images from non-metric cameras mounted in the payload of small UASs. Two calibration procedures to determine the INS-camera calibration (Figure 3.1) are proposed.

The first approach, shows how to perform an in-flight calibration with an open source toolkit for general graph optimization to bypass the need for a commercial bundle adjustment package. In contrast to the second procedure, a precise INS is required

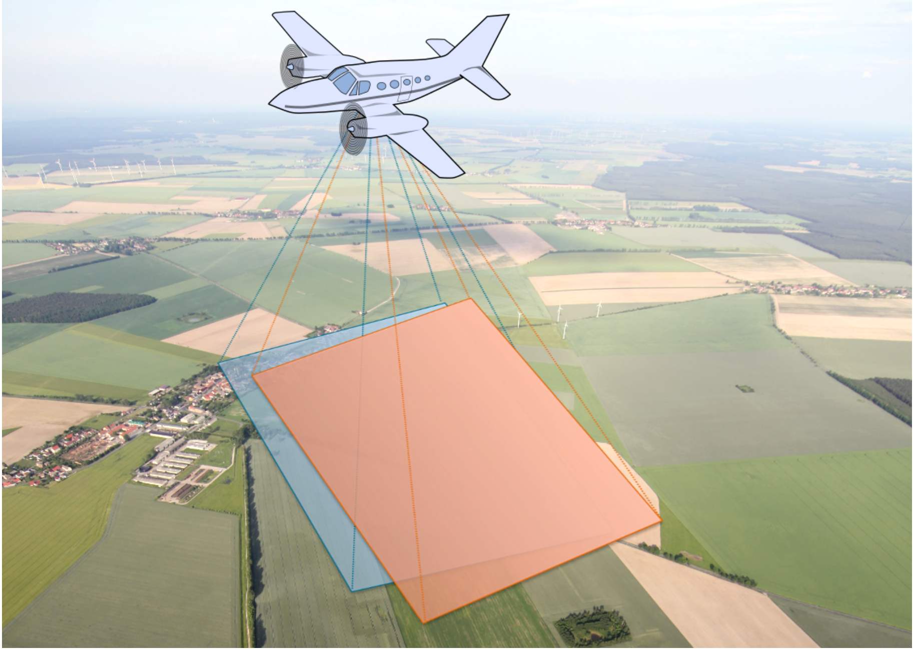


Figure 3.1: Without a system calibration, describing the offsets between the INS and camera coordinate systems, the direct georeferencing of images collected with aerial platforms leads to errors in the ground coordinates (orange). These errors are reduced by using the calibrated mounting parameters during the flight mission (blue).

[BSSC13, BSSC14]. The proposed position free approach can be used for all INS accuracy levels. Input for the calibration is a data set of time synchronized INS poses and camera images observing a planar checkerboard pattern. The latter is in general used to calibrate the intrinsic camera parameters, but allows with the presented approach also the calibration of the angular misalignment of the mounting offsets [BRS+16].

This chapter starts with an overview on related research areas. After a general definition of the INS-camera calibration problem, the two proposed approaches are presented. Both procedures are analyzed in simulated scenarios and finally real-world results achieved with the procedures are evaluated and discussed.

3.2 Related work

A related process of the INS-camera calibration is the so called hand-eye calibration. Given a camera mounted on a robot arm, the rigid-body transformation between the coordinate systems of these devices is estimated. As a result, measurements from the acquired images can be transformed in the robot arm coordinate system. This is necessary to interact with objects recognized and located in the images. The calibration is done by estimating the rigid body transformation from corresponding poses. A direct solution can be computed by firstly optimizing the rotational part and solving the equations for the translation afterwards [TL89]. In contrast, it was shown that the nonlinear optimization for rotation and translation at the same time leads to more robust results in case of noise and measurement errors [HD95]. The motion of the robot arm is typically obtained from encoders, whereas nearly all approaches determine the camera movement by observing a calibration pattern. In contrast to that, the camera poses can be determined by a structure-from-motion approach. As a drawback, the camera movement can only be estimated up to a similarity transformation. This leads to an unknown scale, which in addition has to be estimated during the calibration process [AHE01]. The obtained results were not as accurate as the calibrations from methods using camera calibration patterns. However, the approach has the advantage of being feasible without any additional equipment and therefore allows a recalibration during operation.

The algorithms developed for the hand-eye calibration had a big influence on another related problem: The calibration between an IMU and a camera. These devices can be combined to a vision-aided inertial navigation system. Measurements of the IMU in form of rotational velocities and linear accelerations can be integrated to determine the positions, velocities and attitudes of the device. During this process small estimation errors are summed up over time, which should be corrected by an aided sensor. Typically, this is achieved by exploiting GPS measurements, which is not possible in space, underwater or indoor applications (Chapter 2.4). An alternative is the usage of a camera in addition to the IMU. Prerequisite for exploiting camera-based corrections is the knowledge of the transformation between the two devices. These offsets can be computed with modified hand-eye calibration algorithms [LD07]. The fact that consecutive measurements from shaft encoders are uncorrelated in contrast to data from an IMU results in different noise characteristics. Considering these time correlations leads to a higher accuracy of the estimations [MR08]. The authors utilize an extended Kalman filter (EKF) for estimating the pose transformation, which facilitates the computation of the covariance for the estimated parameters as an indicator of the achieved accuracy. In [WAL$^+$12] the authors describe an approach for navigating an UAS based on measurements from an IMU and a camera. By not using a calibration pattern, they realize the online estimation of the mounting parameters between these sensors with an EKF formulation.

To calibrate the transformation between a GPS-aided INS and a camera, the algorithms for the IMU-camera calibration can be used. This is based on the fact that an IMU is one fundamental components of an INS. The only prerequisite for applying a system calibration with these sensors is that the raw measurements of the IMU is accessible. However this is not the predominantly used method to perform a system calibration between an INS and a camera. The commercially available INSs provide an integrated filtering process, exploiting the GPS measurements to correct the IMU estimations. In conjunction with GPS correction signals from ground control stations an INS accuracy in the range of a few centimeters for the position and a few hundredths of a degree for the attitude is achievable. Thus, in contrast to an IMU, an INS provides a reliable stand alone source describing its own movement. This leads to the estimation of the rigid-body transformation between the INS and the camera with methods similar to the hand-eye calibration. In a first step the camera movement is calculated with a structure from motion (SFM) approach and refined in a bundle adjustment procedure. The observations of ground control points are used to scale the 3D model to real-world coordinates. In a second step, the transformation between the two devices is estimated by relating these absolute camera poses to time synchronized measurements from the INS. This widely used approach is known as two step procedure [CSH00]. The advantage is that each bundle adjustment package can be used without modifications. On the other hand, the integration of the mounting parameters as variables to optimize in the bundle adjustment is possible. This approach is known as single-step calibration and induces a simpler calibration of the mounting offsets due to more flexible flight courses [PF02]. The simultaneous optimization of the rigid-body transformation between the devices and the intrinsic camera parameters should consider correlations between these values. An analysis of the flight pattern influence on the calibration parameters is discussed in [KHB11] with the conclusion that the calibration parameters can be decoupled by performing the data collection at two different altitude levels. The authors also state that at least one GCP is needed for the estimation of the vertical lever-arm and that the addition of further GCPs does not improve the estimation results identifiable. The system calibration has been examined and discussed widely within the european organisation for experimental photogrammetric research (OEEPE) test on integrated sensor orientation [HJW02]. They conclude that it is a serious alternative for applications depending on directly geo-referenced images and flexible flight paths, even though it does not achieve the accuracies of the classical bundle adjustment.

3.3 Problem definition

The objective of this chapter is the estimation of the rigid body transformation between a camera and an INS. The knowledge of this transformation allow a direct geo-referencing of the images captured with the camera with pose measurements from the INS. Precondition for this are a rigid mounting and a hardware synchronization by a cable connecting the trigger jacks of the two devices. If these conditions are fulfilled, the rigid body transformation (Figure 3.2) can be estimated.

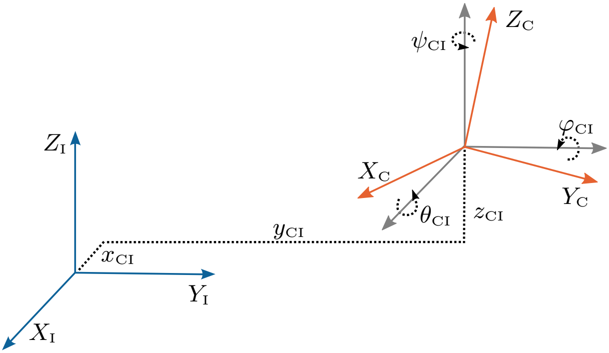


Figure 3.2: The rigid mounting of a camera and an INS induces static offsets between the underlying coordinate systems denoted by C and I. More precisely position offsets $x_{\mathrm{CI}}, y_{\mathrm{CI}}$ and z_{CI} as well as angle misalignments, here visualized in form of Euler angles yaw $= \psi_{\mathrm{CI}}$, pitch $= \theta_{\mathrm{CI}}$ and roll $= \varphi_{\mathrm{CI}}$, arise. These have to be determined in a system calibration.

The INS measures the rotation as Euler angles, namely yaw, pitch and roll, with regard to local navigation systems. The latter are ENU coordinate systems, each with the origin in the current device position described through a GPS measurement in form of latitude, longitude and altitude. This implies tiny differences between the orientations of these coordinate systems and thus also in the measured rotation for successive timestamps. The INS-camera calibration with the presented approach will be performed in an area of only a few hundred square meters, which allows neglecting these differences. A fundamental description of these coordinate systems is given in Chapter 2.1.

The world reference frame W is defined as ENU coordinate system with the origin being approximately in the middle of the calibration area. The rigid body motion $\mathbf{B}_{\mathrm{C}_i\mathrm{W}}$ describes the camera pose at the middle exposure time $t_i, i = 1, 2, ..., n$ with regard to the world frame W. Likewise specifies $\mathbf{B}_{\mathrm{I}_i\mathrm{W}}$ the corresponding configuration of the INS as rigid body motion.

In order to describe the rigid body motion $\mathbf{B}_{\mathrm{C}_i\mathrm{W}}$ of the camera using the measured rigid body motion $\mathbf{B}_{\mathrm{I}_i\mathrm{W}}$ of the INS (Figure 3.3), the devices have to be rigidly mounted. This induces that the offsets between them are static and especially comprises that the rigid body motions describing the movements from the INS to the camera at various exposure times are equal:

$$\forall k, l \in \{1, 2, ..., n\} : \quad \mathbf{B}_{\mathrm{C}_k\mathrm{I}_k} \overset{!}{=} \mathbf{B}_{\mathrm{C}_l\mathrm{I}_l}, \tag{3.1}$$

with $\{1, 2, ..., n\}$ being the middle exposure times of the images. Therefore, the notation can be simplified by omitting the indices for the rigid body motion $\mathbf{B}_{\mathrm{CI}}$ describing

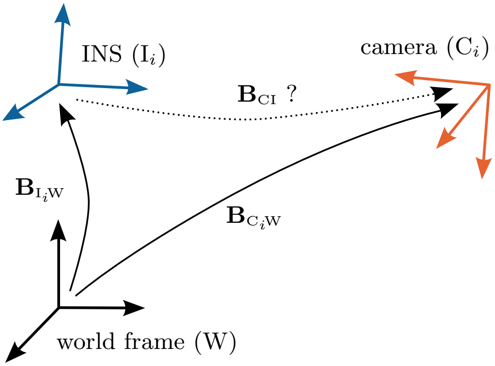

INS (I_i)
camera (C_i)
$\mathbf{B}_{\mathrm{CI}}$?
$\mathbf{B}_{\mathrm{I}_i\mathrm{W}}$
$\mathbf{B}_{\mathrm{C}_i\mathrm{W}}$
world frame (W)

Figure 3.3: Both, the poses of the INS frame I_i and of the camera frame C_i at time i can be described by rigid body motions with regard to the world frame W. The estimation of the static rigid body motion $\mathbf{B}_{\mathrm{CI}}$ between the devices (dotted arrow), enables by composition with the INS measurements $\mathbf{B}_{\mathrm{I}_i\mathrm{W}}$ the description of the camera poses $\mathbf{B}_{\mathrm{C}_i\mathrm{W}}$.

the mounting offsets. The composition with the measured INS movement leads to the camera motion:

$$\mathbf{B}_{\mathrm{C}_i\mathrm{W}} = \mathbf{B}_{\mathrm{CI}}\,\mathbf{B}_{\mathrm{I}_i\mathrm{W}}\,. \tag{3.2}$$

Thus the knowledge of the mounting offsets is required to describe the camera poses with the INS measurements.

Nearly every outdoor application utilizing a camera mounted on an aerial platform induces that the camera focus has been set to infinity. Furthermore, wide-angle lenses are frequently the right choice for a lot of tasks. The minimal distance for a focused image and the big opening angle lead to an impractical size for the calibration pattern (Figure 3.4) and make standard camera calibration procedures (Chapter 2.3) laborious. An easier estimation of these parameters can be achieved by performing an in-flight calibration, which uses a structure from motion (SFM) approach on the images acquired during a flight. Another advantage of this is that the intrinsic camera parameters $\mathbf{k}$ are estimated during real conditions, which leads to slightly different values in comparison to a laboratory calibration due to climate and environmental conditions [Jac01].

The calibration of the mounting offsets $\mathbf{B}_{\mathrm{CI}}$ in conjunction with the intrinsic camera parameters $\mathbf{k}$ out of synchronized data from an INS and a camera is the problem considered in this section.

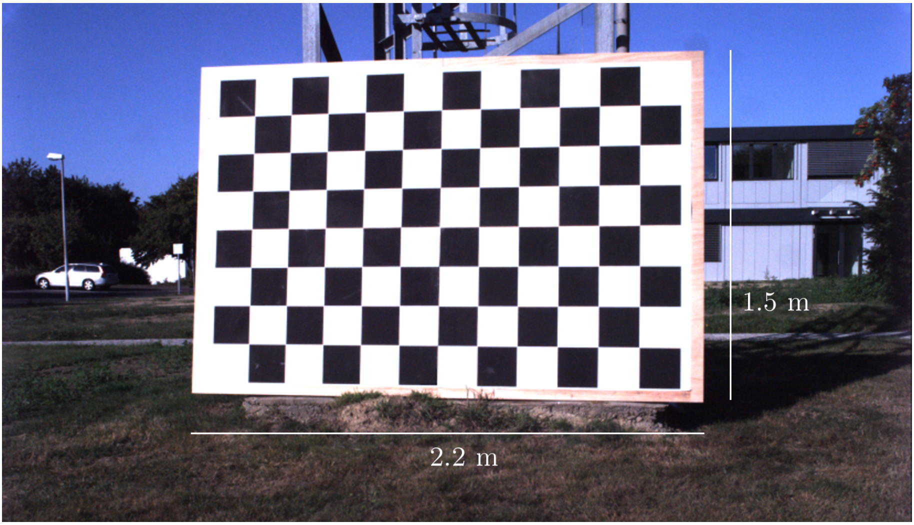


Figure 3.4: The classical checkerboard calibration for a camera equipped with wide-angle lens and focus at infinity requires bulky calibration patterns.

3.4 In-flight calibration

The calibration performed in this section requires a precise INS, which is capable of generating accurate pose information on its own. Devices based on fiber optic gyroscopes, which have a stability up to some hundredths of a degree per hour, can be one choice. In combination with RTK enhanced GPS measurements very accurate pose information are generated. An absolute accuracy of ±2 cm in the position, 0.025° for the yaw angle and 0.01° for the attitude angles pitch and roll are achieved with the system setup [iMA13].

Presently, no freely available bundle adjustment package is able to perform an INS-camera calibration and the adaptation of the bundle adjustment implementations is an error prone and time consuming task. In contrast to this, the general graph optimization (g^2o) framework [KSD$^+$11] is directly designed for the least square optimization of general error functions and can be used for this purpose. The problem has to be embedded in a factor graph by representing the parameters to be optimized as vertices and the observations between them as edges. Further requirements are the definition of error functions for the observations and good initial values for the state variables. The numerical solution of the problem can be computed with an implementation of the Levenberg-Marquardt algorithm.

3.4.1 Objective function

In the classical bundle adjustment all camera poses are optimized according to corresponding pixel observations in the images. Given a number of n images associated with the camera poses $\mathbf{B}_{\mathrm{C}_i\mathrm{W}}$ and m 3D points $\mathbf{p}_j$ with corresponding pixel observations $\mathbf{x}_{ij}$, the optimization minimizes the projection error of the 3D points according to:

$$\mathrm{E}(\mathbf{k}, \mathbf{B}_{\mathrm{C}_i\mathrm{W}}, \mathbf{p}_j) = \sum_i^n \sum_j^m (\boldsymbol{\pi}(\mathbf{k}, \mathbf{b}(\mathbf{B}_{\mathrm{C}_i\mathrm{W}}, \mathbf{p}_j)) - \mathbf{x}_{ij})^2. \tag{3.3}$$

Thereby, the projection $\boldsymbol{\pi}$ projects a 3D point according to the intrinsic camera parameters $\mathbf{k}$ and the camera pose to a pixel coordinate (Chapter 2.3). The difference between the projected and observed pixel observations are summed up to the residuum of the optimization. A disadvantage of this formulation is the independent optimization of the camera poses, which may introduce systematical errors to the set of intrinsic parameters. Therefore, the joint calibration of the intrinsic camera parameters in conjunction with the mounting offsets $\mathbf{B}_{\mathrm{CI}}$, constraining the individual camera poses, promises more accurate results. The latter can be realized by introducing a term which measures how well the parameters of the camera poses $\mathbf{B}_{\mathrm{C}_i\mathrm{W}}$ in composition with the mounting offsets $\mathbf{B}_{\mathrm{CI}}$ satisfy the INS measurements. This is realized by comparing a synthetic measurement generated out of the actual camera pose and mounting offsets with the measured INS pose by using the inverse and the composition of rigid body motions (Chapter 2.2) as follows:

$$\mathrm{E}(\mathbf{B}_{\mathrm{C}_i\mathrm{W}}\, \mathbf{B}_{\mathrm{CI}}) = \sum_i^n \log((\mathbf{B}_{\mathrm{CI}}^{-1} \circ \mathbf{B}_{\mathrm{C}_i\mathrm{W}})^{-1} \circ \mathbf{B}_{\mathrm{I}_i\mathrm{W}})^2. \tag{3.4}$$

The combined minimization of the two energies stated in equation (3.3) and equation (3.4) forms the optimization problem.

3.4.2 Graph optimization

The non linear least square problem defined in the previous section can be optimized by using g^{2}o, a graph optimization framework [KSD$^+$11]. The problem has to be embedded in a factor graph by introducing the variables to optimize as nodes and the observations between them as edges. In the following the restatement of the objective function into the graph-based formulation is presented.

The calibration parameters represented as the intrinsic camera parameters $\mathbf{k}$ and the mounting offsets $\mathbf{B}_{\mathrm{CI}}$ between the INS and the camera are added as nodes to the graph. Furthermore, each 3D point $\mathbf{p}_j$ and each camera pose $\mathbf{B}_{\mathrm{C}_i\mathrm{W}}$ are added as a node to the graph. Thereby, a pixel measurement connects three different nodes, namely: a camera, a 3D point and the intrinsic calibration parameters. This constraint can be

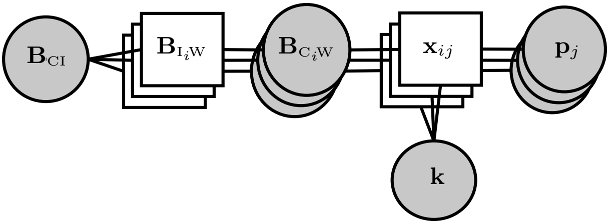

$\mathbf{B}_{\mathrm{CI}}$
$\mathbf{B}_{\mathrm{I}_i\mathrm{W}}$
$\mathbf{B}_{\mathrm{C}_i\mathrm{W}}$
$\mathbf{x}_{ij}$
$\mathbf{p}_j$
$\mathbf{k}$

Figure 3.5: The objective function of the stated problem can be illustrated by a hyper-graph. The measurements (boxes) are presented as links between the nodes concerning each multiple sets of variables (circles). For improved overview multiple state variables and measurements of the same type are visualized in a stacked view unrelated to their number of occurrence.

realized with a hyperedge, which is able to connect an arbitrary number of nodes. The edge of an INS measurement connects the corresponding rigid body motion of the camera with the mounting offsets. A visualization of the graph is given in Figure 3.5.

Further, error functions, which measure how well the measurements are described by the state variables they are connecting, are defined. The first constraint measures the error occurring by the reprojection of a 3D point into the image, in the same form as in equation (3.3). The error function for this constraint can be expressed as:

$$\mathbf{e}_{ij}(\mathbf{k}, \mathbf{B}_{\mathrm{C}_i\mathrm{W}}, \mathbf{p}_j) = \boldsymbol{\pi}(\mathbf{k}, \mathbf{b}(\mathbf{B}_{\mathrm{C}_i\mathrm{W}}, \mathbf{p}_j)) - \mathbf{x}_{ij}. \tag{3.5}$$

The resulting error vector has dimension two and is $\mathbf{0}$ if the pixel observation is perfectly described by the state variables. The second error function states how well the INS measurements can be described by the composition of the camera poses $\mathbf{B}_{\mathrm{C}_i\mathrm{W}}$ and the mounting offsets $\mathbf{B}_{\mathrm{CI}}$ as in equation (3.4):

$$\mathbf{e}_i(\mathbf{B}_{\mathrm{CI}}, \mathbf{B}_{\mathrm{C}_i\mathrm{W}}) = \log((\mathbf{B}_{\mathrm{CI}}^{-1} \circ \mathbf{B}_{\mathrm{C}_i\mathrm{W}})^{-1} \circ \mathbf{B}_{\mathrm{I}_i\mathrm{W}}) \ . \tag{3.6}$$

Using the twist representation for the rigid body motions, results in a 6-dimensional error vector, which is $\mathbf{0}$ if the parameters perfectly satisfy the measurement.

Without limiting the generality, the whole state vector $[\mathbf{k}, \mathbf{B}_{\mathrm{C}_i\mathrm{W}}, \mathbf{p}_j, \mathbf{B}_{\mathrm{CI}}]$ is referred to as $\mathbf{y}$ and the objective function is reformulated as follows:

$$\mathrm{E}(\mathbf{y}) = \sum_i^n \sum_j^m \mathbf{e}_{ij}(\mathbf{y})^\top \boldsymbol{\Sigma}_{ij}^{-1}\ \mathbf{e}_{ij}(\mathbf{y}) + \sum_i^n \mathbf{e}_i(\mathbf{y})^\top \boldsymbol{\Sigma}_i^{-1}\ \mathbf{e}_i(\mathbf{y}) \ , \tag{3.7}$$

where the information matrices are represented by the inverse covariance matrices $\boldsymbol{\Sigma}_{ij}^{-1}$ and $\boldsymbol{\Sigma}_i^{-1}$ of the pixel observations and INS measurements. Within g^{2}o, the Levenberg-Marquardt (LM) algorithm is used to determine the minimum of equation

(3.7) and therefore a good initial guess $\check{\mathbf{y}}$ of the state vector is needed. Iteratively, the first order Taylor expansion around the current guess $\check{\mathbf{y}}$ is used to approximate equation (3.7) and optimize the local increments $\Delta\mathbf{y}$ by solving the resulting sparse linear system. The center for the next iteration is obtained by adding the optimized increments to the current guess. This is done by using the motion composition for the state variables represented by rigid body motions and a simple addition for the 3D points and intrinsic camera parameters. A detailed description of the LM algorithm is stated by Lourakis *et al.* [LA09].

3.4.3 Workflow

The initial camera poses are generated by combining the measured INS poses with faulty mounting offsets determined out of construction drawings. First, feature points are extracted from the images and a matching of the feature descriptions between all image pairs is performed. A prominent approach is the use of scale invariant feature transformation (SIFT) to detect and describe feature points, which are invariant to changes in scale, lightning condition and in some extent even rotation [Low04]. The well known random sample consensus (RANSAC) [FB81] is used to estimate the relative transformation between the images from the matched feature points and eliminate false matches which are not supported by the determined geometry. Afterwards, multiple feature matches observing the same 3D points are combined and a forward intersection, resulting in coordinates of the observed 3D points, is realized. This process is known as Structure from Motion (SfM) and delivers the initial sparse 3D structure of the observed area [HZ04]. Nevertheless, all estimated parameters contain errors, which are minimized by the optimization of a non-linear function, which is phrased as a graph. The entire workflow is depicted in Figure 3.6.

3.4.4 Numerical studies

In this section, results from numerical studies are presented to validate the proposed approach. The main purpose of this is to analyze the influence of various flight configuration and settings on the calibration parameters by comparing the estimations with the known ground truth.

In the simulations the mounting parameters $\mathbf{B}_{\text{CI}}$ consisting of the lever-arm components x_{CI}, y_{CI} and z_{CI} as well as the angle misalignments in form of the Euler angles yaw $= \psi_{\text{CI}}$, pitch $= \theta_{\text{CI}}$ and roll $= \varphi_{\text{CI}}$ (Figure 3.2) are set according to the ground truth defined in Table 3.1. The initialization with $\theta_{\text{CI}} = 180^\circ$ is a result of the definition of the INS frame as ENU coordinate system, while the downward directed camera is modeled with the z-coordinate pointing in its viewing direction. The intrinsic camera parameters in Table 3.1 describe the modeled camera with an image size of 3296 by 2472 pixels using a wide-angle lens, analog to the camera used in the real-world experiments in the next section. The simulated flight courses consist of INS poses sampled along straight lines according to the camera exposure times, determined by the velocity of the UAS and the frames per second of the camera. For a more realistic flight path, the ideal poses were modified by a small random factor and the resulting poses are considered as ground truth. One GCP was defined in the middle of the observed area. A fixed number of additional 3D points were sampled in the environment. The corresponding pixel observations were determined by projecting the 3D points into the images according to equation (2.15) using both, the generated camera poses and the ground truth of the intrinsic camera parameters defined in Table 3.1. A more realistic number of observations was produced by defining a probability of detection of 50 % instead of using all projections of the 3D points which are in the field of view of a camera. Concerning noise, the image observations are simulated with an accuracy of 0.5 pixels. Further, white Gaussian noise is added to the INS poses using a standard deviation of 0.02 m for the positional components and 0.01° for the rotational part represented as Euler angles.

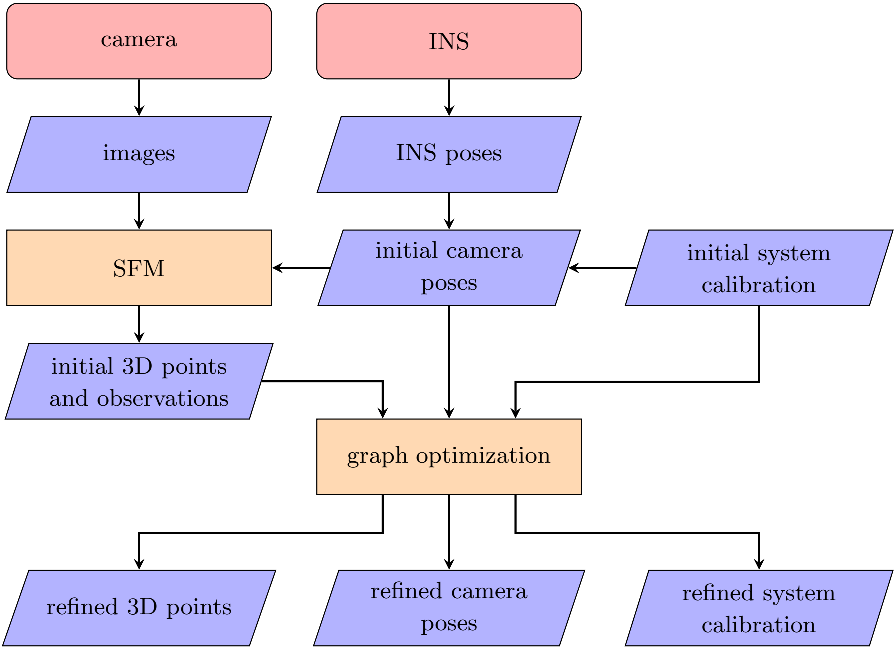


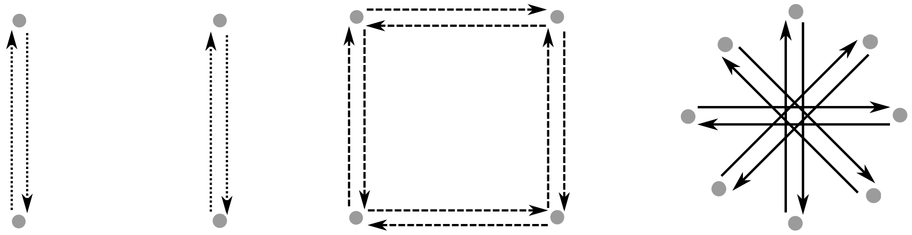
Figure 3.6: Overview of the algorithm. Data from the INS and the camera are processed to determine the system calibration. The latter enables very accurate image georeferencing by concatenation with the measured INS poses for subsequent flights.

Table 3.1: This table states the initialization and ground truth values used in the experiments. For an easier interpretation the difference between the values is also stated.

	x_{CI} [m]	y_{CI} [m]	z_{CI} [m]	ψ_{CI} [°]	θ_{CI} [°]	φ_{CI} [°]
initialization	0.130	0.100	0.100	0.000	180.000	0.000
ground truth	0.132	0.096	0.104	2.344	183.291	−1.937
difference	0.002	−0.004	0.004	2.344	3.291	−1.937
	f_x [pel]	f_y [pel]	o_x [pel]	o_y [pel]	k_1 [pel^2]	k_2 [pel^4]
initialization	1650.00	1650.00	1648.00	1236.00	0.00040	0.00800
ground truth	1663.31	1662.84	1651.52	1234.67	0.00076	0.00908
difference	13.31	12.84	3.52	−1.33	0.00036	0.00108

The principle of the flight configuration stated as optimal in [KHB11] was used to define the first simulated flight course. At an altitude of 20 m two lines with a distance of 20 m are defined (Figure 3.7a). The INS poses are sampled along these lines twice, once in each direction. This results in an image side overlap of 50 % for poses on adjacent lines. The same maneuver was simulated at an altitude of 30 m resulting in a side overlap of about 66 %. Two different altitudes in combination with a GCP were used to decouple the high correlation between the lever-arm and the focal length as well as the principal point. The definition of 20 m strip length in conjunction with a speed of 36 km/h and 5 images per second leads to a total number of 80 images with more than 93 % image overlap in flight direction. This course was used in Simulation 1, 2 and 3 by sampling an increasing number of 3D points in the observed area. Simulation 1 was performed with 1000 points, Simulation 2 with 3000 and for Simulation 3 a total number of 6000 points was sampled in the target area.

As expected, the results in Table 3.2 show that increasing the number of observed 3D points, which results in more constraints, leads to smaller errors for the calibration parameters. Remarkable is that the angle misalignment was nearly the same for Simulation 2 and 3, which shows that increasing this value further will likely have little influence on these parameters. Performing flight course (a) at two different altitudes should decouple the lever-arm components from the correlated intrinsic camera parameters as stated in [KHB11]. But the achieved optimization of the lever-arm components are worse than their initial values representing standard measurements out of constructions drawing with an accuracy in the range of a few millimeters. The error in the lever arm component z_{CI} was for all three simulations about four times bigger than in the other directions. This was also reflected in the higher estimation errors of the focal length parameters f_x and f_y correlated to z_{CI} compared to the



(a) Simulation 1,2,3,4,7 **(b)** Simulation 5 **(c)** Simulation 6

Figure 3.7: The three flight patterns investigated in this study are visualized in a top view. Each line is sampled in both directions, resulting in an image overlap of approximately 100%. The lines of course (a) were sampled at two different altitudes and the flights for (b) and (c) were performed at only one altitude, resulting in a total number of eight flight lines for all courses.

Table 3.2: This table shows the RMSE of the calibration parameters achieved by performing 100 MC runs for flight course (a) with an increasing number of 3D points.

	x_{CI} [m]	y_{CI} [m]	z_{CI} [m]	ψ_{CI} [°]	θ_{CI} [°]	φ_{CI} [°]
Sim. 1	0.00824	0.00728	0.03169	0.01237	0.00030	0.01925
Sim. 2	0.00849	0.00582	0.02167	0.01072	0.00090	0.01719
Sim. 3	0.00800	0.00588	0.01443	0.01052	0.00039	0.01762
	f_x [pel]	f_y [pel]	o_x [pel]	o_y [pel]	k_1 [pel^2]	k_2 [pel^4]
Sim. 1	2.39323	2.35976	0.18362	0.24893	$4.1539e^{-5}$	$5.1066e^{-5}$
Sim. 2	1.57664	1.57361	0.12286	0.16740	$2.3002e^{-5}$	$3.4519e^{-5}$
Sim. 3	1.01410	1.01957	0.07928	0.11551	$1.7486e^{-5}$	$2.2469e^{-5}$

principal point coordinates o_y and o_y correlated to x_{CI} and y_{CI}(Table 3.2). The performed flight course was unsuccessful in decoupling the lever-arm from the instrinsic camera parameters.

Errors in the lever-arm components transfer in their magnitude to the camera poses and the observed 3D points. As stated before, the lever-arm can be measured out of construction drawings with a high accuracy. Due to these observations, the lever-arm components was fixed to the initialization in the following experiments.

In Simulation 4, the same flight and parameter configuration as in Simulation 2 was used. The optimization leads to better results in all estimated parameters (Table 3.3). The fixed lever-arm also removes the requirement of two flight altitudes, which were

Table 3.3: This table shows the RMSE of the calibration parameters achieved by performing 100 MC runs for the optimization with the lever-arm values fixed to their faulty initialization for various flight maneuvers.

	x_{CI} [m]	y_{CI} [m]	z_{CI} [m]	ψ_{CI} [°]	θ_{CI} [°]	φ_{CI} [°]
Sim. 4	0.002	0.004	0.004	0.01110	0.00037	0.00986
Sim. 5	0.002	0.004	0.004	0.01065	0.00123	0.01128
Sim. 6	0.002	0.004	0.004	0.02525	0.00182	0.01028
Sim. 7	0.002	0.004	0.004	0.00357	0.00009	0.00475
	f_x [pel]	f_y [pel]	o_x [pel]	o_y [pel]	k_1 [pel^2]	k_2 [pel^4]
Sim. 4	1.10761	1.11969	0.08950	0.13140	$2.2965e^{-5}$	$2.5092e^{-5}$
Sim. 5	0.78936	0.78776	0.05326	0.07603	$2.8315e^{-5}$	$1.8952e^{-5}$
Sim. 6	1.02642	1.02301	0.05658	0.10738	$3.7536e^{-5}$	$2.3544e^{-5}$
Sim. 7	0.15332	0.23633	0.10430	0.13719	$2.5814e^{-5}$	$1.0906e^{-5}$

established to decouple the lever arm from the intrinsic camera parameters. An investigation of the influence of the flight course was realized by performing simulations of the patterns visualized in (Figure 3.7), whereby for each simulation a total number of 3000 points were sampled in the observed area. By performing the quadratic course in Simulation 5 and the star pattern in Simulation 6 at an altitude of 20 m, the same number of cameras as in all other simulations were created.

Compared with Simulation 4, the errors in Simulation 5 were slightly smaller for the intrinsic camera parameters, but higher for the roll angle. The star pattern performed in Simulation 6 leads to a bigger error in the yaw angle (Table 3.3). Overall the influence of the investigated flight courses on the accuracies of the estimated calibration parameters is small.

Another observation results from performing flight course (a) at altitudes of 200 m and 300 m in Simulation 7, which was ten times higher than in Simulation 4. As expected, the errors in the angle misalignment were smaller due to there bigger influence on the object points at higher altitudes. Only the estimation of the principal point is worse compared to the other simulations. Nevertheless, the difference is in the small subpixel range and most likely occurs on the basis of numerical calculations. It has nearly no effect on the mapping accuracies (Table 3.3).

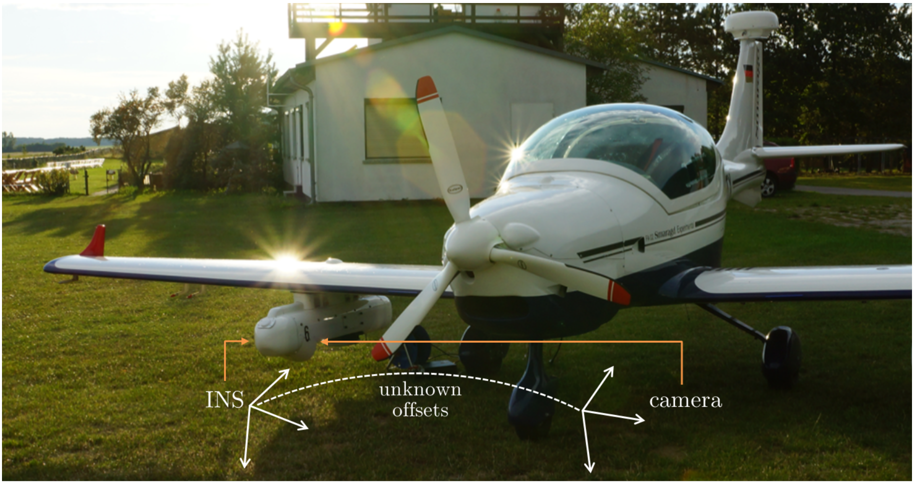


Figure 3.8: The calibration of the static coordinate system offsets between an INS and a camera enables the direct georeferencing of images collected with aerial platforms. For the flight experiments the sensors were integrated into a payload pod, which was mounted beneath one wing of a manned ultra-light airplane.

3.4.5 Real-world experiments

In this section, achieved results from flight experiments are presented. The equipped INS is based on fiber optic gyroscopes, which have a stability up to some hundredths of a degree per hour. In combination with real-time kinematic enhanced GPS measurements very accurate pose information are generated. The optical system consist of a downward looking camera with 3296 × 2472 pixels and a wide-angle lens with a field of view with 54° in the horizontal and 42° in the vertical dimension. The sensors were integrated into a payload pod, which was mounted beneath the wing of a manned ultra light airplane used as UAS demonstrator (Figure 3.8). This platform allows collecting data at a high altitude without the constraints from flight permissions for small UAS. The latter are, at least in Germany, strictly regulated and moreover differ locally. The performed flight course has to introduce measurements that constrain all dimensions of the calibration parameters. For the utilized platform small movements in all axes occur even for straight and level flights, which aim at a constant heading and altitude by accomplishing immediate corrections to unintentional movements.

A total of four flights was performed within two days. At an altitude of 300 m and above, two images per second were captured at a speed of approximately 125 km/h. Based on the results of the numerical studies, the flights were performed as crossing straight lines. To achieve a high image overlap, only images within a circle around

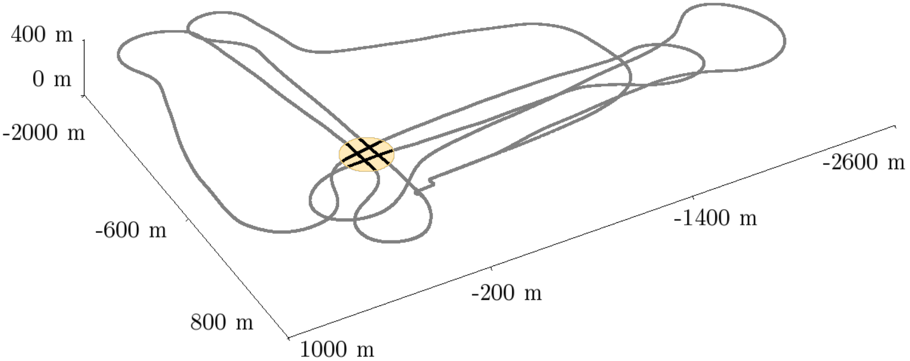


Figure 3.9: Visualization of a flight course performed for the evaluation of the approach. The altitude was 300 m. Only data from the central area, highlighted in the visualization, was used for the calibration.

the central point were used (Figure 3.9). This results in a total number of nearly 700 images for the first two flights and 300 images for the other two.

According to the workflow depicted in Figure 3.6, these images were used to calculate an initial 3D point cloud of the observed area with a SFM approach under consideration of the camera poses [Wu13]. The latter were generated by a concatenation of the INS measurements with the initial mounting offsets determined through terrestrial measurements. The output of the SFM (Figure 3.10) was used as input for the graph optimization. Thereby the pixel observations were introduced as measurements with an accuracy of 1 pixel. Since no quality log files of the INS were available, the manufacturer information of an accuracy of 2 cm in the position, 0.04° for the yaw and 0.01° for the pitch and roll angles [iMA13], were considered. The numerical studies (Chapter 3.4.4) reveal that the optimization of the translational part of the mounting offsets leads to a lower accuracy compared to measurements out of construction drawings. Due to this observation, the lever-arm was fixed in the optimization to the offsets from the construction drawings. Kersting *et al.* stated that the only parameter which can not be estimated without one GCP is the vertical lever-arm, because of its correlation to the focal length of the intrinsic camera parameters [KHB11]. By not optimizing the lever-arm, the proposed system calibration can be performed without GCP. This saves time and money compared to other approaches and furthermore allows an easy recalibration before a critical mission.

The resulting parameters differ from the initial checkerboard calibration and show a high stability for the in-flight calibrations performed with data collected in different flights (Table 3.4). The differences between the optimized boresight angles ψ_{CI}, θ_{CI}

Table 3.4: The optimization results of the boresight angles and the intrinsic camera parameters remain relatively constant between the different flights and differ from the initial checkerboard calibration

	ψ_{CI} [°]	θ_{CI} [°]	φ_{CI} [°]	f_x [pel]	f_y [pel]	o_x [pel]	o_y [pel]
init.	0.000	0.000	0.000	3334.68	3343.50	1744.32	1238.06
Flight 1	0.846	0.215	-0.072	3342.89	3334.88	1730.60	1227.90
Flight 2	0.816	0.205	-0.068	3343.40	3335.44	1724.04	1231.19
Flight 3	0.795	0.205	-0.074	3343.73	3335.95	1725.39	1230.74
Flight 4	0.805	0.193	-0.078	3346.28	3338.36	1730.62	1234.04

	k_1 [pel^2]	k_2 [pel^4]	k_3 [pel^6]	p_1 [pel]	p_2 [pel]
init.	-0.0966519	0.1411390	-0.1055540	0.0000566	0.0019290
Flight 1	-0.0858842	0.0808048	-0.0183501	-0.0001805	0.0002204
Flight 2	-0.0857764	0.0798504	-0.0165184	-0.0001643	0.0002394
Flight 3	-0.0862101	0.0856143	-0.0271507	-0.0001811	0.0002657
Flight 4	-0.0890717	0.0972273	-0.0433398	-0.0002691	0.0002757

and φ_{CI} are very close to the stated accuracy of the INS. Small variations for the intrinsic camera parameters occur most likely due to different climate conditions during flight execution.

To evaluate the accuracy of the achieved results a least-square forward intersection for the pixel observations of five GCP was realized. The image coordinates of these points were measured manually (Figure 3.11) and used to perform the forward intersection once with the initial and once with the optimized camera poses. This leads to 3D coordinates, which were compared to values measured with a mobile GPS-receiver. The latter stated a horizontal accuracy of about 30 cm and a vertical accuracy of about 50 cm. The results from the forward intersection are in the same range, which shows the performance of the approach (Table 3.5). The large initial error of Flight 4 occurred most likely due to the range from 300 to 800 meters for the altitude, which was smaller for the other flights. Nevertheless, the results using the optimized calibration parameters for Flight 4 were in the same range as for the other flights. The generation of the camera poses out of calibration results from Flight 1 for all flights led to a slightly decreasing performance (Table 3.5). Given an altitude of 300 meters and above, the accuracies are high and clearly outperform the terrestrial calibration.

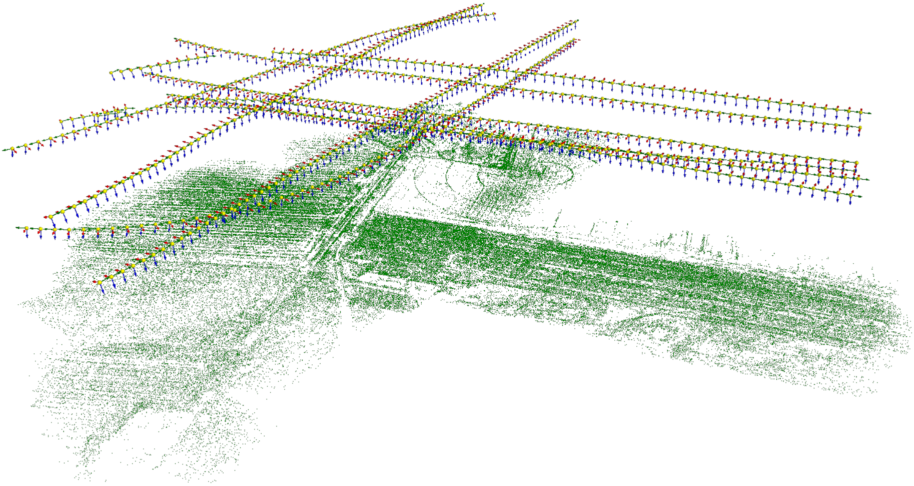


Figure 3.10: Result of the structure from motion preprocessing. Visualization of the camera poses by their coordinate system axes and the observed 3D points (green) introduced as vertices in the graph optimization.

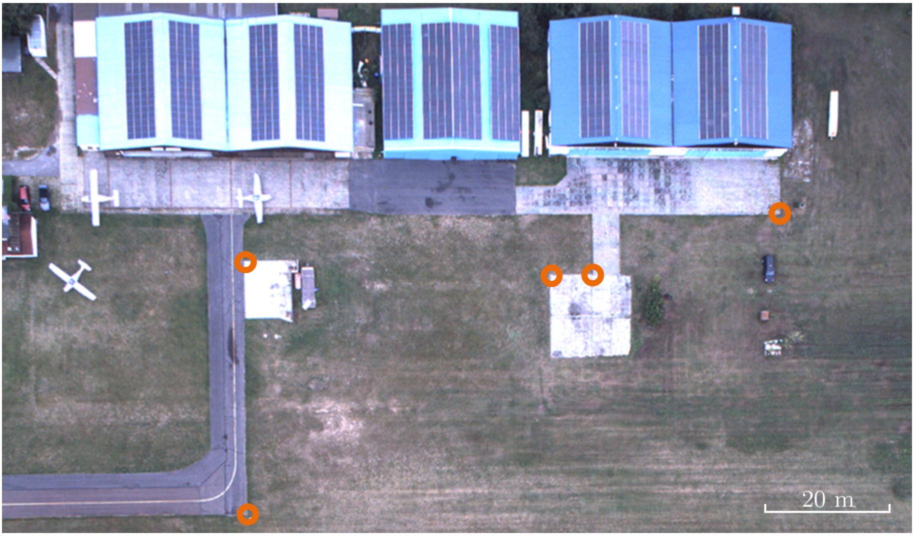


Figure 3.11: Five ground control points were used to verify the accuracy of the approach. These were placed at dominant image corners to allow easy manual measurement of their pixel coordinates (orange circles).

Table 3.5: These tables state the mean Euclidean distance between five ground control points and the forward intersection of their pixel observations. The latter was performed with camera poses generated out of INS measurements and the system calibration. The difference using the initial and the optimized calibration (left) as well as the calibration from Flight 1 for all flights (right) are shown

	init.	opt.	gain
	[m]	[m]	[factor]
Flight 1	3.17	0.53	5.98
Flight 2	2.94	0.47	6.26
Flight 3	2.02	0.37	5.46
Flight 4	4.45	0.44	10.11

	opt.	opt. Flight 1	gain
	[m]	[m]	[factor]
Flight 1	0.53	0.53	1.00
Flight 2	0.47	0.55	0.85
Flight 3	0.37	0.66	0.56
Flight 4	0.44	0.58	0.76

3.5 Position free calibration

The system calibration for a camera and an accurate INS based on fiber optic gyroscopes can be performed with an adapted bundle adjustment as described in the previous section. But inertial sensors based on MEMS technology have become a serious alternative. These devices are small and lightweight at a favorable price. In the last years, the overall performance of these devices was improved by modeling the stochastic error more precisely [NKEES09]. Nevertheless, the error characteristics are worse compared to classical INS and the system calibration should consider that.

In this section, an innovative calibration procedure to determine the angle misalignments, also known as boresight, between the coordinate systems of an INS and a camera is presented. All currently known approaches integrate positional information from the INS in the optimization process. Thereby, the position errors in the range of a few meters of INS devices without RTK negatively influence the accuracy of the boresight estimation in state-of-the-art calibration methods. By using lines, more precisely directions, instead of classical point features within the calibration process, the optimization is performed without positional information and avoids being affected by the corresponding noisy data. For the first time, a reliable calibration for systems with poor positional estimations is possible. The presented approach can be applied to images observing a checkerboard, which allows the calibration of the intrinsic camera parameters and boresight misalignment angles from the same data set. The high performance of the presented procedure is confirmed by evaluating simulated and real-world experiments. Achieved results show the capability to reduce the boresight errors to small sub-degree values.

With the presented approach, the system calibration between an INS and a camera will become as easy as the calibration of the intrinsic camera parameters with the

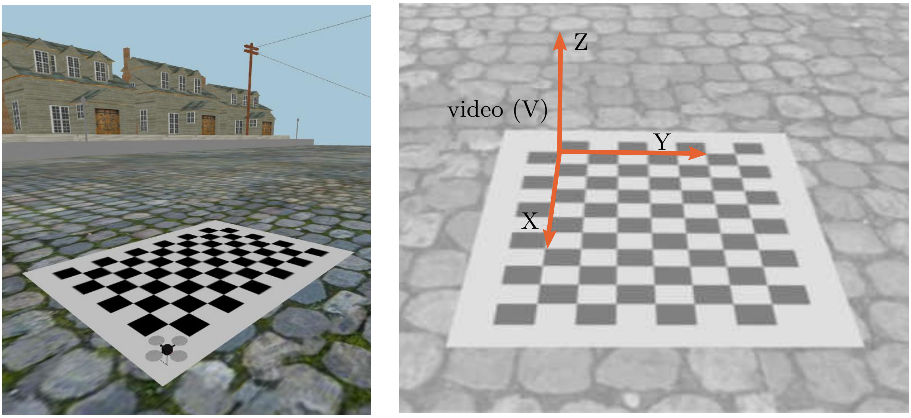


Figure 3.12: Left: The presented approach performs the system calibration with images observing a checkerboard. Therefore, the same data set can be used to calibrate the intrinsic camera parameters and the offsets between the camera and the INS. The approach is evaluated in simulations and real-world experiments, utilizing a small multicopter equipped with both sensors. The picture shows a screenshot of the simulation framework. **Right:** The video coordinate system is a world frame positioned with its origin at the upper left 3D point of the calibration pattern and all corners within the positive xy-plane.

method from Zhang [Zha00]. It will be possible to perform the calibration of all required system parameters from the same data set consisting of images observing a planar calibration pattern and synchronized INS measurements (Figure 3.12).

3.5.1 System setup

The presented method does not impose any special requirements to the INS. Especially the high positional accuracies provided by the use of RTK corrections are not necessary. The INS has to be rigidly mounted to the camera. For the latter there are no special requirements. There is no need for a metric camera. Even rolling shutter or zoom cameras may be used as long as there are methods to calibrate them geometrically. The only known exception for this are some consumer cameras with a moving sensor area.

In contrast to the in-flight calibration described in the previous section, the system calibration presented in the following is performed with images observing a planar checkerboard pattern from various viewpoints.

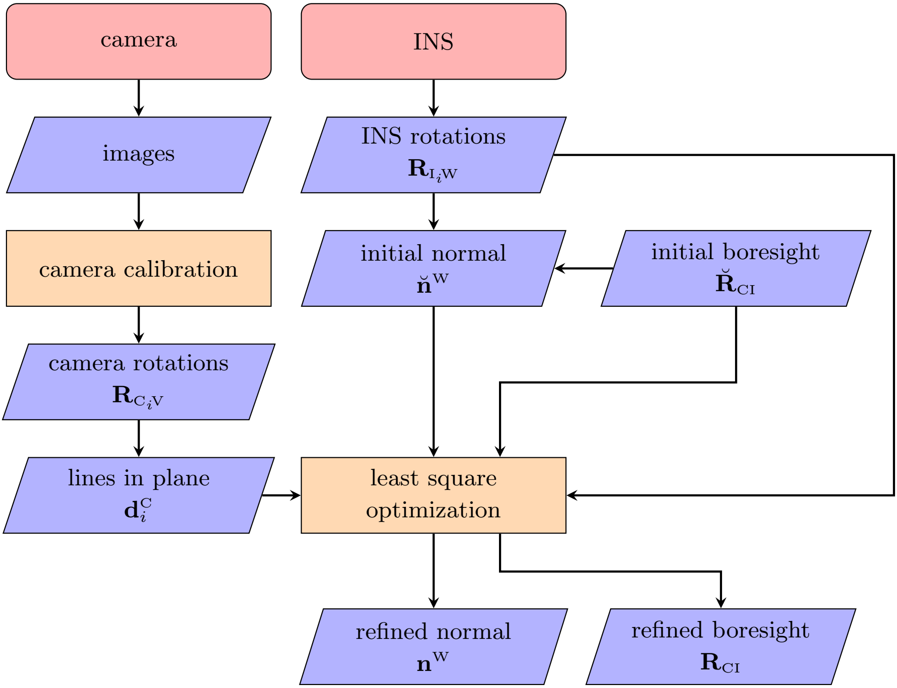

camera
INS
images
INS rotations
$\mathbf{R}_{\mathrm{I}_i\mathrm{W}}$
camera calibration
initial normal
$\breve{\mathbf{n}}^{\mathrm{W}}$
initial boresight
$\breve{\mathbf{R}}_{\mathrm{CI}}$
camera rotations
$\mathbf{R}_{\mathrm{C}_i\mathrm{V}}$
lines in plane
$\mathbf{d}_i^{\mathrm{C}}$
least square optimization
refined normal
$\mathbf{n}^{\mathrm{W}}$
refined boresight
$\mathbf{R}_{\mathrm{CI}}$

Figure 3.13: Overview of the algorithm. Data from the INS and the camera are processed to determine the angle misalignment between the devices, also known as boresight.

3.5.2 Workflow

The proposed boresight calibration requires time synchronized images and INS poses, whereby from the latter only the rotational part is processed. As a first step, a standard camera calibration is performed by processing images observing a planar checkerboard [Zha00]. Besides the camera matrix and distortion parameters, characterizing the mapping of the camera (Chapter 2.3), the extrinsic camera parameters, describing the rigid body motion of the camera in the video coordinate system V (Figure 3.12), are obtained. The rotational part of these poses are used to transform directions within the checkerboard plane from the video coordinate system in the individual cameras. Further, an estimate of the checkerboard plane normal is generated by the concatenation of the measured INS poses and the initial boresight rotation, derived from construction drawings. The initializations of the normal and the boresight will be refined in the least square optimization of the objective function, which is introduced in the following. The entire workflow is depicted in Figure 3.13.

3.5.3 Objective function

The optimization uses the fact that all directions $\mathbf{d}_i^{\mathrm{V}} \in \mathbb{R}^3$ within the checkerboard plane are orthogonal to the normal of the checkerboard $\mathbf{n}^{\mathrm{V}} \in \mathbb{R}^3$ and consequently the scalar products between them are zero:

$$\mathbf{d}_i^{\mathrm{V}} \cdot \mathbf{n}^{\mathrm{V}} = 0 \; . \tag{3.8}$$

Without loss of generality, the corresponding camera time index is used to label a direction, even if there are multiple directions visible in one image which are indistinguishable with this notation. Based on the definition of the video coordinate system (Figure 3.12), the normal $\mathbf{n}^{\mathrm{V}}$ will always be $(0,0,1)^\top$. Moreover, the directions from the video coordinate system V are rotated into the coordinate systems of the cameras C_i, and vice versa with the rotational part of the extrinsic camera parameters:

$$\begin{aligned} \mathbf{d}_i^{\mathrm{C}} &= \mathbf{R}_{\mathrm{C}_i\mathrm{V}} \mathbf{d}_i^{\mathrm{V}} \; , & (3.9)\\ \mathbf{d}_i^{\mathrm{V}} &= \mathbf{R}_{\mathrm{C}_i\mathrm{V}}^{-1} \mathbf{d}_i^{\mathrm{C}} \; . & (3.10) \end{aligned}$$

This allows on the one hand to describe the directions in the camera coordinate systems by applying equation (3.9) and on the other hand to reformulate equation (3.8) as:

$$(\mathbf{R}_{\mathrm{C}_i\mathrm{V}}^{-1} \mathbf{d}_i^{\mathrm{C}}) \cdot \mathbf{n}^{\mathrm{V}} = 0 \; . \tag{3.11}$$

In general the static transformation between the video coordinate system V and the global world frame W, used as reference for the INS measurements, is unknown and difficult to determine precisely. Luckily, this relation is not required for the optimization. The presented approach is based on the fact, that all directions in the checkerboard plane are planar and thus perpendicular to their normal. This relationship has to be fulfilled for all global coordinate systems. Therefore, the video coordinate system V can be exchanged with the global INS reference frame W in equation (3.11) to maintain:

$$(\mathbf{R}_{\mathrm{C}_i\mathrm{W}}^{-1} \mathbf{d}_i^{\mathrm{C}}) \cdot \mathbf{n}^{\mathrm{W}} = 0 \; . \tag{3.12}$$

Based on the pose composition for the rigid body motions equation (2.11), the camera rotation can be described as a combination of the corresponding INS measurement $\mathbf{R}_{\mathrm{I}_i\mathrm{W}}$ and the static boresight misalignment $\mathbf{R}_{\mathrm{CI}}$ by:

$$((\mathbf{R}_{\mathrm{CI}} \mathbf{R}_{\mathrm{I}_i\mathrm{W}})^{-1} \mathbf{d}_i^{\mathrm{C}}) \cdot \mathbf{n}^{\mathrm{W}} \overset{!}{=} 0 \; . \tag{3.13}$$

The INS measurements introduce small errors. Nevertheless, the resulting values in equation (3.13) should be zero for all directions and thus will be the residuals in the optimization process. In contrast to $\mathbf{n}^{\mathrm{V}}$, the normal vector $\mathbf{n}^{\mathrm{W}}$ is unknown and has to be determined within the optimization process. An initial approximation $\check{\mathbf{n}}^{\mathrm{W}}$ can

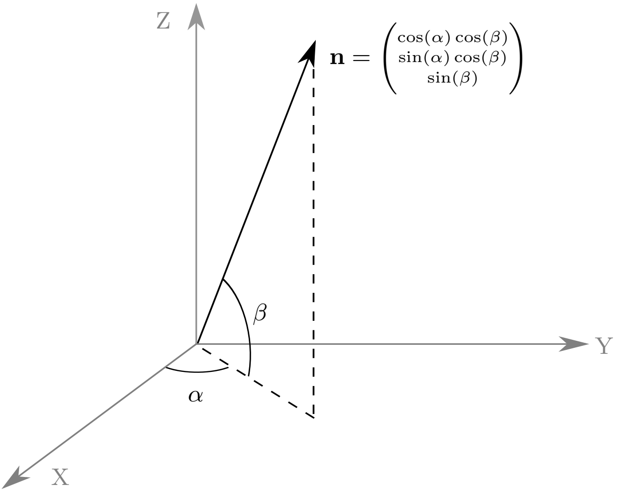

Z
$\mathbf{n} = \begin{pmatrix} \cos(\alpha)\cos(\beta) \\ \sin(\alpha)\cos(\beta) \\ \sin(\beta) \end{pmatrix}$
β
Y
α
X

Figure 3.14: Used configuration to parameterize a normal **n** with only two angles instead of three coordinates. This minimal representation allows formulating an unconstrained objective function for the optimization.

be determined by calculating the cross product of two randomly chosen directions in the checkerboard plane by:

$$\breve{\mathbf{n}}^{\mathrm{W}} = \breve{\mathbf{d}}_i^{\mathrm{W}} \times \breve{\mathbf{d}}_j^{\mathrm{W}} = ((\breve{\mathbf{R}}_{\mathrm{CI}}\mathbf{R}_{\mathrm{I}_i\mathrm{W}})^{-1}\mathbf{d}_i^{\mathrm{C}}) \times ((\breve{\mathbf{R}}_{\mathrm{CI}}\mathbf{R}_{\mathrm{I}_j\mathrm{W}})^{-1}\mathbf{d}_j^{\mathrm{C}}) \; . \tag{3.14}$$

Thereby, the local directions are rotated in the world coordinate system under application of the non-optimized boresight $\breve{\mathbf{R}}_{\mathrm{CI}}$.

The normal vector is described as depicted in Figure 3.14 with two angles (α, β). By parameterizing the boresight rotation $\mathbf{R}_{\mathrm{CI}}$ with Euler angles (ψ, θ, φ) according to equation (2.3) the state vector is composed as:

$$\mathbf{y} = [\psi, \theta, \varphi, \alpha, \beta] \; . \tag{3.15}$$

Finally, the objective function summing up the squared residuals from multiple directions can be formulated as:

$$\mathrm{E}(\mathbf{y}) = \sum_{i}^{n} \Big(((\mathbf{R}_{\mathrm{CI}}\mathbf{R}_{\mathrm{I}_i\mathrm{W}})^{-1}\mathbf{d}_i^{\mathrm{C}}) \cdot \mathbf{n}^{\mathrm{W}} \Big)^2 \; , \tag{3.16}$$

with n representing the total number of direction observations. Thereby multiple observations corresponding to the same image share the same INS orientation, but are otherwise treated independently in the optimization. The minimum of equation (3.16) is computed with a non-linear least square algorithm and therefore a good initial guess $\check{\mathbf{y}}$ of the state vector is needed. The starting values for the boresight angles can be extracted from construction drawings with sufficient accuracies in the range of a few degrees. They are also required to initialized the normal according to equation (3.14).

Iteratively, the first order Taylor expansion around the current guess $\check{\mathbf{y}}$ is used to approximate equation (3.16) and optimize the local increments $\Delta\mathbf{y}$ by solving the resulting sparse linear system. The center for the next iteration is obtained by adding the optimized increments to the current guess.

3.5.4 Simulations

The usage of the robot simulation Gazebo [KH04] allows the validation of the proposed approach by comparing the optimization results to the known ground truth in a realistic scenario and gather valuable information for real-world experiments. A quadcopter UAS was utilized as sensor platform within Gazebo [MSK$^+$12] and steered to observe a checkerboard while collecting synchronized camera images and INS poses (Figure 3.12). The camera simulation provided by Gazebo was used to generate images from a downward facing camera with a resolution of 640×480 pixels at a frame rate of 5 Hz. The chosen opening angle of 100° corresponds to a wide-angle lens.

While performing the camera calibration, an image was only added to the data set if it was sufficiently different from any previous sample. This was realized by comparing the mean of the corner coordinates as well as size and skew of the checkerboard between the current and all already added images. By defining different similarity thresholds, subsets with a varying number of samples were generated from the same video sequence. In the first experiment three different similarity thresholds were chosen which resulted in an ascending number of samples out of the same data set of synchronized INS poses and images. The preprocessing step of calibrating the intrinsic and extrinsic camera parameters led for all three data sets to extraordinary small root-mean-square (RMS) reprojection errors for the 3D checkerboard points of 0.066, 0.064 and 0.062 pixels for an increasing number of images. These values state the achieved subpixel accuracy and also show that using more images only results in very small improvements regarding the camera calibration. However, a high number of images negatively influences the time needed for the preprocessing step of calibrating the camera parameters. The processing time increased very largely from below one minute for the first data set containing 45 images (Simulation 1) to above half an hour for the largest set, processing a total number of 263 images (Simulation 3). This is a consequence of the $O(n^3)$ complexity of the matrix inversion, which is used within the LM algorithm to minimize the equation system build from checkerboard observations. Nevertheless, using more images and their corresponding INS poses should improve

Table 3.6: Setup for the simulations.

	symbol	unit	initialization	ground truth	difference
yaw	ψ	[°]	-88.0	-90.0	-2.0
pitch	θ	[°]	3.0	0.0	-3.0
roll	φ	[°]	178.0	180.0	2.0

the robustness of the boresight calibration, by reducing the influence of INS errors and outliers.

In the first three experiments with an increasing number of images and corresponding INS poses, the INS rotations were modeled by adding white Gaussian noise with a standard deviation of $\sigma_\psi = 0.2°$, $\sigma_\theta = 0.1°$ and $\sigma_\varphi = 0.1°$ to the error free rotation information provided by the simulation. Thereby the error in the yaw angle, describing the rotation around the z-axis, was modeled twice as big as the other angles to match the standard error behavior of an INS. The chosen values represent the accuracies of the device used in the real-world experiments, given hereafter, made of MEMS sensors without any DGPS correction signals or post processing [SBG15]. The boresight angles were initialized with errors in the range of some degrees (Table 3.6), which represents accuracies easily achievable even without access to construction drawings of the sensor setup.

A total number of 1000 Monte Carlo (MC) runs were performed. As expected, the results show that the accuracies of the optimized boresight angles improve with a higher number of input samples (Table 3.7). Even by using only a total of 45 images and corresponding INS rotations in Simulation 1, accuracies which are smaller for all three angles than the simulated standard deviation of the INS were achieved. On closer inspection, however, the root-mean-square error (RMSE) of Simulation 1 shows inconsistencies. The RMSE of the yaw angle should be higher than the results for pitch and roll on the basis of the higher standard deviation. An increasing number of images in Simulation 2 and Simulation 3 shows exactly this relation, but the small number of images in Simulation 1 seems to have a negative influence on the achieved results.

Processing the same images in MC runs with a 20 times higher standard deviation for all INS angles led to the expected RMSE values, which were also around 20 times larger (Table 3.8). This states, that also for INS devices with a poor performance the approach is applicable and reduces the boresight errors from some degrees to sub-degree values.

Table 3.7: Results from 1000 MC runs with small INS errors of $\sigma_\psi = 0.2°$, $\sigma_\theta = 0.1°$ and $\sigma_\varphi = 0.1°$.

	RMSE				STD		
	Sim. 1	Sim. 2	Sim. 3		Sim. 1	Sim. 2	Sim. 3
# img.	45	102	263	# img.	45	102	263
ψ [°]	0.077	0.081	0.043	ψ [°]	0.068	0.048	0.032
θ [°]	0.094	0.056	0.032	θ [°]	0.065	0.045	0.027
φ [°]	0.069	0.050	0.029	φ [°]	0.067	0.046	0.028

Table 3.8: Results from 1000 MC runs with large INS errors of $\sigma_\psi = 4°$, $\sigma_\theta = 2°$ and $\sigma_\varphi = 2°$.

	RMSE				STD		
	Sim. 4	Sim. 5	Sim. 6		Sim. 4	Sim. 5	Sim. 6
# img.	45	102	263	# img.	45	102	263
ψ [°]	1.380	0.992	0.641	ψ [°]	1.375	0.991	0.641
θ [°]	1.297	0.885	0.546	θ [°]	1.296	0.884	0.546
φ [°]	1.270	0.843	0.541	φ [°]	1.270	0.843	0.541

3.5.5 Real-world experiments

The target platform is a small hexacopter equipped with a payload containing an INS based on MEMS and a camera. As INS an Ellipse-D from SBG Systems with standard deviations of $\sigma_\psi = 0.2°$, $\sigma_\theta = 0.1°$ and $\sigma_\varphi = 0.1°$, stated in the technical data sheet [SBG15], is utilized. The camera from XIMEA captures images with a resolution of 2048×1088 pixels and is attached to a wide-angle lens with a focal length of 4.8 mm. The sensors are rigidly coupled and connected with a synchronization cable for hardware trigger signals. The data collection for the system calibration was performed without the hexacopter, by powering the dismounted payload. This allowed an easier data collection of INS synchronized images showing the checkerboard in different perspectives.

The system calibration was performed for seven data sets collected with this setup. The real-world performance was evaluated by checking the consistency of the achieved results. In [BRS+16] it is stated that increasing the samples in the data sets helps to resolve the correlations between parameters of the boresight angles and the checkerboard normal. Based on the results in the simulations, a similarity threshold of 0.15, which extracts up to 300 images from the data sets, was chosen. This value is a trade-

Table 3.9: Camera calibration results from real-world experiments. The similarity thresholds for the checkerboard corners, results in a different number of images in the seven runs. Stated are the resulting camera parameter and the RMS reprojection error of the checkerboard corners.

		Real. 1	Real. 2	Real. 3	Real. 4	Real. 5	Real. 6	Real. 7
# img.		227	287	252	255	224	156	123
f_x	[pel]	899.00	899.71	904.67	899.67	904.68	900.45	900.94
f_y	[pel]	898.14	898.46	903.66	898.89	903.94	899.27	899.96
o_x	[pel]	1007.85	1008.81	1005.81	1008.87	1004.20	1007.55	1002.55
o_y	[pel]	561.99	559.96	557.47	560.75	558.31	556.72	557.41
RMS	[pel]	0.44	1.07	0.39	0.39	0.46	0.74	0.58

Table 3.10: Position free boresight calibration results from real-world experiments. Stated are the resulting boresight angles.

		Real. 1	Real. 2	Real. 3	Real. 4	Real. 5	Real. 6	Real. 7
# img.		227	287	252	255	224	156	123
ψ	[°]	89.326	89.134	88.970	89.359	89.196	89.321	89.054
θ	[°]	0.536	0.346	0.524	0.741	0.548	0.767	0.248
φ	[°]	23.084	23.073	22.686	23.016	23.228	23.143	22.813

off, between the calculation time required for the standard camera calibration and the increase in accuracies by using a bigger sample size for the boresight calibration.

The camera calibration worked properly on all data sets. The results show a high consistency for the intrinsic camera parameters throughout the seven runs. A high accuracy for the individual runs is proved by the small RMS reprojection errors, which for six out of seven runs shows a subpixel accuracy (Table 3.9).

For the boresight calibration the initial angular offsets between the sensors were extracted out of the construction drawing (Figure 3.15). Here again, the achieved results show a high consistency throughout the seven runs (Table 3.10). The initial values as well as the mean values and standard deviations of the boresight angles are given in Table 3.11. The results state on the one hand the already high accuracy achieved out of the construction drawings and on the other hand that these values could still be improved by performing the system calibration with the proposed approach.

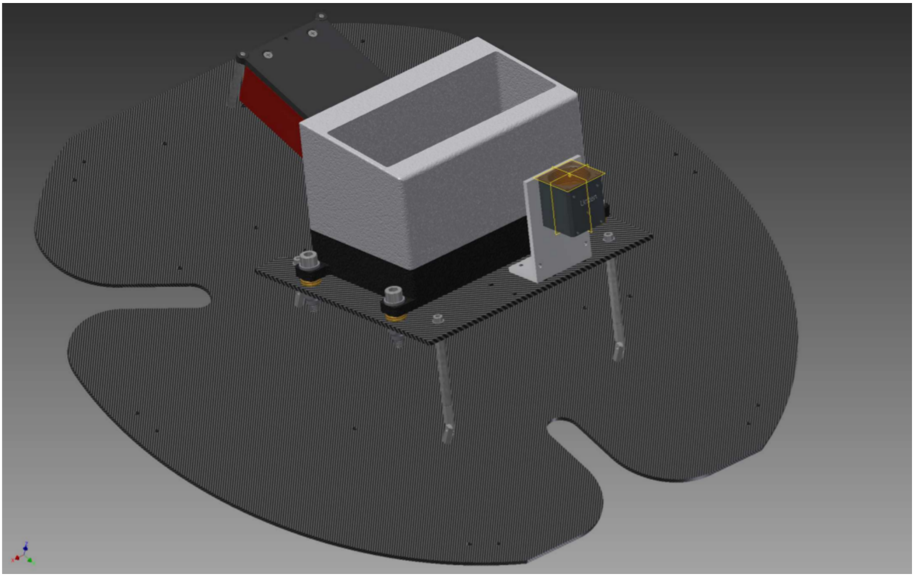


Figure 3.15: Construction drawing of the UAS payload with an INS (red) and a camera (small black cube). Furthermore, there is a laser scanner, which is not used in the presented approach, in the center of the payload.

Table 3.11: Evaluation of position free boresight calibration results from real-world experiments. Stated are the initial and the mean values, along with the standard deviations, of the optimizations from the seven data sets as well as the difference between them.

		initialization	mean value	standard deviation	difference
ψ	$[^\circ]$	90.00	89.19	0.14	-0.81
θ	$[^\circ]$	0.00	0.53	0.17	0.53
φ	$[^\circ]$	22.62	23.01	0.18	0.39

3.6 Summary

In this chapter, two approaches for the system calibration of a sensor suite consisting of a camera and an INS were presented.

The first INS-camera calibration was developed for high precision INS devices, which create reliable pose estimations with precision in the range of a few centimeters for the positional and small sub-degree value for the rotational components. The integration of the calibration parameters in the bundle adjustment equation and a reformulation in a graph structure was presented. Latter was used to estimate both, the mounting offsets between the devices and the intrinsic camera parameters. The high correlation between the lever-arm of the mounting offsets and the intrinsic camera parameters can in theory be decoupled by performing the flight maneuvers at two different altitudes. However, the achieved results showed an accuracy of the lever-arm optimization worse than usual measurements out of construction drawings or terrestrial measurements. The simulations confirmed that the best results are achieved if only the intrinsic camera parameters and the misalignment angles between the devices are optimized in the system calibration. Moreover, this eliminated the constraint of using two different altitudes. In conjunction with the small influence of the flight patterns a recalibration during each operation seems possible.

The second approach was designed for the calibration of low-cost devices, which typically suffer from low position accuracies. It was shown how to estimate the angular offsets of the boresight misalignment by using only the rotational part of the INS measurements. The proposed procedure allows the calibration of the intrinsic camera parameters and the boresight misalignment by processing the same data set observing a planar checkerboard pattern. In contrast to the in-flight calibration, this was performed before the flights. Nevertheless, an in-flight calibration observing the pattern in the starting area after the take-off seems also feasible.

Knowledge of the offsets between the coordinate systems of the sensor allows the fusion of measurements in a unified framework.

CHAPTER IV

Georeferenced maps by INS-based SLAM

"A map does not just chart, it unlocks and formulates meaning; it forms bridges between here and there, between disparate ideas that we did not know were previously connected."

— Reif Larsen

Knowing the current location and orientation in space is essential for a high amount of different tasks in the robotic area. To solve this problem, approaches using various sensors appeared in recent years. For outdoor application the usage of GPS assisted INSs has established as the de facto standard [SCCG93]. Apart from that, processing images captured by cameras for navigation purposes has proved to generate robust results. On the basis on the increased computing power, the algorithm are in these days real-time capable and reveal an alternative to the classical navigation with INS. But instead of making the decision of which sensor to use in a specific scenario, if applicable according to cost, size and weight constraints, it is superior to use both at the same time. In this chapter, the data fusion of a camera and an INS in a framework producing self-localization and georeferenced maps at the same time is shown.

Own publication on this subject: The work presented in this chapter has been partly published in [BRS+16] © 2016 IEEE.

4.1 Introduction

The most used sensor combination for the outdoor navigation of UASs consists of a GPS antenna and an IMU. Their measurements are combined and filtered in a so called INS which results in pose information describing the position and orientation of the device at a high frequency. Nowadays, even small INS devices utilize RTK corrections and therefore have a positional accuracy in the range of some centimeters (Chapter 2.4).

An alternative to the navigation with INS is the usage of camera data (Chapter 2.3). This can be realized by a real-time and robust SLAM system. Camera-based navigation systems have the advantage of low cost and weight, but suffer from the processing power needed for the calculation of pose information out of a huge amount of pixel measurements. Engel *et al.* presented an approach called LSD-SLAM, which runs in real-time on a central processing unit (CPU) [ESC14]. In contrast to previous mono camera-based SLAM systems based on feature points, it uses all image regions with gradients for pixelwise stereo comparisons over multiple frames. This leads to a very robust tracking and detailed maps. However, utilizing only cameras causes a drift in the estimated poses while moving the camera in new, previously not visited, regions. The reason for this is the workflow of all SLAM algorithms, which create pose information out of chained relative transformations to previous estimations. This results in summing up small errors over time, similar to the integration drift for processing the inertial sensors of an IMU.

Consequently, the INS leads to very accurate pose estimations, but provides no information about the environment. The LSD-SLAM generates very detailed maps of the observed area, but the calculated poses have a hight tendency to drift in large scenarios. Furthermore, the created maps have no reference to the metric or any geodetic system. A combination of these two approaches should be able to create very detailed, metric scaled and moreover georeferenced maps of the observed area in real-time (Figure 4.1). The proposed algorithm for the creation of these maps is the main contribution of the work presented in this chapter.

4.2 Related work

Using a sensor combination of an INS and a camera for the self-localization of UAS was examined by various authors. The integration of the optical flow as vision component into the navigation with an INS is one possibility of this type of sensor fusion [WGL$^+$08, KFN09].

Besides the optical flow, various approaches use landmarks to estimate the movement of the camera. For example, [KSS08] combined inertial measurements of an IMU with point landmarks tracked by stereo visual odometry in an extended Kalman filter framework. This enables the navigation of large distances without GPS. Likewise, most approaches utilizing a SLAM algorithm work with a small set of keypoints selected by a feature detector. These points are tracked over time and the corresponding 3D points are optimized in a bundle adjustment. A famous example is the parallel tracking and mapping (PTAM) framework [KM07]. The original purpose was the

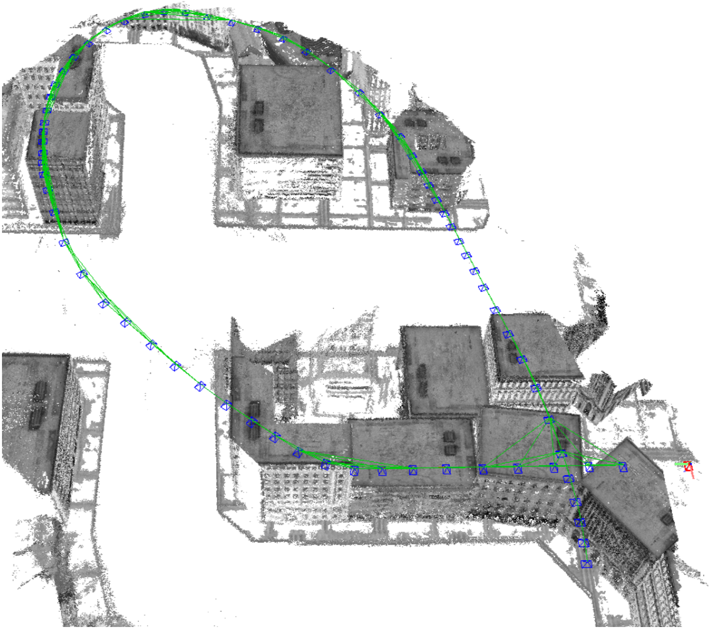


Figure 4.1: The proposed approach creates 3D pointclouds by performing pixelwise multi-stereo comparisons in images captured with a monocular camera. To achieve a metric scale and to minimize the drift measurements from an INS are utilized. The flight path of the UAS is depicted by the blue camera poses and the image constraints between them. The current camera pose with its coordinate system axes is depicted in red in the lower right of the visualization.

realization of a small augmented reality workspace, but the algorithm was already shortly after the publication applied in the area of robotics. There are various approaches using PTAM for the navigation of aerial platforms. Using its pose estimation for a controller-based stabilization of an autonomous UAS allows the navigation without additional sensors in small indoor environments [BWSS10]. A more accurate and robust approach was presented by Weiss *et al.* [WAL$^+$12]. They realized a navigation algorithm combining a camera and an IMU. The focus, besides the integration of PTAM into the navigation framework of the UAS, were various adaptations to work on an on-board processor with limited processing power. One of the main drawbacks of PTAM and its application in the UAS navigation is the scalability for large areas. While the number of keyframes grows, the mapping threads can only perform updates for a subset of keyframes and observed 3D points. The tracking part scales better but is also influenced by the high number of 3D points. To overcome this issue, the multiple map adaption of PTAM from Castle *et al.* [CKM08] was adapted to realize a fully autonomous UAS navigation in large areas [JS11]. They used the pose estimation of the monocular SLAM approach on a horizontally viewing camera to correct the drifting IMU measurements.

As an alternative to these keypoint based approaches, a more robust tracking can be realized by using all image information instead of only a subset of pixels. It was shown, that the convex optimization of a photometric dataterm in an energy functional results in superior tracking performance [NIH$^+$11]. Furthermore, a dense volumetric reconstruction with an UAS was presented by Wendel *et al.* [WMG$^+$12]. Both approaches achieve real-time performance with a graphics processing unit (GPU).

In contrast to this research, where keypoint-based or dense approaches were used, the semi-dense LSD-SLAM, using all image regions with gradients [ESC14], is investigated and adapted in this chapter. This promises a robust tracking and more detailed mapping compared to the keypoint-based approaches, while being less computationally intensive than the dense algorithms. To overcome the drift of the LSD-SLAM, INS measurements are utilized. By integrating filtered INS pose information in the mathematical formulation of the LSD-SLAM, the pose estimations and map quality is improved without the need of additional computational power. Quiet the contrary, some parts like the identification of loop closure becomes less expensive because of a smaller search region on the basis of more accurate pose estimations.

4.3 Definitions

The world frame W is defined as an ENU coordinate system (Chapter 2.1) with its origin being approximately in the middle of the flight area.

According to the general definition of rigid body motions (Chapter 2.2), $\mathbf{B}_{\mathrm{C}_i\mathrm{W}}$ describes the camera pose at the middle exposure time $t_i, i = 1, 2, ..., n$ with regard to the world frame W. Likewise, $\mathbf{B}_{\mathrm{C}_j\mathrm{C}_i}$ specifies the rigid body motion of the camera pose from time t_i to t_j which describes the transformation between the corresponding coordinate frames.

To describe an image, the mapping from the image plane $\mathbf{\Omega} \subseteq \mathbb{R}^2$ to the gray-values as the function $\mathrm{g} : \mathbf{\Omega} \to \mathbb{R}$ is defined. Likewise, the inverse depth map $\mathrm{d} : \mathbf{\Omega} \to \mathbb{R}_+$ and the corresponding depth variance $\mathrm{v} : \mathbf{\Omega} \to \mathbb{R}_+$ are specified.

4.4 LSD-SLAM

This section gives a brief overview of the LSD-SLAM algorithm, which is mandatory to understand the INS-based LSD-SLAM introduced afterwards. A more detailed description is given by the original publication of Engel *et al.* [ESC14].

The LSD-SLAM algorithm computes a real-time estimation of the position and orientation of a moving camera, also known as pose, in an unknown scene. At the same time, the approach generates a map of the observed environment. The algorithm is divided in a map building and a tracking part which are executed on multiple CPU cores to achieve real-time performance. Both parts operate on keyframes, created from new frames based on a weighted combination of distance and angle to the previous keyframes.

4.4.1 Tracking

The tracker estimates the pose transformation $\mathbf{B}_{\mathrm{C}_j\mathrm{C}_i}$ from the latest keyframe g_i to a new camera frame g_j by the direct minimization of the photometric error. This is achieved by evaluating the intensity difference of the keyframe pixels $\mathbf{x} \in \mathbf{\Omega}_i$ and their mapped positions in the image plane $\mathbf{\Omega}_j$ of the new frame:

$$\mathrm{r} : \mathbb{R}^2 \times SE(3) \rightarrow \mathbb{R} \ , \tag{4.1}$$

$$\mathrm{r}(\mathbf{x}, \mathbf{B}_{\mathrm{C}_j\mathrm{C}_i}) = \left(\mathrm{g}_i(\mathbf{x}) - \mathrm{g}_j(\mathrm{w}(\mathbf{x}, \mathrm{d}_i(\mathbf{x}), \mathbf{B}_{\mathrm{C}_j\mathrm{C}_i}))\right)^2 \ , \tag{4.2}$$

whereby the warp function $\mathrm{w} : \mathbb{R}^2 \times \mathbb{R} \times SE(3) \rightarrow \mathbb{R}^2$ performs the mapping of a pixel $\mathbf{x}$ from the keyframe g_i to the new image g_j [ESC14]. This is realized by using the inverse depth estimate of the keyframe $\mathrm{d}_i(\mathbf{x})$ to generate a 3D point $\mathbf{p}$ which is transformed to the coordinate system of the new camera by $\mathbf{b}(\mathbf{B}_{\mathrm{C}_j\mathrm{C}_i}, \mathbf{p})$ and then projected into the image plane g_j.

Thus, the overall energy function that is minimized in the tracking process is given by the sum of all residuals for pixels with an inverse depth:

$$\mathrm{E}(\mathbf{B}_{\mathrm{C}_j\mathrm{C}_i}) = \sum_{\mathbf{x} \in \mathbf{\Omega}_i \,|\, \mathrm{d}_i(\mathbf{x}) > 0} s_{\mathbf{x}} \cdot \mathrm{r}(\mathbf{x}, \mathbf{B}_{\mathrm{C}_j\mathrm{C}_i}) \ , \tag{4.3}$$

where $s_{\mathbf{x}}$ defines a robustness increasing weighting scheme based on the size and smoothness of the residuum $\mathrm{r}(\mathbf{x}, \mathbf{B}_{\mathrm{C}_j\mathrm{C}_i})$ as well as the inverse depth variance of the pixel $\mathrm{v}(\mathbf{x})$ as defined in [ESC14]. This least square problem is solved with an iteratively reweighed Gauss-Newton approach.

The resulting pose estimation is compared to a threshold defined for a weighted combination of position and angle differences to the latest keyframe. If it is lower, the image information of the current frame is used to refine the inverse depth map of the keyframe it was tracked on, otherwise, it will become a new keyframe. In this case its inverse depth map gets initialized by the propagation of information from the previous keyframe.

4.4.2 Mapping

The mapping is executed in a separate thread, independent of the tracking frequency. After a new keyframe is generated by the tracking process, the previous keyframe gets finalized and consequently processed by the mapping. The mapping part only operates on finalized keyframes, which implies that the inverse depth maps of these frames are not updated anymore. To find image constraints between the new finalized keyframe and spatial close keyframes, direct image alignments are performed. In addition to the photometric error defined in equation (4.2), a depth residual, penalizing incompatible inverse depth values, is evaluated. This is possible because, in contrast to the tracking, both frames have an inverse depth map at this stage of the algorithm. As a consequence, the scale between the frames is observable and considered as additional parameter in the direct image alignment. Thus, the pose transformations between keyframes are estimated as elements of $Sim(3)$, which adds the scale as an additional degree of freedom. By an abuse of notation, this $Sim(3)$ transformations are represented by the same symbol $\mathbf{B}$ as for the rigid body motion defined in the $SE(3)$ domain.

The global map containing the pose information of all keyframes is continuously optimized within g^{2}o, the general graph optimization framework [KGS$^+$11]. Therefore, the problem is embedded in a factor graph by introducing the camera poses $\mathbf{B}_{\mathrm{C}_i\mathrm{W}} \in Sim(3)$ as nodes, representing variables to optimize, and the tracked constraints $\mathbf{B}_{\mathrm{C}_j\mathrm{C}_i} \in Sim(3)$ between them as edges. An error function for these constraints connecting two camera poses can be defined as follows:

$$\mathbf{e}_{ji}(\mathbf{B}_{\mathrm{C}_i\mathrm{W}}, \mathbf{B}_{\mathrm{C}_j\mathrm{W}}) = \log(\mathbf{B}_{\mathrm{C}_j\mathrm{W}}^{-1} \circ \mathbf{B}_{\mathrm{C}_j\mathrm{C}_i} \circ \mathbf{B}_{\mathrm{C}_i\mathrm{W}}) \ , \tag{4.4}$$

whereby the logarithmic map transforms the rigid body motion to its twist coordinates. This results in an error vector of dimension seven, which is $\mathbf{0}$ if the image constraint perfectly describes the transformation between the two camera poses. The final energy that is minimized can be stated as follows:

$$\mathrm{E}(\mathbf{B}_{\mathrm{C}_1\mathrm{W}}, ..., \mathbf{B}_{\mathrm{C}_n\mathrm{W}}) = \sum_{\mathbf{B}_{\mathrm{C}_j\mathrm{C}_i}} \mathbf{e}_{ji}^{\top} \, \mathbf{\Sigma}_{ji}^{-1} \, \mathbf{e}_{ji} \ , \tag{4.5}$$

whereby $\mathbf{\Sigma}_{ji}^{-1}$ represents the inverse covariance matrix of the twist coordinates. The g^{2}o framework uses the Levenberg-Marquardt algorithm to compute a numerical solution of equation (4.5) and therefore needs a good initial guess of the state vector.

4.5 INS-based LSD-SLAM

This section describes the adaptations to the LSD-SLAM framework, which utilize measurements from an INS to make the camera-based navigation more robust. The

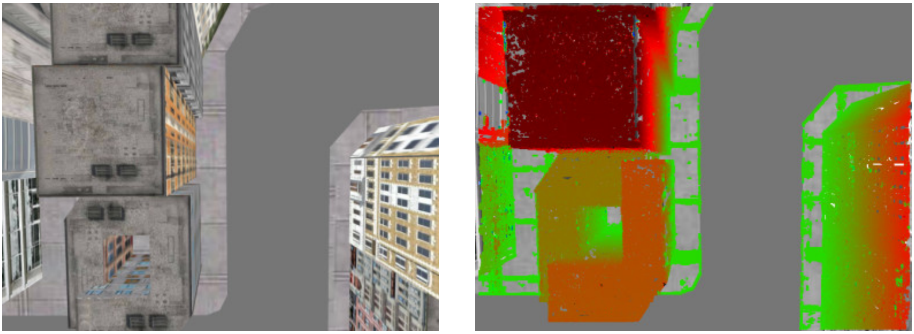


Figure 4.2: The input image on the left is tracked with respect to the inverse depth map of the latest keyframe displayed on the right. The metric scaling of the proposed approach leads to absolute depth information. The applied color coding fades from green (≥30 m) over red (15 m) to black (≤7.5 m). Image regions without gradients are represented in their original greyscale intensity.

original LSD-SLAM implementation offers clear benefits over the feature point approaches in terms of map details and scalability for large regions. Nevertheless, it suffers from the fundamental disadvantage of a monocular SLAM approach, alias the unknown scale drift. Small tracking errors sum up over time and lead to a drift in the calculated camera pose. In comparison to a stereo camera setup, this drift is even bigger because of the unknown scale of the scene. To detect regions already observed and perform a loop closure, Engel *et al.* apply, in addition to the spatial distance between keyframes, the appearance-based image comparison FAB-MAP [CN08]. This approach costs additional computing power and is likely to fail in similar-looking outdoor regions.

The usage of an additional sensor in form of a GPS-based INS leads to independent pose measurements in outdoor environments. To achieve an easy fusion with an INS, the coordinate system of the camera should be adapted to the navigational frame of the INS. Therefore, each camera pose $\mathbf{B}_{\mathrm{C}_i\mathrm{W}}$ is initialized with the composition of the associated INS pose $\mathbf{B}_{\mathrm{I}_i\mathrm{W}}$ and the mounting offsets between the devices $\mathbf{B}_{\mathrm{CI}}$. The latter can be determined in a system calibration beforehand (Chapter 3). The initialization of the original LSD-SLAM algorithm is performed by generating a random depth map with a mean depth of one and a high depth variance for all image regions with gradients larger than a defined threshold. The alternative approach realized in this thesis is based on a stereo initialization using pose information from the associated INS measurements. This allows the generation of an initial depth map in metric scale, by calculating depth estimates for all pixels with gradients above a threshold based on the transformation between the two frames (Figure 4.2).

4.5.1 Modified Tracking

The INS measurements provide an upper bound for the camera-based pose estimation, and thus are able to minimize the drift of the LSD-SLAM poses. Therefore, the pose optimization stated in equation (4.3) is performed while constraining the pose update $\mathbf{B}_{\mathrm{C}_j\mathrm{C}_i}$ to be in the standard deviation range of the corresponding INS pose composition:

$$\mathbf{B}_{\mathrm{I}_j\mathrm{I}_i} = \mathbf{B}_{\mathrm{I}_j\mathrm{W}} \circ (\mathbf{B}_{\mathrm{I}_i\mathrm{W}})^{-1} \;, \tag{4.6}$$

$$\mathbf{B}_{\mathrm{C}_j\mathrm{C}_i} \in [\mathbf{B}_{\mathrm{I}_j\mathrm{I}_i} - 4\boldsymbol{\sigma}_{\mathrm{I}}, \mathbf{B}_{\mathrm{I}_j\mathrm{I}_i} + 4\boldsymbol{\sigma}_{\mathrm{I}}] \;. \tag{4.7}$$

Using $4\boldsymbol{\sigma}_{\mathrm{I}}$ corresponds to the fact that the tracking update compares two independent INS poses and for both, the error is for more than 95% of the samples in the range of two times the standard deviation. Thus, the difference between the pose update and the corresponding INS measurements is limited to be smaller than $2 \cdot 2\boldsymbol{\sigma}_{\mathrm{I}} = 4\boldsymbol{\sigma}_{\mathrm{I}}$. This constraint is in some rare cases too strict. Nevertheless, the produced results are more accurate compared to the usage of $6\boldsymbol{\sigma}_{\mathrm{I}}$ as maximum distance for the calculated update based on empirical data.

4.5.2 Modified Mapping

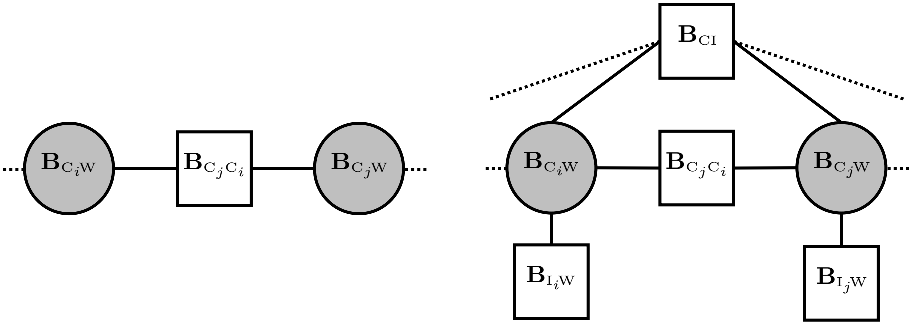
The change in the tracking part can only limit the drift, but not eliminate it. In order to achieve this goal, the corresponding INS measurement and the mounting offsets between the sensors are added in the factor graph as unary edges to each node representing a camera pose. Consequently, new error function constraining the camera poses can be defined as follows:

$$\mathbf{e}_i(\mathbf{B}_{\mathrm{C}_i\mathrm{W}}) = \log((\mathbf{B}_{\mathrm{CI}} \circ \mathbf{B}_{\mathrm{I}_i\mathrm{W}})^{-1} \circ \mathbf{B}_{\mathrm{C}_i\mathrm{W}}) \;. \tag{4.8}$$

This results in an error vector, which is $\mathbf{0}$ if the camera pose estimation is perfectly described by the corresponding composition of the INS measurement and mounting offsets. Adding this constraint to the factor graph extends equation (4.5) to the overall energy:

$$\mathrm{E}(\mathbf{B}_{\mathrm{C}_1\mathrm{W}}, ..., \mathbf{B}_{\mathrm{C}_n\mathrm{W}}) = \sum_{\mathbf{B}_{\mathrm{C}_j\mathrm{C}_i}} \mathbf{e}_{ji}^{\top} \boldsymbol{\Sigma}_{ji}^{-1} \mathbf{e}_{ji} + \sum_{\mathbf{B}_{\mathrm{C}_i\mathrm{W}}} \mathbf{e}_i^{\top} \boldsymbol{\Sigma}_i^{-1} \mathbf{e}_i \;, \tag{4.9}$$

whereby $\boldsymbol{\Sigma}_{ji}^{-1}$ and $\boldsymbol{\Sigma}_i^{-1}$ represent the inverse covariance matrices of the twist coordinates. Adding these edges to the graph eliminates the scale drift and makes the scale estimation performed in the LSD-SLAM algorithm superfluous. This allows using $SE(3)$ instead of $Sim(3)$ constraints in the factor graph (Figure 4.3).


$\mathbf{B}_{\mathrm{C}_i\mathrm{W}}$
$\mathbf{B}_{\mathrm{C}_j\mathrm{C}_i}$
$\mathbf{B}_{\mathrm{C}_j\mathrm{W}}$
$\mathbf{B}_{\mathrm{CI}}$
$\mathbf{B}_{\mathrm{C}_i\mathrm{W}}$
$\mathbf{B}_{\mathrm{C}_j\mathrm{C}_i}$
$\mathbf{B}_{\mathrm{C}_j\mathrm{W}}$
$\mathbf{B}_{\mathrm{I}_i\mathrm{W}}$
$\mathbf{B}_{\mathrm{I}_j\mathrm{W}}$

(a) LSD-SLAM

(b) INS-based

Figure 4.3: The objective function of the mapping stage can be illustrated by a graph. The measurements (boxes) are connected to the variables (circles). Here, the latter are the camera poses. In addition to constraints based on image information between the camera poses, two edges are added to each camera pose representing the corresponding INS measurement and the mounting offsets between the devices.

4.6 Simulations

In this section, results from simulations are presented to validate the proposed approach. The main purpose of this is to analyze the effect of the described integration of INS measurements in the SLAM process by comparing the estimations for the camera poses and the created map with the known ground truth.

4.6.1 Setup

Using the robot simulation Gazebo [KH04] allows testing the approach in a realistic scenario and gather valuable information for real-world experiments. The modeling, control and simulation of a quadcopter UAS within Gazebo were developed by Meyer *et al.* [MSK+12]. The generated pose information are on the one hand used as ground truth in the evaluation and on the other hand the basis for the generation of the noisy INS data. For the latter, white Gaussian noise is added to the poses using a standard deviation of $\boldsymbol{\sigma}_{\mathrm{I}} = (0.02\,\mathrm{m}, 0.02\,\mathrm{m}, 0.04\,\mathrm{m}, 0.1^\circ, 0.1^\circ, 0.2^\circ)^\top$, with the first three values related to the positional components and the last three describing the rotational part of the pose as Euler angles. Thereby the altitude as well as the rotation around the z-axis are chosen twice as big as the other components to model the standard error behavior of an INS. The used values represent the accuracies of a small INS with RTK-corrected satellite navigation measurements [SBG15].

Furthermore, the camera simulation provided by Gazebo is used to generate images from a downward facing camera with 640 × 480 pixels at a frame rate of 30 Hz. The

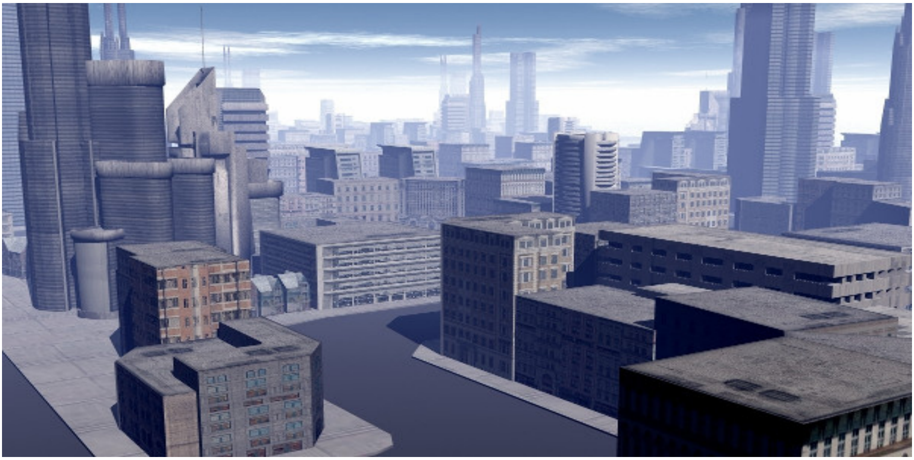


Figure 4.4: Rendered image of the 3D model ’The City’ from Herminio Nieves. The textured mesh is used as environment in the simulations. Captured images of this model contain, a mixture of regions with a lot of gradients (most houses), some gradients (sidewalks) and no gradients (streets).

chosen camera aperture angle of 100° corresponds to a wide-angle lens. The mounting offset between the camera and the INS is known in the simulation. For real-world experiments they can be estimated by performing a system calibration of the sensor setup from data collected in a previous flight (Chapter 3). The 3D model ’The City’ created by Herminio Nieves is used as environment in the evaluation (Figure 4.4). He published this [Nie15] and other models free for commercial and non commercial use.

4.6.2 Performance measures

The results are evaluated in two different ways. On the one hand the pose estimation is compared to the ground truth and on the other hand a matching between the generated point cloud and the 3D mesh, used as environment in the simulation, is performed.

Thereby, the pose comparison is realized separately for the translational and rotational part. Comparing two positions using the Euclidean norm is straightforward by using the scalar product as follows:

$$\Phi : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}_+ \ , \tag{4.10}$$

$$\Phi(\mathbf{t}_i, \mathbf{t}_j) = \sqrt{[\mathbf{t}_i - \mathbf{t}_j] \cdot [\mathbf{t}_i - \mathbf{t}_j]} \ . \tag{4.11}$$

The rotational part of the tracked poses is compared according to the angle difference between the unit quaternions $\mathbf{q}_i$ and $\mathbf{q}_j$ by:

$$\Phi_r : SO(3) \times SO(3) \rightarrow [0, \pi] \; , \tag{4.12}$$

$$\Phi_r(\mathbf{q}_i, \mathbf{q}_j) = 2 \arccos(|\mathbf{q}_i \cdot \mathbf{q}_j|) \; . \tag{4.13}$$

Using the absolute value of the scalar product is necessary due to the ambiguity in the quaternion representation as $\mathbf{q}$ and $-\mathbf{q}$ describe the same rotation [Huy09]. In contrast to comparing Euler angles with the Euclidean norm, the used metric reflects the distance. A smaller value always corresponds to a smaller rotation. Furthermore, the metric is geometrically meaningful as it represents the angle between two orientations, which is exactly twice the angle between the corresponding unit quaternions.

The second performance measure is the difference between the generated point clouds and the 3D mesh used as environment. For each point, the smallest distance to the mesh is determined. The visualization of the point to mesh distances allows an easy interpretation and comparison of the achieved results.

4.6.3 Evaluation

MC simulations with 100 trials were performed for the original LSD-SLAM, versions integrating only the tracking or mapping modification and the presented approach using both modifications. The latter will in the following be referred to as INS-based.

The first evaluated data set was created from a straight flight path at a constant altitude. This simple path without any maneuvers or course corrections shows the main problem of the monocular SLAM approach. Even without large scale changes, the tracked poses show a clear drift. The usage of the proposed approach using INS measurements as additional information prevents the pose drift.

The Euclidean difference of the translation error is visualized over time (Figure 4.5). The results of the LSD-SLAM grow linearly. At the end of the data set, after 60 seconds, the mean error is approximately five meters. Using the proposed tracking modification limits the error, but cannot eliminate the drift. At the end of the data set a mean error roughly half as large as using LSD-SLAM without this modification is stated. The exclusive usage of the mapping modification eliminates the drift, but leads to a sawtooth wave. This pattern is based on the workflow of the mapping modification which only operates on keyframes. Between these, the tracking estimates the current pose relative to the latest keyframe. One unexpected observation is that the combination of the tracking and mapping modification in the INS-based LSD-SLAM seems to eliminate the sawtooth wave completely. Furthermore, this leads to an overall smaller error. The mean values of the MC runs for the proposed approach are smooth and the corresponding standard deviations are small.

The metric defined in equation (4.13) is used to analyze the difference between the rotational part of the pose estimations and the ground truth (Figure 4.6). The mean value of the total angular error for the original LSD-SLAM grows slowly after the first few seconds and is at the end of the sequence still below two degree. This states, that the orientation estimation is more reliable compared to the positional results produced by the monocular SLAM system. Again, the proposed tracking modification reduces the error of the LSD-SLAM. The mapping modification shows in the beginning also the sawtooth pattern, but becomes a lot smoother afterwards. In contrast to the

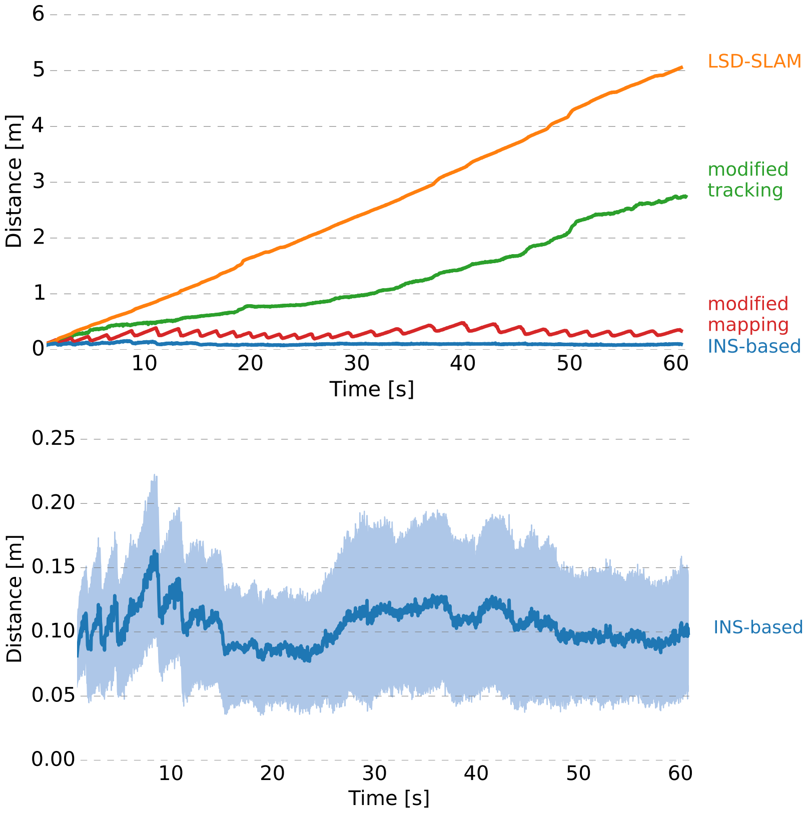


Figure 4.5: Mean Euclidean norm of the translation difference to the ground truth data for a straight flight path from 100 MC runs. The tracking modification limits the position drift of the algorithm to roughly one half of the LSD-SLAM results. The mapping modification leads to a sawtooth wave, which is smoothed in the INS-based LSD-SLAM. The latter combines both modifications. For clarification, its mean error in combination with the standard deviation (colored region) is depicted in the lower plot.

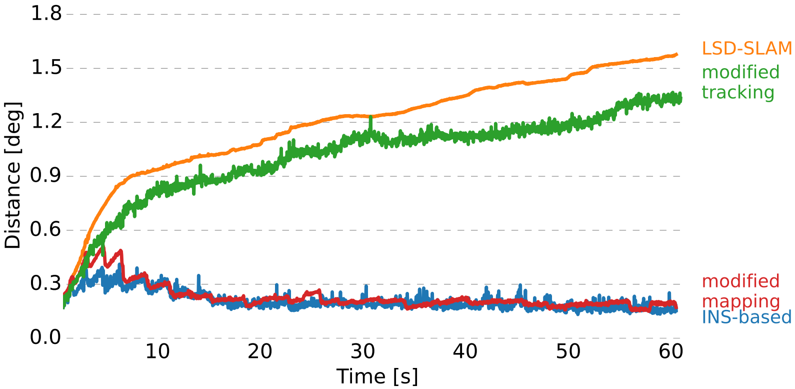
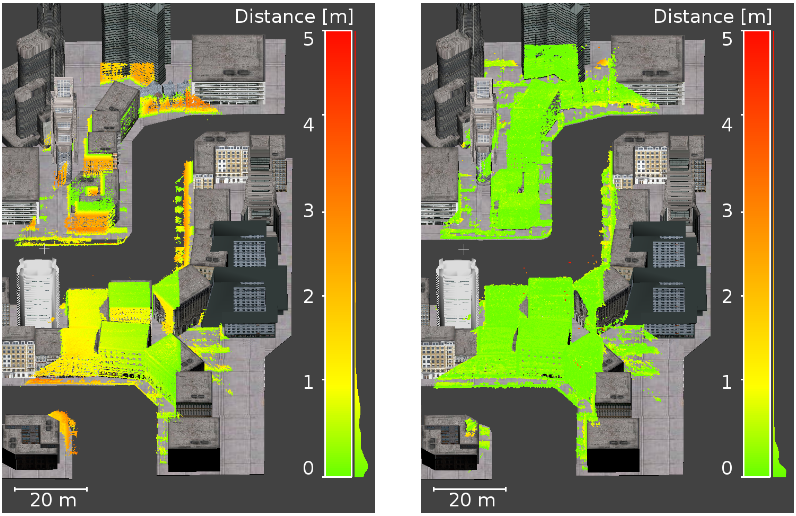


Figure 4.6: Mean angular error between the estimations and the ground truth data for a straight flight path from 100 Monte Carlo runs. The results show that the influence of the tracking modification is rather small. The combination of both modifications in the INS-based LSD-SLAM results only in the first few seconds in more accurate pose estimations compared to the modified mapping.



(a) LSD-SLAM **(b)** INS-based

Figure 4.7: Absolute distance of the generated maps, represented as point clouds, to the ground truth mesh. The straight flight from bottom to the top of the visualized area leads to bigger errors in the end region of the LSD-SLAM results, due to the pose drift (left). Using INS information eliminates the drift, which leads to a much more accurate point cloud (right). Histograms of the distances are displayed to the right of the color bars.

position difference, the combination of these modifications does lead to a smaller error in the first few seconds and is nearly the same afterwards. However, the results show that also the drift in the rotational part of the pose estimation gets eliminated.

The proposed approach limits the position error to the centimeter range and produces also very small errors in the rotational component. This allows the creation of maps in a metric scale with only small deviations from the ground truth model (Figure 4.7). As expected, the LSD-SLAM point cloud shows small differences to the ground truth directly after the initialization and higher differences in the upper region of the

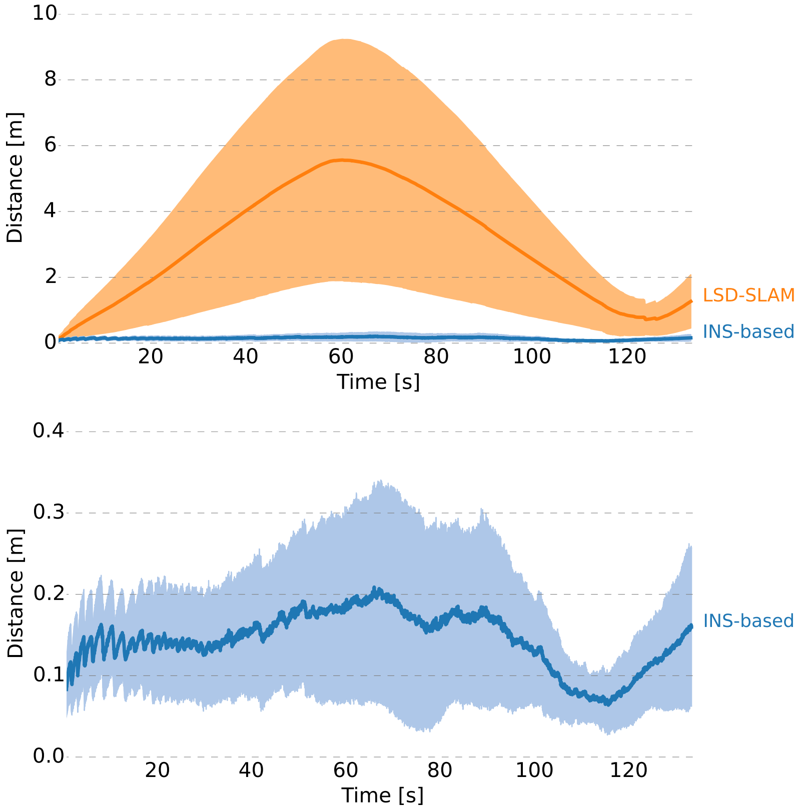


Figure 4.8: Mean Euclidean norm of the translation difference to the ground truth data for a circular flight course from 100 Monte Carlo runs. The mean and the standard deviation (colored region) of the LSD-SLAM pose estimations are improving in second half of the data set and the error starts to grow again after the final loop closure at 120 seconds. The same behavior is, to a lesser extent, also recognizable in the INS-based LSD-SLAM results.

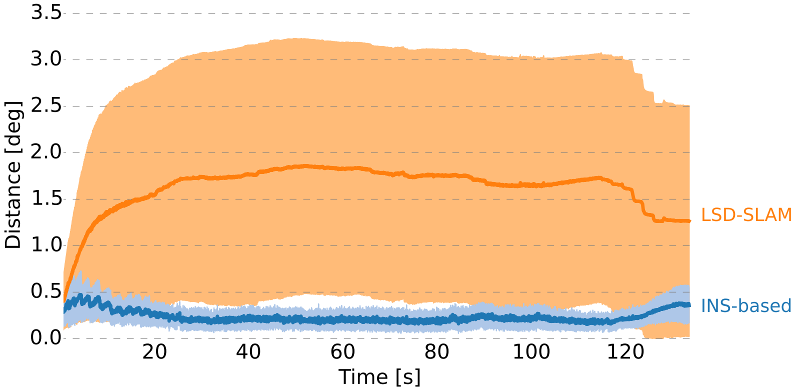
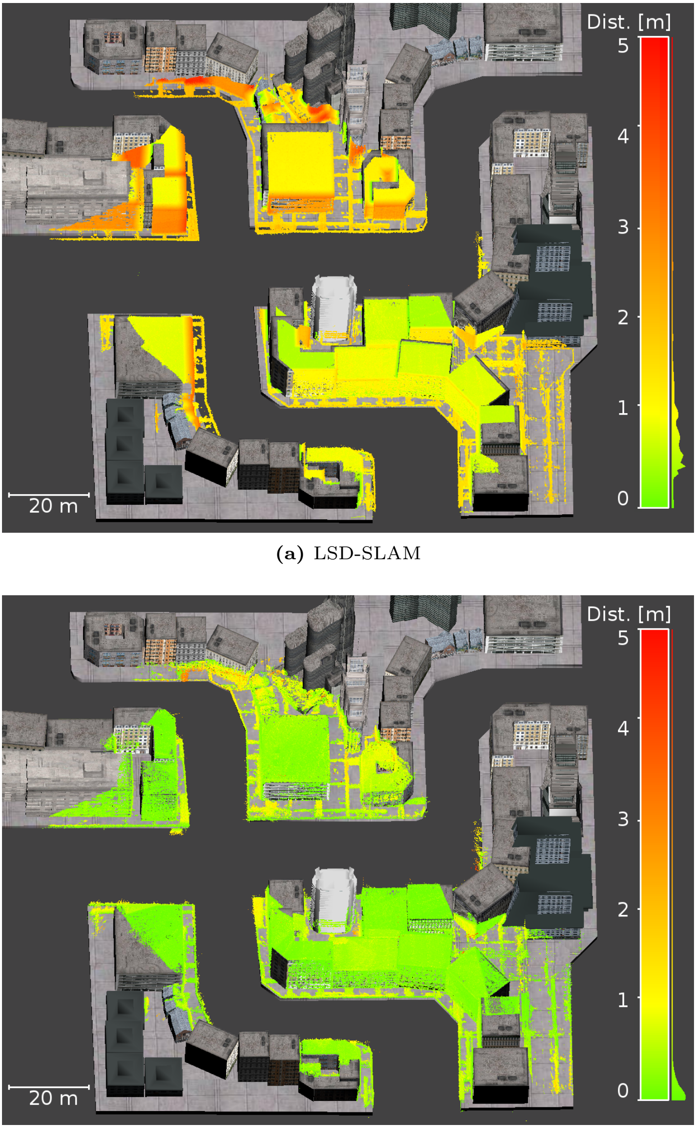


Figure 4.9: Mean angular error between the estimations and the ground truth data for a circular flight course from 100 Monte Carlo runs. The loop closure at 120 seconds reduces the error for the LSD-SLAM estimations, but leads to slightly worse poses for the INS-based LSD-SLAM. The standard deviations (colored regions around the mean curves) shows the great improvement of the INS-based approach regarding the robustness.

visualization, corresponding to the end of the flight path. For the INS-based LSD-SLAM results only some points in the outer region reveal a higher distance to the mesh. These points were most likely created by only one image correspondence and not updated by additional observations. It must be noted that the visualized error of the point cloud seems to be a lot smaller than the positional error of the camera poses produced by the original LSD-SLAM. This inconsistency is explained by the evaluation, which compares the distance of each point to its closest point in the mesh instead of its real correspondence. Especially in the evaluated scenario with mainly flat roofs, this has a positive effect on the distance between the point cloud and the 3D model.

In the second data set, the quadcopter was controlled with a gamepad in the simulation environment. This results in some rough but realistic movements in form of swings while performing course corrections. These movements were tracked without any problems and show the high performance of the LSD-SLAM algorithm. The flight course of this data set is a circular route at nearly constant altitude (Figure 4.1). This course allows the algorithm to perform a loop closure and propagate correction to the preceding pose estimations. However, all plots show the poses at the time they were tracked and not the final estimates of the factor graph optimization. Thus, the slowly decreasing error in the translational part of the pose estimations occurring in the second half of the data set (Figure 4.8) is not the result of the back propagated information from the final loop closure. A closer inspection of the visualization of one realization on this data set (Figure 4.1) reveals many edges in the factor graph



(a) LSD-SLAM

(b) INS-based

Figure 4.10: Absolute distance of the generated maps, represented as point clouds, to the ground truth mesh. The loop course starting in the lower right leads to large errors in the upper left region, where the distance to the starting point is largest. Using INS information eliminates the drift, which leads to a much more accurate point cloud. Histograms of the distances are displayed to the right of the color bars.

between the camera poses in the second half of the data set. These constraints have a strong positive effect on the factor graph optimization and result in decreasing the error. Also, the final loop closure occurs slowly by adding more and more constraints between the camera poses to the graph and, as a consequence, the loop closure does not result in a big jump (Figure 4.8). The mean angular error of the original LSD-SLAM implementation grows in the first 40 seconds to nearly two degree and shows a very high variance in the performed MC runs. In contrast to this, applying the INS-based modifications limit the error to less then half a degree. In the first few seconds the error starts growing slightly and even decreases slightly afterwards. During the last seconds, at the time of the loop closure, the high number of graph edges constraining the relation between camera poses lowers the overall accuracy by some amount. Also the variance of the angular error is limited to smaller values, which shows the increased robustness achieved by integrating INS measurements in the LSD-SLAM (Figure 4.9).

The accurate pose estimations have a direct influence on the generated world map. Visualizing the differences between the point cloud and the ground truth depicts the great results produced in real-time by the proposed approach (Figure 4.10).

4.7 Real-world experiments

The payload described in Chapter 3.5.5 was attached to a hexacopter (Figure 4.11). The collected images are rectified according to the distortion values estimated by the camera calibration and a pixel binning is performed. The latter downscales the images by a factor of four to 512 × 272 pixels. As a result, real-time performance at a framerate of 25 Hz on a standard Laptop with an Intel® Core$^{\text{TM}}$ i7-3632QM CPU was achieved. A downside of the pixel binning is that a lot of details captured in the original images are not used.

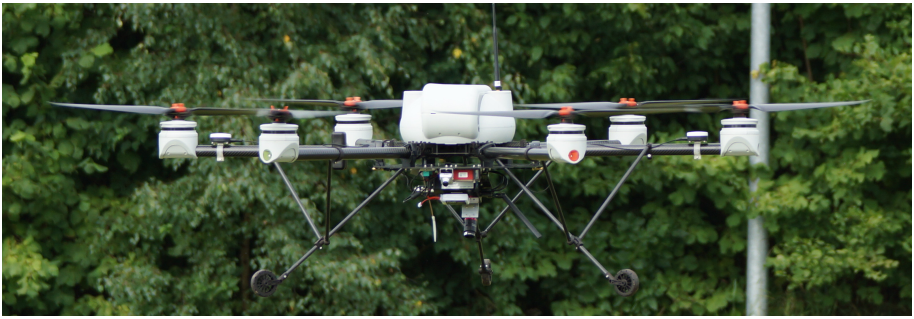


Figure 4.11: Hexacopter of the type AR-200 from AirRobot® GmbH & Co. KG equipped with the multi sensor payload.

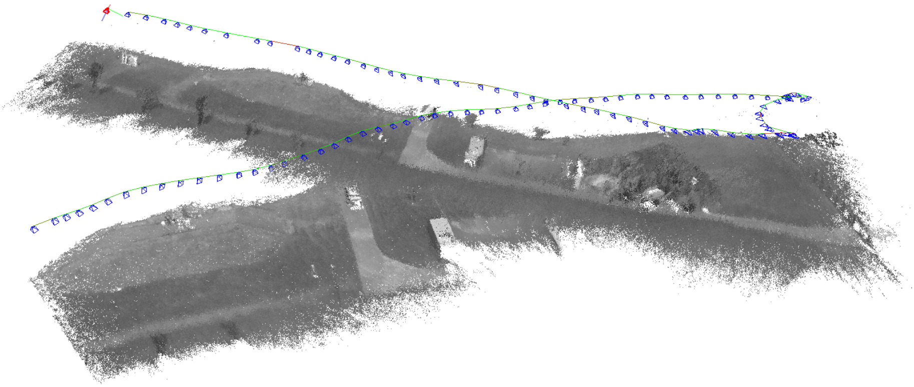


Figure 4.12: Real-world result of the LSD-SLAM. The generated map shows the big drift in the estimated camera position. Loop closure had to be deactivated, because of a high number of incorrect matches due to repetitive structures.

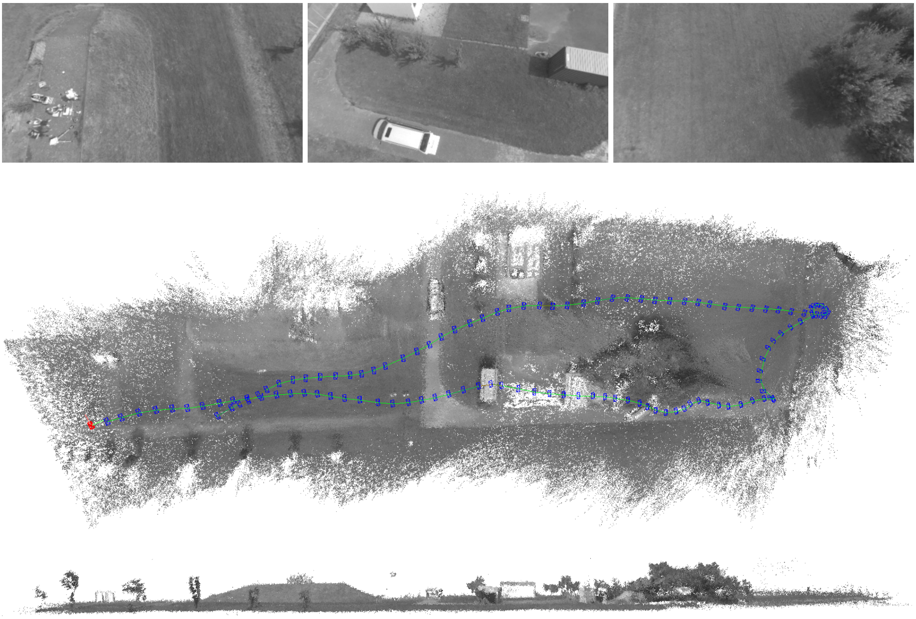


Figure 4.13: Real-world results of the INS-based LSD-SLAM. **Top:** Selection of already rectified and downscaled input images. **Middle:** Top view of the created map with visualization of the flight path and camera poses. **Bottom:** Side view of the created map.

In the evaluated scenario a large amount of the region is covered by grass and only some man-made objects with straight lines and large gradients are in the scene. This makes it quite difficult for the LSD-SLAM. The tracking works fine, but the similarities in the scene leads to a high amount of false loop closure in the map building part. It should be possible to tune the algorithm to perform better in this scenario. However, with the INS in the sensor setup the drift is eliminated. As a result, loop closures are not required to eliminate errors summed up in the tracking process. The parameter controlling the strictness for accepted loop closures was set to prevent there creation. As a drawback the camera only LSD-SLAM is not able to detect already observed regions and has no possibilities to eliminate the drift (Figure 4.12). In contrast, the visualization of the map from the presented INS-based LSD-SLAM shows a consistent point cloud (Figure 4.13).

4.8 Summary

In this chapter an approach which fuses measurements from an INS into a semi-dense monocular SLAM algorithm was presented. It was shown how INS measurements can be used within the tracking and mapping parts of the LSD-SLAM workflow to generate metric depth maps and eliminate the drift.

As a consequence, the generated point clouds of the observed areas are also metric and furthermore, due to the utilized GPS measurements, georeferenced. On the basis of RTK corrections, the created maps are accurate in the centimeter range. In contrast to other computational intensive camera-based algorithms, the results of the presented approach are generated in real-time and thus suitable for the UAS navigation.

The created maps describe the environment with a lot of details and can be used for various tasks. For example they should allow to perform a change detection or can be used as reference data for the robust navigation in scenarios without GPS measurements.

CHAPTER V

Map-based UAS homing using shortcuts

"Now that I'm here, where am I?"

— Janis Joplin

In the previous chapter was shown how to build georeferenced maps by combining the measurements of an INS and a camera. Thereby, the accurate localization is mainly based on the INS measurements, while the camera images allow to build a highly accurate map of the observed environment. If the GPS fails, the performance of the self-localization degrades because of drift characteristics of the inertial sensors and the monocular SLAM approach. This is critical if the INS was used to perform the flight control of the system. The orientation of the platform can as a substitute be extracted from the image stream collected by the camera, but due to the increasing drift it is safer to abort the current mission.

In this chapter, will be shown how to use a georeferenced map and current camera images to return an UAS to its starting position in the case of an GPS outage. Thereby shortcuts, which ensure to reenter areas previously explored, are considered.

Own publication on this subject: The work presented in this chapter has been partly published in [BCK17] © 2017 IEEE.

5.1 Introduction

Two critical aspect of UAS are the flight control and the navigation in large outdoor scenarios. For a long time these tasks were almost exclusively based on readings from inertial sensors corrected by GPS measurements in a so called inertial navigation system (INS). In contrast to the manned aviation, the navigation systems in most UAS are not redundant. A long-lasting GPS outage is for nearly all currently available

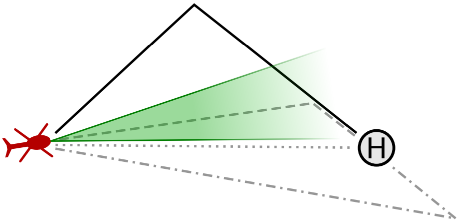


Figure 5.1: The UAS started at the heliport and flew along the solid path. At its current position the GPS fails and the platform starts an autonomous homing. The safest approach would follow the exact path the UAS was using to reach its current location. But this path can be quite long compared to the direct connection (dotted). However, a positional drift may occur, resulting in missing the heliport (dotted / dashed). This can probably be determined and corrected by inspecting the travel distance. A safer and more efficient approach aims to intersect the previous path close to the goal position (dashed). This is the main concept of the proposed path planning approach.

UAS a major problem. Without a pilot takeover, most systems initiate an emergency landing in this situation. The critical aspect is that the maneuver is most likely initiated at the current UAS position, regardless of the suitability of the area for a safe landing procedure.

Nearly all UAS are equipped with a camera for the real-time visualization of the observed area. The proposed approach uses these images in combination with the INS measurements for an efficient homing strategy. During normal operation the data from all sensors is fused to create a metric map of the area with a simultaneous localization and mapping (SLAM) approach [BRS$^+$16]. In the case of a GPS outage, only the camera and the previously generated map are used to return to the take-off position. The path planning procedure uses shortcuts through unobserved areas to minimizes the travel distance. In doing so, it is ensured to reach the goal by taking into consideration the maximal drift of the pure visual navigation (Figure 5.1).

5.2 Related work

Visual homing of an autonomous system describes the process of the system guiding itself to a previous location on the basis of visual sensor inputs. It is a problem worked on since decades and due to the lower processing power the first systems used non-visual sensors like radio beacons or an INS to return to a particular location. The vision-based approaches can be divided according to their operating principle in two categories.

A group of approaches performs a direct association of visual patterns and steering commands without a world model. Examples are the road following by Pomerleau

[Pom90] and the navigation along forest trails by Giusti *et al.* [GGC$^+$16]. Both approaches were realized with a neural net. A homing based on scene familiarity has been proposed by Nelson. The procedure looks for the best match of the current view to a set of previous collected images, saved with associated directions of movement [Nel91].

Other approaches are based on maps that store the position of objects and locations in a common reference frame. Errors are incorporated in the map by noisy sensor measurements and moving objects. The discrepancies between the map and the actual environment may be a problem for the path planning [MF03]. These uncertainties have nicely been covered in the work of Valencia *et al.* [VACP11] by using the Pose SLAM graph directly as belief roadmap to perform a collision free path planning along the route with the lowest accumulated robot pose uncertainty. A method using a graph of poses generated with a bundle adjustment as basis for the path planning has been proposed in [SMRN10]. All these approaches consider only the already traversed trajectory as feasible and obstacle free, which is a valid and useful assumption, especially for the ground based navigation.

In contrast, the in the following processed scenario considers an UAS with a downward looking camera at a fixed altitude without obstacles. As a result, the usage of shortcuts through previously not visited areas is possible. This navigational task has not been covered before in literature. The proposed algorithm works analog to the navigational abilities of dogs [TWBM97]. Chapuis validated in experiments that dogs have a metric representation from previous incomplete explorations. They use shortcuts between known areas, whereby they perform a safety strategy to make a correction in the case of bad direction estimate [Cha88]. The main idea of this concept, already transfered to the UAS context, is visualized in Figure 5.1. One major prerequisite for safe shortcuts by navigating an UAS through previously unexplored areas is the knowledge of the drift for pose estimation from pure visual odometry. A comprehensive theoretical analysis has been performed by Liu *et al.* [LJHW14]. They state that the drift is a random process that will not increase linearly and in some situations even may decrease. However, they declare that the end-point drift of visual odometry algorithms is generally between 1 % and 5 % of the traveled distance.

5.3 Problem description

In this section, the fundamental concept of the proposed UAS homing approach is explained. During normal operation, the INS-based LSD-SLAM (Chapter 4) is used to build a map in the form of a metric and georeferenced 3D point cloud of the observed environment and perform a self-localization at the same time. The integration of measurements from an INS in the LSD-SLAM algorithm eliminates the drift and generates metric depth map estimations for the processed images. As a consequence,

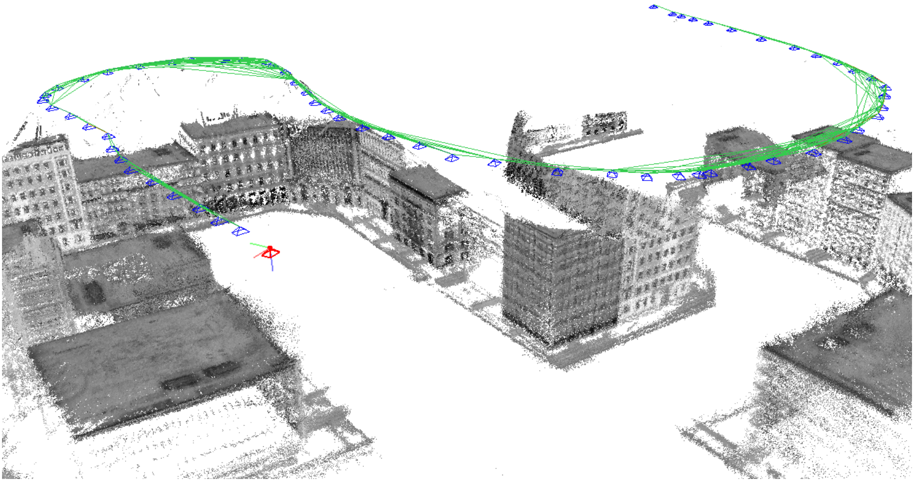


Figure 5.2: Visualization of the LSD-SLAM output. The flight path of the UAS is depicted by the blue keyframes and the image constraints between them. The current camera pose is depicted in red with its coordinate system axes

the generated point clouds of the observed areas are also metric and furthermore, due to the utilized GPS measurements, georeferenced. On the basis of RTK corrections, the created maps are accurate in the centimeter range. In contrast to other computational intensive camera-based algorithms, the results of the approach are generated in real-time and thus suitable for the UAS navigation. The approach generates a factor graph that consists of keyframes and image constraints between them (Figure 5.2).

In the case of an GPS outage, the platform returns to its start position by only using camera images and the previously generated outputs of the metric SLAM. The most secure approach would follow the previously traveled path in reverse. However, this may imply a really large detour, compared with the direct connection to the starting point. By leaving previously exploited areas, the own position is estimated based on visual odometry [ESC13]. Thereby, the relative estimates between the collected images lead to a drift in the self-localization, which in the given scenario can only be corrected by reentering an area observed in the previous map building phase.

The problem considered is the path planning for homing an UAS along a fast and at the same time safe path by only using the current camera images and the previously generated output of a metric SLAM.

5.4 Path planning using shortcuts

By using the output of the INS-based LSD-SLAM as basis for planning, a path back to the starting position is determined under consideration of safe shortcuts. In the following the problem is defined in a more abstract representation and the algorithm to perform the path planning is formulated.

5.4.1 Problem definition

In contrast to most approaches, especially the ones for ground based robots, the area contains no obstacles. Further, the planning is performed in the xy-plane. Dropping the z-coordinate is straightforward for missions which are completely performed at a fixed altitude.

The factor graph, generated by the INS-based LSD-SLAM during normal operation of the UAS, is rephrased as a 2D graph $G = (V, E)$. Thereby, the Cartesian coordinates of the keyframes describe the vertices V and each, except the first, keyframe is connected with its direct predecessor by an edge added to E. If the current position does not coincide with the last keyframe, it is added as new vertex $\mathbf{v}_n$ to V and an edge which connects $\mathbf{v}_n$ with the vertex of the last keyframe to E. By traveling in previously not visited areas visual odometry is used to perform a self-localization. This leads to an integration drift that is bounded by a known factor in relation to the traveled distance. The latter can be transformed to the maximal angular drift α. The graph describes the path already observed during normal operation and any intersection with the edges E allows performing an accurate localization of the UAS by comparing the current camera image with the previously generated map.

By defining the current position as start $\mathbf{v}_s \in V$ and the first keyframe as goal $\mathbf{v}_g \in V$, the problem of UAS homing is reformulated to the search of a path between these two vertices. Thereby, the path is not exclusively bound to the edges E already in the graph, but using a shortcut needs to consider the maximal angular drift α to guarantee an intersection with one of the graph edges E. A solution needs to converge in any scenario and the distance of the traveled path should be close to the distance of the direct connection between the start and goal vertex.

5.4.2 Algorithm description

Each vertex is described by its coordinates $\mathbf{v}_i = [x_i, y_i]^\top$. The shortest path between the start vertex $\mathbf{v}_s$ and the goal vertex $\mathbf{v}_g$ is the direct connection, which is part of the following ray:

$$\mathbf{r} = \mathbf{v}_s + \lambda[\mathbf{v}_g - \mathbf{v}_s] \ , \tag{5.1}$$

with $\lambda \in \mathbb{R}_+$. The maximal angular drift α allows defining a left and a right ray enclosing the drift area:

$$\mathbf{r}_l = \mathbf{v}_s + \lambda \mathbf{R}(\alpha)[\mathbf{v}_g - \mathbf{v}_s] \ , \tag{5.2}$$

$$\mathbf{r}_r = \mathbf{v}_s + \lambda \mathbf{R}(-\alpha)[\mathbf{v}_g - \mathbf{v}_s] \ , \tag{5.3}$$

with the rotation matrix $\mathbf{R} \in SO(2)$. Further, the Euclidean norm is used to describe the distance between two vertices by using the scalar product as follows:

$$\Phi : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}_+ \ , \tag{5.4}$$

$$\Phi(\mathbf{v}_i, \mathbf{v}_j) = \sqrt{[\mathbf{v}_i - \mathbf{v}_j] \cdot [\mathbf{v}_i - \mathbf{v}_j]} \ . \tag{5.5}$$

The path planning is performed shortly after the UAS looses the GPS signal. Therefore, the starting pose can be treated as accurate and the uncertainties from small errors in the position or heading may be covered by increasing the maximal angular drift slightly.

A temporary graph $G_t = (V_t, E_t)$ is created to efficiently work with the graph entities within the drift area. This graph is created from G as follows:

1. Add temporary vertices at the intersections between the maximal drift rays $\mathbf{r}_l$ and $\mathbf{r}_r$ with the edges in E.
2. For each new vertex, add two temporary edges connecting the vertex with the source and target vertex of the original edge it intersects.
3. Remove all vertices and edges outside the drift area. (This can be performed very efficiently by temporarily disabling the entities in a so called graph filtering.)

For each change of the target ray $\mathbf{r}$ the temporary graph G_t is updated. As a first step, the original graph G is recreated by removing the temporary vertices and edges as well as the graph filter. Then the temporary graph G_t is created as described above.

The proposed approach follows the workflow depicted in Figure 5.3. The contained cases are visualized in Figure 5.4 and defined as follows:

1. *Direct connection*: Within the drift area exists a path between $\mathbf{v}_s$ and $\mathbf{v}_g$ which only contains edges in E_t. By following this path the goal position $\mathbf{v}_g$ will be reached (Figure 5.4a).
2. *Ray to ray connection*: There exists a valid sequence of vertices $\mathbf{s} = (\mathbf{v}_l, ..., \mathbf{v}_r)$ connecting a vertex placed on the left ray with a vertex located on the right ray only by edges in E_t. A sequence is considered as valid, if it is either close or

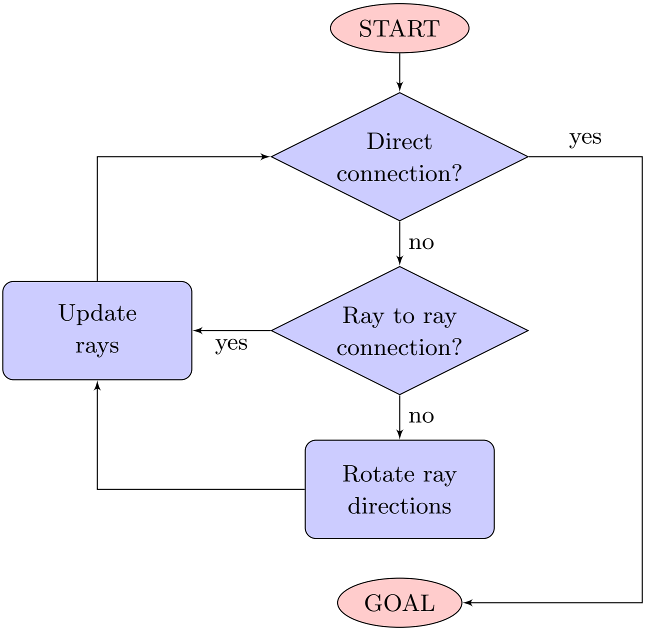
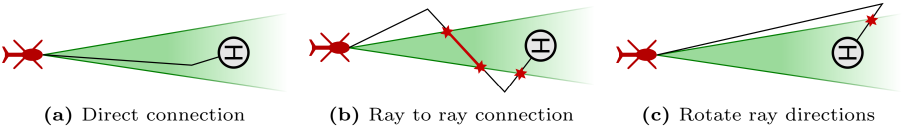


Figure 5.3: Flowchart visualizing the workflow of the proposed algorithm to navigate from a start to a goal position.



(a) Direct connection **(b)** Ray to ray connection **(c)** Rotate ray directions

Figure 5.4: The three cases of the proposed approach. The algorithm investigates the intersections (red stars) between the graph (black lines) and the maximal drift rays (outer green lines, spanning the green area).

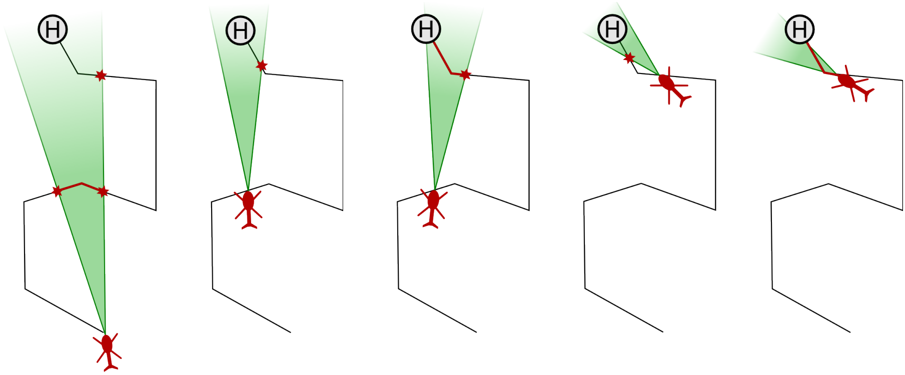
connected to the goal vertex. Regarding the first option, a sequence is considered as close, if for all vertices $\mathbf{v}_i$ the distance to the goal vertex $\mathbf{v}_g$ is less than the distance between the start and the goal vertex (Figure 5.4b):

$$\forall \mathbf{v}_i \in \mathbf{s} : \Phi(\mathbf{v}_i, \mathbf{v}_g) < \Phi(\mathbf{v}_s, \mathbf{v}_g) \ . \tag{5.6}$$

Alternatively, a valid sequence contains or is connected to the goal vertex $\mathbf{v}_g$ by edges in E_t. If a valid sequence has been found it is safe to travel and the random drift determines where the graph is intersected. The intersection becomes the new $\mathbf{v}_s$ and as a consequence the rays defined in equation (5.1), (5.2) and (5.3) have to be updated.

3. *Rotate ray directions*: If neither a direct nor a valid ray to ray connection exists in the temporary graph G_t (Figure 5.4c), the directions of the target ray $\mathbf{r}$ as well as of the rays $\mathbf{r}_l$ and $\mathbf{r}_r$ enclosing the drift area are adapted. By entering this case for the first time with the current start vertex $\mathbf{v}_s$, all rotation angles that would rotate either the left ray $\mathbf{r}_l$ or the right ray $\mathbf{r}_l$ on a vertex in V are determined. Note that this evaluates the vertices in the original graph G. By discarding all angles, with an absolute value larger than the maximal angular drift α, it is guaranteed that the goal vertex is placed in the new drift area after the rotation is applied. The angles are sorted according to their absolute values. Now and in all following iterations of this case with the current start vertex $\mathbf{v}_s$, one angle is removed from the sorted list and the initial directions of the three rays $\mathbf{r}$, $\mathbf{r}_l$ and $\mathbf{r}_r$ are rotated accordingly.

An example of a complete path planning sequence visualizing the iterated cases is given in Figure 5.5.



(a) Ray to ray connection

(b) Rotate ray directions

(c) Ray to ray connection

(d) Rotate ray directions

(e) Direct connection

Figure 5.5: Sequence of steps performed to reach the goal in the presented scenario. For each new start position, the path considered first aims at the goal position and investigates the intersections (red stars) between the graph (black lines) and the maximal drift rays (outer green lines, spanning the green area).

5.4.3 Algorithm convergence

In the following, the proof for the convergence of the proposed algorithm is stated. For reasons of clarity, two Lemmas are formulated and proofed first.

Lemma 1. *After a finite number of ray to ray connections a direct connection will be found.*

Proof. In the previous section two options for a valid *ray to ray connection* were defined:

1. Close: The intersection from traveling, which is used as new start vertex in the next iteration, is always closer to $\mathbf{v}_g$ than $\mathbf{v}_s$ to $\mathbf{v}_g$.

2. Connected to $\mathbf{v}_g$: It is possible to intersect the known path behind the goal vertex. This case has to be investigated further if the new start vertex has a larger distance to $\mathbf{v}_g$ than the distance between the current start vertex $\mathbf{v}_s$ and $\mathbf{v}_g$. The latter will be referred as $\Phi_1 = \Phi(\mathbf{v}_s, \mathbf{v}_g)$ in the following. If this happens only once, the distance to $\mathbf{v}_g$ will always be decreased in the following iterations. After another detour according to the same principle, the distance from the new start vertex to $\mathbf{v}_g$ will be smaller than Φ_1. This is based on the fact, that the connection to $\mathbf{v}_g$ was enclosed by the drift area before the first detour. Therefore, the distance to $\mathbf{v}_g$ will also be decreased in this situation.

By constantly decreasing the distance to $\mathbf{v}_g$ a direct connection will be found, at the latest when the new $\mathbf{v}_s$ is located on the edge directly connected to $\mathbf{v}_g$. □

Lemma 2. *For each start vertex $\mathbf{v}_s$ at least one rotation angle in the rotate ray directions case will result in a direct connection or ray to ray connection.*

Proof. By adding an edge for each new keyframe to its predecessor, there exist exactly one path between the start vertex $\mathbf{v}_s$ and the goal vertex $\mathbf{v}_g$. Both vertices are located in the drift area enclosed by the two rays $\mathbf{r}_l$ and $\mathbf{r}_r$.

There exists one rotation that places $\mathbf{r}_l$ on $\mathbf{v}_g$ and one which does the same for $\mathbf{r}_r$. In one of these two instances, the path directly connected to $\mathbf{v}_g$ will proceed in the drift area. Based on the fact that there exists a connection between $\mathbf{v}_s$ and $\mathbf{v}_g$, an intersection $\mathbf{v}_i$ with $\mathbf{r}_l$ or $\mathbf{r}_r$ will be found by traversing the edges starting from $\mathbf{v}_g$:

1. Intersections with both rays: corresponds to an intersection with $\mathbf{v}_s$ which results in a *direct connection*

2. Intersection with the other ray: *ray to ray connection*

3. Intersection with the same ray: *rotate ray directions*

If the new intersection $\mathbf{v}_i$ occurs with the same ray $\mathbf{v}_g$ is placed on, the path between $\mathbf{v}_g$ and $\mathbf{v}_i$ is traversed and the vertex closest to the other ray is determined. The drift area is rotated in a way that the other ray intersect this vertex, which becomes the new $\mathbf{v}_i$. This results in a connection between $\mathbf{v}_g$ and $\mathbf{v}_i$ within the drift area.

Traversing from $\mathbf{v}_i$ away from $\mathbf{v}_g$ will continue in the drift area. A new intersection $\mathbf{v}_i$ with $\mathbf{r}_l$ or $\mathbf{r}_r$ will be found and investigated as described above. Because there exists a path between $\mathbf{v}_s$ and $\mathbf{v}_g$, iterating this procedure will converge in either a *direct connection* or a *ray to ray connection.* □

Theorem 1. *The proposed algorithm will always converge in the goal position.*

Proof. As depicted in Figure 5.3 and described in the previous section, the algorithm iterates through three cases:

1. *Direct connection*: Follow the connection and reach the goal position.

2. *Ray to ray connection*: Lemma 2 states that after a finite number of *ray to ray connections*, a *direct connection* will be found.

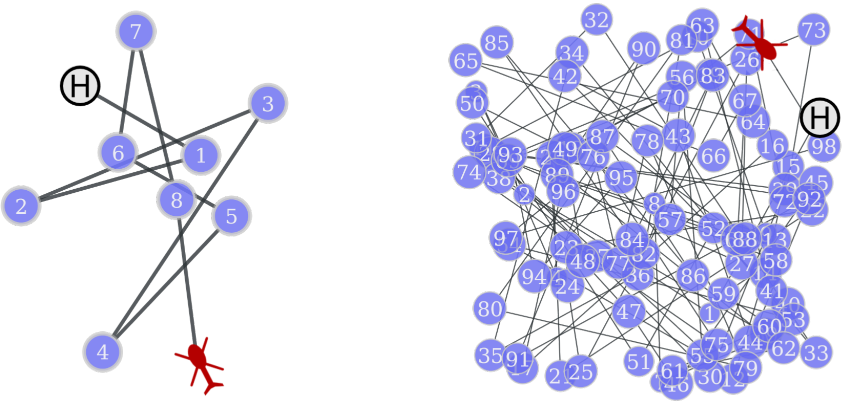
3. *Rotate ray directions*: According to Lemma 1, there will always a rotation be found which results in a *direct connection* or a *ray to ray connection.*

According to these observations a direct connection that leads to the goal position will always be found. □

5.5 Numerical studies

As a first evaluation of the presented approach, MC simulations on random graphs were performed. The latter represent flight courses from the take-off positions to the locations of the GPS loss. To generate a graph, n vertices were sampled at random coordinates in a square area with a side length of x meter. For each new vertex an edge connecting it with its predecessor was added to the graph. This resulted in a random graphs with n vertices and $n-1$ edges. The first vertex was used as take-off position and the last vertex as the start coordinate for the UAS homing. The created graphs look quite chaotic and in most cases are no flight maneuvers an operator would plan for an UAS (Figure 5.6). Nevertheless, the random graphs allow evaluating the approach in a comprehensive number of experiments, which with a high probability detect any problems and weaknesses of the procedure.

Three series of MC runs were performed, each with a different combination of the graph configuration parameters n and x. Further, a maximal drift ratio of 5 % in relation to the traveled distance was defined. For each configuration 1000 random graphs were generated and the path planning for each graph was realized in 100 MC runs. The mean distance used to reach the goal was calculated for each set of 100 MC runs. This value was divided by the optimal solution, characterized by the Euclidean



(a) 10 vertices

(b) 100 vertices

Figure 5.6: The random graphs allow an intensive evaluation of the presented approach. The first vertex is considered as take-off position the UAS wants return to and the last vertex as coordinate of the UAS at the time of the GPS outage.

distance between the start and the goal vertex, to form the travel distance ratio as follows:

$$r_d = \frac{\frac{1}{100}\sum_{i=1}^{100} d_i}{\Phi(\mathbf{v}_s, \mathbf{v}_g)}, \tag{5.7}$$

with d_i stating the travel distance used to reach the goal in the i-th MC run. Thus, a r_d of two states that on average the distance is twice the optimal solution. Further would a value of one point out that the algorithm produced for all runs the optimal solution. The latter is not possible because of the positional drift which has to be considered, but the closer the results are to one, the better. For each of the three configurations the quantiles of the mean traveled distance ratio from the 1000 random graphs were calculated (Table 5.1). The absolute time and distance savings are a lot bigger for data set 2 compared to data set 1 due to the larger size of the area. However,

Table 5.1: Travel distance ratio (achieved mean / optimal). Quantiles of 1000 random graphs for each data set, with 100 MC runs per graph.

data set	vertices	square side length	50 %	95 %	99 %
1	10	500	1.010	1.160	1.479
2	10	5000	1.011	1.139	1.449
3	100	500	1.002	1.021	1.093

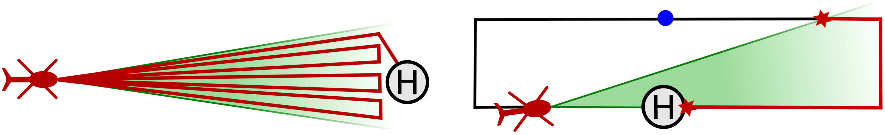


Figure 5.7: Suboptimal planning. **Left:** A connection inside the drift area of the visual odometry (green) may lead to large detours. **Right:** In the visualized scenario, the planning will result in a relocalization at one point on the marked edges (red). An optimal solution would plan a detour using an interim goal (blue).

the evaluation based on the ratio eliminates this difference as expected. The increase in the number of vertices for data set 3 leads to better results. This is based on the higher number of edges resulting in more ray to ray connections while targeting the goal directly. Following this optimal direction more often, avoids detours and decreases the used travel distance.

By investigating the outliers of the numerical studies two special cases were identified. They are depicted in Figure 5.7. The left image shows that a direct connection between the start and goal vertex can be quite long. The visualized instance shows an extreme situation that is very unlikely to happen. Nevertheless, also simpler detours following the same scheme may occur and in all these cases the presented algorithm will traverse the direct connection. Most of these detours could be prevented by traversing the direct connection, but performing a new planing according to the presented approach after a distance threshold is exceeded. This small adaption will very likely lead to new shortcuts. The second type of detours the algorithm will produce is also visualized in the right image of Figure 5.7. By using the heuristic that only movements which contain the goal in the visual odometry drift area are valid, the depicted shortcut using an interim goal will be missed.

5.6 Simulations

The robot simulation Gazebo [KH04] was used to test the approach in a realistic scenario and gather valuable information for real-world experiments. The modeling, control and simulation of a quadcopter UAS within Gazebo were developed by Meyer *et al.* [MSK+12]. The pose information of the quadcopter are on the one hand used as ground truth in the evaluation and on the other hand the basis for the generation of noisy INS data. For the latter, white Gaussian noise was added to the poses using a standard deviation of $\sigma_{\mathrm{t}} = (0.02\,\mathrm{m}, 0.02\,\mathrm{m}, 0.04\,\mathrm{m})$ for the positional components and $\sigma_{\mathrm{r}} = (0.1^\circ, 0.1^\circ, 0.2^\circ)$ for the rotational components represented by Euler angles. Thereby the altitude as well as the rotation around the z-axis are chosen twice as big as the other components to model the standard error behavior of an INS. The

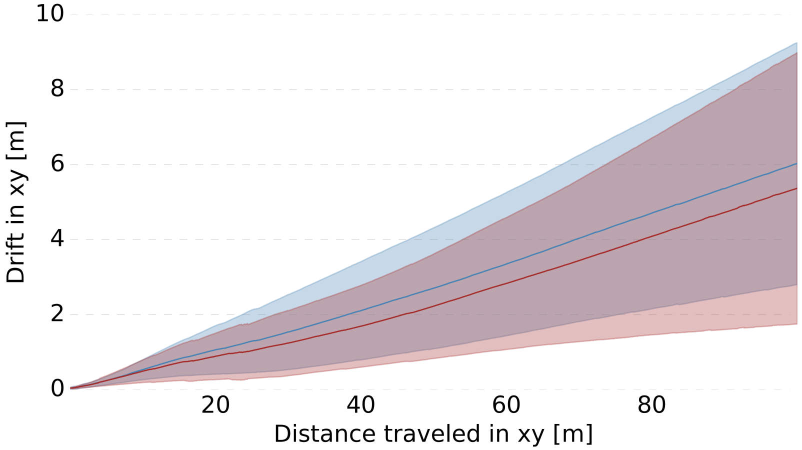


Figure 5.8: Mean and standard deviation of the camera only LSD-SLAM drift from 100 MC runs for two straight level flights courses in Gazebo.

modeled values represent the accuracies of a small INS with RTK-corrected GPS measurements [SBG15]. Furthermore, the camera simulation provided by Gazebo was used to generate images from a downward facing camera with 640 × 480 pixels at a frame rate of 30 Hz. The chosen aperture angle of 100° corresponds to a wide-angle lens. The mounting offsets between the camera and the INS are known. As environment the 3D model 'The City' created by Herminio Nieves was used. He published this [Nie15] and other models free for commercial and non commercial use.

The drift of the LSD-SLAM was analyzed by performing MC runs for two different directions at a straight flight. The error behavior shows a linear increase and is quite similar for both directions (Figure 5.8). At a travel distance of 100 m the drift is characterized by a mean value of $\mu = 6.03\,\mathrm{m}$ and a standard deviation of $\sigma = 3.23\,\mathrm{m}$ for one direction as well as $\mu = 5.37\,\mathrm{m}$ and $\sigma = 3.62\,\mathrm{m}$ for the other. These values are quite high compared to analyses stated in literature [LJHW14] and may be a consequence of the 3D world that is partly without texture (Figure 4.4). However, the large drift can be handled by increasing the upper bound of the angular error to 10°. This value corresponds to an error of 17.63 % in the travel distance, which covers for both evaluated directions more than three times the standard deviation added to the mean. The high value will result in some detours during the planning but guarantees convergence.

The evaluation of the proposed approach was realized with MC runs in Gazebo as follows. The initial path was traveled according to a list of waypoints which describe

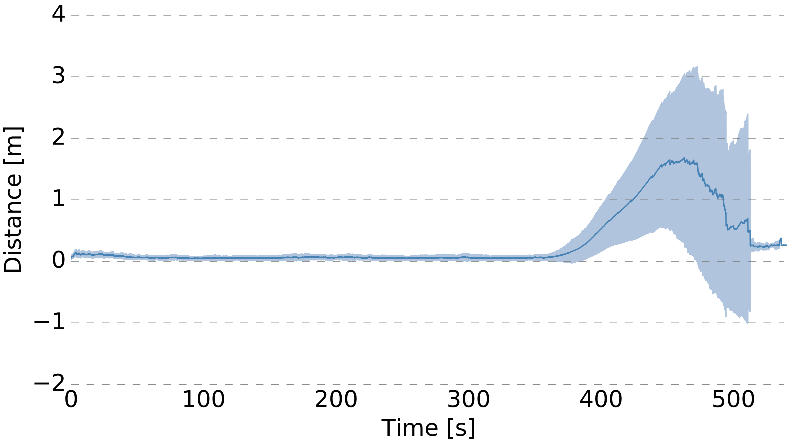


Figure 5.9: Mean and standard deviation of the Euclidean distance between the pose estimation and the ground truth from 100 MC runs performing an UAS homing in Gazebo. INS-based LSD-SLAM was used before the GPS outage around 350 s and the original LSD-SLAM afterwards.

the path depicted in Figure 5.5. This path as well as the map were created for each run from scratch and differ slightly due to the non-deterministic design of the LSD-SLAM. At the last waypoint, the GPS outage is simulated and the UAS uses the proposed algorithm to return autonomously to its take-off position. Thereby the direction of the UAS are updated every second based on the current pose estimate of the LSD-SLAM.

The Euclidean distance between the ground truth and the pose estimations shows the expected increase in uncertainty while traveling in unexplored areas and drops again when the LSD-SLAM creates loop closures by adding constraints between the latest and previously created keyframes (Figure 5.9). In the end of the investigated scenario a mean distance of about 0.25 m is obtained, which is roughly three times higher than the accuracy achieved with the INS-based LSD-SLAM used before the GPS outage. This value depends on the camera resolution as well as the ground sample distance and will differ in other setups. The time at which the runs successfully performed the loop closure differs slightly, which is apparent by the slowly decreasing mean value. The small standard deviations in the last seconds state that the relocalization worked in all runs.

The path information of the MC runs is visualized in Figure 5.10. It shows the change in the course direction at crossing the previous path the first time and how the drift in the visual odometry influences the flight course. After successfully performing a

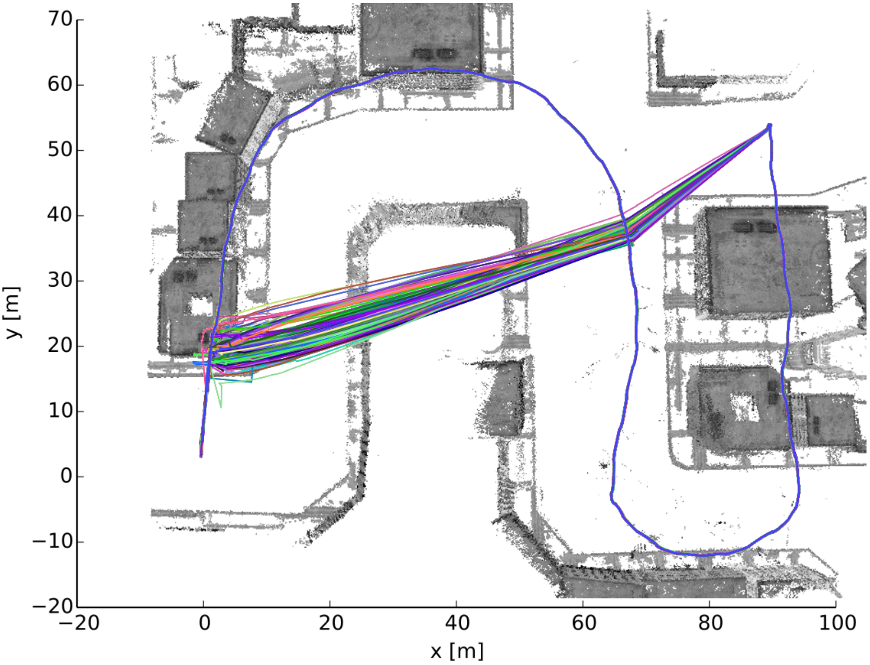


Figure 5.10: The path information of 100 MC runs for returning to the start position at $(0, 0)$ after a simulated GPS outage in Gazebo. The path information is plotted on top of the map created by a single run before the GPS outage.

loop closure to relocalize, the last meters are traversed for each run along the previous path to reach the start position.

Traveling according to waypoints till the GPS outage occurs covered a mean travel distance of 251.85 m. The direct connection back to the take-off position had a mean distance of 100.5 m and the path actually traveled by steering commands from the planning algorithm based on LSD-SLAM pose estimations lead to a mean distance of 108.53 m. This results in a travel distance ratio, according to equation (5.7), of 1.08 which is good, especially under consideration of the high maximal angular error of 10 %.

5.7 Real-world experiments

The same hardware setup as described in the real-world evaluation of the position free system calibration (Chapter 3.5.5) and the INS-based LSD-SLAM (Chapter 4.7) was used.

To estimate the upper bound of the positional drift, MC runs on two data sets recorded while performing straight flights were performed. The nondeterministic LSD-SLAM leads for multiple runs to differing pose estimations (Figure 5.11). The evaluation results in a mean value of $\mu = 8.04\,\mathrm{m}$ and a standard deviation of $\sigma = 4.62\,\mathrm{m}$ for one direction as well as $\mu = 6.4\,\mathrm{m}$ and $\sigma = 4.31\,\mathrm{m}$ for the other (Figure 5.12).

Although nearly the whole area observed by the camera contains gradients large enough to perform pixel-wise stereo matching by the LSD-SLAM, the drift is larger compared to the Gazebo simulation performed in the previous section. This may be the result of repetitive structures in the form of a lot of grass covering the area.

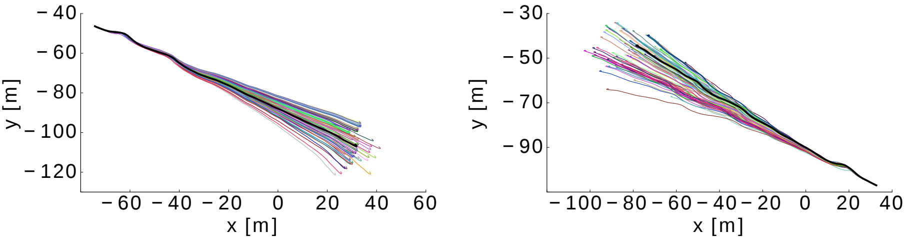


Figure 5.11: Path visualization of the real-world LSD-SLAM drift. The two plots show the INS path with an error in the range of a few centimeters (black) and MC runs performed on the same data set (colors).

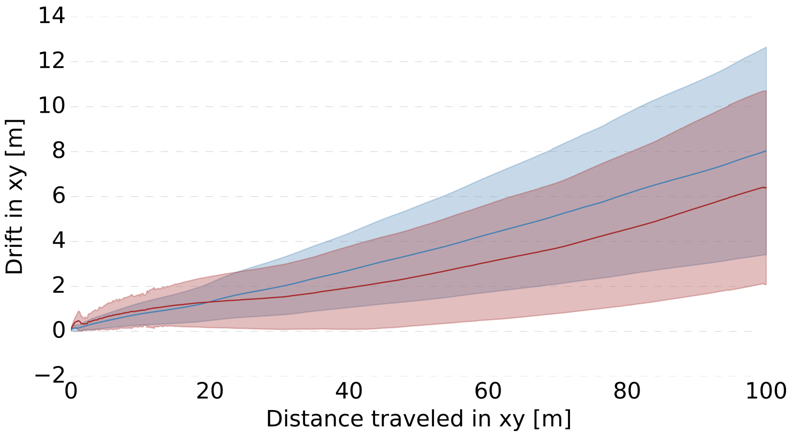


Figure 5.12: Mean and standard deviation of the camera only LSD-SLAM drift from MC runs for two straight level flights courses in the real-world. The INS path, with an error in the range of a few centimeters, is used as ground truth.

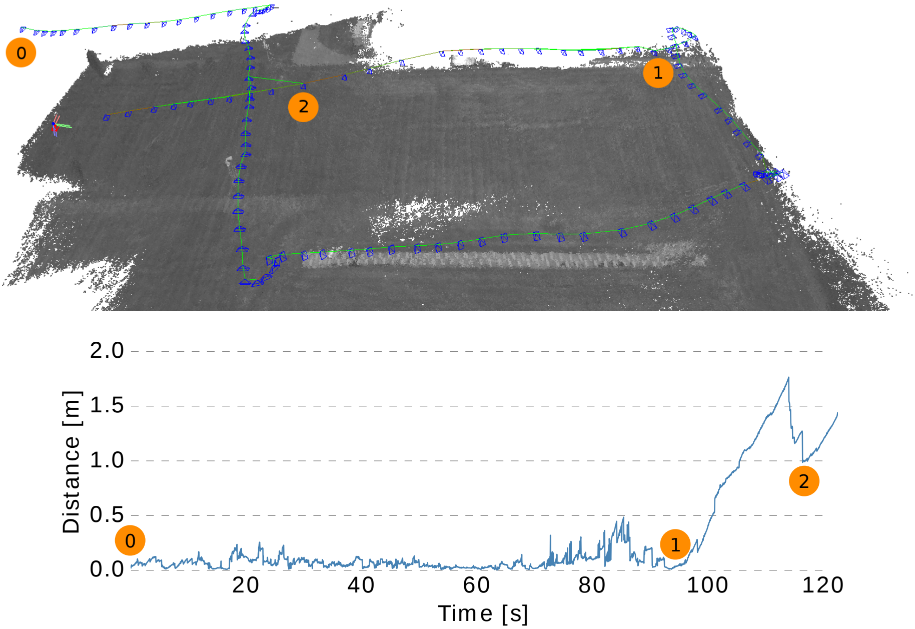


Figure 5.13: Real-world scenario. The copter starts at position ”0” with the INS-based LSD-SLAM and switches to the original LSD-SLAM at ”1”. A loop closure to a previously generated keyframe is performed at position ”2”. **Top:** Visualization of the LSD-SLAM results showing the created map, keyframes and image constraints. **Bottom** Distance between the pose estimation of the LSD-SLAM and the INS measurements.

Nevertheless, the estimation errors can be handled by setting the maximal drift accordingly and the only drawback is that the shortcut algorithm will plan larger detours to ensure reentering previously visited areas. The bigger problem is the number of false loop-closures, which happen because of the repetitive structures in the area. This was already observed in the evaluation of real-world experiments with the LSD-SLAM (Chapter 4.7). The previously used workaround to switch off the loop closures detection and rely on the accurate INS information does not work for the shortcut algorithm. The algorithm requires recognizing previously visited areas in the camera images to perform a relocalization and eliminate the drift. By increasing the strictness for loop closures most of the false loop closure, but at the same time a lot of real loop closures, are eliminated. The detection of previously visited areas did not work reliable in the outdoor area covered by the available flight permission. The flight visualized in Figure 5.13 shows that the detection of only one connection to the previous path will already reduce the error of the pose estimation. More connection would add additional constraints to the mapping part of the LSD-SLAM and decrease the error of the pose estimation even more. This is presented as proof of concept, although it was not possible to perform a successful homing with the presented shortcut algorithm in the test area.

5.8 Summary

An approach which performs a path planning for UAS using a previous build map and actual camera images was presented. The travel distance is minimized by exploiting safe shortcut through unexplored areas. Therefore, a very fast heuristic to perform a local path optimization which produces almost optimal results in nearly all situations is used.

After the proof of convergence, the approach was evaluated in extensive numerical studies. A realistic scenario was realized in the simulation framework Gazebo. The first part of the flights used the INS-based LSD-SLAM and a switch to the original LSD-SLAM was performed after a simulated GPS outtake. To return to the take-off position, the path planning was realized with the presented algorithm and resulted in a relocalization in the previously visited area close to the take-off position. Based on the simulations in Gazebo, real-world experiments were performed. Thereby, the concept was proofed by showing how the loop-closure to previous keyframes reduces the localization error.

CHAPTER VI

Summary and conclusions

"An idea which can be used only once is a trick. If one can use it more than once it becomes a method."

— George Polya and Gabor Szegö

This chapter summarizes the contributions of the thesis and gives an outlook on possible subsequent work and future navigation systems. The contributions can be divided into the methodology developed in the area of system calibration and the presented approach fusing INS and camera data for real-time navigation. Stating subsequent work shows possible developments resulting from the contributions of this thesis. The outlook for future navigation systems will be discussed in a more general context.

6.1 Contributions

This thesis made contributions in the context of multi-sensor navigation with UAS. The proposed approaches combine GPS-assisted INS, used since decades for navigational purposes, and cameras. Latter, or rather the images produces by them, contain lots of information. Nowadays, the increased computing power allow the evaluation of the image data for various task. In this thesis the focus is their usage for navigational purposes. Combining the advantages of these sensors, leads to the development of a powerful framework for the UAS pose estimation.

System calibration

In this part of the thesis the system calibration of a sensor suite consisting of an INS and a camera is investigated. The transformation between the coordinate systems of the sensors can be described as a rigid body motion with six degrees of freedom. Two approaches, which are capable for the calibration of these offsets with sensors of different accuracy levels, are realized. The first considers the calibration of a camera rigidly mounted to a high-precision INS. By formulating a graph based approach, the integration of the mounting offsets in the bundle adjustment is straight forward and allows the calibration of the system with data from a measurement flight. The second approach enables the calibration of systems using a low-cost INS, which especially suffer from bad accuracies in the positional component. The novel contributions are:

1. *Graphical formulation of a single step bundle adjustment.*
   Integrating the calibration parameters in the bundle adjustment is superior to two-step approaches, which first perform a classical bundle adjustment and optimize for the system parameters afterwards. By formulating the problem in a graph based structure, the integration of the system calibration parameters in the bundle adjustment are realized by a clear and descriptive approach.

2. *System calibration without GCPs.*
   Related work states that at least one GCP is needed for the calibration of the mounting offsets between a camera and an INS. Fixing the components of the lever-arm to their initial values extracted from construction drawing allows the calibration without any GCP. This was proved by real-world experiments performing the system calibration of a sensor suite integrated in an ultra-light airplane.

3. *Boresight calibration without position measurements.*
   Low-cost INS suffer from errors in the positional component in the range of a few meters. The presented approach performs the boresight calibration using only the rotational components of the INS.

Navigation

In the second part of this thesis, the semi-dense monocular LSD-SLAM framework is extended by exploiting measurements from an INS. Both sensors are rigidly mounted, which allows the calibration of the mounting offsets with one of the methods described in the first part of the thesis. The integration of INS measurements in the camera-based SLAM approach leads to several superior characteristics. The novel contributions are:

1. *Increased robustness and map quality for the LSD-SLAM.*
   In the evaluated outdoor scenarios, the integration of GPS-aided INS measurements into the LSD-SLAM approach compensates the scale-drift. This results in more accurate pose estimations and maps generated by the algorithm. As

a consequence the search region for possible loop closures gets smaller and the robustness of the overall system increases.

2. *Homing for UAS using shortcuts.*
   The maps created by the INS-based LSD-SLAM are very accurate, which allows using them as a reference for missions with no or disturbed GPS reception. In the case of a GPS failure the maps can be used to navigate the UAS to its start position. An approach which uses the knowledge of the maximal visual odometry drift to perform this homing procedure by taking into consideration shortcuts through unexplored areas was presented.

6.2 The way ahead

The contributions of this thesis show that the multi-sensor navigation has the capability to be the robust framework needed for the save operation of UAS in the controlled airspace. Self-localization of aerial platforms is a basic skill, which is not only needed for the successful navigation from one place to another, but also to avoid collisions by sending its own position to other platforms with a transponder.

Future work should investigate the proposed procedures in more detail. Besides a general evaluation, which aims to adapt the approaches to be suitable for GPS measurements without RTK corrections, each of the presented approaches has its own subsequent works.

The graph-based calibration shows the capability for cost-saving in-flight calibrations without any GCP. A rejection of images, based on a similarity threshold evaluating the INS poses, will accelerate the computations and enable a recalibration during missions. To make use of its full potential, the achievable accuracies and the influence of flight patterns have to be investigated in more detail.

The evaluation of the position free boresight calibration shows that using more data samples increases the robustness of the INS-camera calibration, but leads to a cubic growing processing time for the initial determination of the extrinsic orientations by a standard camera calibration procedure. As an alternative, an incremental SFM approach could be used to calculate the extrinsic orientation of all collected images and not only a subset, which should have a positive effect on the calibration results [Wu13].

The integration of the filtered INS poses in the LSD-SLAM shows a high potential. Instead of the filtered INS pose, it is possible to integrate the inertial measurements in the form of accelerations and angular rates in LSD-SLAM approach without the usage of GPS measurements. This would be a nice application for GPS denied or distorted environments if the produced navigation solutions show less drift compared to INS or camera only approaches and results additionally in detailed semi-dense maps of the environment. An alternative approach for low cost projects, would be the usage of a cheap GPS device to only use GPS measurements to cope with the drift of the LSD-SLAM.

The highly accurate maps created by the presented INS-based LSD-SLAM are the perfect foundation for a pure visual navigation. This was shown by describing a solution for the drone homing task in the case of a GPS failure. Nevertheless, the UAS homing would benefit from a global solution in multiple ways. This should cover the special cases encountered in the numerical studies. In some rare cases a better performance can be expected by allowing travel directions with the goal located outside of the uncertainty area. Furthermore, a global approach allows stating the worst and expected travel distance for returning to the starting position in advance. In the context of UAS homing, this allows the evaluation if returning to the start position is safe under consideration of the current battery level or the search for an alternative landing area is the better option. In the case of a complete INS failure, instead of the here considered GPS outage, the LSD-SLAM may also be considered for UAS flight control.

6.3 Future navigation systems

The usage of GPS or any other GNSS signals for the navigation is the key component in all outdoor applications up to the present day. However, the signal get blocked by any obstacles, or even water, in the direct line of sight to the satellites and as a consequence the approach is unusable for indoor, underground and undersea navigation. Nevertheless, in outdoor scenarios with a clear view of the sky the higher performance compared to all other currently known approaches emphasizes the superior suitability of the GPS for navigation purposes.

A drawback is the low power of the signal at the Earth's surface. The official U.S government information state that the GPS signals are transmitted with enough power to ensure that, after being affected by the free space path loss for a traveling distance of 20 200 km, at least a power level of $-158.5\,\mathrm{dBW}$ for the signal of the L1 band is left at the earth surface [Nat08]. This level is so low that the wanted signal gets lost in the noise and can only be extracted by a correlation exploiting the repetitive structure of the GPS code. Due to this low power level, the signal can become unavailable because of natural interference, a system glitch or intentional jamming.

In nearly all countries it is a violation of the law to use devices that block, jam or interfere with authorized radio communications like phone or GPS signals. Nevertheless, devices which are created for the interference with GPS signals are on the market and can be bought by anyone [Bra10]. According to reports in the international press, GPS malfunctions on a large scale occurred in regions with tense political situations. In the end of March 2016 North Korea started to jam from five regions in the country the GPS signals in South Korea. This violates international agreements and is a threat to civil vessels and airplanes [Nic16]. Another case in October 2016, social media entries from people in central Moscow stated that GPS relying phone apps stopped working or showed a location in the area of the Moscow's Vnukovo airport, which is 29 km away. The relocation to the airport leads to the assumption that the purpose is to prevent drones flying over the Kremlin. Most firmwares of commercial drones integrate no fly zones around airports in their firmware and start an emergency landing because of the relocation [Ass16].

The vulnerability of the GPS principal was known a long time before this serious incidents occurred. Consequently, alternative navigation techniques are an active research area. For example, the Defense Advanced Research Projects Agency (DARPA) of the U.S. Department of Defense runs multiple research programs with the goal to develop a navigation system with the accuracy level GPS achieves already today, but is resistant to jamming. To quote one, the DARPA proposal "Precise Robust Inertial Guidance for Munitions: Advanced Inertial Micro Sensors" has the objective to develop a miniature, high-performance and self-contained navigation system based on inertial sensors [DAR15]. In this context, a contract with a funding of 4.3 million dollar was awarded to the HRL Laboratories LLC, a research and development laboratory owned by The Boeing Company and General Motors. The goal of the research is to combine a MEMS Coriolis vibratory gyroscope and an atomically-stable frequency reference without introducing unintended noise. This would lead to a new class of inertial sensors with a highly improved performance at small weight, size and power requirements [HRL16].

A widely used and also in this thesis considered approach is the usage of images generated by cameras. The contained information content is high and as a result of increasing computing power and new algorithmic approaches the usability for navigational aspects constantly expand. However, the dependence on the weather is a big problem for the reliability, as clouds may block the view in the aerial context or the lack of sunlight is critical for daylight cameras. A possible workaround is the usage of different spectral bands, or even a multispectral setup.

Another promising approach is to integrate new sensor technologies into existing systems. One example is the depth sensor in Google's Tango platforms. Tango is designed to run on a phone or tablet and determine the pose of the device in real-time within an indoor environment. The combination of measurements from inertial sensors and a depth camera in sophisticated algorithms leads to accurate SLAM systems without using GPS or other external sources. The platform leads to applications in the areas of indoor navigation, augmented reality and even size measurements of physical objects [Edd15, DSM$^+$17].

Despite naming the drawbacks of the GPS and also the progress in alternative fields of research, the capabilities of GPS measurements are outstanding and should be used in current systems. In the context of SLAM approaches it leads to the most accurate maps in outdoor scenarios. In non critical mapping tasks it is also common to plan measuring flights at the time with the best possible GPS satellite constellation for the area of interest. The created maps can be used as a reference in the case of GPS failures for following surveys in the same region.

Nevertheless, the vulnerability and the need for the solutions to navigation in area without GPS reception will promote new technologies. The development of alternatives with the same or even a better accuracy compared to GPS will change navigation systems in the long-term perspective. Whatever the future holds in store for us, one thing is certain: The navigation systems of the future will integrate complementary sensors and fuse their information to receive the most accurate and robust solution in each situation.

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Tables

Chapter 2

Chapter 3

Chapter 5

Own Bibliography

[BCK16] Daniel Bender, Daniel Cremers, and Wolfgang Koch. A Position Free Boresight Calibration for INS-Camera Systems. In *Multisensor Fusion and Integration for Intelligent Systems (MFI), 2016 International Conference on*, pages 52–57, 2016.

[BCK17] Daniel Bender, Daniel Cremers, and Wolfgang Koch. Map-Based Drone Homing Using Shortcuts. In *Multisensor Fusion and Integration for Intelligent Systems (MFI), 2017 International Conference on*, pages 505–511, 2017.

[BRS$^+$16] Daniel Bender, Fahmi Rouatbi, Marek Schikora, Daniel Cremers, and Wolfgang Koch. Scaling the World of Monocular SLAM with INS-Measurements for UAS Navigation. In *Information Fusion (FUSION), 2016 19th International Conference on*, pages 1493–1500, 2016.

[BSK11] Daniel Bender, Marek Schikora, and Wolfgang Koch. UAS-Borne Multi-Sensor Surveillance for Military Sensing. In *8th NATO Military Sensing Symposium (SET-169)*, 2011.

[BSSC13] Daniel Bender, Marek Schikora, Jürgen Sturm, and Daniel Cremers. A Graph Based Bundle Adjustment for INS-Camera Calibration. (best paper award). *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, (XL-1/W2):39–44, 2013.

[BSSC14] Daniel Bender, Marek Schikora, Jürgen Sturm, and Daniel Cremers. INS-Camera Calibration Without Ground Control Points. In *Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pages 1–6, 2014.

[SBCK10] Marek Schikora, Daniel Bender, Daniel Cremers, and Wolfgang Koch. Passive Multi-Object Localization and Tracking Using Bearing Data. In *Information Fusion (FUSION), 2010 13th Conference on*, pages 1–7, 2010.

[SBK14] Marek Schikora, Daniel Bender, and Wolfgang Koch. Airborne Emitter Tracking by Fusing Heterogeneous Bearing Data. In *Information Fusion (FUSION), 2014 17th International Conference on*, pages 1–7, 2014.

[SBKC10] Marek Schikora, Daniel Bender, Wolfgang Koch, and Daniel Cremers. Multi-Target, Multi-Sensor Localization and Tracking Using Passive Antennas and Optical Sensors on UAVs. In *Security + Defence*, 2010.

Bibliography

[ACVT04] Leopoldo Armesto, Stefan Chroust, Markus Vincze, and Josep Tornero. Multi-rate Fusion with Vision and Inertial Sensors. In *Robotics and Automation (ICRA), 2004 IEEE International Conference on*, volume 1, pages 193–199, 2004.

[AHE01] Nicolas Andreff, Radu Horaud, and Bernard Espiau. Robot Hand-Eye Calibration Using Structure-from-Motion. *International Journal of Robotics Research*, 20(3):228–248, 2001.

[Ass16] Associated Press. Russians Seek Answers to Central Moscow GPS Anomaly. http://bigstory.ap.org/cc21e7bf77aa48aeba1d7f5faa4d2c4a, 2016. Online, accessed 3-October-2017.

[Bay76] Bryce E. Bayer. Color imaging array, 1976. US patent 3971065.

[Bra10] John Brandon. GPS Jammers Illegal, Dangerous, and Very Easy to Buy. http://www.foxnews.com/tech/2010/03/17/gps-jammers-easily-accessible-potentially-dangerous-risk.html, 2010. Online, accessed 3-October-2017.

[Bro66] Duane C. Brown. Decentering Distortion of Lenses. *Photometric Engineering*, 32(3):444–462, 1966.

[BWSS10] Michael Blösch, Stephan Weiss, Davide Scaramuzza, and Roland Siegwart. Vision Based MAV Navigation in Unknown and Unstructured Environments. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 21–28, 2010.

[Cha88] Nicole Chapuis. *Les opérations structurantes dans la connaissance de l'espace chez les mammifères: détour, raccourci et retour*. PhD thesis, Université Aix-Marseille 2, 1988.

[CKM08] Robert Castle, Georg Klein, and David W. Murray. Video-Rate Localization in Multiple Maps for Wearable Augmented Reality. In *Wearable Computers, 2008 12th IEEE International Symposium on*, pages 15–22, 2008.

[CLD07] Peter Corke, Jorge Lobo, and Jorge Dias. An Introduction to Inertial and Visual Sensing. *The International Journal of Robotics Research*, 26(6):519–535, 2007.

[CN08] Mark Cummins and Paul Newman. FAB-MAP: Probabilistic localization and mapping in the space of appearance. *The International Journal of Robotics Research*, 27(6):647–665, 2008.

[CSH00] Michael Cramer, Dirk Stallmann, and Norbert Haala. Direct Georeferencing Using GPS/Inertial Exterior Orientations for Photogrammetric Applications. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 33(B3/1; PART 3):198–205, 2000.

[DAR15] DARPA. Precise Robust Inertial Guidance for Munitions (PRIGM): Advanced Inertial Micro Sensors (AIMS), 2015.

[Dra02] Samuel Picton Drake. Converting GPS Coordinates [phi, lambda, h] to Navigation Coordinates (ENU), 2002.

[DSM$^+$17] Maksym Dzitsiuk, Jürgen Sturm, Robert Maier, Lingni Ma, and Daniel Cremers. De-noising, stabilizing and completing 3d reconstructions on-the-go using plane priors. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pages 3976–3983, 2017.

[Edd15] Max Eddy. Google: Future Phones Will Understand, See the World. http://uk.pcmag.com/smartphones/42448/news/google-future-phones-will-understand-see-the-world, 2015. Online, accessed 3-October-2017.

[Eis08] Henri Eisenbeiss. The Autonomous Mini Helicopter: A Powerful Platform for Mobile Mapping. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 37 Part B1:977–983, 2008.

[ESC12] Jakob Engel, Jürgen Sturm, and Daniel Cremers. Camera-Based Navigation of a Low-Cost Quadrocopter. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 2815–2821, 2012.

[ESC13] Jakob Engel, Jurgen Sturm, and Daniel Cremers. Semi-Dense Visual Odometry for a Monocular Camera. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1449–1456, 2013.

[ESC14] Jakob Engel, Thomas Schöps, and Daniel Cremers. LSD-SLAM: Large-Scale Direct Monocular SLAM. In *Computer Vision (ECCV), 2014 European Conference on*, pages 834–849. Springer, 2014.

[FB81] Martin A. Fischler and Robert C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[GGC$^+$16] Alessandro Giusti, Jérôme Guzzi, Dan C. Cireşan, Fang-Lin He, Juan P. Rodriguez, Flavio Fontana, Matthias Faessler, Christian Forster, Jürgen Schmidhuber, Gianni Di Caro, Davide Scaramuzza, and Luca M. Gambardella. A Machine Learning Approach to Visual Perception of Forest Trails for Mobile Robots. *IEEE Robotics and Automation Letters*, 1(2):661–667, 2016.

[GWA07] Mohinder S. Grewal, Lawrence R. Weill, and Angus P. Andrews. *Global Positioning Systems, Inertial Navigation, and Integration*. John Wiley & Sons, 2007.

[Hal15] Brian C. Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. Springer, 2015.

[HD95] Radu Horaud and Fadi Dornaika. Hand-Eye Calibration. *The International Journal of Robotics Research*, 14(3):195–210, 1995.

[HJW02] Christian Heipke, Karsten Jacobsen, and Helge Wegmann. Analysis of the Results of the OEEPE Test 'Integrated Sensor Orientation'. *Integrated sensor orientation - Test report and workshop proceedings*, 43:31–49, 2002.

[HRL16] HRL Laboratories. HRL to Receive $4.3 Million DARPA Award to Develop Next-Generation Inertial Sensor Technology. https://www.hrl.com/news/2016/0311/, 2016. Online, accessed 17-November-2016.

[Huy09] Du Q. Huynh. Metrics for 3D Rotations: Comparison and Analysis. *Journal of Mathematical Imaging and Vision*, 35(2):155–164, 2009.

[HZ04] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[iMA13] iMAR. iTraceRT-F200-E: technical data sheet. http://www.imar-navigation.de/downloads/TraceRT-F200-E_en.pdf, 2013.

[Jac01] Karsten Jacobsen. Aspects of Handling Image Orientation by Direct Sensor Orientation. In *Proceedings of the ASPRS Annual Convention*, 2001.

[JS11] Michal Jama and Dale Schinstock. Parallel Tracking and Mapping for Controlling VTOL Airframe. *Journal of Control Science and Engineering*, 2011:26, 2011.

[KD03] Georg Klein and Tom Drummond. Robust Visual Tracking for Non-Instrumental Augmented Reality. In *Mixed and Augmented Reality (ISMAR), 2003 2nd IEEE and ACM International Symposium on*, pages 113–122, 2003.

[KFN09] Farid Kendoul, Isabelle Fantoni, and Kenzo Nonami. Optic Flow-Based Vision System for Autonomous 3D Localization and Control of Small Aerial Vehicles. *Robotics and Autonomous Systems*, 57(6):591–602, 2009.

[KGS$^+$11] Rainer Kümmerle, Giorgio Grisetti, Hauke Strasdat, Kurt Konolige, and Wolfram Burgard. g2o: A General Framework for Graph Optimization. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 3607–3613, 2011.

[KH04] Nathan Koenig and Andrew Howard. Design and Use Paradigms for Gazebo, An Open-Source Multi-Robot Simulator. In *Intelligent Robots and Systems (IROS), 2004 IEEE/RSJ International Conference on*, volume 3, pages 2149–2154, 2004.

[KH05] Elliott Kaplan and Christopher Hegarty. *Understanding GPS: Principles and Applications.* Artech house, 2nd edition, 2005.

[KHB11] Ana Paula Kersting, Ayman Habib, and Ki In Bang. Mounting Parameters Calibration of GPS/INS-Assisted Photogrammetric Systems. In *Multi-Platform/Multi-Sensor Remote Sensing and Mapping (M2RSM), 2011 International Workshop on*, pages 1–6, 2011.

[KM07] Georg Klein and David Murray. Parallel Tracking and Mapping for Small AR Workspaces. In *Mixed and Augmented Reality (ISMAR), 2007 6th IEEE and ACM International Symposium on*, pages 225–234, 2007.

[KM09] Georg Klein and David Murray. Parallel Tracking and Mapping on a Camera Phone. In *Mixed and Augmented Reality (ISMAR), 2009 8th IEEE International Symposium on*, pages 83–86, 2009.

[Koc14] Wolfgang Koch. *Tracking and Sensor Data Fusion: Methodological Framework and Selected Applications.* Springer Science & Business Media, 2014.

[KSD$^+$11] Rainer Kümmerle, Bastian Steder, Christian Dornhege, Alexander Kleiner, Giorgio Grisetti, and Wolfram Burgard. Large Scale Graph-Based SLAM using Aerial Images as Prior Information. *Autonomous Robots*, 30(1):25–39, 2011.

[KSS08] Jonathan Kelly, Srikanth Saripalli, and Gaurav Sukhatme. Combined Visual and Inertial Navigation for an Unmanned Aerial Vehicle. In *Field and Service Robotics*, pages 255–264, 2008.

[LA09] Manolis I. A. Lourakis and Antonis A. Argyros. SBA: A Software Package for Generic Sparse Bundle Adjustment. *ACM Transactions on Mathematical Software*, 36(1):1–30, 2009.

[LD07] Jorge Lobo and Jorge Dias. Relative Pose Calibration Between Visual and Inertial Sensors. *International Journal of Robotics Research*, 26(6):561–575, 2007.

[LJHW14] Hongtao Liu, Ruyi Jiang, Weng Hu, and Shigang Wang. Navigational Drift Analysis for Visual Odometry. *Computing & Informatics*, 33(3):685–706, 2014.

[Low04] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[MF03] Jean-Arcady Meyer and David Filliat. Map-based navigation in mobile robots:: II. A review of map-learning and path-planning strategies. *Cognitive Systems Research*, 4(4):283–317, 2003.

[MR08] Faraz M. Mirzaei and Stergios I. Roumeliotis. A Kalman Filter-Based Algorithm for IMU-Camera Calibration: Observability Analysis and Performance evaluation. *IEEE transactions on robotics*, 24(5):1143–1156, 2008.

[MSK$^+$12] Johannes Meyer, Alexander Sendobry, Stefan Kohlbrecher, Uwe Klingauf, and Oskar von Stryk. Comprehensive simulation of Quadrotor UAVs Using ROS and Gazebo. In *Simulation, Modeling, and Programming for Autonomous Robots*, pages 400–411. Springer, 2012.

[MSKS03] Yi Ma, Stefano Soatto, Jana Kosecka, and S. Shankar Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer, 2003.

[Nat08] National Coordination Office for Space-Based Positioning, Navigation, and Timing. Global Positioning System Standard Positioning Service Performance Standard. http://www.gps.gov/technical/ps/2008-SPS-performance-standard.pdf, 2008.

[Nel91] Randal C. Nelson. Visual homing using an associative memory. *Biological Cybernetics*, 65(4):281–291, 1991.

[Nic16] Michelle Nichols. South Korea Tells U.N. That North Korea GPS Jamming Threatens Boats, Planes. http://www.reuters.com/article/us-northkorea-southkorea-gps-idUSKCN0X81SN, 2016. Online; accessed 3-October-2017.

[Nie15] Herminio Nieves. The City: 3D Model. http://sharecg.com/v/79711/gallery/5/3D-Model/The-City, 2015. Online; accessed 3-October-2017.

[NIH$^+$11] Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. KinectFusion: Real-Time Dense Surface Mapping and Tracking. In *Mixed and Augmented Reality (ISMAR), 2011 IEEE International Symposium on*, pages 127–136, 2011.

[NKEES09] Aboelmagd Noureldin, Tashfeen B. Karamat, Mark D. Eberts, and Ahmed El-Shafie. Performance Enhancement of MEMS-Based INS/GPS

Integration for Low-Cost Navigation Applications. *Vehicular Technology, IEEE Transactions on*, 58(3):1077–1096, 2009.

[PF02] Livio Pinto and Gianfranco Forlani. A Single Step Calibration Procedure for IMU/GPS in Aerial Photogrammetry. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 34 Part B3:210–219, 2002.

[PG09] Uwe Proske and Simon C. Gandevia. The kinaesthetic senses. *The Journal of physiology*, (17):4139–4146, 2009.

[Pom90] Dean A. Pomerleau. Neural network based autonomous navigation. In *Vision and Navigation*, pages 83–93. Springer, 1990.

[QCZ01] Gang Qian, Rama Chellappa, and Qinfen Zheng. Robust Structure from Motion Estimation Using Inertial Data. *Journal of the Optical Society of America A*, 18(12):2982–2997, 2001.

[SBG15] SBG Systems. Ellipse Series: Miniature High Performance Inertial Sensors: technical data sheet. https://www.sbg-systems.com/docs/Ellipse_Series_Leaflet.pdf, 2015. Online; accessed 3-October-2017.

[SCCG93] Klaus-Peter Schwarz, M. A. Chapman, M. W. Cannon, and P. Gong. An integrated INS/GPS approach to the georeferencing of remotely sensed data. *Photogrammetric engineering and remote sensing*, 59(11):1667–1674, 1993.

[SMRN10] Gabe Sibley, Christopher Mei, Ian Reid, and Paul Newman. Vast-scale outdoor navigation using adaptive relative bundle adjustment. *The International Journal of Robotics Research*, 29(8):958–980, 2010.

[TL89] Roger Y. Tsai and Reimar K. Lenz. A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration. *IEEE Transactions on robotics and automation*, 5(3):345–358, 1989.

[TMHF00] Bill Triggs, Philip McLauchlan, Richard Hartley, and Andrew Fitzgibbon. Bundle adjustment—a modern synthesis. *Vision algorithms: theory and practice*, pages 153–177, 2000.

[TWBM97] Olivier Trullier, Sidney I. Wiener, Alain Berthoz, and Jean-Arcady Meyer. Biologically based artificial navigation systems: Review and prospects. *Progress in neurobiology*, 51(5):483–544, 1997.

[VACP11] Rafael Valencia, Juan Andrade-Cetto, and Josep M. Porta. Path planning in belief space with pose SLAM. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 78–83, 2011.

[WAL$^+$12] Stephan Weiss, Markus W. Achtelik, Simon Lynen, Margarita Chli, and Roland Siegwart. Real-Time Onboard Visual-Inertial State Estimation and Self-Calibration of MAVs in Unknown Environments. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 957–964, 2012.

[WGL⁺08] Jinling Wang, Matthew Garratt, Andrew Lambert, Jack Jianguo Wang, S. Han, and David Sinclair. Integration of GPS/INS/Vision Sensors to Navigate Unmanned Aerial Vehicles. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 37:963–970, 2008.

[WMG⁺12] Andreas Wendel, Michael Maurer, Gottfried Graber, Thomas Pock, and Horst Bischof. Dense Reconstruction On-the-Fly. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1450–1457, 2012.

[Woo07] Oliver J. Woodman. An Introduction to Inertial Navigation. *University of Cambridge, Computer Laboratory, Tech. Rep. UCAMCL-TR-696*, 14:15, 2007.

[Wu13] Changchang Wu. Towards Linear-Time Incremental Structure from Motion. In *3D Vision (3DV), 2013 International Conference on*, pages 127–134, 2013.

[Zha00] Zhengyou Zhang. A Flexible New Technique for Camera Calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1330–1334, 2000.