Multi-environment Georeferencing of RGB-D Panoramic Images from Portable Mobile Mapping – a Perspective for Infrastructure Management

Zur Erlangung des akademischen Grades eines

DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)

von der KIT Fakultät für

Bauingenieur-, Geo- und Umweltwissenschaften

des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

MSc. Stefan Blaser

aus Solothurn

Tag der mündlichen Prüfung:

12. Juli 2022

Referent: Prof. Dr.-Ing. Boris Jutzi

Korreferent: Prof. Dr. Stephan Nebiker

Karlsruhe 2022

Abstract

Accurately georeferenced Red, Green, Blue and Depth (RGB-D) imagery with high resolution form the basis for 3D street view web-services, which are already widely used for infrastructure management. Mobile Mapping (MM) and 3D reality capturing techniques enable fast and efficient data acquisition of infrastructures. Most outdoor Mobile Mapping Systems (MMSs) rely on direct georeferencing, enabling absolute accuracies in the cm range in open areas. However, under poor Global Navigation Satellite System (GNSS) reception, the accuracy of direct georeferencing degrades rapidly to the dm- or even to the m-range. By contrast indoor MMS predominately rely on Simultaneous Localization and Mapping (SLAM). As, most SLAM algorithms focus on low-latency and real-time performance they accept trade-offs in accuracy, map quality and maximum extension.

This thesis aims to capture and accurately and reliably georeference high-resolution RGB-D imagery in different environments.

For data acquisition, we developed a high-performance image-focused backpack MMS, featuring a multi-head panoramic camera, two multi-beam Light Detection and Ranging (LiDAR) scanners and a navigation unit combining GNSS with an Inertial Measurement Unit (IMU) with tactical grade performance. All sensors are precisely synchronized and provide access to the raw data. Furthermore, we conducted an overall system calibration in test fields, using bundle-adjustment-based as well as feature-based calibration methods as a prerequisite for kinematic sensor data integration.

For accurate and reliable georeferencing in various environments, we propose a multi-stage georeferencing approach that incorporates various sensor data and georeferencing methods. Direct and LiDAR SLAM-based georeferencing provide initial poses for subsequent image-based georeferencing, using an extended Structure-from-Motion (SfM) pipeline. Image-based georeferencing results in a precise and sparse trajectory, which is suitable for the georeferencing of images. To obtain a dense trajectory, which is suitable for the georeferencing of both panoramic images and LiDAR points, we propose direct georeferencing with pose support from image-based georeferencing.

Comprehensive performance investigations in three demanding large-scale test sites demonstrate the performance and the limitations of our georeferencing approach with the different georeferencing methods. The three test sites in a city center, in a forest and in a public building represent real-world conditions with restricted GNSS reception, poor illumination, moving objects and repetitive geometric patterns.

Image-based georeferencing showed a superior performance, whereby the mean precision ranged from 5 mm to 7 mm between the different test fields and the absolute accuracy ranged from 85 mm to 131 mm. This is an improvement by factor 2–7 compared to direct and LiDAR SLAM-based georeferencing. Direct georeferencing, supported by Coordinate Updates (CUPTs) from image-based georeferencing, resulted in slightly degraded mean precision compared to image-based georeferencing in the range from 13 mm to 16 mm, whereby the mean absolute as well as the relative accuracy did not significantly differ from image-based georeferencing.

Our results within challenging environments confirm previous investigations under optimal conditions as well as work from other groups. They can be used for creating RGB-D image-based 3D geospatial web-services of infrastructures within challenging environments. Moreover, accurately georeferenced RGB-D imagery has great potential for future visual localization and Augmented Reality (AR) applications.

Kurzfassung

Hochaufgelöste, genau georeferenzierte RGB-D-Bilder sind die Grundlage für 3D-Bildräume bzw. 3D Street-View-Webdienste, welche bereits kommerziell für das Infrastrukturmanagement eingesetzt werden. MMS ermöglichen eine schnelle und effiziente Datenerfassung von Infrastrukturen. Die meisten im Aussenraum eingesetzten MMS beruhen auf direkter Georeferenzierung. Diese ermöglicht in offenen Bereichen absolute Genauigkeiten im Zentimeterbereich. Bei GNSS-Abschattung fällt die Genauigkeit der direkten Georeferenzierung jedoch schnell in den Dezimeter- oder sogar in den Meterbereich. In Innenräumen eingesetzte MMS basieren hingegen meist auf SLAM. Die meisten SLAM-Algorithmen wurden jedoch für niedrige Latenzzeiten und für Echtzeitleistung optimiert und nehmen daher Abstriche bei der Genauigkeit, der Kartenqualität und der maximalen Ausdehnung in Kauf.
Das Ziel dieser Arbeit ist, hochaufgelöste RGB-D-Bilder in verschiedenen Umgebungen zu erfassen und diese genau und zuverlässig zu georeferenzieren.
Für die Datenerfassung wurde ein leistungsstarkes, bildfokussiertes und rucksackgetragenes MMS entwickelt. Dieses besteht aus einer Mehrkopf-Panoramakamera, zwei Multi-Beam LiDAR-Scannern und einer GNSS- und IMU-kombinierten Navigationseinheit der taktischen Leistungsklasse. Alle Sensoren sind präzise synchronisiert und ermöglichen Zugriff auf die Rohdaten. Das Gesamtsystem wurde in Testfeldern mit bündelblockbasierten sowie merkmalsbasierten Methoden kalibriert, was eine Voraussetzung für die Integration kinematischer Sensordaten darstellt.
Für eine genaue und zuverlässige Georeferenzierung in verschiedenen Umgebungen wurde ein mehrstufiger Georeferenzierungsansatz entwickelt, welcher verschiedene Sensordaten und Georeferenzierungsmethoden vereint. Direkte und LiDAR SLAM-basierte Georeferenzierung liefern Initialposen für die nachträgliche bildbasierte Georeferenzierung mittels erweiterter SfM-Pipeline. Die bildbasierte Georeferenzierung führt zu einer präzisen aber spärlichen Trajektorie, welche sich für die Georeferenzierung von Bildern eignet. Um eine dichte Trajektorie zu erhalten, die sich auch für die Georeferenzierung von LiDAR-Daten eignet, wurde die direkte Georeferenzierung mit Posen der bildbasierten Georeferenzierung gestützt.
Umfassende Leistungsuntersuchungen in drei weiträumigen anspruchsvollen Testgebieten zeigen die Möglichkeiten und Grenzen unseres Georeferenzierungsansatzes. Die drei Testgebiete im Stadtzentrum, im Wald und im Gebäude repräsentieren reale Bedingungen mit eingeschränktem GNSS-Empfang, schlechter Beleuchtung, sich bewegenden Objekten und sich wiederholenden geometrischen Mustern.
Die bildbasierte Georeferenzierung erzielte die besten Genauigkeiten, wobei die mittlere Präzision im Bereich von 5 mm bis 7 mm lag. Die absolute Genauigkeit betrug 85 mm bis 131 mm, was einer Verbesserung um Faktor 2 bis 7 gegenüber der direkten und LiDAR SLAM-basierten Georeferenzierung entspricht. Die direkte Georeferenzierung mit CUPT-Stützung von Bildposen der bildbasierten Georeferenzierung, führte zu einer leicht verschlechterten mittleren Präzision im Bereich von 13 mm bis 16 mm, wobei sich die mittlere absolute Genauigkeit nicht signifikant von der bildbasierten Georeferenzierung unterschied.
Die in herausfordernden Umgebungen erzielten Genauigkeiten bestätigen frühere Untersuchungen unter optimalen Bedingungen und liegen in derselben Grössenordnung wie die Resultate anderer Forschungsgruppen. Sie können für die Erstellung von Street-View-Services in herausfordernden Umgebungen für das Infrastrukturmanagement verwendet werden. Genau und zuverlässig georeferenzierte RGB-D-Bilder haben ein grosses Potenzial für zukünftige visuelle Lokalisierungs- und AR-Anwendungen.

Contents

Acronyms

ADC Analog to Digital Converter. 34

AR Augmented Reality. i, iii, 22, 33, 103

ATTUPT Attitude Update. 5, 23, 62–64, 72, 79, 80, 82–86, 88, 91, 96, 97, 100, 102, VI

BA Boresight Alignment. 19, 47, 50–52, 54, 59, 60, 99, 101, VI

BIM Building Information Modeling. 1, 2, 99

CCD Charge-Coupled Device. 12, 34

CMOS Complementary Metal Oxide Semiconductor. 12

CP Check Point. 5, 60, 61, 78, 95–97, 100, 102, 103

CPU Central Processing Unit. 95

CUPT Coordinate Update. i, iii, 5, 23, 62–64, 72, 79–88, 91, 96–98, 100–102

DoF Degree of Freedom. 4, 18, 60, 95, 96

EO Exterior Orientation. 21, 25, 50, 51, 53, 60, 61, 78, 100

FOG Fiber Optic Gyroscope. 12, 72

FoV Field of View. 36

GCP Ground Control Point. 3, 4, 17, 23, 25–28, 60, 61, 78, 85, 87, 96, 100, 103

GigE Gigabit Ethernet. 38

GIS Geographic Information System. 2, 7, 8, 99

GLONASS Global Navigation Satellite System. 72

GNSS Global Navigation Satellite System. i, iii, 1, 3, 4, 9–12, 14–20, 22, 23, 26, 28, 32, 33, 36–38, 44–46, 49, 50, 52, 55–60, 62, 63, 66–73, 79, 80, 84, 86, 91, 93–101

GPIO General Purpose Input and Output. 34, 44

GPS Global Positioning System. 72

ICP Iterative Closest Point. 23

IMU Inertial Measurement Unit. i, iii, 1–4, 11, 12, 14–18, 22, 23, 26, 32, 33, 36–38, 44–46, 52, 58–60, 66–72, 75, 79, 84, 93, 94, 98

INS Inertial Navigation System. 4, 9–11, 17, 19, 20, 22, 23, 28, 36, 37, 47, 50, 51, 56–59, 62, 99, 101

IO Interior Orientation. 4, 32, 34, 47, 48, 50, 51, 53, 61, 78, 99, 101

LiDAR Light Detection and Ranging. i, iii, 2–5, 7–10, 13–19, 21–24, 26, 28, 30, 32, 33, 35, 36, 38, 39, 41–43, 45–48, 50–53, 55, 56, 58–61, 66–69, 71, 72, 75, 76, 78, 83–87, 91, 93–102

LOAM LiDAR Odometry and Mapping. 24

MEMS Micro-Electro-Mechanical System. 11–13, 16, 17, 37, 70, 72

MM Mobile Mapping. i, 1, 5, 7–13, 20, 22, 29–31, 38, 41, 60, 70, 71, 99, 102, II

MMS Mobile Mapping System. i, iii, 1–7, 9, 10, 13–22, 29, 30, 32, 33, 35–38, 40–45, 47, 49, 52, 54, 55, 57, 59–62, 65, 69–72, 93, 94, 96–99, 101–103, 107

MOEMS Micro-Opto-Electro-Mechanical System. 11

NMEA National Marine Electronics Association. 36, 45

PPK Post-Processed Kinematic. 23

PPP Precise Point Positioning. 23

PPS Pulse-Per-Second. 36, 45

RAM Random Access Memory. 37, 38, 95

RGB-D Red, Green, Blue and Depth. i, iii, 2, 3, 5, 7–9, 26, 29–32, 93, 94, 99, 101–103

RLG Ring Laser Gyroscope. 12

RO Relative Orientation. 3, 4, 21, 25, 47, 48, 50–54, 59–61, 99–101, VI

ROS Robot Operating System. 4, 41, 45, 46, 99

RTK Real-Time Kinematic. 10, 11, 23, 37

SfM Structure-from-Motion. i, iii, 3–5, 25–28, 32, 56, 60, 61, 78, 96, 100

SIFT Scale Invariant Feature Transform. 25

SLAM Simultaneous Localization and Mapping. i, iii, 2–5, 9, 13, 16–18, 24–28, 32, 39, 41, 43, 46, 55, 56, 58–61, 67–69, 71, 72, 75–78, 83–87, 91, 94–98, 100–102

SPAN Synchronized Position Attitude Navigation. 57, 58, 60

SSD Solid State Disk. 37

TLS Terrestrial Laser Scanner. 35

ToF Time of Flight. 13

UAV Unmanned Aerial Vehicle. 12

USB Universal Serial Bus. 34, 38, 44, 45

UTC Coordinated Universal Time. 45

VR Virtual Reality. 33

VRS Virtual Reference Station. 57

ZUPT Zero Velocity Update. 23

Chapter 1

Introduction

1.1 Image-focused 3D Mobile Mapping for Infrastructure Management

Ongoing digitization trends lead to rapid transformations in infrastructure management, which includes both the private and the public sector. The often-mentioned ambition towards the smart city incorporates novel and more integral concepts in urban planning and in infrastructure management with respect to the entire life cycle, from construction to maintenance to deconstruction. In the construction and real estate industry, Building Information Modeling (BIM) is leading to a paradigm shift that affects and includes planning, construction and operating processes. Such novel processes and paradigms require accurate, detailed and up-to-date 3D geodata as well as intuitive and user-friendly software tools, that enable collaborative work.

MM and 3D reality capturing techniques enable fast and efficient 3D data acquisition from buildings and infrastructures. First experimental stereo image-based outdoor MMS date back to the early 1990ies (Novak, 1991; Schwarz et al., 1993). With the aim of capturing urban environments as comprehensively as possible, image-based MMS have evolved into systems with (multi-) panorama configurations (Meilland et al., 2015). Blaser et al. (2017) present a MMS configuration featuring two tilted panoramic cameras resulting in multiple stereo systems that are directed to the sides and enable the capturing of entire façades of buildings in urban canyons. Such (multi-) stereo image-based MMS hold an accuracy potential for relative measurements (e.g. 3D distances or 3D areas) within the same stereo image ranging from 1 cm to 2 cm and provide an absolute accuracy potential in the cm-range under good conditions (Burkhard et al., 2012; Blaser et al., 2018b).

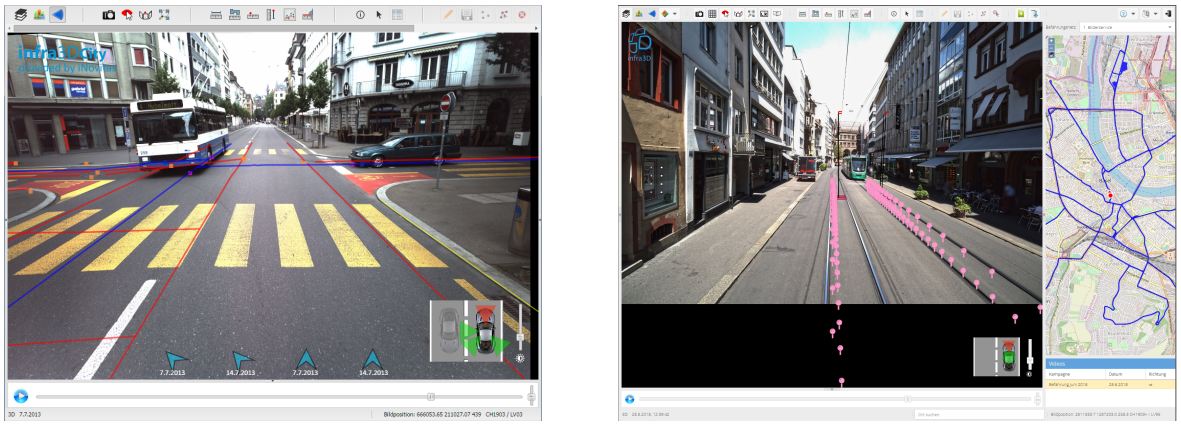
Currently, the majority of outdoor MMS rely on direct georeferencing combining GNSS and IMU data. However, direct georeferencing rapidly degrades under poor GNSS reception from the cm to the dm or even to the m range. Consequently, challenging environments with poor GNSS reception, such as indoor environments, forests or urban canyons, require alternative georeferencing approaches.

In recent years, various smaller and portable MMS for both indoor and outdoor mapping have appeared on the market. They can be divided according to the platform type into trolley-based MMS, such as the *NavVis M6* (NavVis, 2020), backpack-based or personal MMS, such as the *Leica Pegasus Backpack* (Leica Geosystems AG, 2022b), and into hand-held MMS, such as the *ZEB Revo* (Geoslam, 2022). Trolley-based platforms are suitable for mapping of large public buildings e.g. airports or museums, but they fail in obstacle-rich environments e.g. construction sites or staircases. Backpack platforms represent a good compromise in terms

of flexibility and size, which allows the use of high-end sensor technology in obstacle-rich environments, whereas the small form factor of hand-held MMS limits the sensor configuration.

Recent indoor MMS use SLAM for indoor localization and mapping, which originates from the robotics and computer vision community (Stachniss et al., 2016; Thrun, 2002). Lehtola et al. (2017) provide a comparison of numerous state-of-the-art LiDAR-focused indoor MMS based on point clouds.

Cloud-based web services offer the direct using of such 3D data from the web browser and provide interoperability with Geographic Information System (GIS) or BIM databases. Paparoditis et al. (2012) propose image-based street navigation and 3D plotting services for professional users, incorporating both image and LiDAR data. By contrast, Nebiker et al. (2015) introduce cloud-based geospatial 3D image spaces as an urban model for smart city and infrastructure management applications, which represent collections of georeferenced RGB-D imagery (see Figure 1.1).



(a) Street environment with superimposed mains from GIS data.

(b) Tramway environment with visualized track parameters and cable heights.

Figure 1.1: Cloud-based web client for infrastructure management, based on 3D geospatial image spaces, which consist of large RGB-D image data sets. (Nebiker et al., 2015).

By general, LiDAR point clouds are characterized by high precision, which is independent of the object texture and which deteriorates only slightly with increasing measurement distance. However, compared to stereo images, they are limited in density. As a result, corners and edges are not clearly mapped. In addition, images or georeferenced RGB-D images respectively are easier to interpret by humans, while 3D point cloud handling requires expert knowledge.

Since most of the commercially available portable or indoor MMS are LiDAR-focused, they use cameras primarily for the coloring of LiDAR point clouds. Their camera systems often fail to meet the requirements for image-based 3D web services in terms of their configuration or their image quality and size. In addition, such MMS are rather closed and allow neither extensions of the existing sensor configuration nor investigations and improvements of the provided standard georeferencing pipelines. Consequently, in-depth investigations on the georeferencing require the development of a custom MMS in order to access raw sensor data and to customize the configuration to the needs of georeferencing and subsequent mapping.

Tang et al. (2015) conducted performance investigations in forests, using direct georeferencing as well as SLAM and IMU-combined georeferencing with a self-developed MMS on an all-terrain-vehicle. Direct georeferencing showed better accuracy in open forest which was in the dm-range, while SLAM-based georeferencing showed large drifts in the m-range. By contrast, in mature forests, SLAM-based georeferencing showed an accuracy improvement by 38 % compared to direct georeferencing.

Our very first investigations on bundle-adjustment-based georeferencing with an image-based backpack MMS in a forest area failed due to a too large image capturing interval above 2 m and the lack of robust alternative georeferencing methods (Wittmer, 2017). By contrast, Cavegn et al. (2018) successfully bridged GNSS outages of direct georeferencing in urban canyons as well as of SLAM-supported georeferencing in indoor environments with subsequent image-based georeferencing using an extended SfM-pipeline by Ground Control Point (GCP) support and fixed Relative Orientation (RO) constraints of the camera rig. In both environments, their resulting accuracy ranged in the cm-domain.

1.2 Robust and Accurate Georeferencing in Challenging Environments

Creating aforementioned 3D geospatial image spaces for infrastructure management (see Section 1.1) require georeferenced RGB-D image sets, which ideally cover and map respective infrastructure as completely as possible. Creating georeferenced RGB-D images involves a) georeferencing and b) depth reconstruction. When using e.g. single panoramic imagery and LiDAR point clouds, the quality of depth reconstruction strongly relies on a robust and accurate georeferencing.

This thesis focuses on the robust and accurate georeferencing in challenging environments. This provides a basis for a wide range of further investigations, applications and products.

Direct georeferencing strongly depends on the GNSS reception, while LiDAR-based odometry and SLAM algorithm require environments with unique geometric structure to perform robustly. Image-based georeferencing methods, by contrast, depend on the brightness and the texture of environments.

Consequently for georeferencing in multiple environments, we consider the simultaneous use of different sensor technologies and georeferencing methods. First, we expect that this will lead to a higher robustness, since if one technology fails or significantly decreases in accuracy, the other might bridge this. Second, different sensor technologies could also lead to an increase in accuracy by complementing different weaknesses.

However, kinematic multi-sensor data fusion depends on precisely calibrated and synchronized sensors.

The previously stated overall objective of this thesis leads to the following main aspects:

- Conception, design and realization of a prototypical image-focused portable MMS, which ideally a) supports the capture of image and LiDAR data with regard to the creation of high-resolution RGB-D panoramic imagery and b) provides raw data from different navigation and mapping sensors such as GNSS and IMU, cameras and LiDAR scanners, as a basis for georeferencing in demanding environments.

- Development of novel robust and accurate georeferencing methods for challenging environments with different environmental conditions. This is intended to include both classic navigation sensor data such as from GNSS, IMU as well as mapping sensor data such as from LiDAR scanners or from high-resolution cameras.

- Comprehensive investigations on the georeferencing methods in environments which represent real-world conditions. The investigations should prove whether the proposed georeferencing methods achieve similar accuracy in challenging environments as previous image-based MMS on vehicle platformes achieved under ideal conditions (Burkhard et al., 2012; Blaser et al., 2018b).

1.3 Main Contributions

This section contains a compilation of the main contributions of this thesis. They include the development and calibration of the portable, image-focused MMS as a prerequisite for conception of robust and accurate georeferencing methods as well as performance investigations under real-world conditions.

The system development involved:

- the evaluation of suitable navigation and mapping sensors meeting the requirements of infrastructure management and enabling precise sensor synchronization, by considering the limited platform form factor and the maximum payload. Interoperable interfaces and data formats were particularly important.
- the conception of a suitable sensor configuration which has the most comprehensive coverage possible on the one hand and provides sufficient geometric stability for georeferencing on the other hand.
- mechanical and electrical sensor integration of a tactical grade GNSS and Inertial Navigation System (INS)-combined navigation system, a multi-head panoramic camera and two multi-beam LiDAR scanners with precise hardware-based sensor synchronization.
- the development of a modular on-board software, that is based on the Robot Operating System (ROS), used for data acquisition and recording as well as for real-time quality and completeness checks during a campaign.

The overall system calibration comprised:

- the application and extension of a constrained bundle-adjustment-based test-field calibration method (Ellum, El-Sheimy, 2002) for both the multi-head panoramic camera and the GNSS and IMU-based navigation system.
- the extension of the functional model for the constrained bundle-adjustment-based calibration by the equidistant camera model (Abraham, Förstner, 2005). This is required for the Interior Orientation (IO) calibration of the individual panoramic camera heads with wide-angle optics.
- the extension of a point cloud feature-based self-calibration method for calibration of RO parameters between LiDAR scanner and camera (Hillemann, 2020) for the precise calibration of multiple LiDAR scanners in indoor test-fields.

Furthermore, we developed a multi-stage georeferencing approach to increase the accuracy and robustness in different challenging environments. This includes:

- the application of direct georeferencing, using tightly-coupled GNSS and IMU sensor data fusion with multi-pass trajectory processing.
- the extension of a LiDAR-based georeferencing method, based on a state-of-the-art LiDAR SLAM algorithm that operates fully in 3D, supports loop-closure (Hess et al., 2016) and leads to a trajectory based on poses with 6 Degree of Freedom (DoF) .
- the application of subsequent image-based georeferencing, using initial camera poses from direct or LiDAR-based georeferencing, that are based on an extended SfM-pipeline by camera-rig constraints and GCP support (Cavegn et al., 2018).

- the extension of direct georeferencing by external CUPT and/or Attitude Update (ATTUPT) support, whereby the updates originate from image-based georeferencing.

The individual georeferencing methods of our developed multi-stage georeferencing approach can be used separately or in combination. This allows to examine and compare the different georeferencing methods under the same conditions as well as to evaluate our multi-stage georeferencing approach.

Hence, we comprehensively investigated the different georeferencing methods within three challenging and large-scale test sites, the first in a city center, the second in a forest and the third in indoors. They represent real-world conditions for future infrastructure mapping campaigns. Thereby, we investigated:

- the plausibility of the resulting trajectories by assessing the internal accuracy, which may indicate gross errors.
- the precision and the accuracy of coordinate measurements obtained with the forward intersection, using four consecutive image epochs. Thereby, we used Check Points (CPs) with superior accuracy.
- the relative accuracy by comparing 3D distances with ground truth, which indicates the quality of relative measurements (distances, areas, etc.).

1.4 Thesis Outline

The main focus of this thesis lies on robust and accurate georeferencing of a portable image-focused MMS in various challenging environments to subsequently create georeferenced RGB-D images that can be used in a cloud-based web-service e.g. for infrastructure management. The thesis is organized in six chapters.

Chapter 2 provides fundamentals for this thesis, combined with a literature and state-of-the-art review. It begins with the aspect of using 3D data, resulting from MM for infrastructure management, and includes a discussion of different common MM data types and representations, leading to our main objective of obtaining accurately georeferenced RGB-D image sequences. Further, it treats MMS, including navigation and mapping sensors, various sensor configurations, an overview of state-of-the-art terrestrial MMS and the fundamentals and methods of overall system calibration of a MMS. Finally, different georeferencing methods are discussed and compared from a sensing perspective.

Chapter 3 includes the design, development and calibration of a portable image-focused MMS, which was a prerequisite for developing and investigating various georeferencing methods.

Chapter 4 presents our developed multi-stage georeferencing approach, combining various georeferencing methods and different sensor data with the aim of providing robust and accurate georeferencing in challenging environments. First, it outlines adopted direct georeferencing, which is commonly used in outdoor MM. Second, it presents the LiDAR-based georeferencing, which we developed by extending a state-of-the-art 3D LiDAR SLAM algorithm. Third, it treats image-based georeferencing, based on an extended SfM pipeline, which subsequently improves previously estimated camera poses of direct georeferencing or LiDAR SLAM. Finally, it presents direct georeferencing, supported with CUPTs and ATTUPTs, which originate from image-based georeferencing.

Chapter 5 contains our investigations on the georeferencing performance in challenging environments. It presents our three test sites with demanding real-world conditions that

represent scenarios with high practical relevance. In addition to relevant information on data acquisition with the portable MMS, it contains the evaluation of various georeferencing methods with resulting internal accuracy as well as qualitative results. Finally, it includes the investigations on the georeferencing with an evaluation of the precision and the absolute and relative accuracy.

Chapter 6 presents the discussion of our developments and results. It begins with our developed image-focused backpack MMS, continues with the multi-stage georeferencing approach and finally includes a comparison of our results with related work.

Finally, Chapter 7 concludes this thesis with a summary, by showing limitations and providing recommendations, and an outlook for further work.

Chapter 2

Mobile Mapping and Georeferencing

This chapter provides a state-of-the-art review as well as necessary fundamentals, principles and methods related to the following developments and investigations of this thesis.

Section 2.1 treats resulting 3D data from MM, such as point clouds and RGB-D imagery and includes a discussion of different 3D data types and representations with regard to their use in geospatial 3D image spaces for infrastructure management applications.

Section 2.2 provides an introduction in MMS. It ranges from MM sensor technology through various sensor configurations to state-of-the-art vehicle-based MMS as well as portable and indoor MMS.

Section 2.3 contains sensor calibration methods as well as methods for overall MMS calibration, which is essential for precise navigation, georeferencing and mapping. Furthermore, various camera models and LiDAR calibration methods are discussed in detail.

Section 2.4 treats different georeferencing methods from a sensing perspective and evaluates the different georeferencing methods with respect to the development of combined multi-stage georeferencing approaches.

2.1 3D Mobile Mapping Data

By general, MM data recordings result in georeferenced 3D data, whereby data representations depend on the mapping sensors used. Most common mapping sensors of terrestrial MMS are cameras and LiDAR scanners that capture images and point clouds, respectively. Thus, this section focuses on image and LiDAR data from terrestrial MMS.

MM data recordings result in large-scale data sets, and the huge amount of data complicates their handling and efficient use. However, specialized applications and web services provide intuitive tools, so that even non-specialists are able to successfully interact with the 3D MM data and to combine it with existing GIS data.

Verbree et al. (2004) provide the early MM client application *GeoFrame* which was developed for users from the public sector (municipalities, planning departments, etc.). *GeoFrame* supports measurements within panoramic images and allows coordinate observation by forward intersection using neighboring panoramic images. Since the panoramic images are georeferenced, the application also allows the augmentation of existing GIS data (e.g. underground mains).

By contrast, Anguelov et al. (2010) technically describe the popular *Google Street View* service, which was developed for millions of non-professional users. In addition to panoramic images, it provides coarse mesh models originating from LiDAR data in the background. Provided

depth information in this way allows smart navigation functions such as *click to go*, where the web client provides the closest image to a clicked object. Furthermore, it also allows more precise placement of points of interests than in google maps.

Paparoditis et al. (2012) propose image-based street navigation and 3D plotting services for both image and LiDAR data for professional users. Their web application provides 3D measurement and mono-plotting functionality. For this purpose, 2D measurements taken in a single panoramic image are reconstructed in 3D, using the LiDAR point cloud.

Meilland et al. (2015) present a method for dense large-scale mapping of outdoor environments with RGB-D spherical panoramic images with regard to real-time visual localization for autonomous vehicles, while Nebiker et al. (2015) propose cloud-based geospatial 3D image spaces as an urban model for smart city applications. The obtained 3D image spaces represent collections of georeferenced RGB-D imagery. By contrast to Paparoditis et al. (2012) and *Google Street View*, 3D measurements do not rely on additional LiDAR point clouds or mesh models and are completely performed in the georeferenced RGB-D imagery by 3D mono-plotting. Thus, the presented approach allows to follow the demand "what you see is what you get", assuming that the depth information coincides with the RGB image information.

Figure 2.1 illustrates the principle of 3D mono-plotting using solely georeferenced RGB-D imagery. There is a mutual relationship between the RGB-D image and the 3D point cloud, since a point cloud can be derived from an image and, vice versa, depth information can be derived from a 3D point cloud. Consequently, georeferenced RGB-D imagery can be regarded as an alternate representation of a dense and colored 3D point cloud.

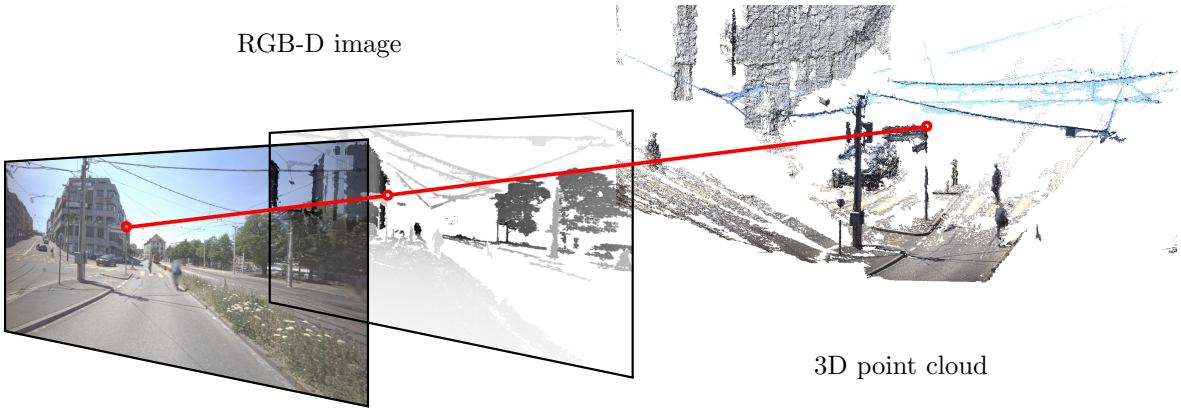


Figure 2.1: Schematic illustration of the relationship between RGB-D images and 3D point clouds according to Nebiker (2017)

Figure 1.1 shows the cloud-based web client, serving as a user interface to 3D geospatial image spaces. It provides numerous navigation and measurement functions, that are useful for infrastructure management, such as multi-view queries that allow objects to be inspected from different sights, an orthogonal distance measurement tool using automatically detected edges, plumb bob measurements as well as functionality for superimposing existing geodata. Nebiker et al. (2015) provide a detailed description of the aforementioned functions.

Such cloud-based web clients, using 3D geospatial image spaces, enable virtual on-site visits and provide an intuitive interface to GIS applications, even for persons not familiar with geodata.

Figure 2.2 exemplarily depicts advantages of image-based or image-focused MM data in terms of human scene interpretation compared to LiDAR-based or LiDAR-focused MM data. By general, images provide higher information density than LiDAR point clouds. By contrast, LiDAR point clouds generally provide a more homogeneous geometric 3D accuracy with less outliers over the scene than 3D reconstructed point clouds from imagery. However,

image-based 3D reconstruction has advantages in the explicit reconstruction of edges and corners, which are often used for measurements in infrastructure management.

In addition to infrastructure management, 3D MM data is essential for numerous current research areas.

Automatic semantic scene interpretation is an active field of research with successful results, using both image and LiDAR data. Nebiker et al. (2021) present an application to gain automatically parking statistics, using RGB-D images from low-cost RGB-D cameras, based on artificial intelligence and region-based convolutional neural networks. Hübner et al. (2021) focus on automatic voxel-based 3D indoor reconstruction and room partitioning and classification from triangle meshes.

Another research and application field with very high potential is visual localization requiring accurate and large-scale dense 3D data. As discussed in Meilland et al. (2015), MM represents an efficient data acquisition method for such large-scale 3D data sets.

(a) RGB-D image (b) Point cloud

Figure 2.2: Comparison of an RGB-D image and a LiDAR point cloud from the same scene. (Nebiker, 2017)

2.2 Mobile Mapping Systems

This section provides an introduction as well as a state-of-the-art review in MMS with a strong focus on terrestrial and portable MMS. Early MM publications clearly divide MMS into platform, navigation sensors and mapping sensors. Recent developments in sensor and navigation technology slightly confused this clear division and interactions as well as overlaps appeared. Figure 2.3 shows a differentiation of the various MMS related terms and components and its mutual relationships. Basically, the platform and the sensor configuration of a MMS are almost independent of each other, since sensor configuration can be mounted on different platforms. However, fundamental properties (e.g. weight and size) of both platforms and sensor configurations might limit each other.

A sensor configuration includes both navigation and mapping sensors. Early MMS showed a clear distinction between navigation and mapping sensors. Navigation sensors usually consisted of GNSS, INS and odometer whereas typical mapping sensors were cameras and LiDAR scanners. However, recent developments towards SLAM, especially in robotics and computer vision, which have been accelerated by autonomous driving in particular, have blurred this clear distinction so that there are overlaps. Techniques such as visual odometry or SLAM increasingly use former mapping sensors for platform navigation as well as for georeferencing.

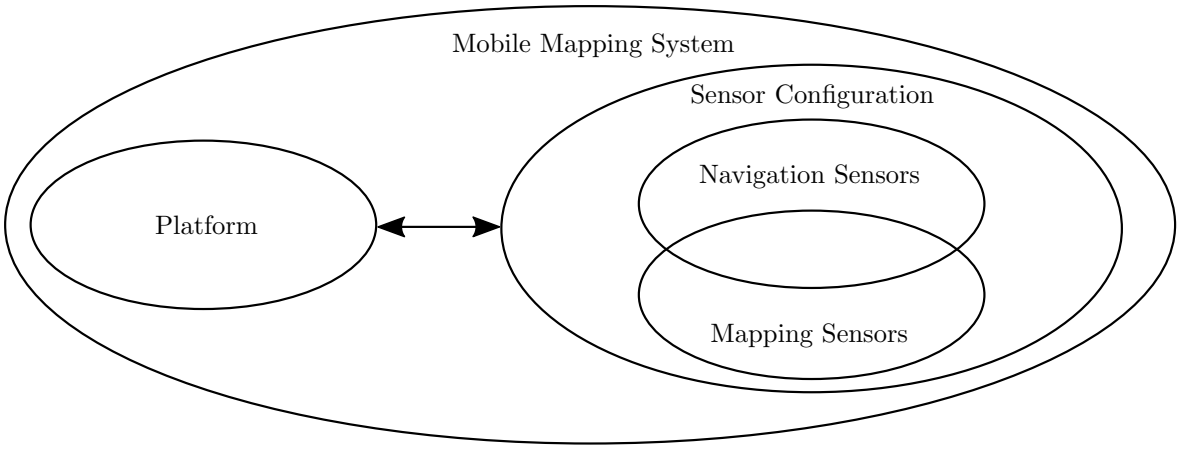


Figure 2.3: Diagram showing the interrelationships between the MMS components.

Subsection 2.2.1 provides a state-of-the-art review of current navigation and mapping sensor technology, while Subsection 2.2.2 gives a brief overview of most common sensor configurations. Subsection 2.2.3 provides an introduction of vehicle-based MMS and addresses early prototypical systems as well as current MMS which are commercially available. Finally, Subsection 2.2.4 strongly focuses on state-of-the-art portable and indoor MMS on different platforms.

2.2.1 Navigation and Mapping Sensors

This subsection includes a brief introduction of most relevant and common navigation and mapping sensor technology. In terms of navigation sensors, mainly the GNSS receiver technology as well as the INS sensor technology are addressed. In terms of mapping sensors, the focus lies on camera and on LiDAR scanner technology. For each sensor technology, the basics are briefly covered, with a focus on MM applications. In addition, recent technological developments and future trends for each sensor type are shown.

GNSS Receivers

GNSS receivers are used for various purposes and applications, that range from precise time estimation up to platform navigation and geodetic measurements. While Kumar et al. (2021) provide a good overview in GNSS technology for geo-scientific applications, Gunawardena, Morton (2020) give an in-depth view of the functionality of GNSS receivers in terms of signal models.

Eugster (2012) subdivides GNSS receivers into navigation receivers and geodetic receivers and further classifies the receiver with regard to the frequencies, the observation types code or carrier phase and the possible evaluation methods pseudo-range, differential, carrier-smoothing of code pseudo-ranges, precise point positioning, kinematic ambiguity resolution and Real-Time Kinematic (RTK), which leads to a typical accuracy.

Differential GNSS measurements require GNSS data from both, a user receiver or a rover as well as a reference receiver or a network of reference receivers. The rover is typically used for measurements at unknown locations and can be mounted on MM platforms, whereby reference receivers or reference stations operate at known and accurately surveyed locations. Gunawardena, Morton (2020) provide several state-of-the-art differential GNSS evaluation algorithms and methods. They further describe that differential GNSS measurements can result in dm accuracy when using code measurements and in cm accuracy when using carrier phase measurements with resolved carrier phase ambiguities.

Rapid developments during the last few years in drone and automotive industries have brought low-cost multi-band GNSS RTK receivers with low energy consumption on the market. In good conditions, GNSS RTK receivers typically provide position accuracy in the range of 0.01 m to 0.02 m and height accuracy in the range of 0.02 m to 0.04 m.

Hamza et al. (2021) compared such low-cost multi-band GNSS RTK receivers with geodetic receivers by performing long-static measurements whereby the low-cost receivers came very close to the geodetic receivers in terms of accuracy. They further show that the accuracy depends significantly on the GNSS antenna used.

Rao et al. (2013) provide a comprehensive treatment of GNSS antennas and important discussions on antenna characteristics and details. Especially noteworthy is the increased use and suitability of multi-frequency quadrifilar helix antennas for geodetic measurements (Wanninger et al., 2022). Their compact design is particularly suitable even for small MM platforms.

Inertial Navigation Systems

An INS is a self-contained system which provides information about its position, attitude and velocity, based on inertial sensors. Thus, an INS contains both a computing unit and an IMU.

IMUs typically contain three orthogonal gyroscopes and three orthogonal accelerometers, measuring angular velocity and linear acceleration respectively. Basically, the following two IMU types are represented:

- Stable platform systems
- Strap-down systems

While the stable platform system is held in alignment with the global frame (e.g. using gimbals), the strap-down system is rigidly mounted on the platform. Depending on the design, different navigation algorithms are necessary. In MM, most IMUs are strap-down systems as they require fewer mechanical components and thereby can be built more compactly. Cramer (2001) gives a comprehensive treatment of the strap-down algorithm based on the principle of dead reckoning.

El-Sheimy, Youssef (2020) provide a recent review on IMU sensor technology and trends, while Passaro et al. (2017) focus on gyroscopes. Table 2.1 provides an overview of different IMU grades in terms of performance, costs and typical applications.

Particularly noteworthy are Micro-Opto-Electro-Mechanical Systems (MOEMSs) as a novel approach which utilizes optical pick-off mechanisms in order to eliminate errors that occur in Micro-Electro-Mechanical System (MEMS)-based sensors. They use similar concepts applied in fiber-optic gyroscopes and are optimized for low-cost manufacturing while achieving high accuracy.

Table 2.1: IMU classification with regard to performance, costs and applications, using information from El-Sheimy, Youssef (2020) and Passaro et al. (2017). Furthermore, they mention the used gyroscope technologies Ring Laser Gyroscope (RLG), Fiber Optic Gyroscope (FOG), and MEMS.

Grade	**Strategic**	**Navigation**	**Tactical**	**Commercial / Automotive**
Positional drift in m/s	0.008–0.028	0.5	5–10	large
Gyroscope drift in °/h	0.0001–0.01	0.01–0.1	0.1–30	30–1000
Gyroscope technology	RLG & FOG	RLG & FOG	FOG & MEMS	MEMS
Costs in $	1 000 000	100 000	1000–10 000	1–1000
Applications	Submarines, intercontinental ballistic missiles	General navigation, high precision georeferencing, mapping	Integrated with GNSS for mapping, short time weapons	Research, Low cost navigation, pedometers, automotive applications

Cameras

Cameras enable full coverage detection of electromagnetic radiation, which is emitted or reflected by an object. However, the majority of cameras operate in the visible spectrum with wave lengths from 380 nm to 780 nm. Their compact and affordable design, and the high information density of the resulting imagery, predestine cameras for the use in MM. While Luhmann et al. (2019) comprehensively cover the fundamentals of camera technology used for close-range photogrammetry, this thesis only briefly addresses a few aspects relevant to MM.

For MM applications, a camera should provide images that were captured at a single point in time (epoch) and that are temporally consistent. Thus, it is essential to avoid opposing effects such as motion blur or rolling shutter. Larger sensor pixel size or larger fill factor (ratio of a pixel's light sensitive area to its total area) increase the light sensitivity of an image sensor, which might result in shorter exposure times (less motion blur) or in less gain amplification (less noise) in darker environments.

Complementary Metal Oxide Semiconductor (CMOS) sensors are becoming more and more widespread and seem to be increasingly replacing Charge-Coupled Device (CCD) sensors. They have lower energy consumption, allow higher frame rates and have significantly reduced blooming and smearing effects (Neumann et al., 2016).

CMOS sensors feature rolling or global shutter, whereby the rolling shutter exposes and reads out the sensor line by line and the global shutter exposes the sensor at once. Rolling shutter sensors have a simpler architecture, are more affordable, and often have a better fill factor than global shutter sensors. However, they introduce significant distortions while moving. Zhou et al. (2020) present a two-step approach for the correction of rolling shutter distortion in Unmanned Aerial Vehicle (UAV) photogrammetry. However, the rolling shutter effect can also be useful. For example, Kim et al. (2020) try to estimate object distances using the rolling shutter effect in terrestrial MM application. By general, global shutter cameras are applied in MM applications in order to avoid rolling shutter effects.

Another important criterion for MM applications is the camera triggering. Particularly critical is the delay that occurs between the trigger command and the physical camera triggering, which leads to errors in the image position. Consequently, the corresponding delay should be minimal.

Further application specific criteria are the frame rate, geometric and radiometric resolution and the optics. For MM applications, optics with wide angles are preferred to cover as much of the scene as possible. Though, wide angle optics or even fish-eye optics lead to significant distortions, which have to be corrected with different distortion or camera models.

LiDAR Scanners

LiDAR scanners are active sensors that use laser to geometrically scan and capture the environment. Shan, Toth (2018) systematically introduce the basic principles of laser ranging, laser profiling and laser scanning. Laser ranging designates the distance measurement between a laser ranger and a ground object using Time of Flight (ToF) of a very short laser pulse or using continuous wave and the phase difference. A laser profile includes a series of laser range measurements along a specific axis (e.g. tilting axis), whereby the measured angle and the range result in 2D coordinates. Finally, a 3D laser scan either requires a series of laser profiles around a further axis (e.g. horizontal axis) for terrestrial laser scanning or a series of laser profiles along a trajectory for mobile laser scanning.

Profile LiDAR scanners are mainly used in MM applications. Important specifications of profile LiDAR scanners are maximum range, range precision, range accuracy, laser pulse repetition rate, scan speed, field of view and the angular resolution.

Increasingly, also multi-beam LiDAR scanners are used in automotive applications and in MM. Since they provide 3D point clouds in real time, they are usable for 3D object detection in real time as well as for SLAM applications.

Another trend, primarily in the consumer sector, is towards solid state LiDAR. Omitting mechanical parts allows a much more compact and cost-effective design, enabling LiDAR technology to even be integrated into smartphones. Nam, Gon-Woo (2021) provide an overview of the most common implementation technologies, such as MEMS, optical phased array or frequency-modulated continuous wave based LiDAR.

2.2.2 Sensor Configurations

There is a wide variety of sensor configurations for MMS. Many of them result from the sensor characteristics and the space or weight limitations of the platform. Nevertheless, certain basic patterns are discernible. This subsection, briefly shows and discusses certain configuration patterns of mapping sensors, using cameras and LiDAR scanners in terms of their characteristics.

Mono or divergent camera configurations (see Figure 2.4a) offer great flexibility and have only few restrictions for the implementation. They are particularly suitable to capture a desired scene area with specific properties based on images. A typical application is the coloring of LiDAR point clouds.

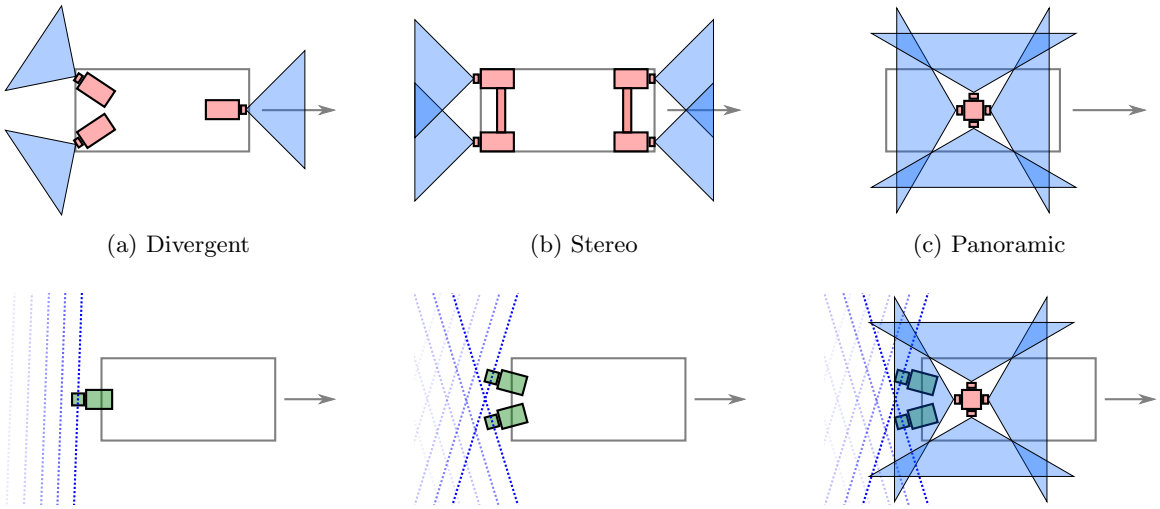
By contrast, stereo camera configurations offer 3D stereo views and also allow 3D measurements within the overlapping and stereoscopic area (see Figure 2.4b). Stereo configurations require a precise calibration in order to obtain rectified stereo images. Furthermore, sufficiently sized stereo bases require large space and are therefore not suitable for every platform. Luhmann et al. (2019) provide comprehensive theory and fundamentals of stereo-photogrammetry.

Panoramic camera configurations (see Figure 2.4c) cover the entire environment and provide a comprehensive overview. They are also suitable for coloring point clouds. Such configurations may range from a single sensor with appropriate optics (e.g. catadioptric or fish-eye), to two or few sensors with fish-eye optics or even to multi-head panoramic cameras. Multi-head panoramic camera configurations require careful design to keep the distances between the

projection centers of each camera head as small as possible, which mitigates or, in the best case, avoids stitching errors.

The single-profile LiDAR scanner configuration (see Figure 2.4d) results in a regular pattern of cross profiles and is for example suitable for road or rail parameter evaluation, while the multi-profile LiDAR scanner configuration (see Figure 2.4e) better captures hidden niches and corners in urban environments thanks to its slightly tilted arrangement.

Multi-sensor configurations enable combinations of different generic sensor configuration patterns, whereby various sensor technologies might complement each other. Figure 2.4f shows a panoramic camera configuration combined with a multi LiDAR profile scanner configuration, which might be useful for street-level city mapping by providing panoramic imagery with corresponding depth information from LiDAR point clouds.



(a) Divergent (b) Stereo (c) Panoramic

(d) Single scanner (e) Multi scanner (f) Multi sensor

Figure 2.4: Top views of typical mapping sensor configurations of MMS. Red filled sensors indicate cameras or multi-camera systems while green filled sensors mark LiDAR scanners. Blue filled triangles represent camera fields of view, where darker blue areas show overlapping. Blue dotted lines represent LiDAR profiles, with profiles recorded at an earlier time shown more transparently.

2.2.3 Vehicle-based Mobile Mapping Systems

First experimental stereo-vision-based MMS were the Ohio State University Highway Mapping System (Novak, 1991) and the VIASAT system (Schwarz et al., 1993). Both systems consist of a GNSS and IMU combined navigation unit and a forward pointing stereo camera system used for mapping. Schwarz et al. (1993) clearly introduced the principle of stereo vision-based MMS and achieved an overall positioning accuracy of 0.3 m and better for all objects within a 50 m corridor on both sides of a highway. Beers (1995) developed an early panoramic-based MMS, using a compass and an odometer for navigation and a single analogous fish-eye camera for mapping.

With the rapid increase in computing power as well as in memory and storage capacity and ongoing developments in sensor technology, stereo vision-based MMS have been evolved to multi-stereo as well as panorama camera-based MMS. The image-based multi stereo MMS *Stereopolis II* (Paparoditis et al., 2012) consists of two stereo systems as well as of a rigid mount with an IMU, three LiDAR scanners and a panorama camera configuration (see Figure 2.5a). The stereo systems pointing to the driving as well as to the reverse direction allow image-based 3D measurements on objects along the track. Blaser et al. (2017) present a completely image-based MMS with a stereo-panoramic 360° configuration (see Figure 2.5b).

The presented system consist of a GNSS and IMU combined navigation unit, of a forward pointing stereo system as well as of two multi-head panorama cameras, which are tilted forward and backward respectively. Thus, two camera heads of each panorama camera form a stereo system to the side, which results in an upward and a downward looking stereo system on both sides.

(a) Stereopolis II (Paparoditis et al., 2012) (b) SmartMMS IGEO (Blaser et al., 2017)

Figure 2.5: Multi-stereo and panoramic MMS configurations

Extensive accuracy analysis showed absolute 3D point accuracy in the range of 2 cm to 8 cm and relative 3D distance accuracy in the range from 1 cm to 5 cm (Blaser et al., 2018b). The accuracy of the panoramic configuration thus lies in the same order of magnitude as the accuracy of a stereo image-based MMS, which is in the range of 4 cm to 5 cm for absolute 3D point measurements, according to prior accuracy investigations (Burkhard et al., 2012).

A much larger number of work has been focused on accuracy evaluation of LiDAR-based MMS. Puente et al. (2013) provide a comprehensive overview and comparison of various commercially available mobile laser scanning systems in terms of their practical application. Barber et al. (2008) performed a geometric validation of their ground-based mobile laser scanning system *StreetMapper*, consisting of a GNSS and IMU-based navigation unit and three LiDAR scanners.

They achieved a measurement precision in elevation in the order of 3 cm and planimetric accuracy of approximately 1 cm. Haala et al. (2008) further performed accuracy investigations with the *StreetMapper* system in challenging urban area. They confirmed an achievable accuracy level below 3 cm with good GNSS coverage.

2.2.4 Portable and Indoor Mobile Mapping Systems

This subsection incorporates state-of-the-art portable and indoor MMS. They can be subdivided, based on their platform, into trolley, backpack and hand-held systems. Furthermore, their sensor configuration is discussed.

Trolley Systems

Trolley platforms are suitable for data recording of large indoor environments. Since they remain on the floor and do not have to be carried, there are fewer limitations in terms of sensor weight. As they can be pushed evenly and the sensors remain constantly at the same height, the result is a smooth trajectory.

Huitl et al. (2012) present a custom-built mapping trolley equipped with two horizontally and vertically mounted LiDAR scanners, a *Ladybug 3* panoramic camera, and two digital single-lens reflex cameras mounted on each side. For 2D navigation, they use odometry as well as a horizontally mounted LiDAR profile scanner. Continued development led to the *NavVis M6* trolley (see Figure 2.6a) which consists of three single profile LiDAR scanners, a

multi-beam LiDAR scanner and a divergent multi-camera configuration on the top. For data processing, they provide a cloud solution, resulting in textured point clouds (NavVis, 2020). By contrast, the *Viametris Trolley iMS3D* features three profile scanners and a *Ladybug 3* panoramic camera configuration (see Figure 2.6c). The *Trimble Indoor Mobile Mapping Solution* (TIMMS) also features a *Ladybug 3* multi-head panoramic camera, which is located directly above the operator's head (see Figure 2.6b) and a high performance LiDAR profile scanner which is mounted across the recording direction.

(a) NavVis M6 Trolley (NavVis, 2020)

(b) TIMMS (Applanix Corporation, 2017)

(c) Viametris iMS3D (Viametris, 2022b)

Figure 2.6: Trolley-based MMS

Backpack Systems

Backpack Systems offer more flexibility than trolley systems and thus have a greater range of applications. However, there are limitations on the maximum sensor payload as well as on the sensor configuration due to limited form factors.

Ellum, El-Sheimy (2000) discuss and simulate various sensor configurations in order to develop a low-cost backpack MMS and Ellum, El-Sheimy (2001) present the developed low-cost backpack MMS. Their system features a *Leica* Digital Magnetic Compass (DMC-SX) and a *NovAtel* dual-frequency RT2 GNSS receiver as navigation sensors and a digital consumer camera *Kodak DC260* for mapping. Their investigations show an absolute accuracy of 25 cm for a 20 m camera to object distance.

Grejner-Brzezinska et al. (2006) present a self-developed backpack MMS. For navigation, they integrate different navigation sensors such as GNSS, IMU, barometer and digital compass within an extended Kalman filter and provide interesting findings with regard to human dynamics-supported navigation.

By contrast, Liu et al. (2010) present a self-developed backpack MMS for indoor localization and visualization, adapted for automated 3D modeling of building interiors. Their backpack incorporates two IMUs and three profile LiDAR scanners that are arranged orthogonal to each other (see Figure 2.7a). The three cameras have a resolution of 1338×987 px and point to the top, to the bottom and to the left. The navigation-grade IMU *Honeywell HG9900* with three ring laser gyros, having a bias stability below $0.003\,°/\mathrm{h}$ and high precision accelerometers, serves as ground truth. The second, MEMS-based IMU is used for navigation in combination with the horizontally mounted "yaw" profile scanner. Their accuracy investigations showed average position errors in the decimeter range. Results presented in Corso, Zakhor (2013) show a mean accuracy of 10 cm within an office environment by using 100 surveyed control points.

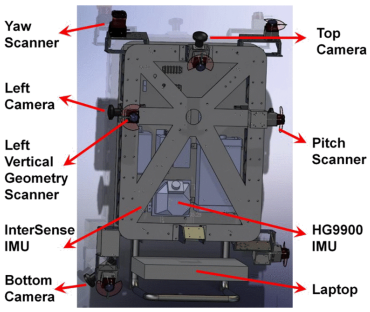
Nüchter et al. (2015) present a completely LiDAR-based backpack MMS without IMU support. The system features a horizontally fixed profile scanner used for 2D mapping, as well as a 3D LiDAR scanner *Riegl VZ-400*, that is able to rotate freely around its vertical axis (see

Figure 2.7b). Lauterbach et al. (2015) provide an evaluation of their developed backpack and their proposed semi-rigid SLAM algorithm. They show maximum pose errors of 25 cm in translation and 7° in rotation, which occurred mainly in places where the operator failed to walk uniformly.

More recent academic backpack developments are mainly LiDAR-focused systems such as the indoor backpack system from Wen et al. (2016) and the ITC-IMMS presented by Karam et al. (2019) (see Figure 2.7c). Both systems feature three profile LiDAR scanners and MEMS-based IMUs.

In recent years, commercial image-focused backpack MMS have appeared on the market. The following three products are available: *Leica Pegasus Backpack* (Leica Geosystems AG, 2022b), *Viametris Backpack BMS3D-HD* (Viametris, 2022a) and *NavVis VLX* (NavVis, 2022). By contrast, the *Vexcel Panther* (Geo Week News, 2017) was recently discontinued.

The LiDAR scanner configuration of the *Leica Pegasus Backpack*, the *Viametris Backpack BMS3D-HD* and the *NavVis VLS* are quite similar, as they all consist of both a horizontally and a vertically mounted multi-beam LiDAR scanner. By contrast, the *Vexcel Panther* features only a single multi-beam LiDAR scanner mounted almost horizontally (see Figure 2.7).



(a) Berkeley Backpack (Liu et al., 2010)

(b) Würzburg Backpack (Nüchter et al., 2015)

(c) ITC-IMMS (Karam et al., 2019)

(d) Leica Pegasus Backpack (Leica Geosystems AG, 2022b)

(e) Vexcel Panther (Geo Week News, 2017)

(f) Viametris Backpack (Viametris, 2022a)

(g) NavVis VLX (NavVis, 2022)

Figure 2.7: Backpack MMS

The four commercial systems differ rather in their camera configuration. The *Vexcel Panther*, the *Viametris Backpack* and the *NavVis VLX* feature panoramic camera configurations. While the *Vexcel Panther* features a multi-head camera with 26 different camera heads and provides a panoramic image with a total resolution of 172 Mpx, the *Viametris Backpack* has a *Ladybug 5* panoramic camera with 6 camera heads and a total resolution of 30 Mpx. The *NavVis VLX* features 4 different cameras, using fish-eye optics with an overall resolution of 80 Mpx. By contrast, the *Leica Pegasus Backpack* uses a divergent and mainly backward-faced camera configuration which consists of five cameras with a resolution of 2046×2046 px.

Finally, the systems also differ in their navigation and georeferencing approaches. The *NavVis VLX* uses a LiDAR SLAM for navigation with GCP support for georeferencing. Both the *Leica Pegasus Backpack* and the *Viametris Backpack* combine a GNSS- and INS-based navigation solution with LiDAR SLAM and the *Vexcel Panther* combines a GNSS and INS-based navigation solution with visual odometry.

Hand-held Systems

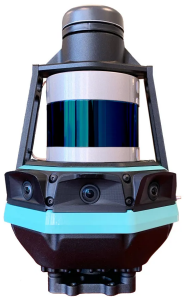
Hand-held platforms offer the greatest flexibility since they can operate in areas that are inaccessible for trolley and backpack systems. Nevertheless, hand-held systems are mainly suitable for capturing smaller areas because a) they become inconvenient to hold after a short time and b) their computing and storage capacity is often limited and insufficient for large data records.

Bosse et al. (2012) present a spring-mounted profile LiDAR scanner combined with an industrial-grade IMU *Zebedee* and a 3D SLAM algorithm for estimating a trajectory with 6 DoF. The system has been continuously improved and industrialized and is marketed under the name *GeoSLAM ZEB-REVO*. The spring was replaced by a motor, which continuously rotates the combined profile LiDAR scanner, and an IMU (see Figure 2.8b). Nocerino et al. (2017) investigated both indoor and outdoor performance of the hand-held LiDAR-based MMS *GeoSLAM ZEB-REVO*.

Another LiDAR-based hand-held MMS is *Kaarta Stencil*, featuring a multi-beam LiDAR scanner as well as a low-grade IMU (Zhang et al., 2016). The successor system *Kaarta Stencil Pro* has been evolved from an only LiDAR-based to a multi-sensor system, featuring a multi-head panoramic camera with 4 camera heads and a total panoramic image resolution of 32 Mpx (see Figure 2.8a). Furthermore, the system is equipped with a GNSS antenna and a multi-frequency and multi-constellation GNSS receiver.

Finally, the former tripod-based LiDAR scanner *Leica BLK360* evolved to a multi-sensor hand-held MMS *Leica BLK2GO*, featuring a highly accurate LiDAR scanner, a 12 Mpx rolling shutter camera for mapping and three fish-eye cameras, arranged to a panoramic configuration for navigation. For navigation, they use a combined IMU-supported LiDAR and visual SLAM.

Leica Geosystems AG (2022a) specify a range noise of 3 mm and an indoor accuracy of 10 mm for the *Leica BLK2GO*, while the *Kaarta Stencil* is specified with a LiDAR accuracy of 3 cm and the *ZEB REVO*'s accuracy is specified from 1 cm to 3 cm (Geoslam, 2022).



(a) Kaarta stencil pro (Kaarta, 2022)

(b) ZEB Revo (Geoslam, 2022)

(c) BLK2GO (Leica Geosystems AG, 2022a)

Figure 2.8: Hand-held MMS

2.3 Mobile Mapping System Calibration

This section treats the MMS calibration. Subsection 2.3.1 shows the functional dependencies between the involved coordinate frames of typical MMS and includes a discussion of various overall system calibration methods. Subsection 2.3.2 treats different camera sensor models. Finally, Subsection 2.3.3 covers the calibration of profile and multi-beam LiDAR scanners and discusses different calibration approaches and procedures.

2.3.1 Overall System Calibration

Figure 2.9 depicts the functional relationships between the different sensors, the body frame and the mapping frame of a typical MMS for direct sensor orientation.

The figure below graphically shows the following equations:

$$ {}^{w}\mathbf{p}_i(t) = {}^{w}\mathbf{H}_b(t) {}^{b}\mathbf{H}_c {}^{c}\mathbf{p}_i(t) \tag{2.1} $$

and

$$ {}^{w}\mathbf{p}_j(t) = {}^{w}\mathbf{H}_b(t) {}^{b}\mathbf{H}_l {}^{l}\mathbf{p}_j(t) \tag{2.2} $$

which transform the homogeneous point coordinates measured, ${}^{c}\mathbf{p}_i(t)$ and ${}^{l}\mathbf{p}_j(t)$ using a stereo camera system and a LiDAR scanner respectively. ${}^{b}\mathbf{H}_c$ and ${}^{b}\mathbf{H}_l$ represent rigid body transformations which are 4×4 matrices, incorporating lever-arm and misalignment of each sensor to the body frame b. In this example, the body frame b is equivalent to the navigation center of the INS. Finally, the body frame pose ${}^{w}\mathbf{H}_b(t)$ from the INS transforms both points into the world coordinate frame w. More details on the homogeneous coordinate transformations can be found in Appendix A.2.

Ellum, El-Sheimy (2002) divide the MMS calibration into three separate steps a) camera calibration, b) lever-arm calibration between the body frame and the GNSS antenna center and c) the boresight calibration between the camera configuration and the body frame. They propose a bundle-adjustment-based self-calibration for the cameras according to Fraser (1997), and discuss both advantages and disadvantages of different techniques for the lever-arm and boresight calibration. Thereby, they consider external measurement of the lever-arm, post-adjustment averaging as well as including the boresight parameters into the bundle-adjustment.

Rau et al. (2011) extend the functional model of the bundle-adjustment-based camera calibration by the boresight alignment. This allows to simultaneously calibrate the cameras with the boresight alignment. Kersting et al. (2012) further extend the functional model by relative orientations between camera and reference camera coordinate systems enabling the single-step calibration of multi-camera MMS.

Khoramshahi et al. (2019) present an accurate calibration scheme for a multi-camera MMS. Their system features a *Ladybug 5+* multi-head panoramic camera and a GNSS and INS-combined navigation system. By contrast, Blaser et al. (2017) present an overall system calibration procedure for stereo-panoramic MMS, featuring two tilted *Ladybug 5* multi-head panoramic cameras that build stereo camera configurations to the sides. The presented MMS additionally features a stereo system to the front as well as a GNSS and INS-combined navigation system.

Khoramshahi et al. (2019) and Rau et al. (2016) use the pinhole camera model for the individual camera heads of the *Ladybug 5+* and the *Ladybug 5* panoramic camera, respectively.

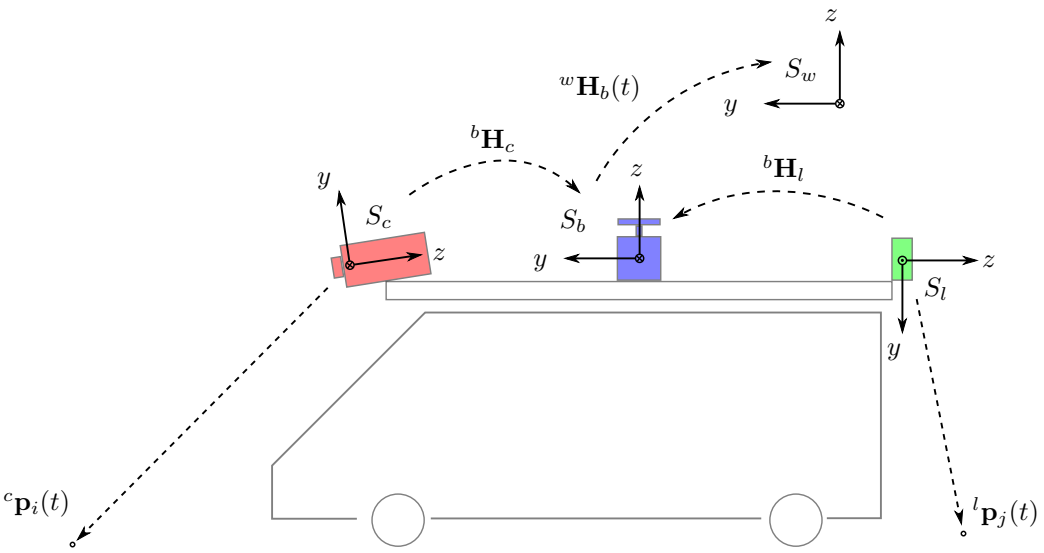


Figure 2.9: Outline of a MMS with a typical sensor configuration (a forward-pointing stereo camera system, a profile LiDAR scanner in the rear and a GNSS- and INS-combined navigation unit on the top). It shows the different coordinate frames involved, their relationships and the principle of direct sensor orientation. Mapping sensors (e.g. stereo camera systems (red) or LiDAR scanners (green)) capture 3D points ${}^c\mathbf{p}_i(t)$ and ${}^l\mathbf{p}_j(t)$ at time t within their local sensor coordinate frames S_c and S_l. Furthermore, the navigation unit (blue) continuously estimates poses ${}^w\mathbf{H}_b(t)$ between the body frame S_b and the world coordinate frame S_w at the same time t. With pre-calibrated Boresight Alignment (BA) (${}^b\mathbf{H}_c$, ${}^b\mathbf{H}_l$) and ${}^w\mathbf{H}_b(t)$, the locally mapped 3D points can be transformed into the world coordinate frame S_w.

By contrast, Blaser et al. (2017) use the equidistant fish-eye camera model to cope with the significant distortions of the individual *Ladybug 5* panoramic camera heads.

2.3.2 Camera Models

The increasing number of camera types used in close-range as well as in MM requires new approaches in sensor modeling. Luhmann et al. (2016) provide a comprehensive review of recent camera developments, state-of-the-art camera calibration methods and models, and extended and generic calibration models.

Most camera calibration models assume a central perspective camera model and use additional parameters for modeling distortions between the ideal mathematical camera model and the physical camera reality. The distortion parameters provided by Brown (1971) are widely used in close-range photogrammetry and have become a quasi-standard. They work well for cameras with narrow-angle or wide-angle lenses. However, they are not suitable for fish-eye lenses with large fields of view. Abraham, Förstner (2005) suggest using an ideal projection of the imaging sphere onto the image plane and then applying subsequent corrections.

They review existing camera and fish-eye projection models, such as

- the perspective: $r = f \tan(\theta)$,
- the stereo-graphic: $r = 2f \tan(\theta/2)$,
- the equidistant: $r = f\theta$,
- the orthogonal: $r = f \sin(\theta)$
- and the equi-solid-angle: $r = 2f \sin(\theta/2)$,

where θ is the angle between the principal axis and the incoming ray, r denotes the distance between the image point and the principal point and f represents the focal length. Furthermore, they propose two rectification models for fish-eye stereo images, providing parallel epipolar lines: a) the epipolar equidistant and b) the epipolar stereo-graphic model. Both models allow fast depth reconstruction of fish-eye stereo images, using standard stereo matching algorithms.

By contrast, Kannala, Brandt (2006) propose a generic camera model for fish-eye cameras, which is based on Taylor polynomials and approximates both the best-fitting fish-eye projection model and the physical distortions. Heuvel et al. (2006) present a calibration procedure for MMS with a fish-eye-based panoramic camera, which is based on the generic camera model presented in Kannala, Brandt (2006). Luber (2015) demonstrates that generic camera models are suitable for a wide range of fish-eye optics and stereo systems and achieve maximal calibration errors below 1 px. However, generic camera models used for stereo systems result in epipolar curves that are not supported by standard stereo matching algorithms.

2.3.3 LiDAR Calibration

LiDAR calibration can be divided into a) the intrinsic sensor calibration, b) the extrinsic calibration (boresight alignment), and c) the relative orientation between another sensor. In particular, sensor data fusion of LiDAR and image data requires precisely calibrated relative orientations between cameras and LiDAR scanners.

Chan et al. (2013) provide an overview of state-of-the-art boresight alignment calibration methods for only LiDAR-based MMS and propose a feature-based approach for boresight self-calibration. By contrast, Karam et al. (2019) present a plane-based calibration procedure for calibrating their profile LiDAR scanner-based backpack indoor MMS, which relies on a rectangular-shaped calibration room.

Similarly, the kinematic self-calibration procedure of a backpack-based multi-beam LiDAR MMS, proposed by Kim et al. (2021), is based on plane constraints. They additionally consider intrinsic sensor parameters. Thereby, their functional model includes the range offset, the vertical and horizontal angular offset for each LiDAR beam, and the Exterior Orientation (EO) parameters for each LiDAR scanner. Furthermore, they investigated the correlation coefficients between intrinsic and extrinsic calibration parameters as well as the long-term stability of the calibration. With their approach, they achieved an improvement of planar misclosure in the range of 35 % to 81 %, while the internal calibration of Glennie et al. (2016) improves the planar misclosure by 20 %.

There exist different calibration methods for the relative orientation between LiDAR scanners and cameras.

Scaramuzza et al. (2007) use a cost function that minimizes the difference of the bearing angles of the camera points and the bearing angles of the re-projected LiDAR points and use the Levenberg-Marquardt algorithm for non-linear optimization. They propose manually measured corresponding points by using a visualization technique which facilitates the measurement process.

By contrast, Alismail et al. (2012) present an automatic calibration procedure for the RO between a LiDAR scanner and a camera system, using a single calibration target. Ilci, Toth (2020) use LiDAR targets to calibrate the misalignment between LiDAR scanners and camera systems, while estimating the lever-arm externally. Ravi et al. (2018) propose a simultaneous system calibration of multi-LiDAR and multi-camera MMS, using planar targets which are measurable from LiDAR data as well as from image data.

Hillemann (2020) proposes an approach for self-calibration of LiDAR profile scanners and cameras that relies on LiDAR point cloud features and does not depend on targets or

corresponding points. The approach is based on the optimization of the following cost function

$$\mathcal{K} = s(\mathcal{F}) \sum_{i=1}^{N} (f_{\mathcal{F},i}({}^{w}\mathbf{p}_i, \mathcal{N}_i))^2 = s(\mathcal{F}) \sum_{i=1}^{N} (f_{\mathcal{F},i}({}^{w}\mathbf{H}_{c(i)}\,{}^{c}\mathbf{H}_l\,{}^{l}\mathbf{p}_i, \mathcal{N}_i))^2, \tag{2.3}$$

where the function $f_{\mathcal{F},i}$ calculates a feature, using a LiDAR point ${}^{w}\mathbf{p}_i$ that refers to the global coordinate frame w and a neighborhood $\mathcal{N}_i$ which incorporates a predefined number of points that are closest to ${}^{w}\mathbf{p}_i$. Since the LiDAR point ${}^{l}\mathbf{p}_i$ refers to the local frame l of the LiDAR scanner, the relative orientation ${}^{c}\mathbf{H}_l$ between the LiDAR coordinate frame l and the camera coordinate frame c as well as the current camera pose ${}^{w}\mathbf{H}_{c(i)}$ are required to transform the LiDAR point into the global coordinate frame w.

Hillemann (2020) evaluated different cost functions using different geometric 3D LiDAR features whereby the omnivariance showed the best results. The omnivariance is calculated as follows

$$f_{\mathcal{O},i}({}^{w}\mathbf{p}_i, \mathcal{N}_i) = \sqrt[3]{\lambda_{1,i}\lambda_{2,i}\lambda_{3,i}}, \tag{2.4}$$

where $\lambda_{1,i}$, $\lambda_{2,i}$ and $\lambda_{3,i}$ are the eigenvalues of the structure tensor of the local neighborhood $\mathcal{N}_i$ of the LiDAR point ${}^{w}\mathbf{p}_i$.

2.4 Georeferencing Methods

Achieving homogeneous accurate and reliable georeferencing across different environments is one of the most important unsolved research topics in MM. Historically, the problem of georeferencing has been tackled mainly by the geomatics community. Due to the rapid developments in emerging technologies, especially in the field of AR or autonomous vehicles and platforms, other disciplines such as computer vision or robotics are increasingly addressing georeferencing issues.

In addition to conventional georeferencing sensors such as GNSS and IMU, environmental sensors such as cameras or LiDAR scanners are increasingly used for pose estimation and georeferencing.

This section briefly treats various georeferencing methods from a sensing perspective, whereby fundamentals as well as recent and novel developments are addressed. Subsection 2.4.1 includes the direct georeferencing which mainly involves GNSS and IMU sensor data. Subsection 2.4.2 treats different LiDAR-based georeferencing methods and Subsection 2.4.3 includes various image-based or visual methods for georeferencing. Subsection 2.4.4 summarizes, discusses and evaluates the different georeferencing methods and provides a qualitative overview. Finally, Subsection 2.4.5 briefly introduces the quality measures used for the quantitative evaluation.

2.4.1 Direct Georeferencing

By general, direct georeferencing refers to estimating the pose of the platform and sensors using on-board sensors, that are independent from the mapping sensors. Since the early days of MMS however, the sensor combination of GNSS and IMU has been widely used for direct georeferencing.

Schwarz et al. (1993) present a GNSS and IMU integration for an early image-based MMS using a Kalman Filter and describe in detail the advantages of direct georeferencing. Angrisano (2010) provide a comprehensive treatment of different GNSS and IMU integration strategies.

In this context, the two predominant strategies are the loosely and the tightly coupled sensor integration. A loosely coupled integration processes raw data from GNSS and IMU separately and integrates pre-processed positions and velocities from both sensors in combination with attitudes from IMU within a Kalman Filter. By contrast, a tightly coupled sensor integration combines raw observations from GNSS with predicted measurements from INS.

A loosely coupled sensor integration requires smaller filters, which results in lower processing times. Furthermore, it is more robust since the raw data of both sensors are processed separately and provide independent navigation solutions. However, tightly coupled sensor integration results in a lower process noise and is capable of integrating partial GNSS satellite constellations. Tightly coupled sensor integration proves beneficial in areas of disturbed GNSS reception by reducing drifts caused by the INS.

By general, direct georeferencing strongly depends on the GNSS reception. In outdoor environments with good GNSS reception, it enables cm-accuracy by using Post-Processed Kinematic (PPK) or RTK GNSS combined with a tactical grade IMU.

However, partial or complete GNSS signal loss may reduce the accuracy to the dm or even to the m-range (Cavegn et al., 2018).

There exist several methods and strategies for bridging GNSS outages or reducing accuracy decrease. Grejner-Brzezinska et al. (2001) investigate the effects and the benefits of Zero Velocity Updates (ZUPTs) on navigation accuracy in urban areas. Since a ZUPT sets the velocity state of the Kalman filter to zero when the platform comes to a standstill, this also has a beneficial effect on the accuracy of the correlated position and attitude.

By contrast, Eugster et al. (2012) propose to support a loosely-coupled integration of GNSS and INS during GNSS outages by CUPTs from external sources. They investigate on CUPTs that originate from either a stereo-image sequence obtained with a bundle-adjustment or from individual GCP measurements in stereo images. Stereo image sequences processed with bundle-adjustment can additionally be used to create ATTUPTs, whereas individual GCP measurements are more efficient. Eugster et al. (2012) show improvements in standard deviation from the dm- to the cm-range.

In case of post-processing, significant improvements in accuracy are possible by processing trajectories two-sided and multiple times as well as by smoothing. Chen et al. (2021a) present a new triple filtering algorithm, which performs a forward-backward-forward extended Kalman Filter for trajectory processing. They show that the third forward processing further reduces inaccuracies in the first forward processing when the filter has not reached convergence.

Recent developments in direct georeferencing focus on the tightly coupled GNSS and INS integration using Precise Point Positioning (PPP). Gu et al. (2021) propose this approach for vehicle navigation since PPP GNSS is cost-effective. They achieved horizontal positioning errors in the sub-m-range in urban environments.

2.4.2 LiDAR-based Georeferencing

This subsection discusses various georeferencing methods that rely mainly on LiDAR sensors.

LiDAR Odometry

In general, LiDAR odometry leads back to a scan registration problem, whereby estimated position and orientation differences between local LiDAR scans result in a local trajectory. Most existing LiDAR odometry algorithms are based on variants of the Iterative Closest Point (ICP) algorithm (Besl, McKay, 1992), such as point-to-plane ICP, plane-to-plane ICP, or

combined methods developed to increase processing speed and robustness.

When using rotating multi-beam LiDAR scanners, the ongoing platform movements distort the local LiDAR profiles. Thus, LiDAR odometry using rotating profile scanners includes the following tasks:

1. Undistort or de-wrap the local LiDAR (multi-) profiles.
2. Estimate the rigid-body transformations to previous LiDAR (multi-) profiles or to the global coordinate frame respectively.

De-wrapping local LiDAR (multi-) profiles requires high-frequency motion information, while precision is less important. By contrast, rigid body transformation estimation requires the opposite.

In literature, the term LiDAR Odometry and Mapping (LOAM) refers to algorithms that consider both tasks simultaneously, whereby odometry refers to the estimation of high-frequency motions, and mapping relates to the estimation of the rigid-body transformation. Zhang, Singh (2017) present an edge and planar feature-based LOAM algorithm, showing drift rates around 1 % in indoor environments and 2.5 % in outdoor environments.

Recent developments use deep convolutional neural network pipelines such as *LO-Net* (Li et al., 2019). They show promising results with similar accuracy to state-of-the-art geometry-based approaches. However, such pipelines are often trained with ground truth data, which significantly limits the application scenarios.

By contrast, Zheng, Zhu (2021) significantly improve efficiency and accuracy with a geometry-based LiDAR odometry approach by using both projected spherical and ground images. Their algorithm shows drift rates around 0.7 % and requires a computational time of about 27 ms on a *NVIDIA Jetson AGX* computer.

LiDAR SLAM

This subsection primarily deals with the general aspects of SLAM as well as LiDAR SLAM, while the aspects of visual SLAM are treated in Subsection 2.4.3.

SLAM algorithms allow to simultaneously estimate sensor poses and create a map. This technique was developed to realize autonomous robots and platforms. In contrast to LiDAR odometry and LOAM algorithms, LiDAR SLAM additionally considers long term drifts by loop closures.

Durrant-Whyte, Bailey (2006) describe the history and the formal structure of the SLAM problem. Cadena et al. (2016) give a general comprehensive review on SLAM and Alsadik, Karam (2021) provide a geomatics-related overview of various aspects of SLAM.

By general, a SLAM algorithm includes the following core components:

1. The **map**, which may represent a sparse map, a dense map or a grid-based map.
2. The **front-end** (also called local SLAM), which performs real-time tasks at a high frequency (e.g. feature extraction and matching or local point cloud matching).
3. The **back-end** (also called global SLAM), which performs optimization tasks at a low frequency (e.g. graph-optimization, particle filtering, loop closure, re-localization or map estimation)

Traditional LiDAR SLAM algorithms use grid-based maps, which either consist of 2D occupancy grids based on pixels or 3D occupancy grids based on voxels. Each grid cell can indicate its state, whether it is free, occupied or unknown. Corso, Zakhor (2013) present a 2D offline LiDAR SLAM algorithm with automatic loop-closure detection from a 2D occupancy grid, while Hess et al. (2016) propose the real-time 2D LiDAR SLAM algorithm *Cartographer*, based on 2D occupancy grids, which they subsequently extended to a real-time 3D SLAM algorithm.

More recent LiDAR and multi-sensor SLAM algorithms such as Behley, Stachniss (2018) or Park et al. (2021) focus on sparse or dense map representation. Especially SLAM with dense map representation requires a high level of computing power as well as a large amount of memory. However, geometric primitives derived from the point cloud (e.g. planes, cylinders or surfels) that are used for scan registration increase computational efficiency.

2.4.3 Image-based Georeferencing

Image-based georeferencing denotes the pose estimation of a platform or a camera by using imagery. There exist various techniques and approaches, which are differently suited, depending on the situation or system.

Structure-from-Motion and Bundle-Adjustment

Image-based georeferencing using bundle-adjustments dates back several decades and was first applied in aerial photogrammetry. Kruck (1983) extended the functional model for bundle-adjustment with additional parameters used for terrestrial photogrammetry.

Advances in automatic detection and matching of robust image features, such as the Scale Invariant Feature Transform (SIFT) introduced by Lowe (2004), led to novel methods for image orientation and 3D reconstruction such as SfM. Similar to traditional photogrammetry and bundle-adjustment, SfM uses the parallax between different images to reconstruct the depth. By general, SfM lacks scale and global reference frame because it does not involve GCPs. Consequently, camera poses and 3D points are estimated in a local image space, using automatically detected and matched image features. However, subsequent 3D similarity transformations allow the transformation of such local scenes into the desired global reference frame.

There exist various SfM and image matching strategies to handle larger image data sets, which may consist of thousands or even millions of images. Schönberger, Frahm (2016) mention incremental, hierarchical and global approaches, whereby incremental SfM is the most popular. Their presented open-source SfM implementation *COLMAP* basically follows the incremental reconstruction strategy.

Cavegn (2020) extended the *COLMAP* SfM pipeline by a) support for GCPs, which enables georeferencing directly in the SfM pipeline, b) support of prior EO parameters, which enable spatial feature matching and which are used for bundle-adjustment and c) support for RO parameters that enable to constrain multi-camera configurations and significantly reduce the number of unknown parameters, thereby increasing the robustness of the bundle-adjustment. By contrast to the original implementation, the extension of Cavegn (2020) follows the global reconstruction strategy. It relies on prior EO parameters and allows to directly triangulate the tie points for all images. Using the extended *COLMAP* SfM pipeline, Cavegn (2020) significantly improved directly georeferenced poses within urban and indoor environments from distortions in the dm-range to the cm-range. However, the processing time significantly exceeded the data acquisition time. Thus, such SfM pipelines do not perform in real-time.

Visual Odometry

Visual odometry originates from robotics and is primarily used for real-time platform navigation. Nister et al. (2004) present a visual odometry system that supports single as well as stereo visual odometry and estimates the camera motion in real-time. An overview on visual odometry research can be found in Aqel et al. (2016). While SfM focuses on precise 3D reconstruction and requires high-resolution imagery captured with a low frame rate, visual odometry focuses on real-time performance with high frame rates and low latency and thus low-resolution.

Visual SLAM

Visual SLAM techniques originate from robotics and are closely related to visual odometry. Visual SLAM algorithms allow to simultaneously estimate the pose of the sensor and create a map. While visual odometry algorithms focus on local optimizations, visual SLAM algorithms also consider global consistency e.g. employing loop closure detection to compensate drifts accumulated at the end of the trajectory. Macario Barros et al. (2022) provide a recent review on visual SLAM algorithms, while they focus on visual-only, visual-inertial and RGB-D SLAM as the three main visual-based SLAM approaches.

2.4.4 Qualitative Comparison of Different Georeferencing Methods

This subsection provides a qualitative comparison of the georeferencing methods discussed in the previous subsections with respect to terrestrial applications.

Basically, each sensor technology has its own characteristic with strengths and weaknesses in various points, affecting the robustness and defining the range of possible applications. Furthermore, environmental conditions required for successful georeferencing differ due to the different physical principles of the measurement sensor technology.

Direct georeferencing using GNSS and IMU measurements strongly depends on the GNSS reception. Even with only short GNSS absences, quite strong drifts occur, which can degrade position accuracy from the cm- to the dm- or m-range – even with high-end equipment. Therefore, direct georeferencing without further support is only suitable for outdoor environments with good GNSS reception.

Since LiDAR scanners are active sensors, LiDAR-based georeferencing methods are less dependent on environmental conditions such as lighting or surface texture than image-based georeferencing methods. However, LiDAR-based georeferencing methods rely on environments with unique geometric structures that provide a sufficient number of constraints, since they mainly use geometric features (e.g. lines, planes or surfels) for pose estimation (see Table 2.2).

In addition, the georeferencing methods differ in the reference frame used. Direct georeferencing operates in a global geodetic reference frame, whereas LiDAR or visual odometry generally use local 3D coordinate frames. Also most SLAM and SfM algorithms operate in local 3D coordinate frames, which usually start from the first sensor position. There exist also implementations with GCP support, enabling the use of existing local or global reference frames.

SLAM and SfM provide global map consistency since SLAM algorithms include loop-closure detection and re-localization techniques for global optimization, while SfM uses common features and tie points that are optimized within a bundle-adjustment. Since the direct georeferencing depends on GNSS updates, it also provides global consistency. By contrast, visual as well as LiDAR odometry algorithms are not globally consistent since they focus on local optimizations and do not respect long-term deviations and drifts.

By general, the local accuracy potential for SfM and visual and LiDAR-based odometry depends on sensor resolution and accuracy, reconstruction, feature quality and feature distribution. Whereas SfM generally requires high-resolution images and uses sophisticated image features, its accuracy potential lies in the cm- or even mm-domain. By contrast, visual odometry algorithms tend to use low-resolution images and simple features, resulting in low computational demand but also in a lower local accuracy, which typically lies in the dm- or cm-domain.

Furthermore, the global accuracy of visual and LiDAR-based SLAM is additionally affected by the map resolution and by local distortions that are introduced by loop-closures.

Generally, LiDAR-based methods have lower drifts than image based methods due to higher measuring distances. Loop-closures (SLAM) and the use of GCPs (SfM) significantly reduce drifts. In odometry and GNSS-absent direct georeferencing, drifts extend along the trajectory, while in SLAM and SfM without reference points they predominantly occur as a scale across the entire scene. In case of GCP supported SLAM or SfM, drifts tend to occur between GCPs.

In addition, the georeferencing methods differ in the required sensor reception frequency, which is strongly correlated with the resulting trajectory density. While direct georeferencing as well as SLAM and odometry algorithms require high reception frequency, SfM algorithms cope with a medium to lower frame rate.

Odometry and SLAM algorithms are optimized for real-time applications, such as platform navigation, and require low computational time. Direct georeferencing is real-time capable and is also suitable for platform navigation. By contrast, SfM pipelines focus on precision and accuracy, usually processing high-resolution imagery. Hence, their processing time significantly exceeds the time required for data acquisition, so that they are not real-time capable.

Table 2.2: Qualitative comparison of the georeferencing methods discussed. Pluses mark positive effects and minuses indicate negative effects.

Sensors	**GNSS & IMU**	**LiDAR-based**		**Image-based**		
Methods	Direct	Odometry	SLAM	SfM	Odometry	SLAM
Environmental requirements	GNSS reception	Geometric variety		Radiometric texture		
Robustness	−−	+	+	++	+	+
Reference frame	global	local	local or global	local or global	local	local or global
Consistency	+	−	+	+	−	+
Local accuracy range	dm–cm	cm	dm–cm	cm–mm	dm–cm	dm–cm
Drifts	0, (INS only: − − −−)	−−	−	−	− − −	−−
Required reception frequency	GNSS: low, IMU: high	high	high	low – medium	high	high
Processing t_p vs. acquisition t_a time	$t_p < t_a$	$t_p \leq t_a$	$t_p \leq t_a$	$t_p > t_a$	$t_p \leq t_a$	$t_p \leq t_a$

As the different sensors and georeferencing methods have partly complementary characteristics, strengths, and weaknesses, there is great potential to improve accuracy, robustness, and reliability when combined. Zhang, Singh (2018) propose a LiDAR-visual-inertial-based odometry and mapping algorithm with high robustness and low drifts, while Chen et al. (2021b) provide a LiDAR and visual-combined SLAM algorithm. Both use multi-profile LiDAR scanners combined with low-resolution, high-frame-rate cameras to generate point clouds in real-time.

Cavegn et al. (2018) improve resulting camera poses from direct georeferencing and from LiDAR SLAM by subsequent image-based georeferencing with an extended SfM pipeline. By contrast, Eugster et al. (2012) support direct georeferencing during GNSS outages by external updates from stereo-based bundle-adjustment and GCP measurements.

However, there is still great potential for novel robust and accurate georeferencing methods combining GNSS, INS, multi-profile LiDAR and high-resolution imagery to create an image-focused map, which is similar to geospatial image-spaces proposed for infrastructure management.

2.4.5 Quality Measures for Quantitative Comparision

This subsection briefly introduces essential quality measures, which we use for the quantitative evaluation of different georeferencing methods. While Luhmann et al. (2019) defines precision and accuracy, we additionally distinguish between absolute and relative accuracy.

Precision defines the closeness of independent measurements to each other. Thus, it denotes a measure of statistical variability. Luhmann et al. (2019) further defines precision as *internal accuracy* which mainly indicates random errors.

Absolute Accuracy defines the closeness of measurements to the true value or to ground truth, which is often estimated with superior accuracy and thus regarded as the true value. Luhmann et al. (2019) further defines (absolute) accuracy as *external accuracy* that additionally indicates systematic errors.

Relative Accuracy defines the local consistency of positions. Thus, it denotes a measure of positional consistency between for example a point and another near point. The relative accuracy could also be defined as *local accuracy*. This indicator has a practical relevance, e.g. when the accuracy of local distances or area measurements has to be quantified, which are not affected by locally occurring systematic errors.

Chapter 3

Design and Calibration of a Portable Image-focused Mobile Mapping System

This chapter includes the design, the development and the calibration of our portable image-focused MMS. It combines and complements content concerning the system development from the following conference papers:

- Blaser, S., Cavegn, S., Nebiker, S., 2018a. Development of a Portable High Performance Mobile Mapping System Using the Robot Operating System. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1/1, 13–20. DOI: 10.5194/isprs-annals-IV-1-13-2018.
- Blaser, S., Meyer, J., Nebiker, S., Fricker, L., Weber, D., 2020. Centimetre-Accuracy in Forests and Urban Canyons – Combining a High-Performance Image-Based Mobile Mapping Backpack with new Georeferencing Methods. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-1-2020, 333–341. DOI: 10.5194/isprs-annals-V-1-2020-333-2020.

The overall system calibration is mainly based on the methodology of the following journal article:

- Blaser, S., Nebiker, S., Cavegn, S., 2018b. On a Novel 360° Stereo Panoramic Stereo Mobile Mapping System. *Photogrammetric Engineering & Remote Sensing*, 84(6), 347–356. DOI: 10.14358/PERS.84.6.347.

The design and calibration of a portable image-focused MMS is a key component and a prerequisite for developing and investigating novel georeferencing approaches that are robust and accurate across different environments. The design and functionality of the complex multi-sensor system strongly affects both possible applications and the feasibility of various data evaluation capabilities.

Section 3.1 discusses the requirements on the MM platform, the sensors and the configuration for novel georeferencing approaches with high accuracy and robustness across multiple environments as well as for subsequent mapping with respect to creating geospatial 3D image spaces with georeferenced RGB-D image sequences.

Section 3.2 treats the sensor evaluation for our portable image-focused MMS. This includes on the one hand, the degree of fulfilment for each sensor and on the other technical data. The

specifications and relevant technical data of each sensor used as well as their potentials and limitations are presented.

Section 3.3 describes and discusses the finally implemented hardware and sensor configuration in detail. This concerns the platform, the fixation and orientation of the sensors and also the coverage and the sensor footprints.

Section 3.4 covers the sensor synchronization, which is essential for kinematic data acquisition with MMS, while Section 3.5 briefly shows the modular and graph-based design of our data acquisition software.

Finally, Section 3.6 includes the overall system calibration of the portable image-focused MMS in test fields.

3.1 Portable Mobile Mapping System Requirements

This section presents and discusses in detail the requirements for a portable and image-focused MMS for infrastructure management in challenging, non-drivable areas. As shown and discussed in Section 2.1, cloud-based web services that provide 3D geospatial image-spaces from MM data for infrastructure management have been proven and are already widely used. However, due to the system size of image-based MMS in use, data acquisition was so far limited to drivable areas. Therefore, our aim is to extend data acquisition to non-drivable areas, while achieving comparable georeferencing as well as mapping quality in order to enable seamless extension of existing infrastructure services.

Some academic and commercially available portable MMS already exist. However, most of them are LiDAR-focused and their cameras or camera configurations do not meet the requirements of high-resolution RGB-D images. In addition, the few commercial image-focused MMS tend to be closed systems, not allowing changes to the sensor configuration or the acquisition and post-processing software, nor providing access to raw sensor data. Thus, further investigations and improvements on the sensor configuration as well as on georeferencing and post processing algorithms are very limited or impossible. For these reasons, the in-house development of an image-based backpack MMS prototype was inevitable.

The developed MMS should be image-focused and ideally provide high-resolution and panoramic RGB-D image data to meet the mapping requirements for infrastructure management. Furthermore, it should ideally provide similar overall accuracy in different and in challenging environments as state-of-the-art image-based MMS on vehicle platforms, lying in the cm-domain. Ideally, it should offer the same accuracy and data quality across different environments. The developed MMS should have a compact platform to cover non-drivable areas, which may be expanded flexibly with sensors as needed.

Subsection 3.1.1 discusses the platform requirements in detail, while Subsection 3.1.2 includes the requirements for navigation and georeferencing. Finally, Subsection 3.1.3 treats the requirements for mapping.

3.1.1 Platform Requirements

There are several demands on the MM platform in terms of flexibility, versatility and ergonomics. However, there are also conflicting requirements that demand compromises.

Investigations on the sensor configuration and rapid advances in sensor technology require a flexible platform design that allows the rearrangement of sensors as well as the integration of additional sensors and components. This requires a modular hardware design.

The MM platform has to be versatile, enabling data acquisition in different environments. This should include, as far as possible, all places accessible to humans and should also include obstacle-rich environments. The platform therefore needs to have a small and convenient shape and size, which ideally does not exceed the size of a person.

Furthermore, the platform has to be stable and appropriately sized, with enough payload for high-end sensors. However, the MM platform should be as ergonomic as possible and convenient enough for larger campaigns. Thus, a high system weight and a large form factor are rather disadvantageous.

As already stated in Subsection 2.2.4, one may distinguish the following portable MM platform types:

- trolley
- backpack
- hand-held

Obviously, hand-held MM platforms allow the greatest degree of freedom in data acquisition. The large degree of freedom allows the acquisition of narrow areas which are inaccessible to other platform types. However, they have severely limited space and payloads which often excludes the integration of high-end sensors.

By contrast, trolley MM platforms are less restrictive on sensor size or sensor weight, so that high-end sensors can be physically installed. The facts, that the trolley is usually moved in a regular motion along the floor and the degree of freedom in height is eliminated, support the uniform data acquisition in terms of data density as well as data quality. Trolley platforms enable highly efficient acquisition of high quality data at the same time. They are suitable for capturing large-area objects with few obstacles (e.g. airports, museums or other public areas). However, obstacle-rich environments (e.g. construction sites or stair cases) are inaccessible.

Backpack MM platforms allow both the integration of high-end sensors and flexible data acquisition. However, they are limited in the possible sensor configurations as the height of the sensor array is not visible to the operator. Thus, if the height of the sensor assembly exceeds the head height of the operator, this could become critical and prone to damage when capturing indoor environments. In addition, the operator obscures the field of view to the front for lower sensor assemblies. Nevertheless, the backpack platform fulfills our requirements best in terms of flexibility in sensor configuration, versatility and ability to be equipped with high-end sensors as well as ergonomics.

3.1.2 Navigation and Georeferencing Requirements

Creating georeferenced RGB-D images for infrastructure management does not require real-time georeferencing. Therefore we focus in this thesis on post-processed georeferencing methods. The navigation solution is only necessary for platform control.

From this point of view, the requirements for platform navigation are quite different from the requirements for georeferencing. Intended applications relying on a real-time platform navigation solution are geometry-based camera triggering and real-time progress control. Both applications require high pose frequency and low latency, while the accuracy is secondary and should preferably be in the sub-m range.

By contrast, post-processed georeferencing should be robust across various environments and provide accuracy in the cm-domain. In order to achieve a high level of robustness, different sensor types for different georeferencing methods are required to complement each

other's strengths and weaknesses in different environments. To achieve high accuracy with MMS, precise sensor synchronization is essential. It should be sufficiently precise that the remaining delays have a negligible effect on the accuracy, which strongly depends on the planned platform operation speed. Furthermore, it is essential that all raw sensor data is stored during acquisition and is accessible for evaluation.

For navigation and georeferencing in outdoor environments with good GNSS reception, we consider to use direct georeferencing, combining GNSS and IMU sensor data. To achieve an overall accuracy in the cm range under optimal conditions in the outdoor area, a (multi-) phase GNSS receiver and a tactical grade IMU are required.

For indoor navigation and georeferencing, we propose using LiDAR SLAM. LiDAR SLAM seems to be promising to overcome the lack of GNSS. Furthermore, since LiDAR sensors directly provide 3D geometry, LiDAR SLAM tends to require significantly less computational resources than visual SLAM and performs in real-time. Consequently, LiDAR SLAM may be used for platform navigation in indoor and in GNSS-denied areas.

We propose using the high-resolution images for both mapping and georeferencing. High-quality sensor technology used for mapping performs in darker environments than low cost cameras, which are commonly used for visual SLAM. Thus, image sensor may be better suited for high-precision and computationally intensive SfM with bundle-adjustment.

3.1.3 Mapping Requirements

As previously introduced, we intended to develop an image-focused MMS in order to obtain georeferenced RGB-D imagery. Thus, the cameras are the primary mapping sensors, which may be supplemented by other secondary mapping sensors.

The cameras should cover the environment as completely as possible and simultaneously deliver accurate and temporally coherent depth information. In a purely image-based implementation, these two requirements conflict with each other.

For example, the use of multiple pre-calibrated stereo-bases could provide accurate and temporally coherent depth information. However, multiple stereo camera configurations are not suitable for backpack platforms due to limited platform space and operator occlusions. In addition, the limited platform space severely limits or precludes necessary base lengths for sufficiently accurate depth reconstruction. Besides, multiple stereo cameras may lead to configurations with very different projection centers, which in the case of a panoramic projection leads to large stitching errors.

Another example is the use of a panoramic camera, which is the most comprehensive way to capture the environment. However, a single panoramic camera does not provide additional depth information. Although, depth information could be obtained with SfM, using parallax from the movement, or by using additional LiDAR scanners. However, the depth information would not be temporally coherent with the source images. Nevertheless, LiDAR and image sensors could be ideal sensor combinations for robust depth reconstruction due to their complementary characteristics.

Successful image-based depth reconstruction and precise image measurements rely on precisely calibrated and stable IO parameters as well as on stable image geometry. Thus, the cameras used should feature fixed focal length and global shutter. In addition, they should not apply automatic image corrections that affect the IO, such as image stabilization.

On the one hand, the cameras should capture the environment as detailed as possible in high resolution with small ground sampling distances. One the other hand, the cameras must also operate under low light conditions for example in indoor and underground campaigns. These

two requirements are opposite, as a low light sensor requires a certain pixel size to capture enough photons in low light conditions, which stands in contrast to fine sensor subdivisions of high resolution sensors. Another option would be to use a larger sensor, resulting however in larger optics and a higher system weight. Consequently, an optimal balance between the two requirements has to be found.

3.2 Sensor Evaluation

This section contains the evaluation of the sensors and the main hardware components for the portable image-focused MMS. This concerns the sensors and components for navigation, mapping, data registration and data preprocessing.

Subsection 3.2.1 treats the evaluation of the multi-head panoramic camera *FLIR Ladybug 5* as the primary mapping sensor and presents its relevant specifications.

Subsection 3.2.2 discusses the evaluation and shows specifications of the multi-beam LiDAR scanner *Velodyne VLP-16 Puck* which is used as a navigation as well as a mapping sensor.

Subsection 3.2.3 the GNSS- and IMU-combined navigation unit *NovAtel SPAN CPT 7*, which is used as the principal navigation sensor and as the reference clock.

Finally, Subsection 3.2.4 shortly describes the high-performance mini computer used for data registration and preprocessing.

3.2.1 Panoramic Camera

The panoramic camera serves as the primary mapping sensor in our portable MMS. As discussed in Section 3.1.3, the image-based primary mapping sensor should meet the following requirements:

1. Cover the environment as completely as possible (desirable)
2. Provide a high image resolution (desirable)
3. Deliver accurate and temporally coherent depth information (desirable)
4. Cover a wide range of different light conditions (desirable)
5. Stable interior orientation (mandatory)
6. Feature precise external sensor synchronization (mandatory)
7. Suitable size and weight for a backpack platform (mandatory)

Numerous cameras and camera configurations partially fulfill our requirements, while panoramic cameras meet the requirements best. By general, panoramic cameras almost completely cover the entire environment (criterion 1), whereas they do not provide any depth information (criterion 3). Multi-head panoramic cameras provide a large overall sensor area, which is composed of multiple sensors. Consequently, multi-head panoramic cameras offer the highest resolution (criterion 2) and often provide a large dynamic range (criterion 4) thanks to large sensor pixel sizes.

Most multi-head panoramic cameras available on the market are in the consumer range and have been specifically designed for AR and Virtual Reality (VR) applications. However, they do not offer interfaces for precise external sensor synchronization (criterion 6).

By contrast, industrial multi-head panoramic cameras provide numerous interfaces for precise triggering and sensor synchronization and feature stable IO.

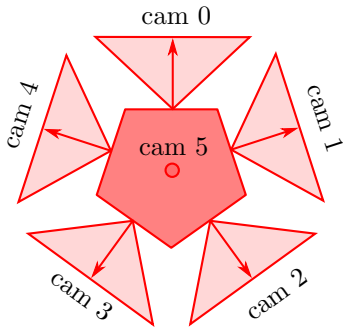
However, there are only few industrial multi-head cameras available. Apart from its successor model *FLIR Ladybug 5+*, the *FLIR Ladybug 5* multi-head panoramic camera currently is the only industrial grade high-resolution multi-head panoramic camera on the market. Since the *Ladybug 5+* does not offer significant technological advances, we decided to use the *Ladybug 5*.

The *FLIR Ladybug 5* consists of six camera heads, whereby five of them point sideways and the sixth camera head points upward (Figure 3.1a & b). All individual camera heads feature *Sony ICX655* CCD sensors with a resolution of 5 MP and a sensor pixel size of 3.45 µm (see Table 3.1).

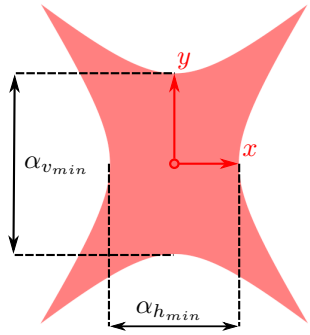
The camera heads feature wide-angle optics with a focal length of 4.4 mm. The wide-angle optics strongly distort the image, so that the pinhole camera model in combination with the Conrady-Brown distortion model (Brown, 1971) cannot adequately approximate the distortions. Nevertheless, previous internal research has shown that the equidistant fish-eye camera model (Abraham, Förstner, 2005) approximates the strong distortions the *Ladybug 5* camera heads best.

In contrast to the perspective camera model, the horizontal and vertical opening angles are not constant (see Figure 3.1c). Appendix B.1 includes derivations on the opening angles of the equidistant camera. The diagonal viewing angle α_d amounts to 143°. The minimal horizontal viewing angle $\alpha_{h_{min}}$ amounts to 92° and the minimal vertical viewing angle $\alpha_{v_{min}}$ amounts to 110° (see Table 3.1).

(a) Oblique view of the *FLIR Ladybug5* multi-head panoramic camera (FLIR Inc., 2017).



(b) Horizontal outline of the *FLIR Ladybug5* panoramic camera with labeled camera heads. The bright red triangles show the view frustum of the horizontally pointing camera heads and the red arrows show their viewing direction.



(c) Equidistant image shape from an individual camera head that is projected to the perspective camera model. $\alpha_{h_{min}}$ and $\alpha_{v_{min}}$ mark the minimal horizontal and vertical viewing angle.

Figure 3.1: Multi-head panoramic camera *FLIR Ladybug 5*

In total, the *Ladybug 5* panoramic camera has a resolution of 30 Mpx, while a single panoramic camera head's resolution amounts to 5 Mpx. The overall field of view amounts to 90 % of a full sphere. The focus distance is 200 cm and FLIR Inc. (2017) specify an acceptable sharpness from 60 cm to infinity. Furthermore, the camera has a 12-bit Analog to Digital Converter (ADC), which allows a higher dynamic range in sunlight as well as in shadows.

The General Purpose Input and Output (GPIO) interface has several hardware trigger modes, with a trigger latency of about 8 µs. By contrast, the latency of the software trigger mode, using the Universal Serial Bus (USB) 3.0 interface, amounts to 2 ms (FLIR Inc., 2017).

Table 3.1: Technical data of an individual *Ladybug 5* camera head (FLIR Inc., 2017)

Sensor type	Sony ICX655 CCD
Shutter type	Global shutter
Sensor dimension $u \times v$ in px	2048×2448
Pixel size in µm	3.45
Focal length c in mm	4.4
Diagonal viewing angle α_d in °	143
Min. horizontal viewing angle $\alpha_{h_{min}}$ in °	92
Min. vertical viewing angle $\alpha_{v_{min}}$ in °	113

The camera has a diameter of 197 mm and a height of 160 mm, whereby the weight amounts to 3 kg.

In the overall view, the *FLIR Ladybug 5* panoramic camera fulfills the requirements for our portable MMS. Particularly advantageous are the field of view, the high dynamic range as well as the hardware trigger interface which provides numerous options.

3.2.2 Laser Scanner

The LiDAR scanner serves as a navigation sensor as well as a secondary mapping sensor in our portable MMS. As discussed in Sections 3.1.2 and 3.1.3, the laser scanner should meet the following requirements:

1. Cover the environment as completely as possible (desirable)
2. Provide a high resolution (desirable)
3. Provide a high data acquisition rate for navigation purposes (mandatory)
4. Provide an accuracy in the cm-range (mandatory)
5. Provide raw data access (mandatory)
6. Provide a precise external sensor synchronization (mandatory)
7. Suitable size and weight for a backpack platform (mandatory)

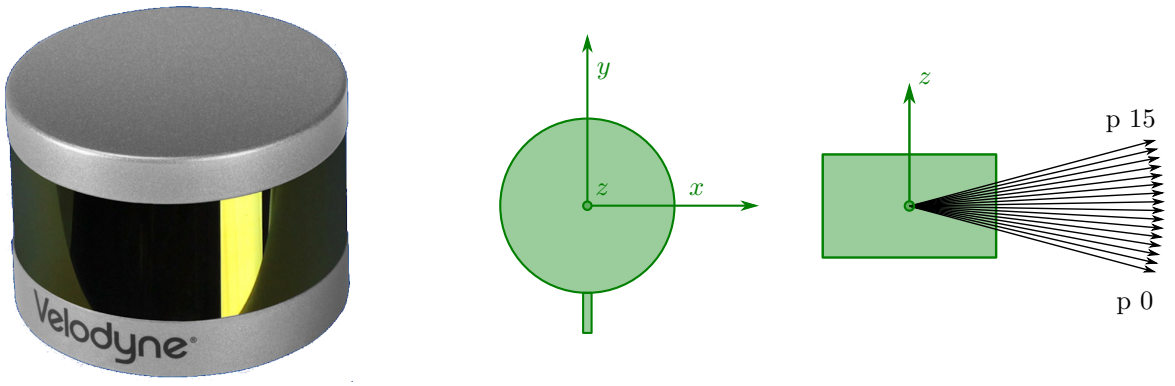
There exist numerous LiDAR scanners, which partially fulfill our requirements. For example, high-end Terrestrial Laser Scanner (TLS) usually provide high resolution as well as high precision. In a static application, they usually cover the entire environment. However, their size and weight exceed the backpack platform limitations (criterion 7). Since they are designed for static application, they usually do not provide any interfaces for external sensor synchronization (criterion 6). Furthermore, TLS often deny raw data access and only provide the resulting point clouds (criterion 5).

Profile LiDAR scanners and multi-profile LiDAR scanners are often used in robotics and for autonomous platforms. They are primarily designed for mobile applications and thus have a small size and low weight (criterion 7). In addition, they offer a high data acquisition rate (criterion 3), an external sensor synchronization, and provide access to raw data. However, the field of view is often limited by the number of beams and by the high acquisition rate.

The multi-profile LiDAR scanner *Velodyne VLP-16 Puck* meets our requirements best (see Figure 3.2). It was primarily conceived for collision avoidance applications in the automotive industry, and thus has a compact and rugged design and ranges from the medium- to low-cost price segment.

The LiDAR scanner *Velodyne VLP-16 Puck* consists of 16 different LiDAR modules, which are fixed on a rotating module, resulting in 16 profiles (see Figure 3.2c). The module rotates with an adjustable speed between 5 and 20 revolutions per second (Velodyne, 2016). This results in a full horizontal Field of View (FoV) (see Table 3.2). The horizontal resolution is limited by the detection rate of the LiDAR modules and by the rotation rate used. By contrast, the vertical FoV and the vertical resolution are limited by the arrangement of the 16 LiDAR modules.

The *Velodyne VLP-16* supports precise synchronization from an external source, using a Pulse-Per-Second (PPS) signal in conjunction with a one-per-second National Marine Electronics Association (NMEA) sentence (Velodyne, 2016).



(a) Oblique view (Velodyne, 2016).

(b) Horizontal outline with the sensor coordinate frame.

(c) Outline with the LiDAR profiles.

Figure 3.2: Multi-beam LiDAR scanner *Velodyne VLP-16 PUCK*

Table 3.2: Specifications of the LiDAR scanner *Velodyne VLP-16 Puck* (Velodyne, 2016)

Max. range in m	100
Typical accuracy in cm	3.0
Number of channels	16
Angular resolution (H × V) in °	0.1 to 0.4 × 2.0
FoV (H × V) in °	360 × 30
Rotation rates in Hz	5 to 20
Max. points per s	300 000

In the overall view, the *Velodyne VLP-16 Puck* multi-profile LiDAR fulfills the requirements for our portable MMS. Its high data acquisition rate, compact dimensions and low weight are particularly advantageous. Furthermore, precise external sensor synchronization and raw sensor data access enable mobile applications.

However, the specified typical accuracy of 3 cm is at the lower limit of the acceptable. In addition, a higher vertical resolution as well as an extended vertical field-of-view would be desirable.

3.2.3 Inertial Navigation System

The INS serves as a navigation sensor in our portable MMS. As discussed in Section 3.1.2, the INS should meet the following requirements:

1. Provide direct georeferencing, combining GNSS and IMU (mandatory)
2. Provide tactical grade performance (desirable)

3. Allow external position and attitude updates for sensor data fusion (mandatory)

4. Provide precise sensor synchronization (mandatory)

5. Provide access to sensor raw data (mandatory)

6. Provide sensor raw data without system initialization (desirable)

7. Suitable size and weight for a backpack platform (mandatory)

The GNSS and IMU-combined INS *NovAtel SPAN CPT 7* meets our requirements best. The system includes the MEMS-based IMU *Honeywell HG4930P* with tactical grade performance and high robustness as well as a GNSS dual antenna receiver.

The specifications state a position accuracy of 10 mm horizontally and 20 mm vertically under good GNSS coverage and after post-processing (NovAtel Inc., 2020). The accuracy of the attitude angles roll Φ and pitch Θ is specified as 0.005° and the heading Ψ as 0.010° (see Table 3.3). A GNSS outage of 60 s degrades the horizontal accuracy to 150 mm, the vertical accuracy to 50 mm, the roll Φ and pitch Θ attitude accuracy to 0.007° and the heading accuracy to 0.012°. The system use in the navigation mode with single GNSS position and without correction data provides a position accuracy in the m-range and an attitude accuracy in the range from 0.010° to 0.055°.

Table 3.3: Performance specifications during GNSS outages of the INS *NovAtel SPAN CPT7* (NovAtel Inc., 2020)

Positioning mode	**Outage duration** in s	**Position RMSE H** in m	**Position RMSE V** in m	**Attitude RMSE Φ or Θ** in °	**Attitude RMSE Ψ** in °
Single position	0	1.20	0.60	0.010	0.030
	10	1.30	0.65	0.020	0.040
	60	5.10	1.30	0.030	0.055
Post processing	0	0.01	0.02	0.005	0.010
	10	0.01	0.02	0.005	0.010
	60	0.15	0.05	0.007	0.012

In addition, the *NovAtel SPAN CPT 7* provides GNSS RTK which significantly improves the accuracy in the navigation mode. In our case, the single GNSS position accuracy meets the requirements for the navigation solution. For this reason, RTK accuracy is not listed in Table 3.3 but is specified in NovAtel Inc. (2020).

Furthermore, the *NovAtel SPAN CPT 7* provides raw data access without any prior system initialization, since the system consists of an internal as well as a precise system clock. Moreover, numerous sensor synchronization interfaces are available.

The connected GNSS dual quadrifilar helix antenna *HX-CHX600A* enables low noise signal reception of satellites even with a low elevation angle. Furthermore, the manufacturer promises exceptional coverage in areas with weak signals and signal reflections.

Finally, the *NovAtel SPAN CPT 7* fulfils our requirements on a primary navigation sensor, since it is compact and lightweight (450 g) and its performance is comparable to tactical-grade navigation systems used in MMS vehicles with fiber-optic gyroscopes (cf. *UIMU-LCI* in Cavegn et al. (2018)).

3.2.4 Computer

Our MMS includes the powerful mini computer *Prime Mini Pro* with fast data storage capacities above 2 TB on Solid State Disk (SSD) as well as high Random Access Memory (RAM) resources of 32 GB (see Table 3.4). In addition, the computer has a low energy consumption and no mechanical moving parts.

The Intel processor with the widely used *x64* or *amd64* architecture is advantageous in terms of the available hardware drivers and libraries, compared to the *ARM* processor architecture of most embedded computers.

The standard high-performance interfaces USB 3.0 and Gigabit Ethernet (GigE) allow transmission and reception of large amounts of sensor data. However, the used mini computer lacks on manually configurable pins, which could be useful for hardware-based camera triggering and sensor synchronization.

Table 3.4: Specifications of the portable MMS computer *Prime Mini Pro* (Prime Computer AG, 2017)

Processor	Intel Core i5-5300U vPro
Graphic card	Intel HD Graphics 5500
RAM	32 GB DDR3L – 1600 MHz
Data Storage	1 × 500 GB SSD 1 × 2000 GB SSD
Interfaces	4 × USB 3.0 1 × GigE 2 × Mini-DP 1.2

3.3 Sensor Configuration

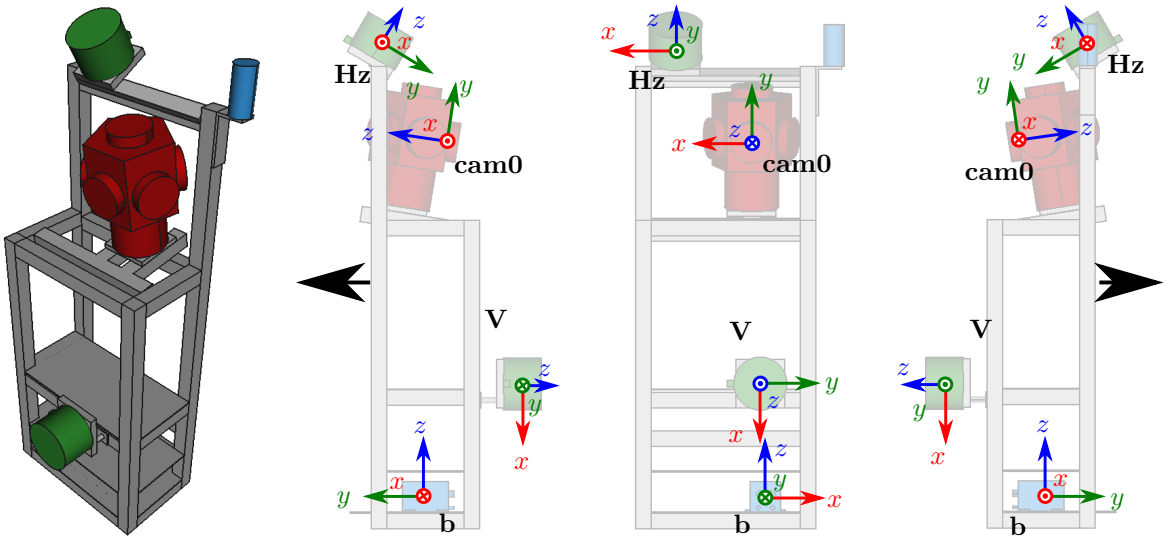
This section focuses on the sensor configuration of our backpack MMS by describing the MM platform and the sensor configuration in general. Subsection 3.3.1 discusses the realized camera configuration and Subsection 3.3.2 treats the realized LiDAR scanner configuration in detail.

The MM platform consists of a robust aluminium frame which is fixed on a load carrier backpack. The aluminium frame comprises super lightweight 30 mm aluminium extrusions that can be flexibly arranged. It provides a modular and rugged design for stable sensor mounting, by enabling fast and efficient incremental sensor configuration improvements (see figure 3.3).

We fixed the navigation system *NovAtel SPAN CPT7* with IMU and GNSS receiver on the bottom of the frame, where the sensor can be fixed in a stable and protected way. The GNSS antenna sits on top of the frame, providing optimal GNSS coverage and reception.

The multi-head panoramic camera *FLIR Ladybug 5* is mounted on the top so that it slightly overlooks the operator's head to provide forward visibility. At the same time, the camera should not be attached too high to avoid collisions with overhead obstacles such as door frames. In addition, we tilted the panorama camera slightly backwards by a few degrees to compensate the operator's forward bending. This results in a roughly horizontal camera plane (see Figure 3.3).

One of the two LiDAR scanners sits on top of the frame and is tilted by 30° to cover mainly the walls and some parts of the floor and ceiling. By contrast, the second vertically mounted LiDAR scanner mainly covers the floor and ceiling.



(a) View from the top back right (b) View from the left (c) View from the back (d) View from the right

Figure 3.3: CAD outline of the *BIMAGE Backpack*, showing the sensor configuration. Big black arrows mark the moving direction. Bold labels **b** (body frame), **Hz** (horizontal LiDAR), **V** (vertical LiDAR) and **cam0** (panoramic camera) represent the coordinate frames. Point symbols in the coordinate frame origin represent backward-pointing axes, while cross symbols mark forward-pointing axes.

Figure 3.4 shows the final sensor configuration, including all components. Generally, heavy components, such as the battery, are located at the bottom and nearest to the operator's back. This keeps the center of gravity as low and as close to the back as possible and improves ergonomics.

The backpack's overall weight amounts to 25 kg. Significant weight reductions could be achieved in future by simplifying and adapting the sensor frame to the final configuration as well as by optimizing hardware interfaces and by reducing unnecessary cable lengths.

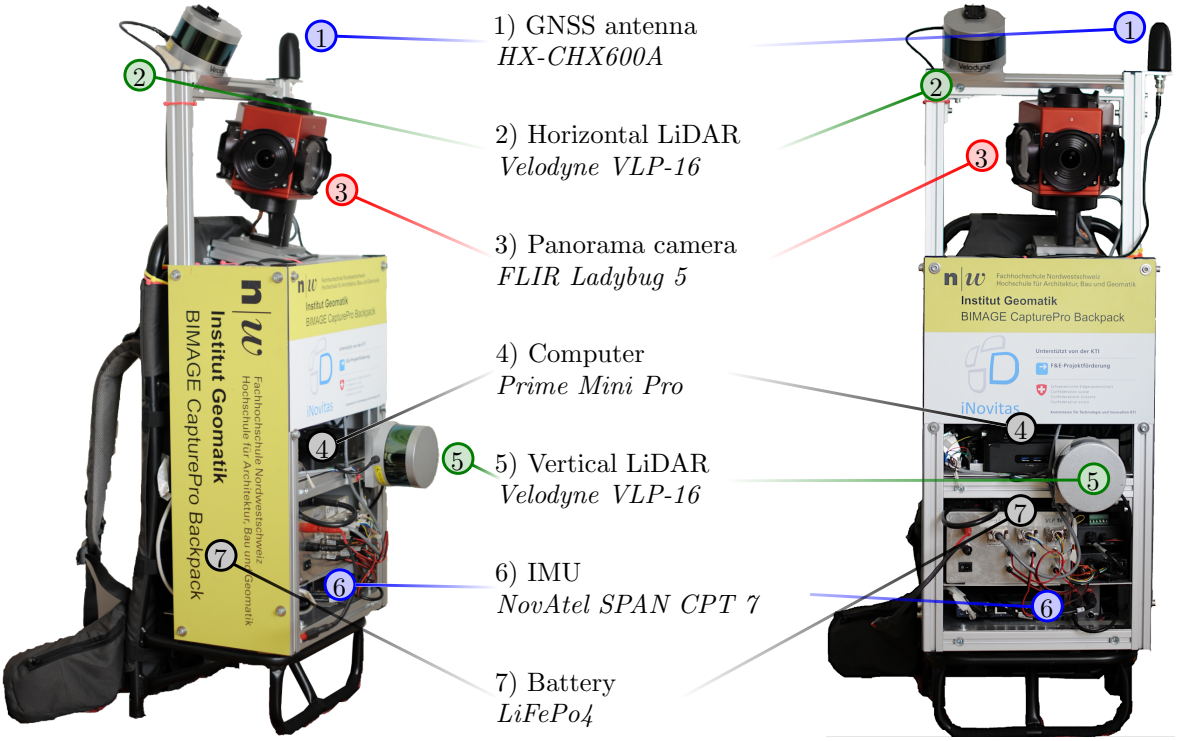


Figure 3.4: *BIMAGE Backpack* system setup for indoor and outdoor applications. Sensors marked with blue lines are mainly used for direct georeferencing, green lines show sensors used for LiDAR SLAM-based georeferencing and red lines components used for image-based georeferencing. Black lines indicate supporting components.

3.3.1 Camera Configuration

The multi-head panoramic camera *FLIR Ladybug 5* is located on top of the backpack frame, so that the camera slightly overlooks the operator's head.

The main panoramic camera head *cam0* points opposite to the walking direction. The sub-ordered panoramic camera heads *cam1 – cam4* are horizontally arranged, so that they form a regular pentagon with *cam0*. By contrast, the fifth sub-ordered panoramic camera head *cam5* points upwards (see Chapter 3.2.1 and Figure 3.1). Figure 3.5 shows a simulated panoramic camera image epoch, consisting of six images from the individual panoramic camera heads.

The system weight of 25 kg causes a natural forward-bending of the operator's upper body. Therefore, we slightly backwards-tilted the panoramic camera by approximately 8°, which almost compensates the operator's forward-bending, so that the panoramic camera comes to bear horizontally (see Figure 3.4, no. 3).

The overlapping images of the individual panoramic camera heads achieve an almost complete coverage. The backpack frame causes only minor occlusions, while most occlusions are caused by the operator's head (see Figure 3.6).

Mounting the panoramic camera at a higher level, for example directly above the operator's head, would reduce occlusions and provide even more favorable image coverage. However, this would restrict the flexibility of the backpack MMS and cause collision risks with overhead obstacles (e.g. door frames, elevators, etc.).

A significant advantage of the panoramic camera configuration we implemented is that the panoramic camera covers the environment almost completely with only minor occlusions by the frame and the operator. The overlapping images from the individual panoramic camera heads allow stitching them together to a panoramic image. Other advantages are the almost horizontal camera alignment and the camera placement nearby the operator's head. This creates a familiar image perspective from a pedestrian's point of view, which is ideal for creating image-based web services similar to Street-View.

However, the only mono-panoramic configuration does not provide depth information from the same image epoch. Furthermore, the front view is partially occluded by the operator's head. In addition, the panoramic camera mounted on the top limits the ground coverage. However, the ground would be of great interest for many applications, for example for road state assessments

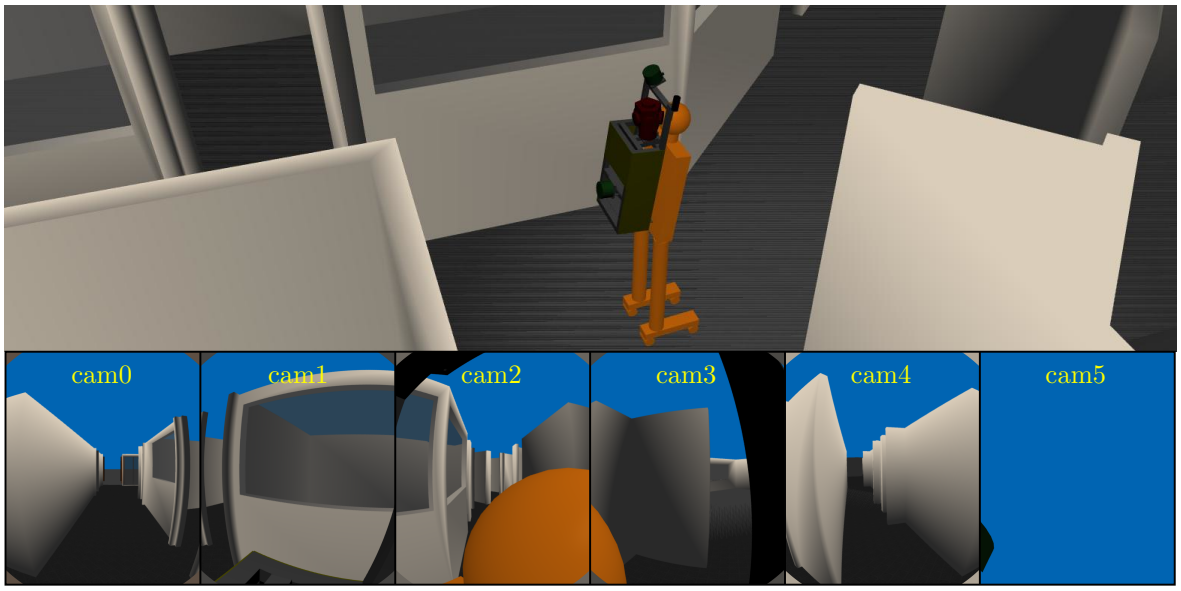


Figure 3.5: Backpack model worn by a dummy person in a virtual environment (top) and the camera views from the individual camera heads (cam0 - cam5) of the *Ladybug 5* panoramic camera (bottom), assuming the equidistant camera model. The simulation was carried out with ROS and with the multi-robot simulator *gazebo* within the *Willow Garage World*.

3.3.2 Laser Scanner Configuration

The backpack MMS consists of two *Velodyne VLP-16 Puck* LiDAR scanners.

The first LiDAR scanner, which we label as "horizontal LiDAR scanner", sits on the top of the MM sensor frame (see Figure 3.4, no. 1). It is tilted backwards by approximately 30° so that it mainly covers the walls and partially covers the ceiling and the floor.

The second LiDAR scanner, which we label as "vertical LiDAR scanner" is vertically mounted on the center of the MM sensor frame (see Figure 3.4, no. 5). It complements the first LiDAR scanner and mainly covers the floor and the ceiling but also some parts of the walls.

In an early stage of development, we only used the horizontal LiDAR scanner. However, combined with LiDAR SLAM, this resulted in geometrical singularities. When using only the horizontal LiDAR scanner within a room with a small plot combined with a heigh ceiling (e.g. staircase), it only scans the walls without the floor and the ceiling (see Figure 3.6, green profiles). Assuming a room with even walls and no other objects, the backpack pose cannot be estimated without LiDAR points from the floor or ceiling. Hence, the vertical LiDAR scanner brings additional geometric stability to the LiDAR slam.

Otherwise, if only the vertical LiDAR scanner was used for LiDAR SLAM, the same phenomenon would occur, but this time in the walking direction. By contrast, for pure mapping applications, it is uncritical to use only the vertical scanner. In addition, this would even be advantageous, as the amount of data generated could be halved.

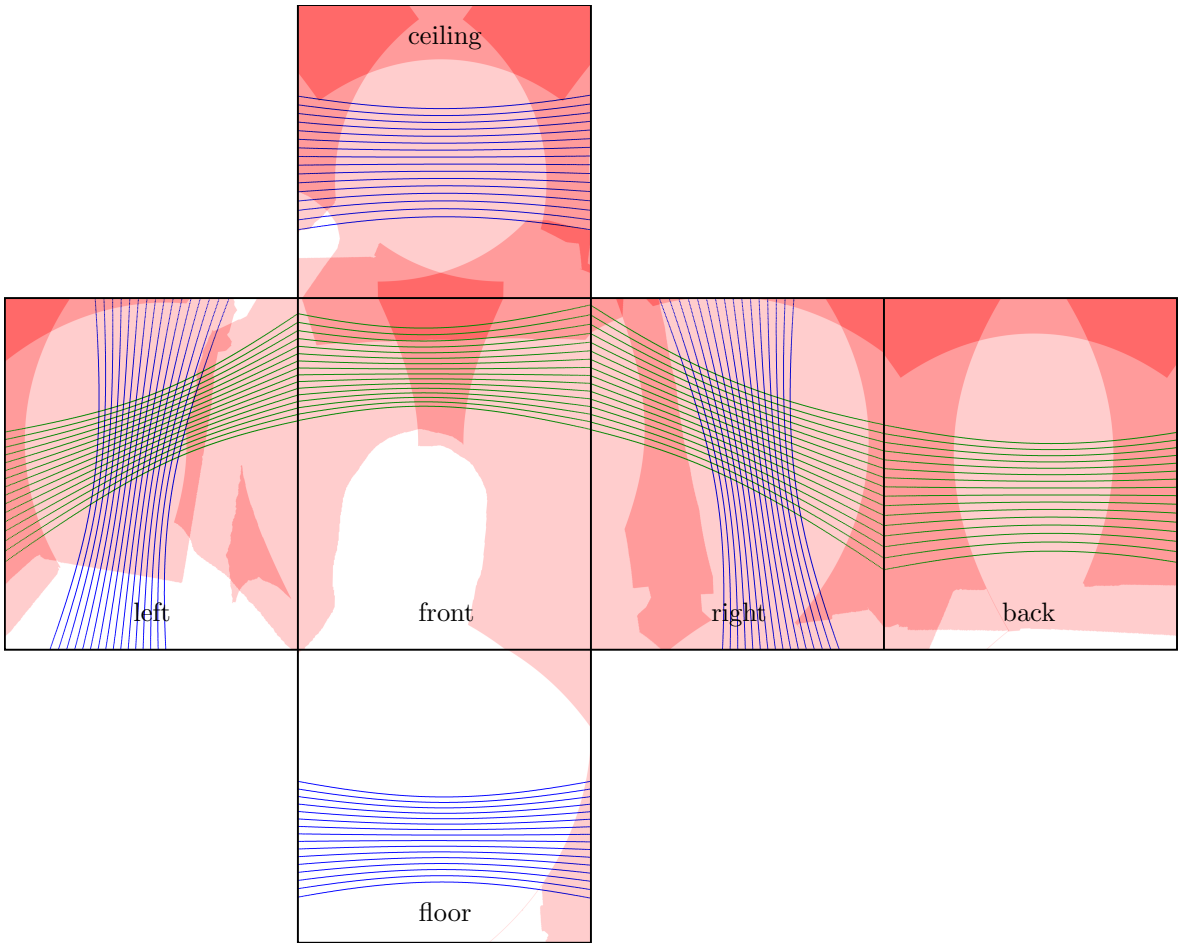


Figure 3.6: Footprint of the *BIMAGE Backpack* MMS sensor configuration within a quadric shaped room with a side length of 2.5 m and with a height of 3 m that is shown unfolded. The body frame of the backpack MMS is horizontally placed in the center of the floor and vertically placed 1.08 m over the floor. The orientation $\omega = -8°$, $\varphi = 0°$ and $\kappa = 0°$ simulates the natural forward bending during carrying. Red areas show the panoramic camera footprint, whereby occlusions by the frame and the operator are respected. Blue lines mark the horizontal LiDAR scanner profiles and green lines show the vertical LiDAR scanner profiles.

3.4 Sensor Synchronization

In order to realize an appropriate sensor synchronization, it is essential to identify time-critical processes that may have a direct influence on the resulting accuracy. However, the distinction between time-critical and non-time-critical processes depends on the application.

In our case, we kinematically capture data with our backpack MMS but we primarily use the post-processed data. Hence, the actions in the field are less time-critical than the captured sensor data.

This section describes our requirements for the sensor synchronization of our system and shows and discusses the implemented solution for every sensor component. Subsection 3.4.1 describes the reference clock of our system, while Subsection 3.4.2 contains the panoramic camera synchronization and Subsection 3.4.3 treats the synchronization of the LiDAR scanners.

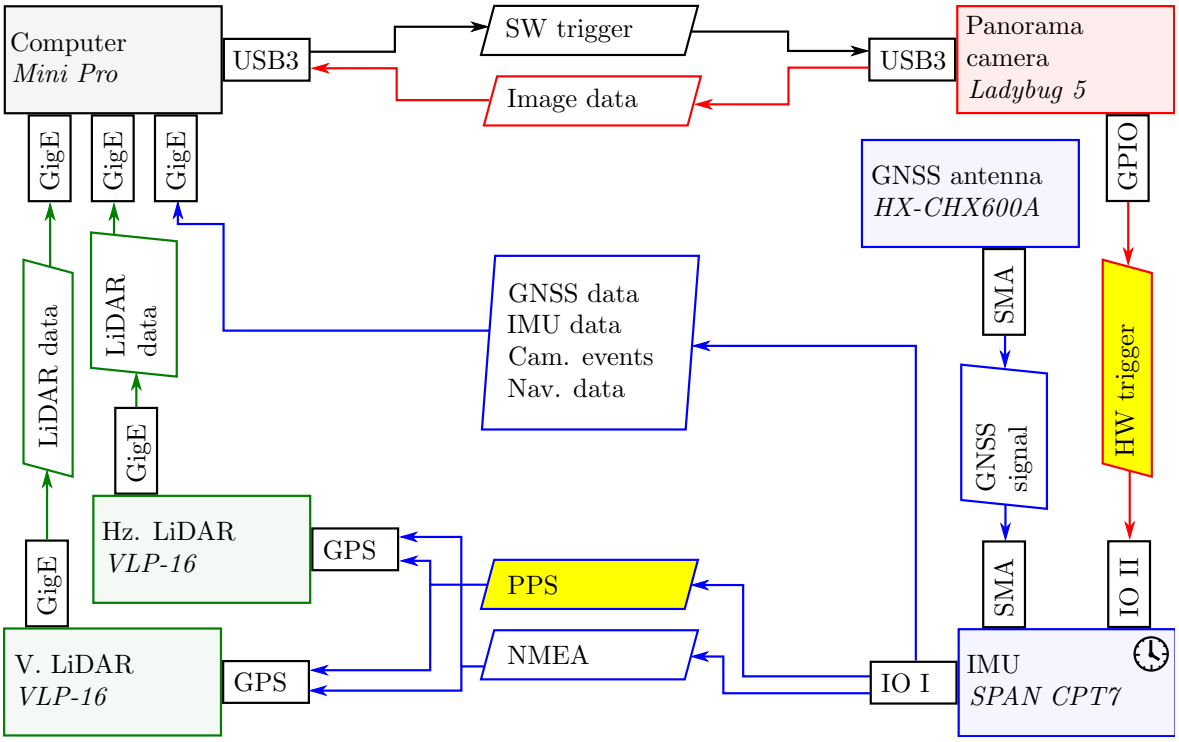


Figure 3.7: Diagram, showing the electronic design of the *BIMAGE Backpack*. The lightly colored squares mark the sensors. Their interfaces are marked as small attached squares outlined in black. Rhombi represent the data exchanged between the components and yellow filled rhombuses mark time-critical data. Data flow and connections are represented with arrows. Blue symbols refer to direct georeferencing, green symbols relate to LiDAR SLAM-based georeferencing, and red symbols refer to image-based georeferencing.

3.4.1 Reference Clock

To synchronize all sensors uniformly, we use a system-wide reference clock. In our system, the navigation unit *NovAtel SPAN CPT7* supplies the system's reference time (see Figure 3.7, bottom right). The reference time provided is calculated as follows:

$$t_\delta = t_{GNSS} - t_{int} \tag{3.1}$$

$$t_{clock} = t_{int} + t_\delta \tag{3.2}$$

where:

t_{int} = Internal receiver time
t_{GNSS} = GNSS time (coarse or fine)
t_δ = Clock offset
t_{clock} = IMU time used as system reference time

The navigation unit continuously estimates the clock offset using the best available GNSS time reference. Then, the estimated clock offset is used to correct the internal receiver time.

The accuracy of the estimated clock offset depends on the GNSS reception. A *coarse time* can be calculated using the GNSS navigation message, which is transmitted every six seconds. NovAtel Inc. (2020) specify the *coarse time* accuracy to the GNSS time with 30 ms.

When a complete GNSS constellation with at least four satellites is available, the so-called *fine time* can be calculated, which is a more precise clock offset with internal compensation. NovAtel Inc. (2020) specify the *fine time* accuracy, using the standard position service, with 20 ns.

3.4.2 Panoramic Camera Synchronization

To use post-processed images of a panoramic camera, kinematically captured by a portable MMS, we need to determine exact triggering timestamps of the panoramic camera in the system reference time. Based on this objective, the entire camera triggering process can be separated into time-critical and non-time-critical operations.

We divide the panoramic triggering process into three different operations:

- sending the triggering command to the camera system
- triggering all cameras of the camera system
- estimating the precise time of the camera system triggering

Time delays during the first operation will affect the distance between two image epochs, but do basically not affect the georeferencing accuracy of the image pose. Thus, we regard the first operation as non-time-critical. By contrast, time delays during the second operation result in non-synchronous single images of the camera system and the third operation obviously directly affects the accuracy of the image poses. Consequently, the second as well as the third operation are time-critical.

We implemented the panoramic camera triggering as follows: First, the computer sends a software-based camera trigger command over the USB 3.0 interface (see Figure 3.7, top). Then, the panoramic camera *Ladybug 5* internally triggers all panoramic camera heads simultaneously. At the time all panoramic camera heads are simultaneously triggered, the

panoramic camera generates an electronic pulse over the GPIO interface to the navigation unit IMU *SPAN CPT7* and triggers the creation of a timestamp (see Figure 3.7, left). Finally, the navigation unit sends the created timestamp to the computer, where the data is stored.

However, in our implementation, we accept that time delays during the trigger command transmission over the USB 3.0 interface from the computer to the panoramic camera might occur. Nevertheless, this might only affect the execution of the trigger settings, so that the chosen temporal or spatial differences between image epochs can vary. The accuracy of the image pose will however not be affected.

3.4.3 Laser Scanner Synchronization

Post-processing of LiDAR data, using direct georeferencing, requires exact firing timestamps for each LiDAR point.

Both *Velodyne VLP-16 Puck* LiDARs contain two counters, which represent the number of µs since the top of the hour. The first counter represents the s since the top of the hour and the second counter shows the µs since the top of the second. Basically, the internal oscillators of the scanners drive the µs counter. If an external PPS source is available, then the LiDAR scanner adjusts the second counter that represents the µs since top of the s. In order to align the first counter with Coordinated Universal Time (UTC) time, the LiDAR scanner requires an additional external NMEA sentence, which provides a valid GNSS time stamp (Velodyne, 2016).

In our backpack MMS, the navigation unit *NovAtel SPAN CPT7* provides both a PPS signal to align the µs counter and a related NMEA sentence to adjust the s counter of the LiDAR scanners. The navigation unit transmits both elements from its *IO I* interface to the *GPS* interfaces of the LiDAR scanners (see Figure 3.7).

Physically, each data block from the LiDAR scanner contains a precise timestamp from both counters, whereby a data block consists of 384 LiDAR points in the single return mode, which were captured within 1306.37 µs (Velodyne, 2016). Within one data block, we interpolate the timestamps of the individual LiDAR points, using the specified firing times.

Consequently, when the LiDAR scanner is synchronized with a PPS and a NMEA sentence, the clock drift depends on the PPS source. By contrast, the clock offset additionally depends on the transmission and reception delay of the PPS signal as well as on the quality of the timestamp of the NMEA sentence. Physically, a clock drift correction is applied discretely at the beginning of each data packet with the length of 384 LiDAR points.

3.5 Acquisition Software

Our acquisition software runs on a mini computer with a Linux Ubuntu 16.04 installation. Each external device with Virtual Network Computing (VNC) support is able to control the mini computer via remote desktop over a WLAN connection.

The acquisition software is based on the ROS framework. Quigley et al. (2009) give an introduction into the principles, paradigms and functionality of ROS. Thanks to the open source philosophy and the wide distribution in robotics, there are numerous existing tools and a comprehensive hardware support. Particularly noteworthy is the flexible graph-based communication concept, consisting of nodes, messages, topics and services. A node describes a software module, which we represent as a circle in our software schema (see Figure 3.8). Nodes can communicate with each other by passing messages in a strictly predefined data structure through given topics. In our software schema, we represent topics as rectangles (see

Figure 3.8). Either a node can publish messages to one or more topics (red arrows) or it can subscribe to one or more topics to receive messages (blue arrows). By contrast, services offer synchronous communication. A node can advertise a service with predefined data structures of both request and response, similar to a web service.

For both LiDAR scanners, we use the *Velodyne driver* from Withley (2016) with minor modifications. The driver converts the sensors raw data into ROS *pointcloud2* messages and publishes them.

For the support of the navigation unit *NovAtel SPAN CPT7* we used the ROS *novatel_span_driver* (Purvis, 2020). We subsequently extended the driver to record raw GNSS and IMU data as well as camera timestamps for post-processing.

The ROS node *Bag* stores the IMU messages as well as the messages from both LiDAR scanners in a so-called ROS bag file. The ROS bag file can be replayed for post-processing.

By contrast, our implemented ROS panoramic camera driver writes image raw data in so-called *pointgrey image stream* files, which might be post-processed for example in the software *LadybugCapPro*.

The 3D LiDAR SLAM *Cartographer* combines LiDAR and IMU data and provides a system pose and a map, which shows the acquisition progress. Hess et al. (2016) introduce the functionality of *Cartographer* exemplarily with their 2D LiDAR SLAM.

Furthermore, we implemented the *Conditional Trigger Node* which supports spatial trigger criteria, such as distance and orientation differences as well as temporal trigger criteria, such as time differences. The Conditional Trigger Node uses the system pose to calculate spatial differences. The system pose comes either from the navigation unit *NovAtel SPAN CPT7* or in indoor use from the LiDAR SLAM *Cartographer*.

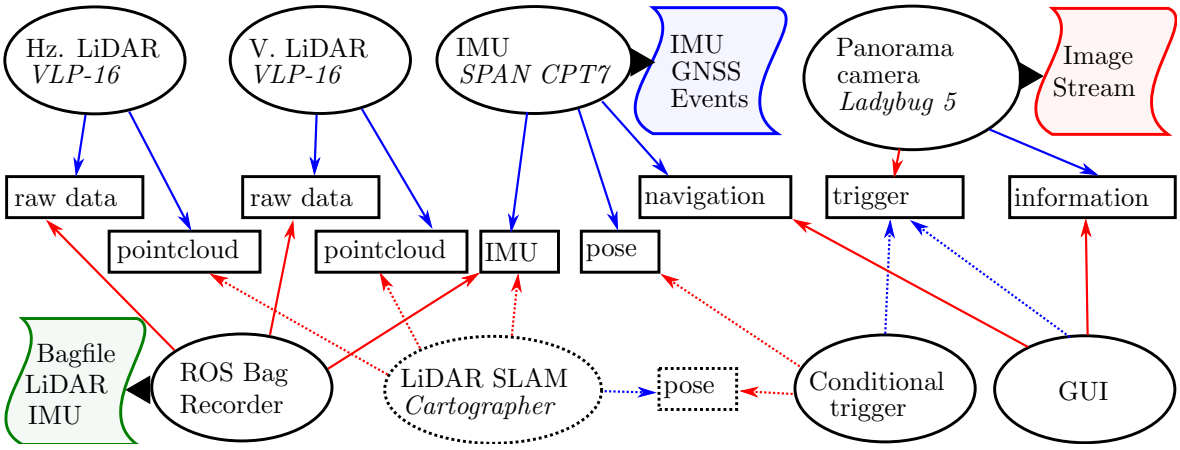


Figure 3.8: Diagram showing the graph-based design of our acquisition software, which is based on ROS, Ellipse-shaped items represent ROS nodes and squares indicate ROS topics. ROS publishers are marked as blue and ROS subscribers as red arrows.

3.6 Overall System Calibration

This section includes the overall system calibration that we performed to estimate the following parameters:

- IO parameters of the individual panoramic camera heads.
- RO parameters that describe the transformation from the sub-ordered panoramic camera heads to the reference panoramic camera head.
- BA parameters, which consist of the lever-arm as well as the misalignment from an individual sensor (e.g. panoramic camera or LiDAR scanner) to the body frame.

Since our backpack MMS incorporates high-end sensors that are mounted on a stable aluminium frame, we do not expect parameter interference, occurring during a campaign, to significantly affect the overall system accuracy.

Thus, we intend to pre-calibrate the aforementioned parameters in test fields. As test field calibrations are performed in well-known and controlled environments, their results are comprehensible and reproducible. The time and effort required for test field calibrations is considerably lower than for laboratory calibrations, although it is possible to achieve superior accuracy with a high degree of reliability, which is not guaranteed with self-calibrations.

However, test field calibrations do not necessarily consider real environmental conditions (e.g. dimensions, temperature, etc.) of a campaign, which could lead to systematic deviations. Therefore, it is crucial to design the test fields as close to reality as possible in order to avoid extrapolations or additional biases during a campaign.

Our overall system calibration procedure in test fields is based on the calibration approach for stereo image-based MMS which was presented in Ellum, El-Sheimy (2002) using bundle-adjustment. Blaser et al. (2018b) advanced the approach in order to add support for multiple stereo systems and cameras with fish-eye optics.

In this thesis, we extended the method to support a) a single multi-head panoramic camera with fish-eye optics and b) multiple LiDAR scanners.

Thereby, we considered extending and integrating a geometric 3D feature-based calibration approach, such as presented by Hillemann (2020), in order to precisely calibrate the RO between the panoramic camera heads and both LiDAR scanners. With regard to future image and LiDAR sensor data fusion, a high relative accuracy between the LiDAR scanners and the panoramic camera heads is essential.

Subsection 3.6.1 treats the functional model of the overall system calibration in detail. Subsection 3.6.2 describes our calibration fields used, whereas Subsection 3.6.3 comprises the overall system calibration procedure and Subsection 3.6.4 presents the calibration results.

3.6.1 Functional Calibration Model

The functional calibration model describes the relationship between the estimated parameters and the reference measurements.

Figure 3.11 provides an overview on the functional model of the overall system calibration. The navigation center of the INS corresponds with the body frame, which we regard as the origin of our backpack MMS.

First, we define the BA parameters to the panoramic camera and to both LiDAR scanners. They mathematically equal with three rigid body transformations ${}^{b}\mathbf{H}_{cam\,0}$ from the reference

panoramic camera head $cam\,0$ to the body frame b, ${}^{b}\mathbf{H}_{Hz}$ from the origin of the horizontal LiDAR scanner Hz to the body frame b, and ${}^{b}\mathbf{H}_{V}$ from the origin of the vertical LiDAR scanner V to the body frame b.

Second, we define five RO parameter sets from the sub-ordered panoramic camera heads $cam\,1$–$cam\,5$ to the reference panoramic camera head $cam\,0$. They also mathematically equal with the following rigid body transformation: ${}^{cam\,0}\mathbf{H}_{cam\,n}$.

Third, we define six IO parameter sets for the individual panoramic camera heads $cam\,0$ – $cam\,5$. Each IO parameter set consists of the principal distance $z' = -c$, and the principal point offsets h_x and h_y. Luber (2015) shows, that generic calibration models are able to calibrate wide-angle fish-eye lenses to around 1 pixel precision. In order to achieve higher precision, we decided to use an explicit wide-angle fish-eye camera model. Our previous investigations have shown, that the equidistant camera model fits best to the *FLIR Ladybug 5* panoramic camera. Thus, we extend the collinearity equations

$$x' = h_x + z'\frac{k_x}{k_z} + \Delta x' \tag{3.3}$$

$$y' = h_y + z'\frac{k_y}{k_z} + \Delta y' \tag{3.4}$$

by the equidistant camera model, which results in

$$x' = h_x + z'\frac{k_x}{\sqrt{k_x^2 + k_y^2}} \arctan\left[\frac{\sqrt{k_x^2 + k_y^2}}{k_z}\right] + \Delta x' \tag{3.5}$$

$$y' = h_y + z'\frac{k_y}{\sqrt{k_x^2 + k_y^2}} \arctan\left[\frac{\sqrt{k_x^2 + k_y^2}}{k_z}\right] + \Delta y', \tag{3.6}$$

where x' and y' denote image coordinates, k_x, k_y and k_z represent coordinate components from $\mathbf{R}^T(\mathbf{p} - \mathbf{t})$ where $\mathbf{R}$ and $\mathbf{t}$ represent the rotational and the translational component of the image pose and $\mathbf{p}$ represents a point in the object coordinate frame (more details see Appendix B.2). In addition, $\Delta x'$ and $\Delta y'$ represent image distortions.

Furthermore, we model the distortions $\Delta x'$ and $\Delta y'$ according to the distortion model in Brown (1971) with the radial distortion parameters A_1, A_2 and A_3 as well as the decentring distortion parameters B_1 and B_2. We use the following equations to compensate the radial distortions:

$$\Delta r'_{rad} = A_1 r'(r'^2 - r_0^2) + A_2 r'(r'^4 - r_0^4) + A_3 r'(r'^6 - r_0^6) \tag{3.7}$$

$$\Delta x'_{rad} = x'\frac{\Delta r'_{rad}}{r'} \tag{3.8}$$

$$\Delta y'_{rad} = y'\frac{\Delta r'_{rad}}{r'} \tag{3.9}$$

where $r' = \sqrt{x'^2 + y'^2}$ and $r_0 = \frac{2}{3} r_{max}$.

For compensating the decentring distortions, we use the following equations:

$$\Delta x'_{dec} = B_1(r'^2 + 2x'^2) + 2B_2 x'y' \tag{3.10}$$

$$\Delta y'_{dec} = B_2(r'^2 + 2y'^2) + 2B_1 x'y' \tag{3.11}$$

3.6.2 Calibration Fields

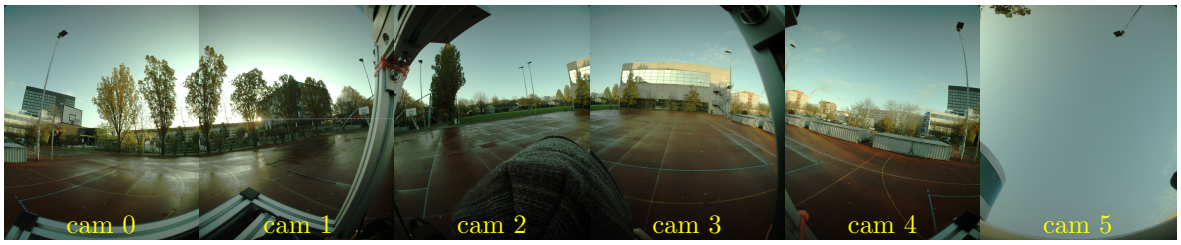
For the overall system calibration, we use both an indoor and an outdoor calibration field.

The indoor calibration field is located in the building of our industrial partner *iNovitas AG* in Baden-Dättwil (Switzerland). Originally, the calibration field was designed for MMS vehicles. It is located in a room with a length of 12.50 m, a width of 6.90 m and a height of 3.25 m. By total, the indoor calibration field contains 329 well-distributed photogrammetric targets. Most of the photogrammetric targets were measured by a contact-less high-precision industrial measurement system with a superior accuracy below 0.3 mm. Coordinates of subsequently extended target points were measured photogrammetrically with the same precision.

Figure 3.9 a provides a good overview of the indoor calibration field. High contrasts between the photogrammetric targets and the environment allow precise image measurements. However, difficult illumination with strong variations in the brightness, which are mainly caused by the glass door, complicates or hinders automatic target detection. Therefore, all target points have to be measured semi-automatically or even manually, which is very time-consuming.

The outdoor calibration field is located on a basketball court near the *FHNW* campus in Muttenz (Switzerland). By total, it contains 98 targets. Most of them are natural target points, such as edges of lines on the ground, façade elements or window corners. By contrast, some of them are marked with photogrammetric targets on pillars of basketball hoops as well as on an artwork. We measured the target points by a total station and achieved an accuracy below 5 mm. Figure 3.9 b provides a good overview of the outdoor calibration field, which provides good GNSS reception as well as clearly defined target points. However, the outdoor calibration field requires manual point measurements, since natural points are not automatically detected.

(a) Indoor calibration field



(b) Outdoor calibration field

Figure 3.9: Overall views of our calibration fields captured with the *FLIR Ladybug 5* panoramic camera. Images from the individual panoramic camera heads are stitched together.

3.6.3 Calibration Procedure

The proposed overall system calibration procedure comprises data from both the indoor and outdoor calibration field and consists of three separate evaluation steps in order to finally determine a) the IO parameters for each panoramic camera head $cam\,0$ - $cam\,5$, b) the RO parameters between the sub-ordered panoramic camera heads $cam\,1$ - $cam\,5$ and the principal panoramic camera head $cam\,0$, and finally c) the BA parameters pointing from the principal panoramic camera head $cam\,0$ and from both LiDAR scanners, Hz and V, respectively, to the body frame b (see Figure 3.11). The body frame b equals with the INS navigation center.

We then distinguished all steps that do not rely on GNSS reception, so that we could perform them in the indoor calibration field with superior target point distribution and accuracy.

This led to the following three evaluation stages:

1. Calibration of the IO parameters and the RO parameters of the panoramic camera in the indoor calibration field using bundle-adjustment.
2. Calibration of the partial BA parameters pointing from both LiDAR scanners Hz and V to the principal panoramic camera head $cam\,0$ in the indoor calibration field, using 3D feature-based calibration.
3. Calibration of the BA parameters pointing from the principal panoramic camera head $cam\,0$ to the body frame b in the outdoor calibration field using bundle-adjustment.

Figure 3.10 gives an overview of the entire calibration procedure. It begins with data acquisition in the indoor calibration field by capturing multiple epochs with both the panoramic cameras and the LiDAR scanners. Each epoch comprises a total of six images from the individual panoramic camera heads and a total of two multi-profile scans from both LiDAR scanners. In order to avoid motion blur in the images on the one hand and movements in the LiDAR multi-profiles on the other hand, we recorded all epochs statically. For each epoch, we put the backpack on a certain height so that the distance between the vertical LiDAR scanner and the floor did not fall below the scanner's minimum distance. In the calibration field, we captured a total of 16 epochs at two different locations in the calibration field. At each location, we captured 8 epochs, while we rotated the backpack 45° around the z axis between each epoch. The 45° rotations ensure that the calibration field has been completely scanned. In addition, the data acquisition from two different locations results in an ideal intersection geometry for the image measurements.

We semi-automatically performed a total of 4985 image measurements using our Python-based self-developed image measurement toolbox, which supports the equidistant fish-eye camera model. The toolbox features an ellipse fit function for precise target point measurements. In addition, it provides direct linear transformation for approximate image orientation, using a few manually measured target points. This allows to re-project the remaining target points into the image and to perform the remaining point measurements semi-automatically. During semi-automatic point measurement, it is possible to manually discard erroneous measurements, which are mainly caused by difficult light conditions.

The first evaluation stage ends with a first bundle-adjustment-based calibration. Thereby, we used the image measurements as well as the initial poses, which resulted both from the semi-automatic image measurement tool. Furthermore, we set the BA parameters ${}^{b}\mathbf{H}_{cam\,0}$ to zero, so that $cam\,0$ equaled with b. The calibration results in IO and RO parameters as described above and additionally in the EO parameters for each epoch. Figure 3.11a gives an overview of the functional model used.

The second step includes the feature-based calibration of the first part of the BA parameters of both LiDAR scanners ${}^{cam\,0}\mathbf{H}_{Hz}$ and ${}^{cam\,0}\mathbf{H}_{V}$. Figure 3.11b depicts its functional model used. The feature-based calibration estimates the BA parameters between the camera and the LiDAR scanners by minimizing the omnivariance of the resulting point cloud (Hillemann, 2020). We extended the original functional calibration model from Hillemann (2020) by multiple LiDAR scanner support.

For feature-based calibration, we used the resulting EO parameters from the first calibration step as camera poses with superior precision. Furthermore, we estimated initial values for ${}^{cam\,0}\mathbf{H}_{Hz}$ and ${}^{cam\,0}\mathbf{H}_{V}$ with a gauge for the translation. For the initial attitude values, we obtained the initial values from the backpack construction plan.

Since we acquired the multi-profiles from both LiDAR scanners and the camera poses as temporally coherent data pairs without any movements, the calibration results are only affected by deviations of the image poses and of the LiDAR profiles used. However, the feature-based calibration did not converge as we estimated all parameters of ${}^{cam\,0}\mathbf{H}_{Hz}$ and ${}^{cam\,0}\mathbf{H}_{V}$.

The initial translation parameters measured using a gauge have a precision in the mm-range. While the initial translation parameters are below the range of our expected overall accuracy, the initial attitude values that originate from the backpack construction plan might show significant deviations.

In order to mitigate the non-convergence of the feature-based calibration, we reduced the model by the translation parameters. Thus, we only estimated the attitude parameters of ${}^{cam\,0}\mathbf{H}_{Hz}$ and ${}^{cam\,0}\mathbf{H}_{V}$, using the feature-based calibration.

The third calibration stage begins with data acquisition in the outdoor calibration field in order to obtain multiple image epochs with six images from the individual panoramic camera heads as well as directly georeferenced INS poses, which are temporally associated to the corresponding image epochs (see Figure 3.10, right). In the outdoor calibration field, we performed a kinematic data acquisition. It resulted in a set of image epochs from the panoramic camera and in associated, directly georeferenced body frame b poses.

By total, we used 6 measurement epochs, located in the center of the outdoor calibration field and pointing in different directions. Furthermore, we manually performed a total of 519 image measurements by using our Python-based image measurement tool.

Figure 3.11c shows the functional model of the third calibration stage, based on bundle-adjustment. In this step, we fixed the previously estimated IO and RO parameters as well as the directly georeferenced poses from the body frame b and used the image measurements in order to obtain the BA parameters ${}^{b}\mathbf{H}_{cam\,0}$ from $cam\,0$ to b.

Finally, we obtained the BA parameters of both LiDAR scanners by concatenating both BA parts as follows:

$$\begin{aligned} {}^{b}\mathbf{H}_{Hz} &= {}^{b}\mathbf{H}_{cam\,0}\,{}^{cam\,0}\mathbf{H}_{Hz} \\ {}^{b}\mathbf{H}_{V} &= {}^{b}\mathbf{H}_{cam\,0}\,{}^{cam\,0}\mathbf{H}_{V} \end{aligned} \tag{3.12}$$

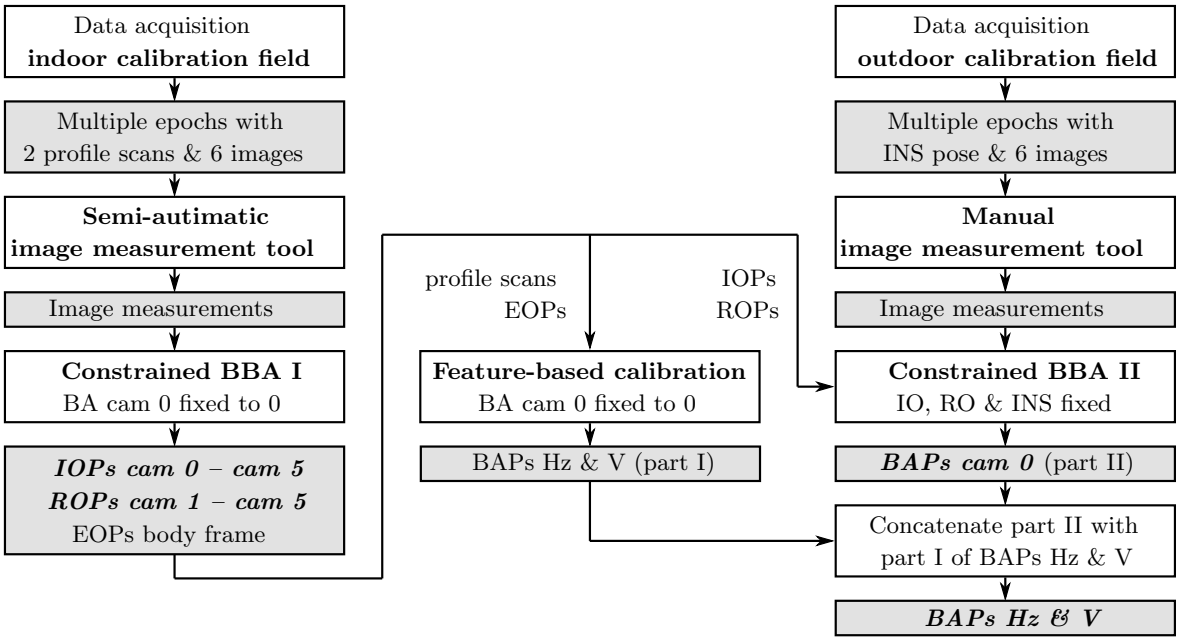
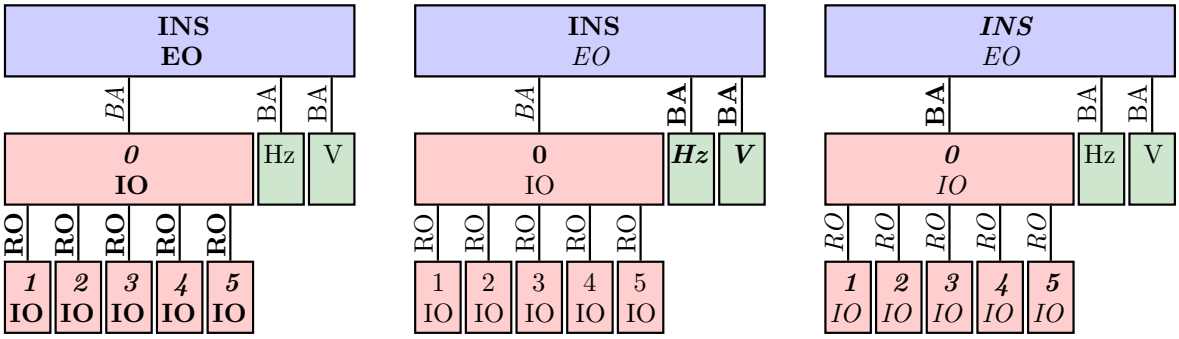


Figure 3.10: Sequence of the developed overall system calibration for the portable image-focused MMS. The left column describes the bundle-adjustment-based panoramic camera calibration and whereas the middle column outlines the feature-based RO calibration of both LiDAR scanners in the indoor test field. The right column describes the BA calibration in the outdoor test field. Boxes representing data and results are filled in gray, while boxes representing processes are filled in white. All final results are highlighted in bold italics.



(a) Indoor calibration of the camera configuration

(b) Indoor calibration of the LiDAR scanners

(c) Outdoor calibration of the boresight alignment

Figure 3.11: Functional models of the three overall system calibration stages. Colored rectangles represent sensors (red: panoramic camera heads, green: scanners, blue: GNSS/IMU navigation unit) and the top label shows the sensor name, while the bottom label indicates sensor-related parameters. Lines mark the relationships between sensors whereas the labels next to the lines represent their associated parameters. Bold parameters are estimated, while parameters with italic labels are fixed and parameters with normal font are not involved. Bold sensor labels denote involved sensors, bold italicized sensors denote involved sensors with measurements used, and normal sensor labels denote non-involved sensors.

3.6.4 Calibration Results

The first, bundle-adjustment-based calibration step, using 4985 image measurements, resulted in a mean standard deviation of 0.29 px for an image measurement. Table 3.5 contains the calibrated IO parameters for each individual panoramic camera head. Furthermore, table 3.6 shows the calibrated RO parameters of the panoramic camera and shows its standard deviations. While the standard deviation of the translation components varied between 0.2 mm and 0.3 mm, the standard deviation of the rotation components ranged from 0.0044° to 0.0233°. The standard deviations of the EO parameters of the image epochs were in the same range as the standard deviations of the RO parameters.

The second, feature-based calibration step was performed using 2×16 LiDAR multi-profiles results in a total of 921.6×10^3 LiDAR points. Since we only calibrated the orientations of ${}^{cam\,0}\mathbf{H}_{Hz}$ and ${}^{cam\,0}\mathbf{H}_{V}$, the translation parameters originated from the initial measurements, using a gauge. For the translation parameters, we assumed a standard deviation for each component of 2 mm. By contrast, the standard deviations of the rotation components varied between 0.0001° and 0.0162° (see Table 3.6) and lay in the same range as the respective standard deviations of the RO parameters.

The third, bundle-adjustment-based calibration step, using 519 image measurements, resulted in a mean standard deviation of 0.62 px for an image measurement. The translation parameters of ${}^{b}\mathbf{H}_{cam\,0}$ resulted in a standard deviation of the translation parameters of 0.8 mm and the standard deviations of the rotation components ranged from 0.0173° to 0.0177°.

Thus, the standard deviations of the rotation components of all three steps are in the same range. By contrast, the standard deviations of the translation components varied due to the different measurement methods. Nevertheless, all standard deviations are significantly below the expected georeferencing accuracy and within the targeted absolute measurement accuracy.

Table 3.5: Calibrated IO parameters of the individual panoramic camera heads. The sensor pixel size amounts to 3.45 µm and the sensor dimension is 2048×2248 px.

Camera head	$\mathbf{c_k}$ mm	$\mathbf{h_x}$ mm	$\mathbf{h_y}$ mm	$\mathbf{A_1}$ 10^{-4}	$\mathbf{A_2}$ 10^{-5}	$\mathbf{A_3}$ 10^{-8}	$\mathbf{B_1}$ 10^{-5}	$\mathbf{B_2}$ 10^{-5}
cam 0	4.259	0.000	−0.067	4.389	−1.660	3.831	2.604	−5.276
cam 1	4.267	−0.094	0.033	4.545	−1.978	14.16	7.655	−4.494
cam 2	4.269	−0.058	−0.067	3.072	−0.575	−29.48	3.255	−7.806
cam 3	4.259	−0.106	0.008	4.330	−1.552	2.647	4.292	0.070
cam 4	4.249	−0.092	−0.007	5.148	−2.237	18.45	−4.647	−7.302
cam 5	4.265	−0.038	−0.050	4.341	−1.725	7.070	−10.26	0.896

Table 3.6: Calibrated BA and RO parameters of the backpack MMS. While the first lines contain the calibrated parameters, the second lines show their standard deviations.

Parameters	**Translation** in m			**Rotation** in °		
	x	**y**	**z**	ω	φ	κ
${}^{b}\mathbf{H}_{Hz}$	−0.174	0.075	0.882	31.410	−0.011	−179.514
$s({}^{cam\,0}\mathbf{H}_{Hz})$	0.0020	0.0020	0.0020	0.0047	0.0059	0.0001
${}^{b}\mathbf{H}_{V}$	−0.005	−0.208	0.216	89.510	0.100	−90.068
$s({}^{cam\,0}\mathbf{H}_{V})$	0.0020	0.0020	0.0020	0.0001	0.0001	0.0162
${}^{b}\mathbf{H}_{cam\,0}$	−0.028	−0.033	0.685	−78.903	0.494	179.001
$s({}^{b}\mathbf{H}_{cam\,0})$	0.0008	0.0008	0.0008	0.0174	0.0173	0.0177
${}^{cam\,0}\mathbf{H}_{cam\,1}$	0.058	0.000	0.042	−0.636	−72.066	−0.246
$s({}^{cam\,0}\mathbf{H}_{cam\,1})$	0.0002	0.0002	0.0002	0.0233	0.0103	0.0216
${}^{cam\,0}\mathbf{H}_{cam\,2}$	0.035	0.000	0.110	0.535	−144.047	0.413
$s({}^{cam\,0}\mathbf{H}_{cam\,2})$	0.0002	0.0002	0.0002	0.0089	0.0103	0.0044
${}^{cam\,0}\mathbf{H}_{cam\,3}$	−0.036	0.000	0.110	0.027	−216.104	0.239
$s({}^{cam\,0}\mathbf{H}_{cam\,3})$	0.0002	0.0002	0.0003	0.0098	0.0103	0.0050
${}^{cam\,0}\mathbf{H}_{cam\,4}$	−0.058	0.000	0.041	0.554	−288.253	−0.235
$s({}^{cam\,0}\mathbf{H}_{cam\,4})$	0.0002	0.0002	0.0002	0.0182	0.0089	0.0166
${}^{cam\,0}\mathbf{H}_{cam\,5}$	0.000	0.074	0.061	90.071	0.041	180.652
$s({}^{cam\,0}\mathbf{H}_{cam\,5})$	0.0002	0.0003	0.0002	0.0075	0.0053	0.0069

Chapter 4

Robust and Accurate Georeferencing

This chapter presents our methods for robust and accurate georeferencing in challenging environments, by using backpack MMS sensor data. It extends and complements content on georeferencing methods from the following conference paper:

- Blaser, S., Meyer, J., Nebiker, S., Fricker, L., Weber, D., 2020. Centimetre-Accuracy in Forests and Urban Canyons – Combining a High-Performance Image-Based Mobile Mapping Backpack with new Georeferencing Methods. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-1-2020, 333–341. DOI: 10.5194/isprs-annals-V-1-2020-333-2020.

Robust and accurate georeferencing in a variety of vehicle-denied areas is a challenging task. Vastly varying environmental conditions make it difficult or impossible to develop a universal georeferencing method that can cover all different areas.

Due to this, we propose a modular multi-stage georeferencing approach, which combines different georeferencing methods, using various sensor technologies. Figure 4.1 provides an overview of the different georeferencing methods that we use in combination to achieve a robust and accurate georeferencing. It also shows dependencies between the different methods.

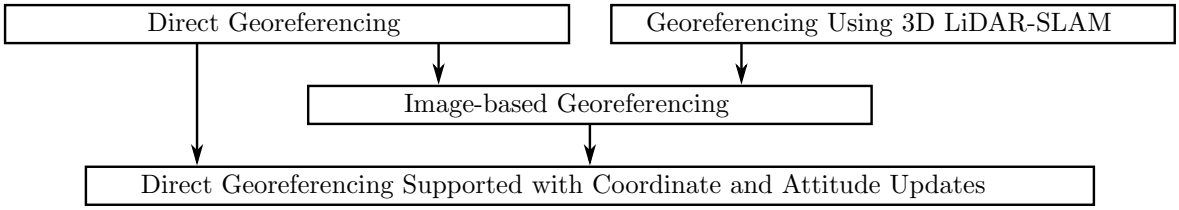


Figure 4.1: Flow of our proposed multi-level georeferencing approach incorporating different georeferencing methods and sensor data.

Direct georeferencing and georeferencing using 3D LiDAR SLAM are independent of each other. While direct georeferencing performs well in outdoor environments with good GNSS reception, 3D LiDAR SLAM suits best for indoor environments with varied geometry. Both methods perform in real-time and thus can be used for platform navigation.

Direct georeferencing provides cm-accuracy in environments with good GNSS reception, results in a dense trajectory and is not limited by the size of a measurement campaign. However, in GNSS obstructed or denied environments the accuracy rapidly degrades to the dm- and m-range.

By contrast, 3D LiDAR SLAM provides dm- to cm accuracy in closed or obstructed environments with objects and walls at a short distance. However, uniform geometry or repetitive geometric patterns may cause large errors. Moreover, 3D LiDAR SLAM shows an increasing demand for memory and computing power with increasing map size. This limits the size of a measurement campaign or requires a subdivision into multiple maps.

Both methods, direct georeferencing and 3D LiDAR SLAM, provide a dense trajectory, which is required to georeference LiDAR points. While direct georeferencing operates in a global coordinate frame, 3D LiDAR SLAM generally uses a local coordinate frame with the origin at the starting point of the map or measurement campaign. The use of 3D LiDAR SLAM as georeferencing method requires for each map a subsequent transformation from its local coordinate frame to the global coordinate frame.

By contrast, image-based georeferencing aims at the subsequent improvement of camera poses, obtained by direct georeferencing or by LiDAR SLAM. For image-based georeferencing, we use an extended SfM pipeline, based on bundle-adjustment. As we use high-resolution images, captured by the multi-head panoramic camera and taken at a low frame rate, the result of image-based georeferencing is a sparse trajectory. By contrast to direct georeferencing as well as to 3D LiDAR SLAM, the processing time exceeds the acquisition time by a multiple. Thus, image-based georeferencing is not real-time capable with the currently available computing capacity. The use of initial camera poses accelerates feature matching as well as the bundle-adjustment. For this reason, we propose to use directly georeferenced poses or georeferenced poses with 3D LiDAR SLAM as initial values.

Finally, we aim to obtain a dense trajectory with high accuracy and robustness. For this, we support direct georeferencing with coordinate and attitude updates from image-based georeferencing, which significantly reduces drifts during GNSS outages.

The following sections describe the georeferencing methods in detail and include extensions and further developments that are necessary for the combined use. Section 4.1 contains direct georeferencing, while Section 4.2 treats georeferencing using 3D LiDAR SLAM. A detailed treatment of image-based georeferencing can be found in Section 4.3 and finally, Section 4.4 includes the improvement of direct georeferencing supported with coordinate and attitude updates.

4.1 Direct Georeferencing

GNSS and INS sensors allow to directly estimate complete sensor poses in a global coordinate frame without any additional measurements. The GNSS and INS sensor data fusion based on Kalman Filters utilizes complementary properties from both sensor technologies and leads to an accurate and reliable georeferencing solution. While the INS provides measurements at a high rate but drifts strongly with time, the GNSS, having a low measurement rate but a high absolute accuracy, corrects drift. Finally, the result is a dense trajectory with a positional accuracy in the cm-range. However, GNSS occlusions and GNSS absence significantly degrade the accuracy of direct georeferencing from the cm- to the m-range.

Nonetheless, different integration depths of sensor data fusion deal with GNSS absences in a different way. Loosely-coupled GNSS and INS sensor data fusion introduce pre-processed GNSS data (position and velocity) by a separate Kalman Filter into the principal Kalman Filter. In the pre-processing of GNSS data, gross errors can be detected and eliminated before the data support the INS-based trajectory. By contrast, tightly-coupled GNSS and INS sensor data fusion introduces raw GNSS observations (pseudo-range measurements, phase measurements or Doppler observations) into the principal Kalman Filter. Hence, even a GNSS satellite constellation with less than four satellites is able to support the INS-based trajectory

and to reduce drifts (Cramer, 2001). Thus, tightly-coupled GNSS and INS sensor data fusion possibly deliver better results than the loosely-coupled integration even in GNSS obscured environments, such as urban canyons or forests.

4.1.1 GNSS- and INS-based Platform Navigation

The GNSS/INS-combined navigation system *NovAtel Synchronized Position Attitude Navigation (SPAN) CPT7* provides both a real-time navigation solution and accessible sensor raw data, which can be used for post-processing.

We use the real-time navigation data for initialization as well as for several on-board applications. The azimuth standard deviation provides a good indication of how the system has been aligned. We also use the navigation solution during the mission for geometrically constrained camera triggering. Furthermore, we plot the navigation trajectory on a map during the mission, so that the operator can track and check the data acquisition progress. Since meter position accuracy is sufficient for both system initialization and on-board applications, we operate the *SPAN* system without GNSS reference data.

However, the navigation solution rapidly degrades in large GNSS denied areas (e.g. buildings, tunnels, etc.). In addition, during long GNSS absences, the velocity value may drift and exceed the regulatory velocity limit of 515 m/s. This will automatically block all navigation-related raw data. Hence, GNSS- and INS-based platform navigation is only suitable for outdoor missions with partial GNSS obscuration or only short time periods with complete GNSS absence. For missions with large GNSS denied areas we recommend to disable the real-time navigation and only record raw sensor data.

4.1.2 Post-processed GNSS and INS Sensor Data Fusion

The main advantage of post-processed GNSS- and INS-based sensor data fusion, using an extended Kalman filter, are that all observations are available for state estimation at any epoch. By contrast to real-time applications, also later recorded sensor data is available for processing. This allows multi-sided trajectory processing, which has the potential to improve the accuracy in GNSS denied areas.

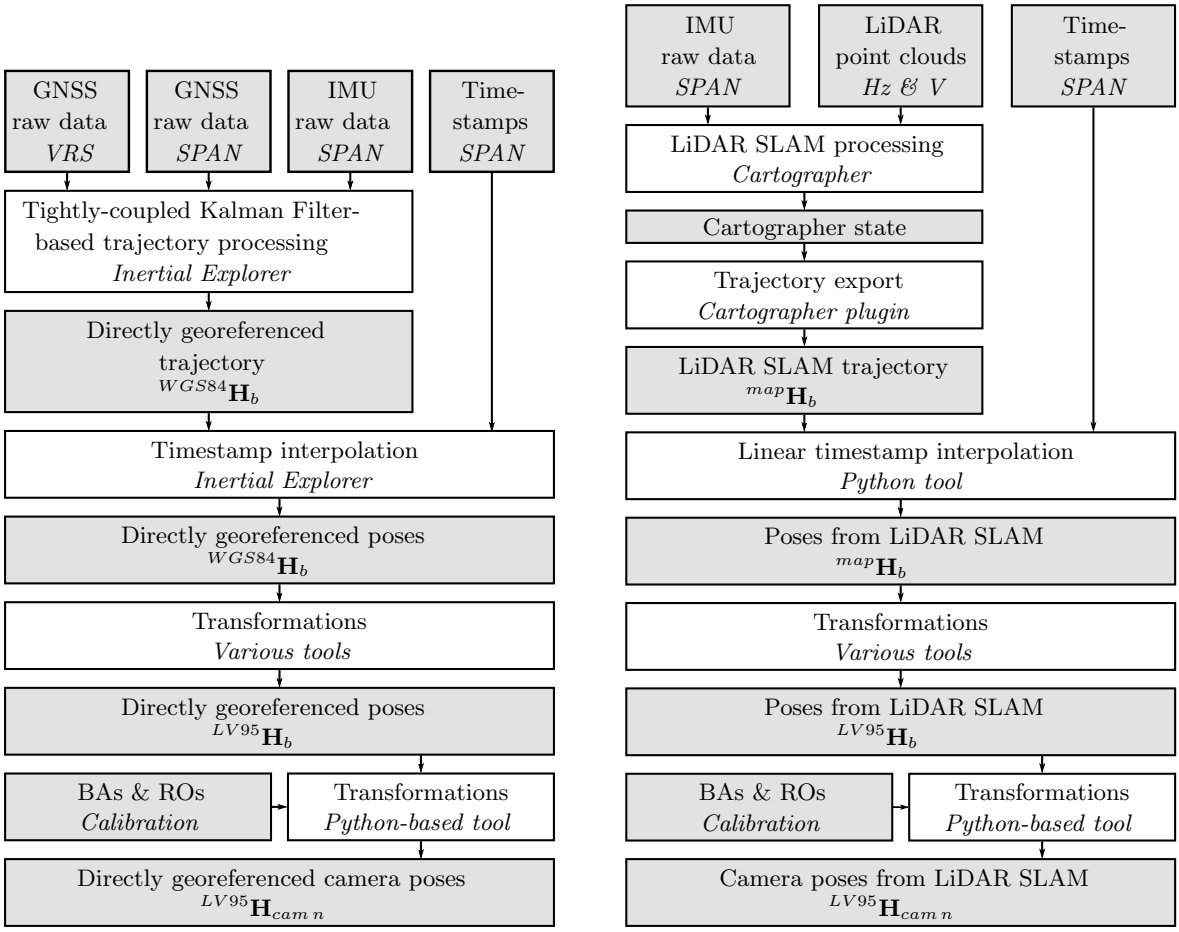
Elbahnasawy et al. (2018) graphically show the error behaviour during a GNSS outage over time for a) the forward processed trajectory, b) the backward processed trajectory and c) the forward and backward processed and smoothed trajectory.

In addition, Chen et al. (2021a) discuss the advantages of multi-pass trajectory processing. They achieved a more advantageous convergence of the extended Kalman filter with forward-backward-forward processing, since the first forward processing is only used to provide initialization. Finally, they only use the second backward and the third forward processing for smoothing.

As we use sensor data from a backpack MMS with low dynamics, we aim to achieve a more rapid azimuth convergence during initialization, using multi-pass trajectory processing. In addition, since we expect numerous areas of large GNSS outages in non-drivable environments, we intend to use combined forward and backward processing and smoothing to improve error performance.

For trajectory processing, we use the *Inertial Explorer* software *version 8.9* from *Waypoint*. By contrast to real-time platform navigation (see Section 4.1.1), we use differential GNSS, incorporating raw data from the backpack MMS as well as from a Virtual Reference Station (VRS) provided by a network of permanently installed GNSS antennas that is operated by the Swiss Federal Office of Topography *swisstopo*. For GNSS and INS sensor data fusion, we

use tightly-coupled processing that performs two-sided and multi-pass trajectory processing with subsequent smoothing (see Figure 4.2 a).



(a) Direct georeferencing

(b) Georeferencing using LiDAR SLAM

Figure 4.2: Flow charts from georeferencing methods for outdoor a and indoor b environments. The first two labels in the grey-filled data fields describe the data, whereas in the white-filled process fields they describe the process. In data fields either the data source or a pose with the coordinate frames involved is indicated below, whereas in process fields the software used is denoted. A pose from the body frame b to the global reference frame $WGS84$ is expressed as ${}^{WGS84}\mathbf{H}_b$. $LV95$ denotes the Swiss national projected reference frame, while *map* shows a local map coordinate frame.

The algorithm uses the static initialization phase to perform an initial IMU bias estimate using GNSS observations, to initialize roll and pitch values using the gravity and to start with zero velocities. By contrast, the dynamic initialization phase is mainly used for aligning the heading component of the IMU attitude with the navigation coordinate frame, using the GNSS heading (NovAtel Inc., 2020).

Since the trajectory processing is two-sided, it requires static and kinematic initialization phases at the beginning and at the end of the mission.

The resulting directly georeferenced trajectory consists of numerous poses ${}^{WGS84}\mathbf{H}_b$ of the body frame b in the global coordinate frame $WGS84$ with a resolution of the INS frequency of 200 Hz.

In order to obtain directly georeferenced sensor poses, we interpolate sensor timestamps, using the *Inertial Explorer* software. In the case of camera poses, we use the timestamps from camera triggering that were registered by the *NovAtel SPAN* unit. Then, we transform the directly georeferenced poses from the global coordinate frame $WGS84$ into the projected Swiss

national coordinate frame *LV*95 using the *REFRAME* software from *swisstopo*. Appendix A.4 treats the additionally required transformation for orientations in detail. Finally, we apply the BA and the RO parameters with a self-developed python tool in order to obtain directly georeferenced poses from the individual panoramic camera heads.

4.2 Georeferencing Using 3D LiDAR SLAM

SLAM is widely used in robotics as well as in autonomous driving. By general, it simultaneously creates a map, using sensor data from the mapping sensor and estimates the current pose, based on the created map data.

Both LiDAR scanners of our backpack MMS provide point clouds at a high data frequency. Hence, they are suitable for 3D LiDAR SLAM.

For 3D LiDAR SLAM-based georeferencing, we use the real-time SLAM solution *Google Cartographer*. Previous investigations have shown that the graph-based 3D LiDAR SLAM algorithm *Cartographer* is robust and suitable even for larger indoor environments. Hess et al. (2016) give a system overview of their first developed 2D approach, which they have subsequently extended to the full 3D SLAM *Cartographer*.

Their SLAM architecture includes a local as well as a global part. The local SLAM matches each consecutive scan against a submap, using a non-linear optimization, whereby the submap represents a probability grid with a given resolution (e.g. 5 cm). However, local scan matching requires good initial values, for example from an IMU, in order to estimate the rotational component between scan matches. Usually, the scan matching provides better precision than the grid resolution. A submap only consists of a predefined number of consecutive scans because the local scan matching accumulates error over time.

Hence, the global SLAM prevents and compensates error accumulation over time. An additional global scan matcher, using branch and bound, runs in the background. In case of a good match, the global scan matcher returns a constraint, which represents a relative pose with its associated covariance matrix. The *Cartographer* distinguishes two different types of constraints. *Intra* constraints extend over the current node and two further submaps, whereas *inter* constraints or loop closures extend beyond this. Every few seconds, the sparse pose adjustment optimizes submap poses as well as scan poses by using previously estimated constraints. The sparse pose adjustment uses a loss function to prevent outliers due to false matches, which can occur in symmetrical environments (e.g. office cubicles).

4.2.1 LiDAR SLAM-based Platform Navigation

The 3D LiDAR SLAM *Cartographer* is real-time capable, so that we can use the SLAM algorithm for platform navigation.

By contrast to the GNSS and INS-based platform navigation, the LiDAR SLAM operates in a local 3D coordinate frame. Its origin is located at the starting point of the algorithm.

As with the GNSS and INS-based platform navigation, we use the local coordinate pose for geometric camera triggering. In addition, the LiDAR SLAM provides a 2D map view of the probability grid, which we use for progress monitoring in real time.

However, LiDAR SLAM-based platform navigation has increasing computational power requirements as map size increases. This can lead to a situation where the LiDAR SLAM algorithm no longer runs smoothly and the data is only partially processed or even to other processes being disrupted.

Consequently, LiDAR SLAM-based platform navigation is suitable for indoor environments or for smaller GNSS-denied outdoor environments.

4.2.2 Post-processed 3D LiDAR SLAM

To obtain sensor poses, we use post-processed 3D LiDAR SLAM.

While real-time SLAM mainly operates with the default parameter set, post-processed SLAM allows subsequent parameter tuning in order to achieve a more precise performance. The *Cartographer* provides numerous parameters and different configurations, which are partially documented in Hess et al. (2021). Hess et al. (2021) give an introduction in the basic principles of parameter tuning. Rechsteiner, Wisler (2018) show that adjustments in the number of accumulated LiDAR data, translation and rotation weights for scan matching and the number of scans per submap have the greatest impact on local SLAM. The sparse pose adjustment rate, the maximum constraint distance and the minimum score parameters of the constraint builder have greatest impact on the global SLAM.

The processed LiDAR SLAM, incorporating LiDAR point clouds from both scanners and raw IMU data, results in a so-called *cartographer state* which mainly contains the trajectory (see Figure 4.1 b). In a subsequent step, the *Cartographer* uses raw sensor data, the boresight alignment, and the *cartographer state* to create a dense 3D point cloud. There are different asset writers available to export the 3D point cloud into various data formats.

We developed an additional asset writer as a *Cartographer* plug-in that exports the timestamped trajectory in a readable text format. The LiDAR SLAM trajectory has a temporal resolution of ten poses per second and refers to the local map frame *map* that denotes a cartesian coordinate frame with its origin at the starting position. To obtain camera poses, we first linearly interpolate camera timestamps recorded by the *SPAN* unit. We then transform the poses from the local coordinate frame into a global coordinate frame, using a self-developed *Python*-based 6 DoF coordinate transformation tool.

Finally, we apply the BA and RO parameters, using the same tools as for direct georeferencing, to obtain SLAM-based camera poses.

4.3 Image-based Georeferencing

With image-based georeferencing we aim to improve directly georeferenced or LiDAR SLAM-based camera poses.

Our proposed method of sub-sequent image-based georeferencing is based on the fundamentals of integrated georeferencing, which is widely used in aerial photogrammetry. We extended this method so that it is also applicable for image-based MM with a multi-head panoramic camera.

SfM pipelines and toolboxes enable camera pose estimation without external tie point measurements and finally provide 3D information at least as a sparse point cloud. The algorithms of SfM pipelines are developed for the close range use.

Cavegn et al. (2018) extended the open source SfM toolbox *COLMAP* by RO, EO, GCP and CP support, so that it supports integrated georeferencing for image-based MMS with multi-camera systems. Basically, the *COLMAP* pipeline incorporates two different processing steps. In the first step of correspondence searching, the algorithm extracts features and performs a spatial feature matching, using prior EO parameters. Furthermore, the first step also incorporates a geometric verification. By contrast, the second step of global reconstruction comprises scene point triangulation and local or global bundle-adjustment, which additionally incorporates

EO, RO parameters, GCPs and CPs. The iterative progress of global reconstruction also includes an outlier filter.

Our method does not rely on the extended *COLMAP* SfM toolbox, but is also supported by other SfM software such as *Agisoft Metashape* or any other SfM software which supports

- the equidistant camera model,
- RO constraints,
- prior EO parameters,
- GCPs and CPs.

First, we undistort the raw images from each individual panoramic camera head to the equidistant camera projection model, based on pre-calibrated IO parameters, by using our self-developed Python-based rectification tool *Undistorter*. Since we use undistorted images for image-based georeferencing, we fix the distortion parameters in the SfM pipeline to zero and disable self-calibration of any IO parameters (see Figure 4.3).

Second, we define and fix the RO constraints in the bundle-adjustment between all sub-ordered panoramic camera heads $cam\,n$ and the principal panoramic camera head $cam\,0$. RO constraints define the camera rig of a MMS. Predefined RO constraints significantly reduce the number of unknown parameters in the bundle-adjustment. Thus, for each image epoch only the EOs of the principal panoramic camera head $cam\,0$ remains to be determined. Consequently, using pre-calibrated RO constraints strengthens the robustness of image-based georeferencing so that even images with an insufficient number of features can be oriented and support the pose estimation of $cam\,0$.

Third, we use prior estimated EO parameters as initial values for the bundle-adjustment. For indoor environments, we use previously estimated $cam\,0$ poses from LiDAR SLAM. By contrast, for outdoor environments, we use directly georeferenced $cam\,0$ poses. Initial EO parameters enable spatial feature matching that significantly reduces processing time, while increasing the robustness.

Finally, we introduce both GCPs and CPs as well as the corresponding image measurements into the bundle-adjustment. By general, GCPs are used for the absolute orientation. In our case, we have a single panoramic camera configuration, where the image scale is not fixed. Using GCPs in a bundle-adjustment fixes the image scale of the 3D reconstruction and transforms the 3D reconstruction into the reference frame of the GCPs.

Image-based georeferencing results in subsequently improved camera poses as well as in a sparse 3D point cloud.

However, the resulting trajectory from image-based georeferencing has the same temporal resolution as the frame rate of the backpack MMS. Since our backpack MMS acquires image epochs with a frame rate between 0.5 and 2 frames per second, the temporal resolution of the resulting image-based georeferencing is too sparse to be able to register 3D LiDAR points over the trajectory.

Moreover, the processing time of image-based georeferencing exceeds the acquisition by far. Thus, this method is currently not real-time capable with the currently available processing power.

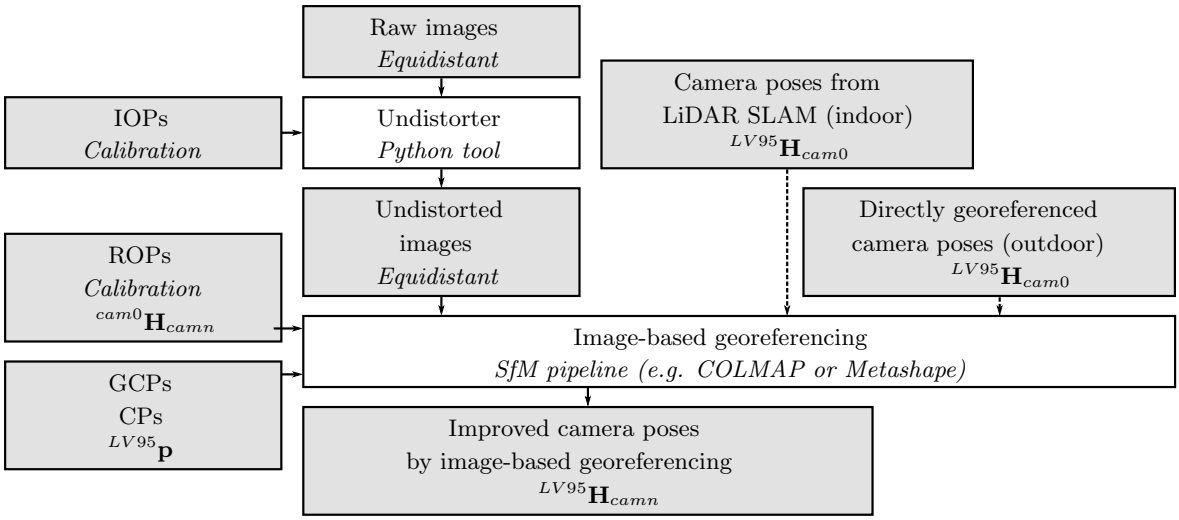


Figure 4.3: Flow chart from image-based georeferencing. Labels with normal font represent data or processes, while italicized labels indicate data sources, camera models or software tools. The left upper label from a pose represented as a homogeneous transformation $\mathbf{H}$ mark the target coordinate frame, while the right bottom label marks the origin coordinate frame.

4.4 Direct Georeferencing Supported with Coordinate and Attitude Updates

With direct georeferencing supported with coordinate and attitude updates from image-based georeferencing, we aim to combine the advantages of direct and image-based georeferencing. While direct georeferencing provides a continuous trajectory with high resolution, image based georeferencing results in a sparse trajectory with highly accurate image poses which is independent from GNSS coverage.

Angrisano (2010) provides a comprehensive review on GNSS and INS sensor data fusion using both loosely coupled and tightly coupled integration depth as well as on supporting the heading by measurements from external sources. In addition, Eugster et al. (2012) mainly focus on supporting a loosely coupled GNSS and INS sensor data fusion by CUPTs and ATTUPTs from external source. They investigated both, CUPTs and ATTUPTs originating from bundle adjustment as well as CUPTs originating from single control point measurements in the images.

In our method, we update the tightly coupled GNSS and INS sensor data fusion and trajectory processing with both CUPTs and ATTUPTs which originate from image-based georeferencing. Since the frame rate of the panoramic camera in our portable backpack MMS corresponds with the GNSS frequency, CUPTs and ATTUPTs from image-based georeferencing hold the potential to both improve the overall trajectory and to support and bridge the trajectory in GNSS denied areas. CUPTs may reduce trajectory displacements and ATTUPTs may primarily support the critical heading component, which is degrading due to multipath-effects or due to the lack of GNSS reception.

We developed a workflow for direct georeferencing supported with CUPTs and ATTUPTs by using an undocumented legacy feature of *Inertial Explorer* (see Figure 4.4).

The improved camera poses from $cam\,0$ by image-based georeferencing refer to the Swiss projected national coordinate frame *LV95* and to the local Swiss height levelling system *LN02*. Before they can be used for tightly coupled trajectory processing, they have to be transformed so that they originate from the body frame b and refer to the global coordinate frame $WGS84$. Various tools perform the necessary transformations. In a first step, a

self-developed Python-based tool appends the boresight alignment as follows,

$$^{LV95}\mathbf{H}_b = {}^{LV95}\mathbf{H}_{cam\,0}{}^{cam\,0}\mathbf{H}_b, \tag{4.1}$$

so that the poses originate from the body frame b and refer to the Swiss projected national coordinate frame $LV95$. The second transformation step concerns the transformation from $LV95$ to $WGS84$ and to the navigation coordinate frame respectively,

$$^{WGS84}\mathbf{H}_b = {}^{WGS84}\mathbf{H}_{LV95}{}^{LV95}\mathbf{H}_b, \tag{4.2}$$

while we use the *swisstopo* software tool *REFRAME* for the translation part of $^{WGS84}\mathbf{H}_{LV95}$. The rotation part from $LV95$ to the navigation coordinate frame is performed in a self-developed Python tool. Appendix A.5 contains detailed formulas for the rotation part of the datum transformation.

In addition, a Python-based tool creates the updates, while CUPTs or ATTUPTs as well as combined updates are supported. Moreover, each individual component can be weighted for trajectory processing (see Figure 4.4), which allows further tuning of tightly coupled trajectory processing.

In contrast to image-based georeferencing, which only results in improved camera poses, direct georeferencing supported with CUPTs and ATTUPTs provides an improved dense trajectory. Since this method is based on direct georeferencing, it is only available for campaigns with outdoor system initialization at the beginning and at the end and with good GNSS reception. Thus, indoor-only campaigns are excluded. Moreover, erroneous image-based updates will distort the resulting trajectory. Ideally, outliers should be previously detected and eliminated. For this, it is reasonable to select the updates based on quality indicators from image-based georeferencing.

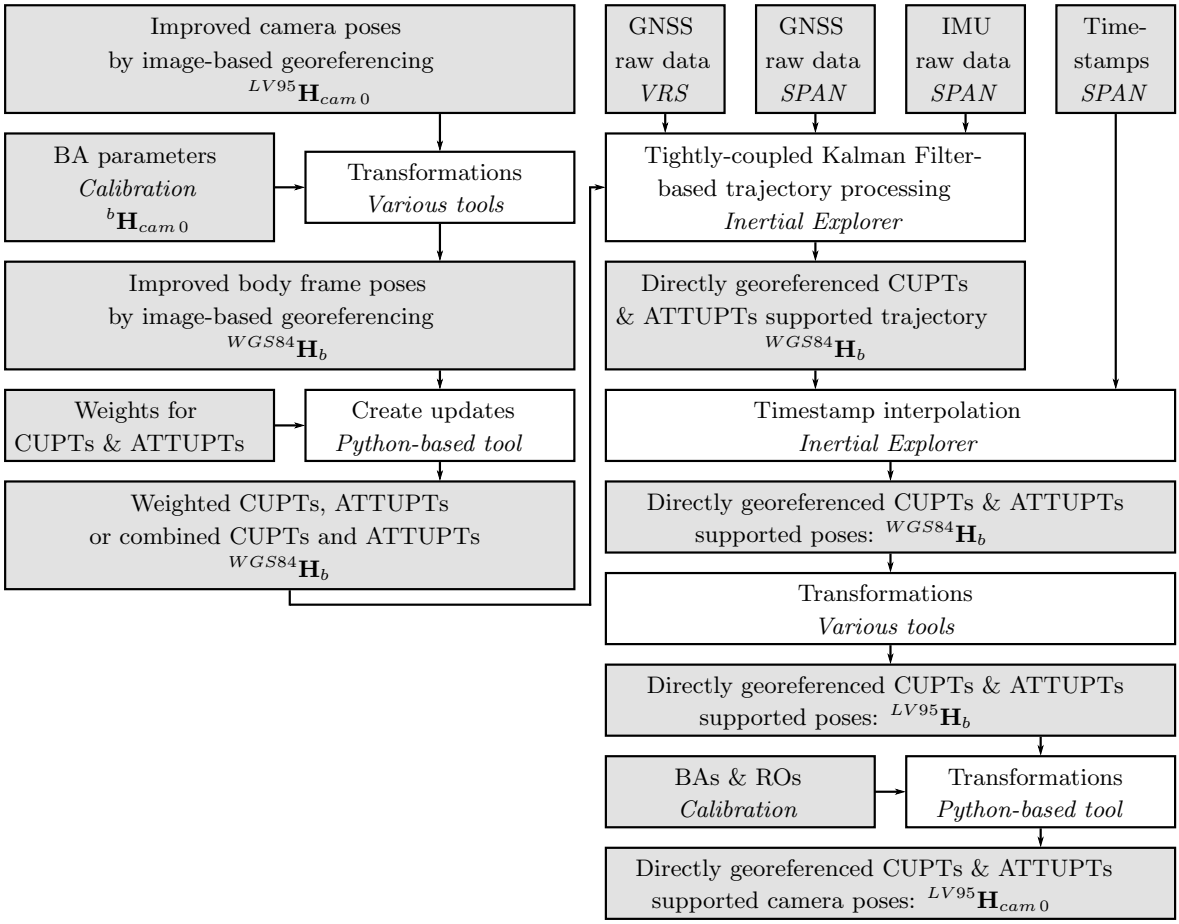


Figure 4.4: Flow chart showing the direct georeferencing workflow with CUPT and ATTUPT support. The left side represents the CUPT and ATTUPT preparation, while the right side depicts the direct georeferencing. In white-filled process fields, labels with normal font represent process names and italicized labels show the software tools used. In grey-filled data fields, labels with normal font represent the data description, while italicized labels depict data sources. The left upper label of a pose represented as a homogeneous transformation $\mathbf{H}$ mark the target coordinate frame, while the right bottom label marks the origin coordinate frame.

Chapter 5

Georeferencing Performance in Challenging Environments

In this chapter, we investigate the performance and suitability of our backpack MMS under real world conditions. As our backpack MMS is a multi-sensor system with well-calibrated sensors, it allows a direct comparison of different georeferencing methods.

We conducted our extended investigations in three test sites within challenging environments in the outdoor as well as in the indoor. The first test site is located in a city center, whereas the second test site is located in a forest. Both test sites include areas inaccessible to vehicles for which the backpack platform is particularly suitable. The third test field is located in the FHNW campus in Muttenz, which represents a typical public building. Section 5.1 gives a detailed overview of the test fields and the data sets collected. The data sets from the city center and the forest test site are publicly available as open data sets [1] (Blaser et al., 2021), while the indoor data set is unpublished.

Section 5.2 addresses various aspects of data acquisition with our backpack MMS that are relevant for further data processing and evaluation.

The evaluation of different georeferencing methods is described in Section 5.3. In each case, internal results of the respective georeferencing method are presented and discussed. The internal accuracy is a first indicator of the performance of a georeferencing method, which may be sufficient to reject a method before further investigation is conducted.

Finally, Section 5.4 compares different georeferencing methods within three different test sites in terms of precision, absolute accuracy and relative accuracy. Since we compared point measurements taken from the images with reference points of superior accuracy, we are able to obtain meaningful statements on the overall system performance.

5.1 Test Sites

This section shows the test sites where we recorded the data sets for our research and describes the captured data sets. In total, we conducted three campaigns within three different test sites. All test sites are located in vehicle denied areas and include challenging environments for all sensors. Two test sites are located outdoors, one in the city center and the other in the forest. We published both outdoor data sets in Blaser et al. (2021), so that they are publicly available for research purposes. By contrast, the third test site is located in indoor environment. The data set from the indoor campaign has not been published yet.

[1] `https://www.fhnw.ch/habg/bimage-datasets`

Table 5.1 shows a summary with key figures relating to the test sites and data recordings.

All test sites include GNSS-denied areas. However, they significantly differ in GNSS availability. All three data sets contain imagery from the panoramic camera, LiDAR raw data from both LiDAR scanners, raw GNSS data from both backpack and virtual reference station as well as raw IMU data.

Table 5.1: Key figures on test sites and data recordings.

Test site	**City centre**	**Forest**	**Indoor**
Number of targets (CPs, GCPs)	69 (54, 15)	72 (57, 15)	112 (104, 8)
Number of panoramic images	725	843	575
Number of single images	4350	5058	3450
Number of LiDAR points	840×10^6	850×10^6	1240×10^6
Number of IMU epochs	288×10^3	300×10^3	432×10^3
Trajectory length in m	800	740	573
Duration of data recording in s	1440	1500	2160

5.1.1 City Center

The first test site is located in the city center of Basel in Switzerland and covers an area of 150 m × 200 m. It covers a large square as well as different types of roads and alleys. Image 1 in Figure 5.1 shows the large square with good GNSS reception that is well suited for system initialization. By contrast, the narrow alley in image 2 (Figure 5.1) has significant GNSS signal blockage. It has steps and slopes up to 16 % and is only accessible to pedestrians. The test site includes a main traffic axis through the city center with busy tram and bicycle traffic (see Figure 5.1, image 3), as well as wide promenade with stores on both sides that is very crowded with pedestrians (see Figure 5.1, image 4).

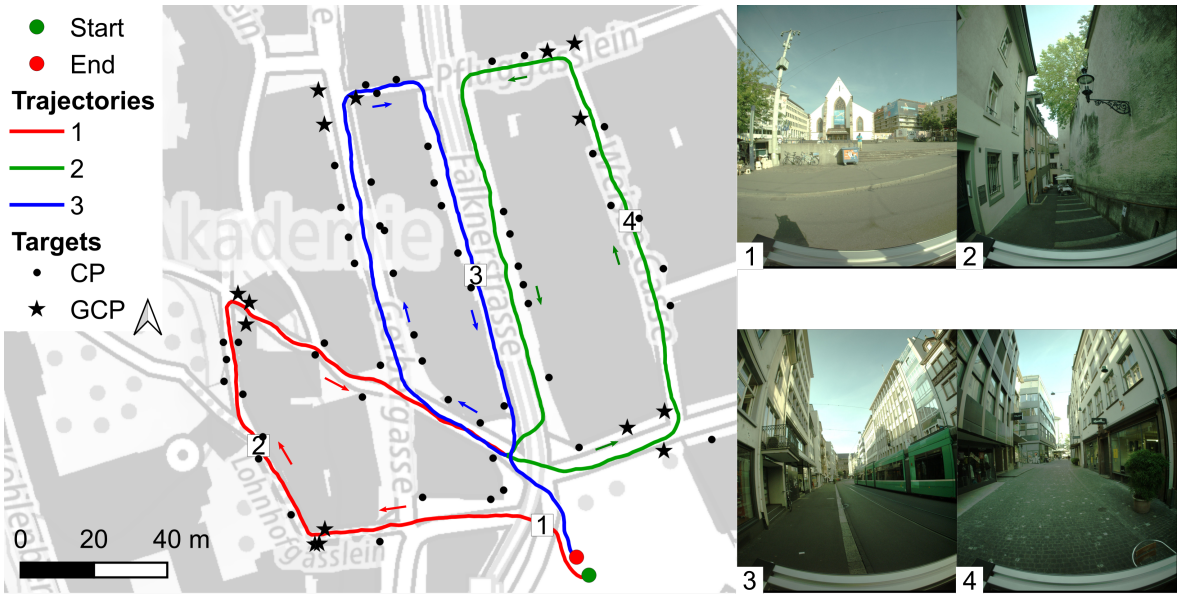


Figure 5.1: Map of the test site located in the city centre of Basel with images showing environmental conditions. Arrows in combination with different colored trajectory parts show the moving direction.

The city center test site comprises a total of 69 reference points of which we published 33 in the open data set and retained 36. Most of them are well-defined natural reference points whereas some are marked with photogrammetric targets. Fricker, Weber (2019) provide a

detailed description of reference point measurements by tachymetry and show a 3D standard deviation below 5 mm.

Our captured city data set extends along an 800 m loop-shaped trajectory that starts and ends on the large square for system initialization. For our performance investigations, we divide the overall trajectory into three parts, based on the loops (see Figure 5.1, Trajectories 1-3). This allows the measurements to be clearly assigned in the case of duplicate passes and further increases the oversight.

The data set was captured within 24 min and contains 721 panoramic images, approx. 840×10^6 LiDAR points, GNSS data as far as available as well as approx. 288×10^3 IMU epochs (see Table 5.1).

The city center test site is challenging for all georeferencing methods used. Narrow alleys as well as urban canyons with high buildings on both sides can cause GNSS signal blockage and multi-path effects, affecting the accuracy and the robustness of the direct georeferencing. Even non-GNSS-dependent georeferencing methods will reach their limits. The three trajectory loops have a length between 200 m and 300 m, which is significantly longer than even in large indoor environments where LiDAR SLAM is typically used. Consequently, the drifts at the end of a loop will be higher so that the automatic loop closure detection could be more difficult and the subsequent trajectory correction could also be subject to larger uncertainties. Furthermore, poor and homogeneous textured façades with repetitive patterns might be challenging for feature detection and matching of the image-based georeferencing.

5.1.2 Forest

The second test site is located in a forest in Münchenstein nearby Basel in Switzerland and covers an area of approximately 100 m × 200 m. The test site incorporates both narrow paths within dense vegetation at ground level that are only accessible by pedestrians (see Figure 5.2, images 3 and 5) and drivable forest roads with less dense vegetation (see Figure 5.2, Images 4 and 6). Furthermore, the forest path in image 2 of Figure 5.2 leads trough a road underpass. In addition to various forest paths, the test field also includes a highway exit where good GNSS reception is available, allowing system initialization (see Figure 5.2, image 1).

Figure 5.2: Map of the test site located in a forest nearby Basel with images showing the environmental conditions. Arrows in combination with different coloured trajectory parts show the moving direction.

The forest test site comprises a total of 72 reference points of which we published 23 in the open data set and retained 49. Most of them are marked with photogrammetric targets fixed on trees or driven-in pillars, whereas only few of them are well-defined natural target points. Fricker, Weber (2019) provide a detailed description of reference point measurements by tachymetry and show a 3D standard deviation below 5 mm.

Our captured forest data set extends along a 740 m trajectory that starts and ends at the highway exit. The start and end of the trajectory are linear shaped (see Figure 5.2, Trajectories 1 and 4), whereas trajectories 2 and 3 are loop-shaped (see Figure 5.2, Trajectories 2 and 3). This partition allows clearly assigned measurements even in the case of duplicate passes and further increases the oversight.

The data set was captured within 25 min and contains 843 panoramic images, approx. 850×10^6 LiDAR points, approx. 300×10^3 IMU epochs as well as GNSS observation as available (see Table 5.1).

The forest data set is challenging for all georeferencing methods used. The forest leads to significant GNSS signal blockage, which affects the direct georeferencing. In addition to the long loop length discussed in Section 5.1.1, ambiguous geometry and moving boughs and leafs can affect the LiDAR SLAM. Dark environments in the narrow forest paths with dense vegetation makes feature detection and feature matching difficult or even impossible, and thus affects the image-based georeferencing.

5.1.3 Indoor

The third test site is located in the *FHNW* campus in Muttenz nearby Basel in Switzerland and covers an area of approximately 60 m × 80 m in the eastern part of the building. The indoor test site represents a typical public building and incorporates dimly lit core zones with staircases and elevators (see Figure 5.3, image 1), long corridors with large window fronts on the one side and numerous doors on the other (see Figure 5.3, images 2 and 3) as well as a modern open space office with flexible workplaces (see Figure 5.3, image 4).

Figure 5.3: Map of the indoor test site located in the FHNW campus nearby Basel with images showing environmental conditions. Arrows in combination with different colored trajectory parts show the moving direction

The indoor test site comprises a total of 112 well-defined natural target points. We previously measured all target points by tachymetry with a superior accuracy, achieving a 3D standard deviation below 5 mm.

We initialized our system on a survey platform on top of the building on the 13^{th} floor and then descended the stairs to our test site on the 10^{th} floor. The trajectory is loop-shaped and leads the first time counter-clockwise and the second time clockwise through the test site. In total, the trajectory has a length of 573 m. The data set was recorded within 36 min and contains 574 panoramic images, approx. 1240×10^6 LiDAR points and approx. 432×10^3 IMU epochs. GNSS measurements were only recorded during the system initialization on the survey platform as there occurred a total GNSS signal loss in the building. The indoor recordings took longer because the images were taken manually while the backpack was stationary in order to avoid motion blur due to long exposure times.

The indoor test site is challenging for all georeferencing methods used. The complete GNSS absence in the building causes large drifts in the meter range for the direct georeferencing. Challenges for SLAM-based georeferencing in such buildings are repetitive geometric patterns (e.g. modularized room size) that may cause wrong loop closures, resulting in gross errors. Moreover, narrow-shaped stairwells and doors are challenging for SLAM-based georeferencing and may cause geometric singularities and gross errors. Main challenges for image-based georeferencing are poorly textured walls and repetitive patterns, reflective objects and large glass walls as well as changing light conditions (e.g. when a motion detector switches on the light or a room or corridor is significantly brighter than the other). The above-mentioned occurrences make feature detection and feature matching difficult or even impossible.

5.2 Mobile Data Acquisition

This section shows different aspects of mobile data acquisition with our self-developed backpack MMS, that are important or essential for further data processing and investigations. This ranges from different kinds of system initialization in various environments, described in Subsection 5.2.1, to the amount of recorded data from each sensor in Subsection 5.2.2.

5.2.1 System Initialization

Our backpack MMS supports system initialization in indoor or other environments without GNSS reception as well as in outdoor environments with good GNSS coverage.

Indoor initialization

When the backpack MMS is only used in indoor environments or in other environments without GNSS reception, the data acquisition is completely performed in a local system. This concerns on the one hand the coordinate frame which is a local 3D coordinate frame and on the other hand the system reference time, which completely uses the local sensor time without any GNSS offset so that it completely depends on the oscillator from the navigation unit *NovAtel SPAN CPT7*.

In indoor use, the mapping frame is equal to the local coordinate frame and no GNSS synchronization is performed. Consequently, there is no initial alignment routine required. Thus, the backpack MMS is ready for data acquisition just after the system start-up.

However, in indoor use, the backpack MMS only acquires raw IMU data, raw LiDAR data and raw image data from the panoramic camera. Because of the lack of GNSS, the navigation unit

does not provide any navigation solution during data acquisition and direct georeferencing as well as integrated georeferencing methods that are based on the direct georeferencing are not available. Furthermore, there is no correction of the time drift by GNSS synchronization.

Outdoor initialization

For outdoor use, the backpack MMS requires a initialization process so that the real-time navigation solution from the navigation unit is available and the trajectory for direct georeferencing can be post-processed.

The initialization process of the navigation unit fulfills the following tasks:

1. to synchronize the navigation unit with the GNSS time
2. to estimate initial velocity, position and attitude information

The first task is performed within the static system initialization and requires good GNSS reception. The most critical part of the second task is to align the IMU with the mapping frames, which requires additional external information or separate alignment routines.

The MEMS-based IMU of the *NovAtel SPAN CPT7* navigation unit can be aligned with a dynamic alignment routine that estimates the heading from the GNSS velocity vector. NovAtel Inc. (2020) note that the platform roll angle should remain below 10° during the dynamic alignment routine. Furthermore, the unit only performs the dynamic alignment when a minimum velocity above 0.5 m/s is reached. The duration of the alignment routine strongly depends on the GNSS reception and the GNSS coverage. Our experience has shown that the dynamic alignment usually takes between 2 min and 3 min under good GNSS conditions.

In order to enable forward as well as backward trajectory post-processing, using a tightly-coupled Kalman filter, the system needs to be initialized at the beginning as well at the end of the campaign in the following order:

1. static initialization (approx. 3 min)
2. dynamic alignment (approx. 2 min – 3 min)
3. outdoor MM campaign
4. dynamic alignment (approx. 2 min – 3 min)
5. static initialization (approx. 3 min)

Initialization for both indoor and outdoor

For combined indoor and outdoor campaigns, the same initialization procedure as for outdoor campaigns can be used. However, longer GNSS absence in indoors will probably cause strong drifts so that the velocity of the Kalman filter exceeds the regulated speed limits. In case that the regulated speed limit is reached, the navigation unit restricts data processing and recording.

For this reason, we disable the Kalman filter on the navigation unit and only record raw data for combined indoor and outdoor campaigns. As a result, the real-time navigation solution using GNSS and IMU data is not available and no quality indicators are available to directly assess the system initialization in the field.

5.2.2 Data Acquisition and Recording

During the MM campaign, the panoramic camera can be triggered a) manually or b) time-based or c) based on geometric conditions (e.g. distance-based or angle-based). An image epoch consists of six images from the individual panoramic camera heads. Each image has a resolution of 5 Mpx. Consequently, the total resolution of an image epoch amounts to 30 Mpx. The computer's hardware limits the maximum frame rate of the *Ladybug 5* panoramic camera to approximately 2 Hz (see Figure 5.4).

The backpack MMS continuously captures LiDAR data with a rate of 10 multi-profiles per second from both *Velodyne VLP-16 Puck* LiDAR scanners. A multi-profile comprises in total 32 profiles, whereby each of the two LiDAR scanners includes 16 profiles. In addition, one LiDAR detector scans a profile with a rate of 18.08 kHz (see Figure 5.4). Consequently, both scanners of the backpack MMS detect a maximum of approximately 576 000 LiDAR points per second.

The navigation unit *NovAtel SPAN CPT7* provides IMU raw data with 200 Hz, while the frequency of GNSS epochs amounts to 1 Hz.

The data acquisition rates shown in Figure 5.4 give an indication of the amount of captured data. Thereby, the data type as well as the size of the data of the individual data epochs are considered.

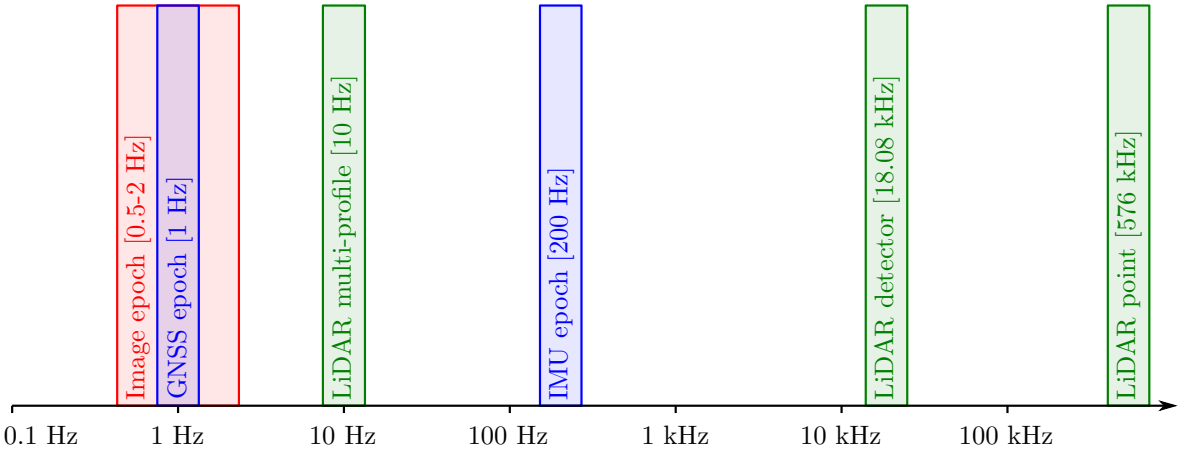


Figure 5.4: Frequency spectrum of the data recordings from the *BIMAGE Backpack*. *Image epoch* denotes a high-resolution multi-head panoramic image consisting of 6 individual images. *GNSS and IMU epoch* declare one sensor reading. *LiDAR multi-profile* represents the revolution rate of the multi-beam LiDAR scanners, while *LiDAR detector* represents the point rate of one LiDAR beam and *LiDAR point* indicates the point rate of both LiDAR scanners involved. Blue indicates data used for direct georeferencing, green indicates data used for SLAM-based georeferencing and red indicates data used for image-based georeferencing.

5.3 Evaluation of Georeferencing Methods

This section contains the evaluation of four different georeferencing methods. Thereby, internal results and the accuracy of each georeferencing method are presented and discussed.

While the first two methods, direct georeferencing (Subsection 5.3.1) and LiDAR SLAM-based georeferencing (Subsection 5.3.2), are almost independent from each other, the other two methods image-based georeferencing (Subsection 5.3.3) and direct georeferencing supported with CUPTs and ATTUPTs originating from image poses, are more integrated (Subsection 5.3.4).

5.3.1 Direct Georeferencing

For all test sites, we used the following data for direct georeferencing:

- GNSS raw observations from a reference station
- GNSS raw observations from the rover (backpack MMS)
- IMU raw observations from the rover (backpack MMS)

We obtained GNSS reference data from the network of continuously operating GNSS reference stations *AGNES*, which is provided by the Swiss Federal Office of Topography *swisstopo*. Beside raw data from a single reference station, they also provide raw data from virtual reference stations, which are interpolated using surrounding reference stations. For both city center and forest data sets, we used virtual reference stations close to the test sites to minimize lengths of base lines. For the indoor data set, we used the GNSS reference station which is directly installed on top of the FHNW campus building in Muttenz.

For the processing of the direct georeferencing, we used two-sided tightly-coupled multi-pass trajectory processing, using differential GNSS, provided by the *Inertial Explorer* software *version 8.9* from *Waypoint*. Furthermore, we used the error model for the IMU type *IMU HG4930* that is provided by the manufacturer.

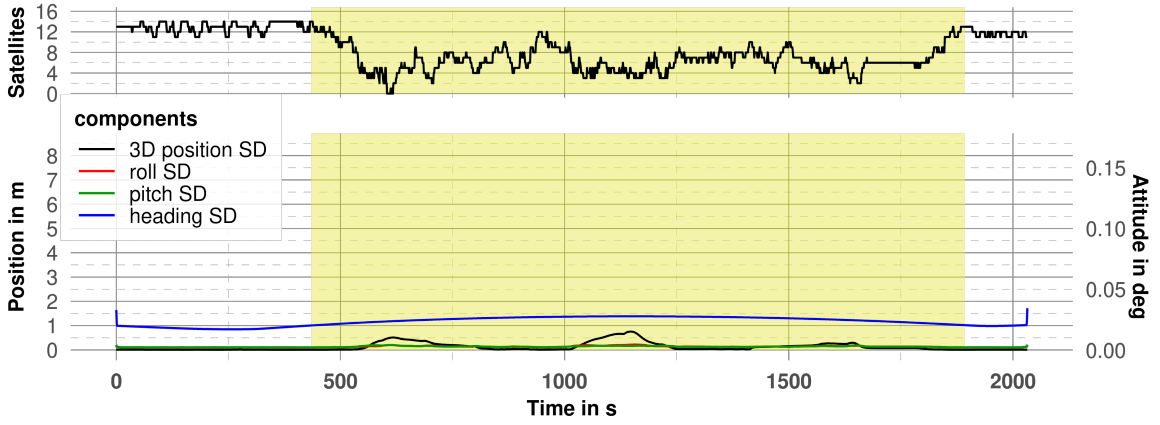
The IMU used cannot detect the earth rotation accurate enough as required for static alignment of the heading angle, because it contains MEMS-based gyroscopes with lower sensitivity than FOG-based gyroscopes. Thus, for all test we used sites kinematic alignment only with a minimum speed threshold of 0.5 m/s and with a maximum heading standard deviation tolerance of 45°.

Trajectory processing successfully performed without failures in all test sites. In the city center and in the forest test site 12 or even more GNSS satellites were available (see Figure 5.5a and b) whereas in the indoor test site the number of GNSS satellites used varied between 5 and 8 (see Figure 5.5c). The lower number of satellites is due to the fact that Global Navigation Satellite System (GLONASS) was deactivated in the indoor test field and only Global Positioning System (GPS) was used, while in both outdoor test sites both satellite navigation systems, GPS as well as GLONASS, were activated.

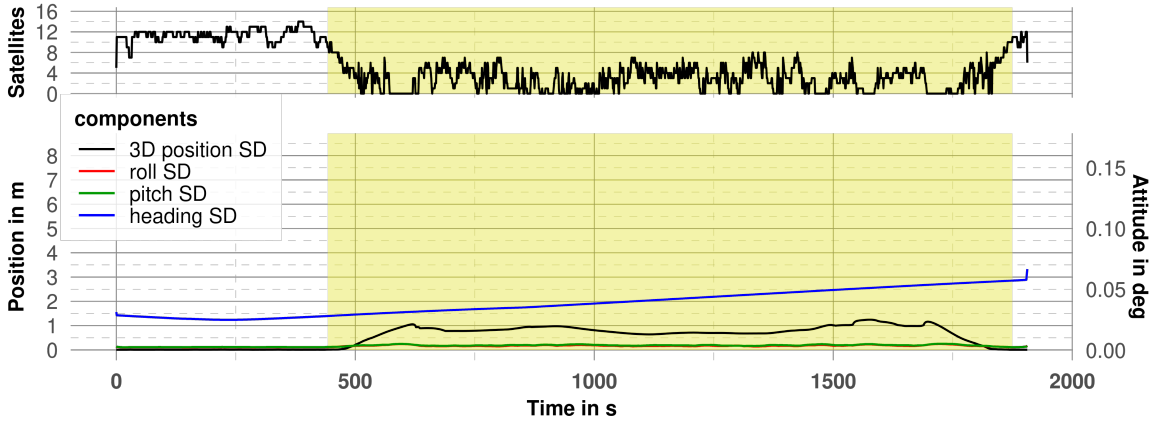
The mean standard deviations of the roll and pitch angles varied in all test sites between 0.003° and 0.004° and the standard deviations of the heading angle amounted to 0.023° in the city center test site, 0.039° in the forest test site and 0.029° in the indoor test site. The higher mean heading standard deviation in the forest test site might come from the shorter initialization phase at the end of the campaign. This is also apparent in Figure 5.5b where the heading standard deviation curve increases towards the end of the campaign. By contrast,

in the other test sites with longer initialization periods, the heading standard deviation curves decrease to the end.

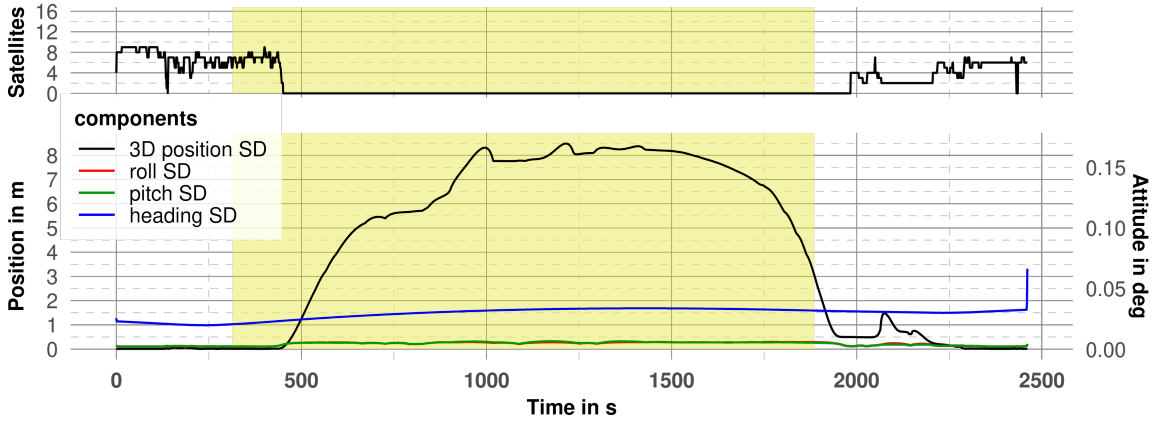
The mean standard deviation of a 3D position varied the most between the different test fields. In the city center test field it amounted to 0.124 m, in the forest test field to 0.567 m and in the indoor test field to 3.930 m. Figure 5.5 shows a strong correlation between the GNSS availability and the standard deviation of a 3D position. Furthermore, it clearly shows that the standard deviation of a 3D position increases from the cm to the dm or even to the sub-m range during short-term GNSS outages of a few s. During long-term GNSS outages, it clearly drifted to the m range (see Figure 5.5c).



(a) City centre



(b) Forest



(c) Indoor

Figure 5.5: Results from the direct georeferencing for each test site. The upper part of each diagram shows the satellite availability whereas the bottom of each diagram contains the internal standard deviation, estimated by the Kalman filter. Panoramic images were captured during the yellow marked time period.

5.3.2 LiDAR SLAM-based Georeferencing

For the evaluation of the LiDAR SLAM-based georeferencing, we used the following data:

- LiDAR raw data from the horizontally mounted LiDAR scanner
- LiDAR raw data from the vertically mounted LiDAR scanner
- IMU raw data

We tuned the SLAM algorithm for our large-scale test fields by adjusting some parameters. First, we adjusted the IMU gravity time constant to 9.81 $\frac{\mathrm{m}}{\mathrm{s}^2}$ in order to achieve higher precision. Second, we increased the maximum constraint distance of the constraint builder, which is part of the global SLAM. This causes poses that are further away from the sub-map to be considered for constraints as well. The additional overlapping constraints might further stabilize the trajectory. Third, we aim to prevent wrong constraints which may caused by repetitive patterns (e.g. similar room shapes) by the following parameter adjustments. We increased the minimum score threshold as well as the minimum global localization score of the constraint builder. Finally, we used the same configurations for all test sites.

Internal quality indicators of the LiDAR SLAM are the resulting point cloud, which is a first visual and qualitative indicator, and the rotation and translation deviations from the constraints of the global SLAM, which are numeric indicators.

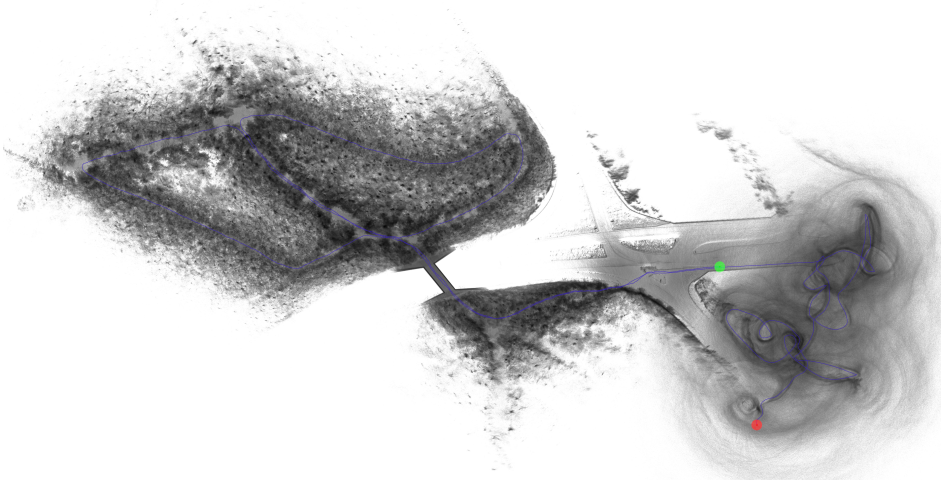
On the one hand, top views of the resulting point clouds (see Figure 5.6) show the completeness. On the other hand, possible gross errors, caused for example by wrong loop-closures, can be identified.

Thanks to the prior SLAM parameter tuning, such gross errors can be prevented. In the city center test site (see Figure 5.6a) also interiors of stores which were visible through store windows, were partially mapped. In the indoor test site, some rooms and corridors with large window walls were reflected (see Figure 5.6c). In the forest test site, a point cloud mess and a circling trajectory occurred at the end of the trajectory (see Figure 5.6b). The reason for this was an IMU data outage caused by an IMU failure, which occurred shortly after the initialization at the end of the campaign. During this time the SLAM lacked the IMU data for correct point cloud registration. Nevertheless, the recordings in the forest were not affected by this and are still usable for our investigations.

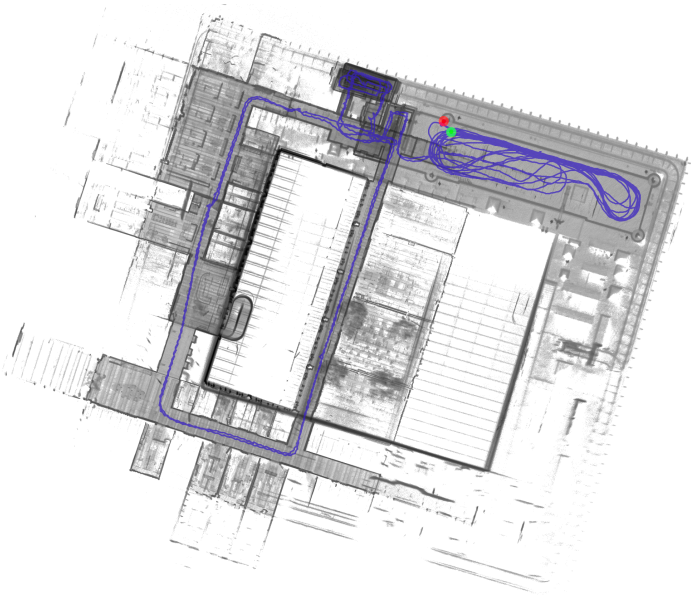
In the city center, the SLAM trajectory consists by a total of 14 935 poses and 29 557 global observations (poses, local constraints and loop-closures) with a mean translational error of 0.053 m and a mean rotational error of 0.200°. In the forest, the SLAM trajectory comprises 15 655 poses and 31 150 global observations with a mean translational error of 0.061 m and a mean rotational error of 0.523°. The SLAM trajectory of the indoor test site comprises 16 054 poses and 31 948 observations with a mean translational error of 0.066 m and a mean rotational error of 0.412°. The cumulative frequency plots (see Figure 5.7) show the error distribution which behaves similarly in the translational component a across all test sites.

In the rotational component b, the city centre test site showed lower residuals than the forest and indoor test sites. In the forest test site, the failure of IMU may also have been the cause of the larger rotational residuals, since the error distribution equals almost that of the indoor test site.

(a) City centre

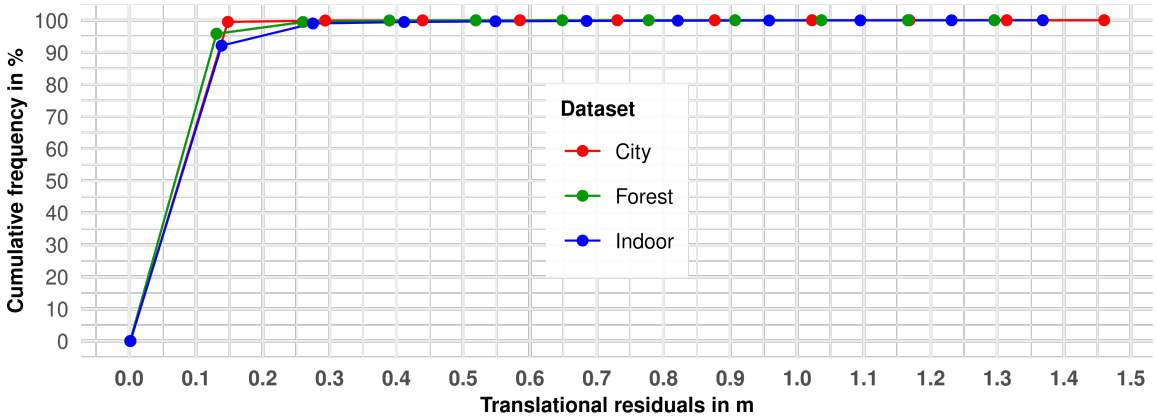


(b) Forest

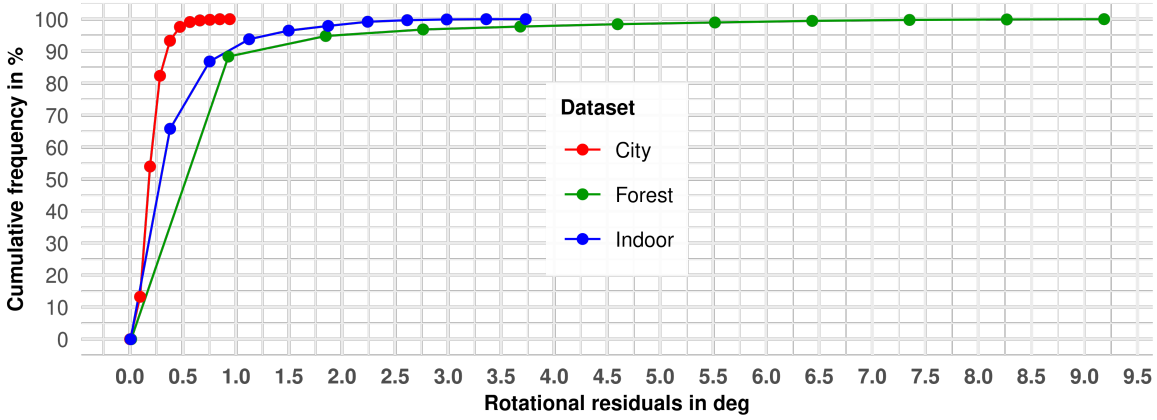


(c) Indoor

Figure 5.6: Top views of the resulting point clouds from LiDAR SLAM. The blue lines represent the trajectory, whereby the start point is marked in green and the end point is marked in red.



(a) Translational residuals



(b) Rotational residuals

Figure 5.7: Cumulative frequency plots with translational and rotational residuals after the final optimization of the global SLAM. For each test site, the range of residuals was divided into ten equal parts, for which the cumulative frequency was estimated.

5.3.3 Image-based Georeferencing

For the image-based georeferencing, we used the following data:

- Single images from the panoramic camera
- Initial poses for each panoramic camera epoch
- GCPs

Since we previously undistorted the single images from the panoramic camera according the equidistant camera model, we fix the IO parameters, by using calibrated focal lengths, setting the distortions of the principal point and the distortion parameters to zero. We previously defined the relative orientations, using calibrated values. By contrast to Cavegn et al. (2018) which use fixed weights for the relative orientation constraints, we used the standard deviations resulting from the calibration as weights.

Furthermore, we used prior estimated EO parameters from *cam*0 as initial poses for each panoramic camera epoch. To process the city center and the forest test site, we used initial poses from direct georeferencing whereas to process the indoor test site, we used initial poses from LiDAR SLAM-based georeferencing.

Finally, we introduced and measured the GCPs. In contrast to Blaser et al. (2020), we only measured the GCPs in four images next to the points but did not measure any CPs.

We processed the image-based georeferencing using the SfM software *Agisoft Metashape* on a desktop workstation with the following specifications:

- CPU: Intel Xeon CPU E5-2623 v3 with 3.0 GHz and 8 cores
- GPU: NVIDIA GeForce GTX 1080
- RAM: 32 GB
- OS: 64-bit Ubuntu 18.04 LTS
- Agisoft Metashape v1.7.1 build 11797

For the processing, we used the full image resolution for feature extraction and limited the number of features to 80 000 per image. Furthermore, we disabled the tie point limit.

Table 5.2 shows the most important internal results of image-based georeferencing. With the developed workflow and chosen settings, all images in all test fields were successfully registered. In the indoor test site, the mean reprojection error amounted to 2.25 px, which was significantly higher than in both outdoor test sites with 0.425 px in the city center and 0.555 px in the forest. One reason could be the shorter object distance and the resulting higher object resolution. This, in addition to the higher exposure time in indoors, favors additional motion blur. Another reason could be that in certain cases the minimum object distance of the *Ladybug 5* panoramic camera was missed, which resulted in image blur.

By general, the processing time of image-based georeferencing exceeds the acquisition time by far. In our test sites, the ratio between processing time and acquisition time varied between 1.9 in indoor environments and 4.4 in the forest.

Table 5.2: Table with the results from the image-based georeferencing.

Test site	**City centre**	**Forest**	**Indoor**
Number of single images	4350	5058	3450
Registered single images	4350	5058	3450
Number of projections	1029×10^3	1777×10^3	2063×10^3
Number of tie points	315×10^3	628×10^3	683×10^3
Mean reprojection error in px	0.425	0.555	2.25
Matching time in s	1655	3600	2664
Alignment time in s	3141	3057	1432
Total processing time in s	4796	6657	4096
Ratio processing time / recording time	3.3	4.4	1.9

5.3.4 Direct Georeferencing Supported with Coordinate and Attitude Updates

For the evaluation of the direct georeferencing supported with CUPTs and ATTUPTs, we used the following data:

- GNSS raw observations from a reference station
- GNSS raw observations from the rover
- IMU raw observations from the rover
- Poses from image-based georeferencing for creating CUPTs and ATTUPTs

We introduced external updates to the direct georeferencing evaluation to improve the trajectory in GNSS denied areas. In our test sites, GNSS obscured periods correspond almost with periods where panoramic images were captured (see Figure 5.9).

Evaluations for Obtaining Optimal Update Settings

For obtaining optimal weights and combinations of CUPTs and ATTUPTs, we performed in a first step six evaluation samples with different settings. In a second step, we assessed the internal accuracy of each setting, in order to identify the most promising settings for further investigation.

The first evaluation sample concerned all test sites with CUPTs and global weights. We used equal weights of 5 mm for all coordinate components, which corresponds to the standard deviation of the reference points.

By contrast, for the second evaluation sample, we used the resulting standard deviation from image-based georeferencing. Both weighting approaches succeeded for all test sites. Since the individual weights did not produce any significant improvement, we used only the evaluation with the globally weighted updates for the further investigations.

The third evaluation sample concerned all test sites with ATTUPTs and global weights. For the roll and pitch angles, we used low weights of 0.1°, because they are not significantly affected by GNSS absence. By contrast, we used higher weights of 0.005° for the heading component.

In the fourth evaluation sample, we weighted the ATTUPTs individually, based on the resulting standard deviations from the image-based georeferencing. Both evaluations resulted in standard deviations of the 3D position in the m-range.

In the fifth evaluation sample, we used combined CUPTs and ATTUPTs with global weights and in the sixth evaluation sample, we used combined CUPTs and ATTUPTs with individual weights from the image-based georeferencing. While the fifth evaluation sample succeeded for all test sites, the sixth evaluation sample partially failed and resulted standard deviations of the 3D position were in the m-range.

Finally, we used only globally weighted CUPTs (first evaluation) and globally weighted CUPTs combined with ATTUPTs (fifth evaluation) for further investigations.

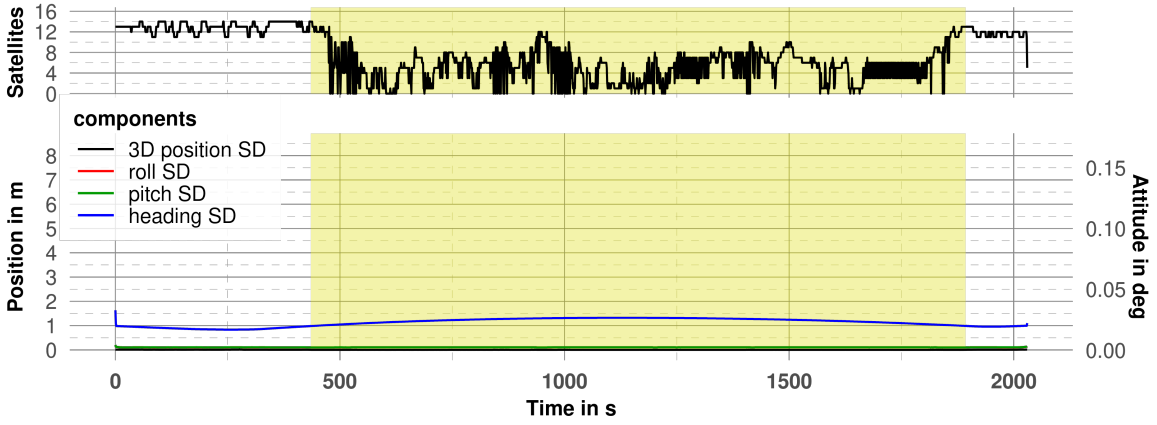
Evaluation of the Internal Accuracy

As far as two settings were identified for further investigation, we evaluate their internal accuracy in detail.

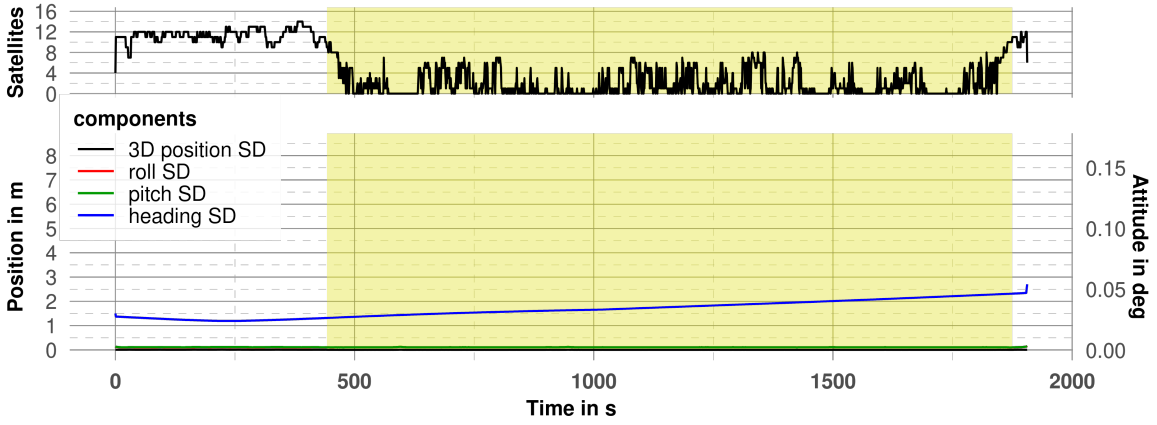
The first setting included globally weighted CUPTs with equal weights of 5 mm for all coordinate components. The second setting included globally weighted CUPTs and ATTUPTs with equal weights of 5 mm for all coordinate components of the CUPTs. For the ATTUPTs we weighted the roll and pitch angles with 0.1°, whereby we used with 0.005° higher weights for the heading component.

Direct georeferencing supported with globally weighted CUPTs showed a significant decrease of the internal mean standard deviation of the 3D position to 0.006 m in the city center, to 0.005 m in the forest and to 0.067 m in the indoor test site. Figure 5.8 shows the significant improvement of the mean standard deviation of the 3D position, even in GNSS denied areas. Moreover, the mean standard deviation of the roll and pitch angles slightly decreased in all test sites to 0.002° and the standard deviation of the heading angle remained constant.

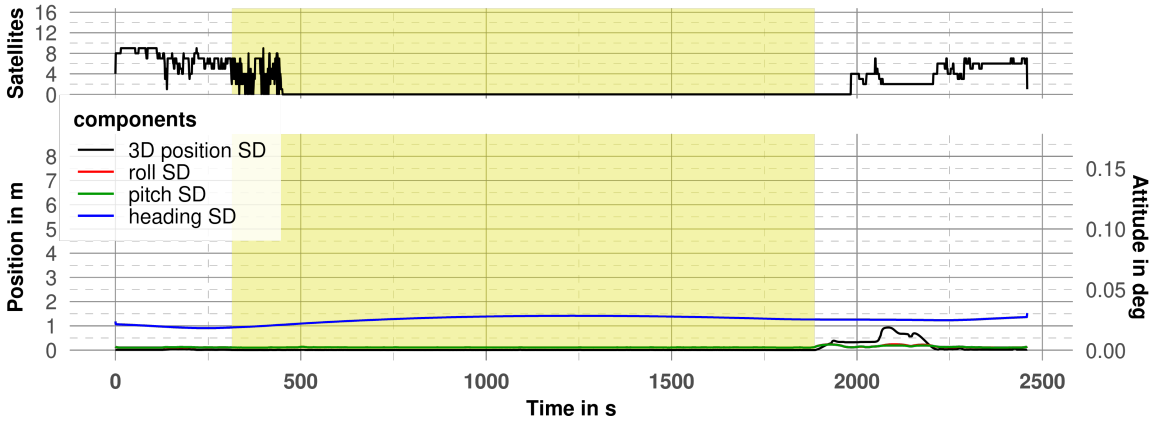
By contrast, direct georeferencing supported with globally weighted combined CUPTs and ATTUPTs, resulted in a significantly decreased mean heading standard deviation of 0.004° in the city center and the forest test sites and 0.006° in the indoor test site. These standard deviations are by an order of magnitude lower than without the use of ATTUPTs. Figure 5.9 shows the significant decrease of the heading standard deviation during the time range of the introduced updates, which equals with the time period when panoramic images were captured.



(a) City centre

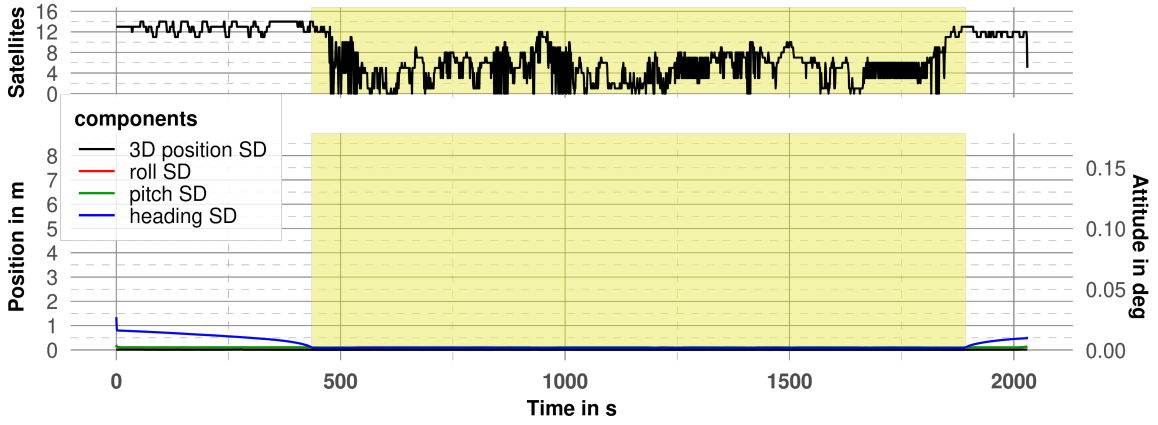


(b) Forest

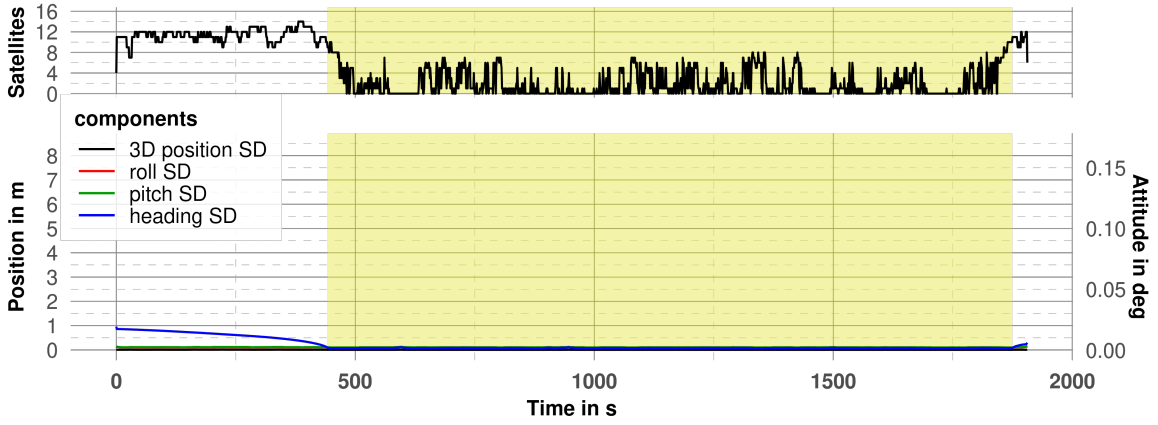


(c) Indoor

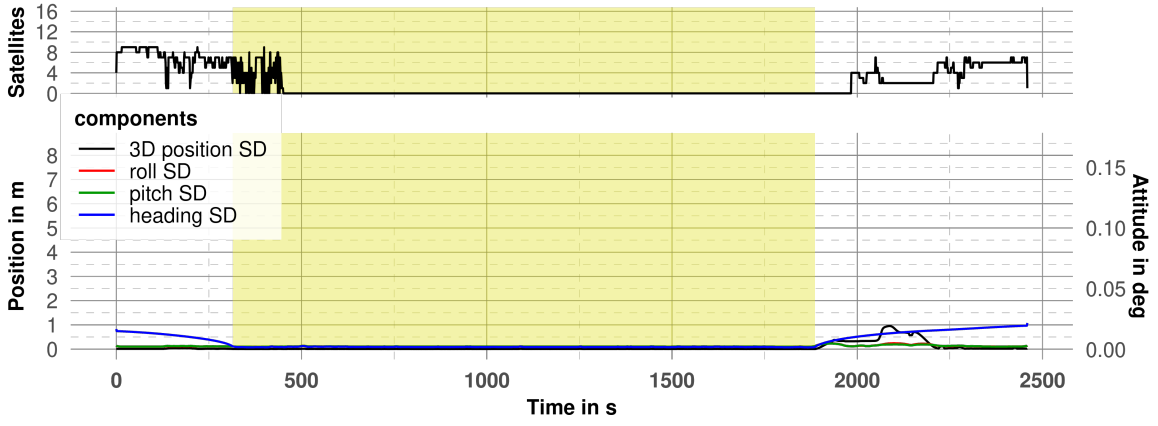
Figure 5.8: Results from the direct georeferencing supported with CUPTs for each test site. The upper part of each diagram shows the satellite availability whereas the bottom of each diagram contains the internal standard deviation, estimated by the Kalman filter. Panoramic images were captured during the yellow marked time period.



(a) City center



(b) Forest



(c) Indoor

Figure 5.9: Results from the direct georeferencing supported with CUPTs and ATTUPTs for each test site. The upper part of each diagram shows the satellite availability whereas the bottom of each diagram contains the internal standard deviation, estimated by the Kalman filter. Panoramic images were captured during the yellow marked time period.

5.4 Investigations on the Georeferencing Methods

While the previous Section 5.3 considered the separate evaluation of the different georeferencing methods and addressed their internal accuracy, this section focuses on the investigation of their precision, absolute accuracy as well as their relative accuracy in different test sites. Thereby we aim to obtain a meaningful performance comparison of the different georeferencing methods under a variety of conditions.

The investigation structure includes the five evaluated georeferencing methods a) direct georeferencing, b) LiDAR SLAM-based georeferencing, c) image-based georeferencing, d) direct georeferencing supported with CUPTs and e) direct georeferencing supported with both CUPTs and ATTUPTs in different test sites 1) city center, 2) forest and 3) indoor. Thereby, we compare the different georeferencing methods in the three test sites in terms of precision and absolute and relative accuracy.

To obtain a meaningful evaluation of the accuracy, we compared image-based point measurements with reference points. Therefore, we conducted image-based point measurements, using forward intersection of four consecutive image poses in the closest possible proximity to the point.

Subsection 5.4.1 describes the investigation on the precision, whereby we used the standard deviation of the forward intersection. The precision indicates the amount of local relative deviations between the four consecutive image poses used. By using the same image measurements for all georeferencing methods, the results become comparable since other error influences, such as the image measurement accuracy or influences of the intersection geometry, occur in equal measure.

Subsection 5.4.2 concerns the investigation on the absolute accuracy, whereby we investigated the amount of the deltas between image-measured points and reference coordinates.

Finally, Subsection 5.4.3 includes the investigation on the relative accuracy. Thereby, we defined 3D distances, using neighboring reference points, with various lengths. This indication is relevant in practice, when, for example, a high 3D distance accuracy is required and the absolute coordinate accuracy is not important.

5.4.1 Precision

The precision shows the closeness of measurements by each other and is a measure of statistical variability. Furthermore, precision is also a requirement to achieve a certain accuracy and can therefore also be part of the absolute accuracy.

In our case, local relative deviations between the image poses, used for forward intersection affect primarily the precision. We investigated the precision for all evaluated georeferencing methods in order to quantify the magnitude of local deviations. Whereas, we used the same image measurements for all georeferencing methods, we obviously used different camera poses. This avoids differential effects of image measurements on the accuracy between the different methods.

The standard deviation of the point measurement with forward intersection from four consecutive images represents the precision. We use the euclidean norm of the resulting covariance matrix $\mathbf{K}_{XX}$

$$m_{3D} = \sqrt{K_{11} + K_{22} + K_{33}} \tag{5.1}$$

as a scalar measure to compare the precision.

Figure 5.10 shows the distribution of the precision for all georeferencing methods and test sites separately in a box-plot diagram. All test sites and measurement methods are subject to outliers. Thus, we used the median in addition to the mean as a statistical parameter which is more robust to outliers.

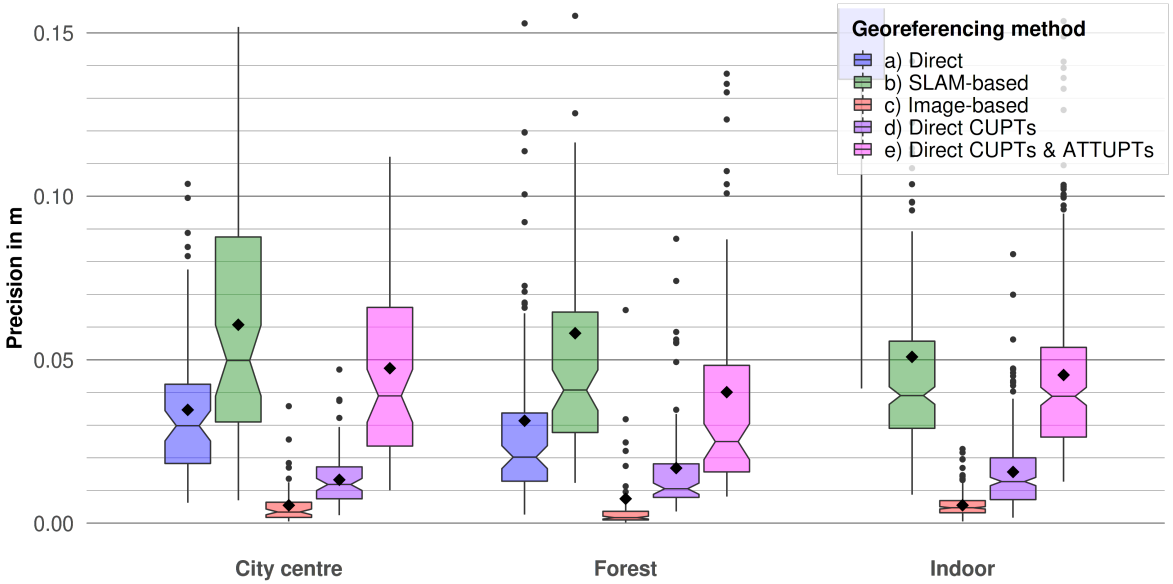


Figure 5.10: Boxplot with the precision of the 3D coordinate observations, arranged by test field. The boxplot colors represent the georeferencing method, whereas the black diamond symbols indicate the mean precision. By contrast, the lines in the center of the boxes show the median precision and the upper and the lower bounds of boxes mark the upper and the lower quartiles.

In both outdoor test sites (city center and forest) direct georeferencing showed a lower median precision (30 mm and 20 mm, respectively) than LiDAR SLAM-based georeferencing (49 mm and 41 mm, respectively). By contrast, in the indoor test site the median precision of SLAM-based georeferencing (39 mm) was significantly lower than the median precision from the direct georeferencing (235 mm, see Table 5.4). Thus, the precision seems to be robust to some minor GNSS degradation but still seems to increase significantly in the complete absence of GNSS. This could be related to the high relative accuracy of the IMU, which only decreases with strong drift after a long absence of GNSS.

In all test sites, the median precision of the image-based georeferencing was significantly lower (by factor 10 or even more) compared to LiDAR SLAM-based georeferencing and direct georeferencing. In the city center, it amounted to 3 mm, in the forest to 2 mm and in the indoor test site to 5 mm.

Furthermore, the support of the direct georeferencing with CUPTs from image-based georeferencing improved the median precision by factors 2-3 in outdoor environments and by factor 20 in the indoor test sites and achieved a median precision in a range of 11 mm to 13 mm.

However, the support of the direct georeferencing with both CUPTs and ATTUPTs did not improve but even slightly degrade the median precision. Thus, contrary to our expectations, additional ATTUPTs had rather a counter-productive effect on the local relative accuracy. External alignment updates possibly introduced discontinuities into the attitude estimation, which is less affected by missing GNSS-updates.

5.4.2 Absolute Accuracy

The absolute accuracy incorporates systematic or gross deviations as well as random errors and is a measure of truthfulness.

In advance, we had conducted rigid 3D body transformations to all georeferencing methods, using the same GCPs in order to guarantee that all measurement methods will be evaluated in the same global reference framework. For the transformation, we used a rigid 3D body transformation, where we first eliminated the shifts by reducing the GCPs to the center of gravity. We then estimated the rotation parameters ω, φ and κ. Thus, the rotational points equal with the centers of gravity. Check points or poses must therefore be transformed as follows:

$$ {}^{w}\mathbf{p}_n = {}^{w}\mathbf{H}_{c_t}\, {}^{c_t}\mathbf{H}_{c_o}\, {}^{c_o}\mathbf{H}_{l}\, {}^{l}\mathbf{p}_n, \tag{5.2} $$

where ${}^{l}\mathbf{p}_n$ is a point in a local coordinate frame, that is being transformed into the global coordinate frame ${}^{w}\mathbf{p}_n$, whereas ${}^{c_o}\mathbf{H}_l$ is the translation from the local coordinate frame into the center of gravity, ${}^{w}\mathbf{H}_{c_t}$ is the translation from the center of gravity to the global coordinate frame, and ${}^{c_t}\mathbf{H}_{c_o}$ represents the rotation.

The translation values x, y and z listed in Table 5.3 designate the concatenated translation related to the global coordinate frame. The translation parameters of the direct georeferencing resulted in the dm-range for the outdoor test sites and in the m-range for the indoor test site. By contrast, the translation as well as the rotation parameters of the LiDAR SLAM-based georeferencing additionally incorporate the transformation from the local to the global reference frame. Translation parameters from the other georeferencing methods resulted in the cm-range.

Table 5.3: Parameters of the rigid 3D body transformations for each test site and georeferencing method. The letter a marks direct georeferencing, b denotes LiDAR SLAM-based georeferencing, c represents image-based georeferencing, d marks direct georeferencing with CUPT support, and e represents direct georeferencing with CUPT and ATTUPT support.

Test site	**Method**	**x** in m	**y** in m	**z** in m	ω in °	φ in °	κ in °
City	a	-0.119	0.381	0.529	-0.233	-0.496	0.113
centre	b	2611329.767	1267111.837	259.357	-0.068	-0.051	26.591
	c	-0.007	-0.016	-0.009	0.012	0.012	-0.011
	d	-0.004	-0.022	-0.018	0.012	0.012	-0.010
	e	0.002	0.028	-0.020	0.019	0.028	-0.032
Forest	a	0.256	-0.148	-0.226	-0.116	1.556	-0.266
	b	2612903.332	1261958.365	281.846	-0.220	-0.031	-75.406
	c	0.019	-0.047	0.080	-0.037	-0.112	-0.004
	d	0.005	-0.046	0.063	-0.019	-0.105	-0.016
	e	-0.016	-0.067	0.062	-0.015	-0.099	-0.035
Indoor	a	1.395	19.296	1.930	-9.747	-15.242	1.045
	b	2615247.016	1264936.289	342.652	-0.097	-0.129	166.553
	c	-0.002	0.001	-0.001	-0.007	-0.004	0.001
	d	-0.005	-0.009	-0.113	-0.013	0.037	0.012
	e	-0.003	0.041	-0.140	-0.016	0.039	-0.045

In order to evaluate the absolute accuracy, we estimated the 3D coordinate deviation that is euclidean norm δ_{3D} between the image-measured point coordinates $\mathbf{i}$ and the corresponding reference points coordinates $\mathbf{r}$, which were measured by tachymetry:

$$\delta_{3d} = \|\mathbf{i} - \mathbf{r}\|_2. \tag{5.3}$$

Thereby, we calculated the image-measured point coordinates $\mathbf{i}$ for each georeferencing method with the transformed image poses by using the same image measurements as for the evaluation of the precision.

While Table 5.4 lists mean as well as median 3D coordinate deviations from different georeferencing methods and test sites, Figure 5.11 shows the error distribution. The direct georeferencing resulted in a median accuracy of 484 mm in the city center, 898 mm in the forest and 7513 mm in the indoor test site. Obviously, the absolute accuracy of the direct georeferencing correlates with GNSS availability. The LiDAR SLAM-based georeferencing resulted in lower median 3D coordinate deviations than direct georeferencing and amounting of 272 mm and 351 mm in both outdoor test sites and to 136 mm in the indoor test site. The absolute accuracy of image-based georeferencing was by factor 1.5 (indoor test site) to 6 (city test site) better than the absolute accuracy of the LiDAR SLAM. The median accuracy amounted to 47 mm in the city center, to 140 mm in the forest and to 78 mm in the indoor test site. Direct georeferencing supported with CUPTs resulted in similar mean 3D coordinate deviations than in image-based georeferencing, while the median 3D coordinate deviations slightly increased. Finally, the direct georeferencing supported by both CUPTs and ATTUPTs was by factor 1.5 to 2 worse than the image-based georeferencing.

Table 5.4: Summary of precision and accuracy of the 3D coordinate observations using different georeferencing methods for all test sites. The table contains both the mean and the median precision and accuracy values. The precision represents the standard deviation of the forward intersection of a single point measurement, whereas the accuracy shows the 3D coordinate deviations to the ground truth.

Method	**Test Site**	**Points**	**Precision** in mm		**Accuracy** in mm	
			mean	median	mean	median
Direct	City center	68	34	30	522	484
	Forest	88	31	20	957	898
	Indoor	246	1225	235	N.A.	7513
SLAM-based	City center	68	61	49	334	315
	Forest	88	58	41	321	272
	Indoor	246	51	39	150	136
Image-based	City center	68	5	3	85	47
	Forest	88	7	2	131	140
	Indoor	246	5	5	87	78
Direct	City center	68	13	12	88	70
CUPTs	Forest	88	17	11	129	125
	Indoor	246	16	13	90	76
Direct	City center	68	47	39	162	149
CUPTs	Forest	88	40	25	168	166
ATTUPTs	Indoor	246	45	39	138	127

Figure 5.12 depicts the error distribution in four error vector maps of the city center, resulting from different georeferencing methods. Figure 5.13 shows the resulting error distribution in the forest and Figure 5.14 only shows three error maps of the indoor test site, where the error map for the direct georeferencing was omitted due to gross errors.

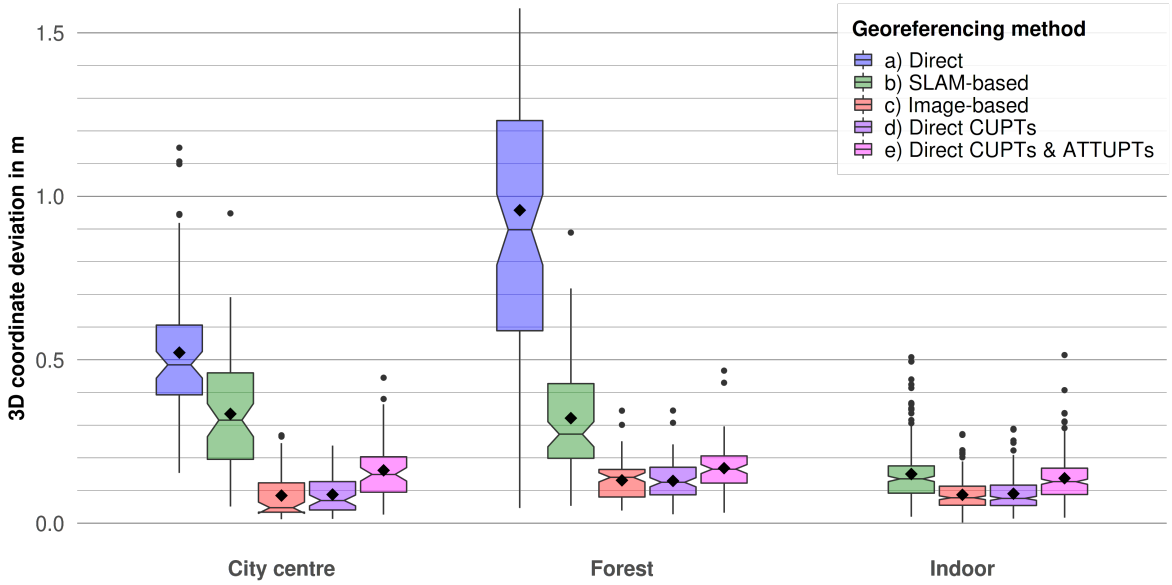


Figure 5.11: Boxplot with 3D coordinate deviations to the ground truth, representing the absolute accuracy. All samples are grouped by test site while the boxplot colors represent the georeferencing method. The black diamond symbols indicate the mean 3D coordinate deviation.

Considering the direct georeferencing, largest 3D coordinate deviations occurred in the center of the entire trajectory which are the farthest in time from the initialization phases at the beginning and end of the campaign. This phenomenon can be recognized in trajectory 2 of the city center (see Figure 5.12a) and in the trajectories 2 and 3 of the forest test site (see Figure 5.13a). Since the error vectors in the indoor test field clearly exceeded the meter range, no further analysis with error vectors was necessary.

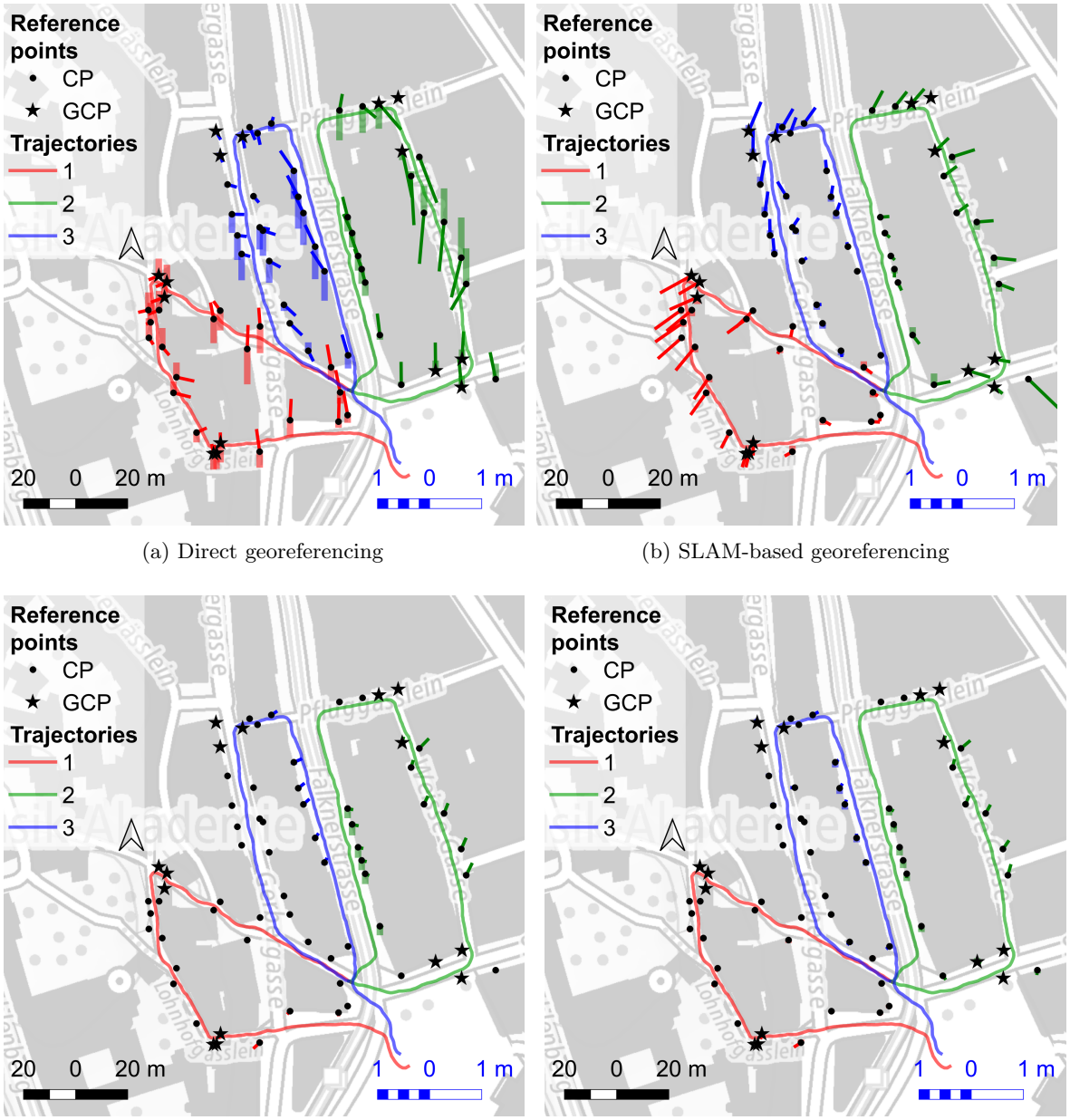
By contrast, regarding the error distribution of LiDAR SLAM-based georeferencing, larger 3D coordinate deviations occurred at the periphery of the test sites, whereas the deviations in the center were much smaller. This phenomenon can be clearly recognized in the city center test site (see Figure 5.12b) and partially in the forest test site when regarding both ends in the north and in the south (see Figure 5.13b). This indicates a scaling whereby the SLAM-based georeferencing had mapped the environment rather too small in extent. By contrast, this error behavior is not apparent in the indoor test site as there were no points in the center of the test site. In the indoor test site gross errors did occur in the south of the test site. The passing through a glass door might have caused systematic errors. While the error vectors of the same points, which were measured in different runs, point in various directions in the SLAM-based georeferencing, they almost point in the same direction in the image-based georeferencing (see Figure 5.14a and b).

With image-based georeferencing, the much smaller errors on the vector plot are hardly visible any more in the city center and in the forest (see Figures 5.12c and 5.13c). However, larger errors in the dm-range occurred in areas with poor initial values or within poorly illuminated areas, such as trajectory 2 of the city center (see Figure 5.12c) and the south-eastern part of the indoor test site (see Figure 5.14b). In addition, deviations in the dm-range occurred in the center of long corridors without GCP support in the indoor test site. In the western corridor, a systematic error behavior in height occurred, while the error vectors in the eastern corridor indicated a systematic drift.

The error behavior of direct georeferencing with CUPT support equals almost with the image-based georeferencing. In some cases with larger systematic deviations of the image-based georeferencing, small improvements can be seen (see Figure 5.12d, trajectories 2 and 3 as well

as Figure 5.13d, trajectory 2). However, direct georeferencing with CUPT support distributed error vectors from the same point which were measured from different runs, which can be observed in the indoor test site (see Figure 5.14c).

Since the support of the direct georeferencing with both CUPTs and ATTUPTs did decrease the accuracy, we refrained from a detailed analysis and from creating an error vector map.



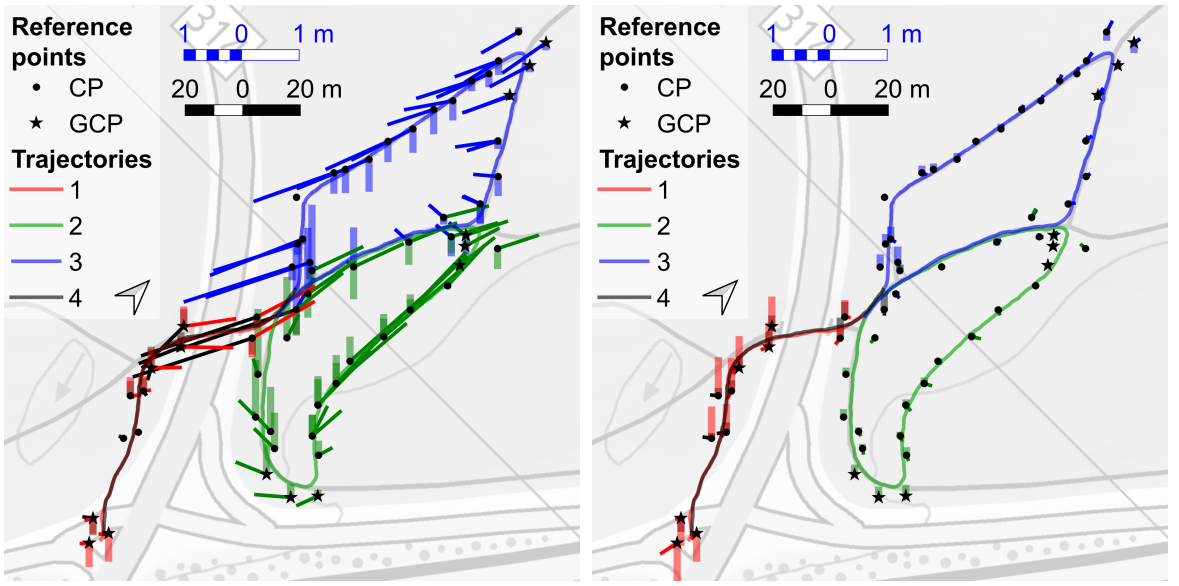
(a) Direct georeferencing

(b) SLAM-based georeferencing

(c) Image-based georeferencing

(d) Direct georeferencing supported with CUPTs

Figure 5.12: Error vector maps of the city test site for different georeferencing methods. The error vectors are colored according to the trajectory section. The thin and opaque error vectors represent the position error while the thicker and transparent vectors represent the elevation error. The map scale is black, whereas the twenty times larger error vector scale is colored in blue.



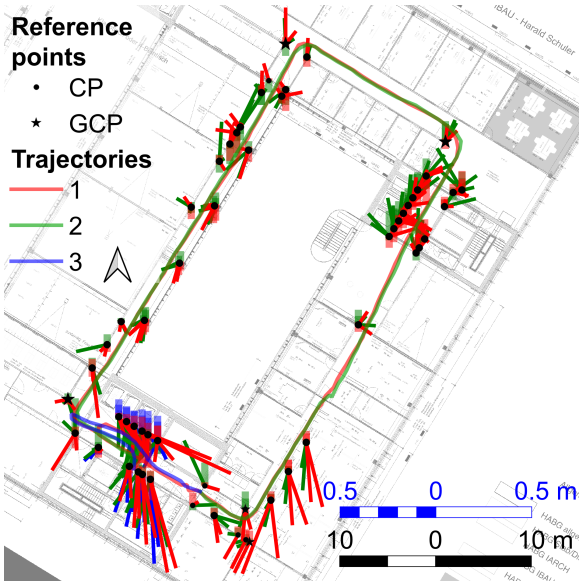
(a) Direct georeferencing

(b) SLAM-based georeferencing

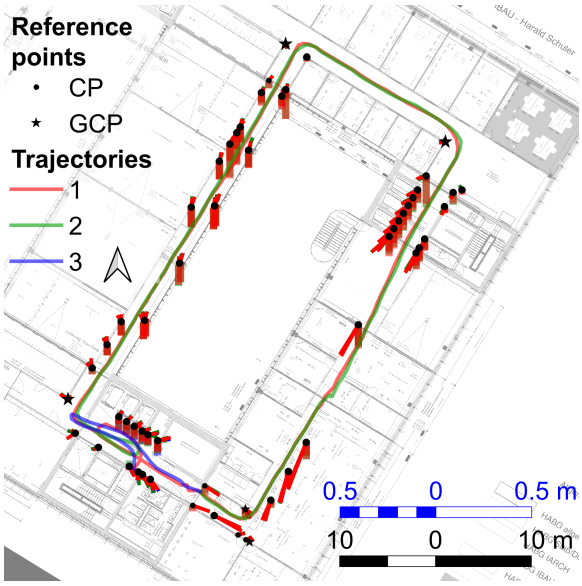
(c) Image-based georeferencing

(d) Direct georeferencing supported with CUPTs

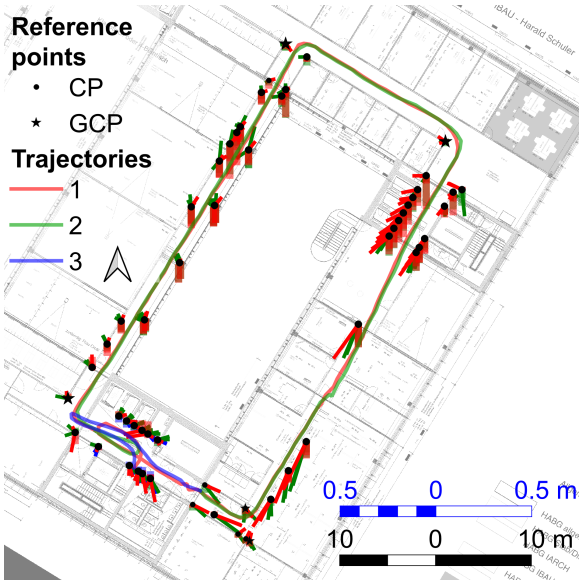
Figure 5.13: Error vector maps of the forest test site for different georeferencing methods. The error vectors are coloured according to the trajectory section. The thin and opaque error vectors represent the position error while the thicker and transparent vectors represent the elevation error. The map scale is black, whereas the twenty times larger error vector scale is colored in blue.



(a) SLAM-based georeferencing



(b) Image-based georeferencing



(c) Direct georeferencing with CUPTs support

Figure 5.14: Error vector maps of the indoor test site for different georeferencing methods. The error vectors are coloured according to the trajectory section. The thin and opaque error vectors represent the position error while the thicker and transparent vectors represent the elevation error. The map scale is black, whereas the twenty times larger error vector scale is colored in blue.

5.4.3 Relative Accuracy

In addition to the precision and the absolute accuracy, we also investigated the relative accuracy. The relative accuracy has a practical relevance when a high accuracy for 3D distance measurements or area measurements is required. Relative accuracy might be possibly less affected by some systematic errors regarding the absolute 3D coordinates.

We investigated the relative accuracy by forming 3D distances with neighboring reference points. Then, we estimated the length deviation δ_l between the 3D distance formed with image-measured point coordinates $\mathbf{i_1}$ and $\mathbf{i_2}$ and the 3D distance formed with reference point coordinates $\mathbf{r_1}$ and $\mathbf{r_2}$ that were measured by tachymetry:

$$\delta_l = \|\mathbf{i_1} - \mathbf{i_2}\|_2 - \|\mathbf{r_1} - \mathbf{r_2}\|_2. \tag{5.4}$$

Thereby, we used the same image measurements as for the precision and accuracy investigations.

Figure 5.15 shows increasing length deviations in test sites with poorer GNSS reception for direct georeferencing. The LiDAR SLAM-based georeferencing resulted in the city center and in the indoor test site in similar median deviations of 55 mm and 39 mm, respectively, while the median deviations in the forest test site were slightly higher and amounted to 76 mm. Image-based georeferencing showed a significantly better relative accuracy (by factor 2–5) whereby the median deviations amounted to 10 mm in the city center, to 23 mm in the forest and to 17 mm in the indoor test site. The support of the direct georeferencing with CUPTs slightly decreases the median relative accuracy in the city center to 20 mm, in the forest to 41 mm and in the indoor test site to 21 mm, while a further support with ATTUPTs leads to a further decrease by factor 1.5–3.0 (see Table 5.5).

Table 5.5: Summary of accuracies of 3D distance observations using different georeferencing methods for all three test sites. The table contains both the mean and the median accuracy values. Accuracy represents the length deviation to the ground truth.

Method	**Test Site**	**Distances**	**Accuracy** in mm	
			mean	median
Direct	City centre	69	90	50
	Forest	85	169	113
	Indoor	140	N.A.	506
SLAM-based	City centre	69	84	55
	Forest	85	93	76
	Indoor	140	56	39
Image-based	City centre	69	27	10
	Forest	85	42	23
	Indoor	140	24	17
Direct CUPTs	City centre	69	35	20
	Forest	85	51	41
	Indoor	140	29	21
Direct CUPTs ATTUPTs	City centre	69	82	56
	Forest	85	96	77
	Indoor	140	44	34

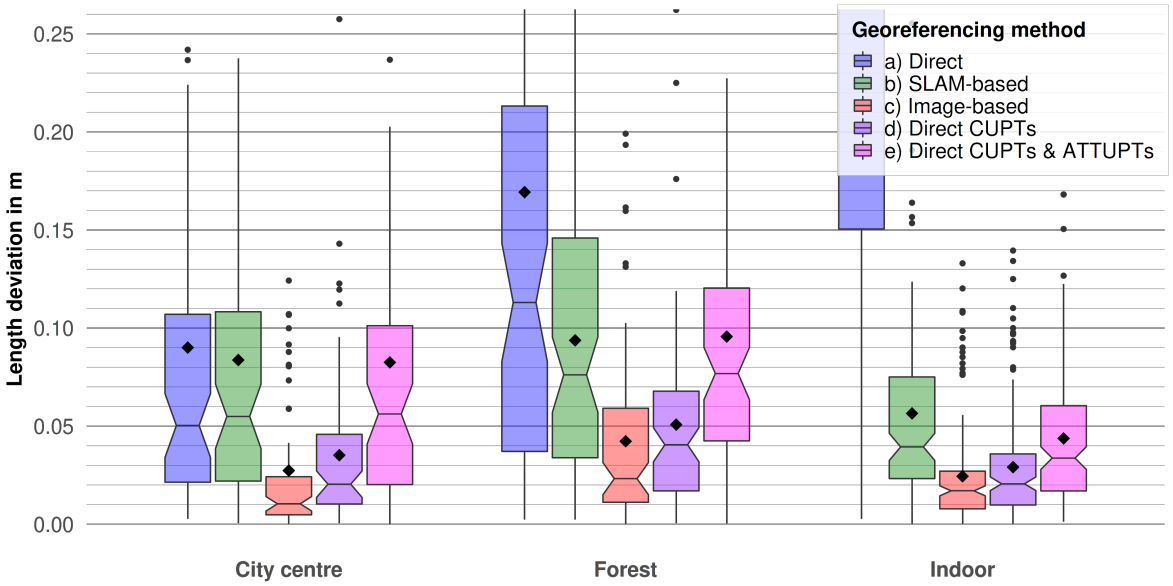


Figure 5.15: Box-plot with length deviations between measured 3D distances and ground truth, representing the relative accuracy. All samples are listed from left to right by test site, while the box-plot colors represent different georeferencing methods. The black diamond symbols indicate the mean 3D distance deviation.

Chapter 6

Discussion and Evaluation

This chapter contains the discussion of our developments and results. It begins with our developed image-focused portable backpack MMS (Section 6.1), continues with the georeferencing methods (Section 6.2) and finally includes a comparison of our results with related work (Section 6.3).

6.1 Developed Image-focused Portable Backpack MMS

This section contains a discussion of the strengths and shortcomings of our developed portable MMS with regard to practical use and application in infrastructure management.

Platform

The backpack platform used allows to cover a wide range of different areas that are inaccessible to other platform types such as trolley or vehicle platforms. This thesis successfully demonstrates its application in the city center, in forests and in indoor environments.

Furthermore, the developed backpack platform allows sufficiently large payload and enough space to use high-end sensors and implement different sensor configurations. For our prototypical implementation, we used a stable aluminum frame with a modular design, that allows changes on the sensor configuration in a straightforward manner. This enabled the incremental development from an only indoor MMS (Blaser et al., 2018a) to a MMS suited for both indoor and outdoor environments (Blaser et al., 2020), that is extended by a second vertical LiDAR scanner and holds a GNSS and IMU-combined navigation system with tactical grade performance instead of the formerly used industrial grade IMU.

However, the modular and flexible platform design resulted in an overall weight of 25 kg where the frame already weights 11 kg. Hence, there is great potential for weight reduction through redesign and industrialization.

Sensor Configuration

Since the backpack platform does not offer enough space for a stereo panoramic camera configuration with reasonable base lines, we propose a combined panoramic camera and LiDAR configuration. In contrast to numerous LiDAR-focused portable MMS with comparable sensor configurations that provide colored LiDAR point clouds, we focused on the resulting panoramic images with the aim of using LiDAR point clouds for depth reconstruction, yielding RGB-D panoramic images. A stereo panoramic camera configuration provides images with temporally

coherent RGB-D channels. By contrast, using a combined configuration with (panoramic) camera and LiDAR scanners provides LiDAR points that are not acquired at the same time as the images so that moving objects cannot be captured with the correct depth information. This limits the field of application to static objects.

As described in Section 3.3, the sensor configuration of our portable MMS incorporates a multi-head panoramic camera *Ladybug 5*, two multi-beam LiDAR scanners *Velodyne VLP-16* and a GNSS and IMU-combined navigation unit *SPAN CPT7*.

The panoramic camera and the LiDAR scanners provide a large coverage of the environment. However, the operator's head obscures a significant area of the panoramic camera of the front and the horizontally mounted LiDAR scanner, which sits on top of the frame, obscures a small part of the top left (see Figure 3.6). In addition, the area towards the top tends to be well imaged, while the ground is only captured by the camera from a certain distance. Within buildings, coverage towards the top is essential, since the ceiling often contains objects and installations that are important for infrastructure management. By contrast, in outdoor environments, the top view often captures only the sky, while the ground with road or rail infrastructure is of higher importance. Therefore, it could be useful to supplement the camera configuration with additional cameras pointing downwards.

Sensors

The used panoramic camera *Ladybug 5* satisfied the requirements for robust and precise georeferencing under difficult conditions. Both a higher camera resolution and better low light performance would nevertheless be desirable for infrastructure management in order to recognize critical scene details such as labels and to allow a more efficient data acquisition in indoor environments without halting during image captures. Both improvement options conflict with each other to some extent and require balancing to the physical optimum. Furthermore, the lack of hardware triggering interfaces on the onboard computer demanded the triggering of the panoramic camera by software. This limited the possible frame rate of 5 frames per second to 2 frames per second.

The *Velodyne VLP-16* LiDAR scanners used proved themselves suitable for initial SLAM-based georeferencing. However, the limited field of view due to the small vertical angle of 30° required the use of a second vertically mounted scanner in order to cover the ceiling and ground in indoor environments.

Since the *NovAtel SPAN CPT7* GNSS and IMU-combined navigation unit provide raw data access, it was possible to use IMU raw data for both direct georeferencing and SLAM-based georeferencing.

6.2 Georeferencing Methods

This section contains a discussion of strengths and shortcomings of our developed georeferencing methods.

Direct Georeferencing

Direct georeferencing using GNSS and IMU with tactical grade performance may provide cm accuracy under good GNSS reception. In areas with partial GNSS obscurity, the position accuracy rapidly decreases to the dm range and in areas with complete GNSS absence it decreases to the m range or even more. Thus, direct georeferencing is suitable for areas with

good GNSS reception and also provide suitable initial values for image-based georeferencing in areas with partial GNSS obscurity.

Our investigations in the city center and in the forest test site resulted in mean accuracies of 52.2 cm and 95.7 cm, respectively, which is sufficient to provide suitable initial values for image-based georeferencing with spatial feature matching. In contrast, directly georeferenced image poses in the interior proved insufficient as initial values due to the strong drifts in the clear m range.

Nevertheless, direct georeferencing is rapidly processed and results in a trajectory with high pose density that is suitable for georeferencing of both images and LiDAR points.

LiDAR SLAM-based Georeferencing

Our extended LiDAR SLAM algorithm fully operates in 3D and provides poses with 6 DoF.

Since point cloud matching, local constraints as well as loop closures base on a 3D occupancy grid map, the grid resolution strongly affects the precision and the relative and absolute accuracy. For our investigations, we used a voxel size of 5 cm. This manifests itself in the reported internal accuracy, which shows deviations between poses, local constraints as well as loop closures. The mean translational errors varied from 5.3 cm to 6.6 cm, while mean rotational errors varied in the range from 0.2° to 0.5°. The mean precision ranged in the same order of magnitude from 5.1 cm to 6.1 cm, mainly showing very local deviations between the four consecutive poses used for the forward intersection. The mean accuracy of 3D distance observations between neighboring CPs ranged from 5.6 cm to 9.3 cm and showed slightly higher deviations than the mean precision.

The mean accuracy of absolute coordinate deviations amounted in indoors to 15 cm whereas it was twice as high in both the city center (33 cm) and in the forest (32 cm). This significant difference could be related to the size of test sites. While the test sites city center and forest have a similar size, the indoor test site is significantly smaller. This behavior could indicate a scale effect over the entire map. Another indication for a global scale factor is the typical systematic error behavior, which is especially visible in the error vector plot of the city center test site (see Figure 5.12b). In the future, a global scaling factor could be considered in the transformation from the local to the global coordinate system.

In our investigations, our SLAM-based georeferencing pipeline based on *Cartographer* proved to be robust in different environments. Basically, our LiDAR SLAM-based georeferencing is real-time capable but claims high RAM and Central Processing Unit (CPU) resources which increase with growing map size. Hence, the map size of LiDAR SLAM is limited.

Chen et al. (2021b) performed investigations on the *Cartographer* within an indoor test site with equally shaped rooms, which resulted in gross errors in the m range. They point out, that the *Cartographer* tends to erroneous loop closures within buildings with equally shaped rooms, because the *Cartographer* detects loop closures based on the point cloud similarity. Consequently, they recommend either to deactivate loop closures or to use another SLAM algorithm for campaigns within buildings with equally shaped rooms.

All in all, LiDAR SLAM-based georeferencing is suitable for various environments and results in dm accuracy. LiDAR SLAM clearly outperformed direct georeferencing in GNSS-denied areas, in terms of accuracy and robustness and proves itself as an alternative method to obtain initial values for image-based georeferencing. Although, LiDAR SLAM results in a dense trajectory and the high local and relative pose deviations cause a noisy point cloud when used for LiDAR georeferencing.

Image-based Georeferencing

Subsequent image-based georeferencing results in a highly accurate but sparse trajectory which consists of image poses with 6 DoF.

Image-based georeferencing using the SfM pipeline *Agisoft Metashape* resulted in a mean re-projection error in outdoor environments of 0.4 px and 0.5 px, respectively, and in the indoor test site of 2.2 px. As already discussed in 5.3.3, reasons for the significantly higher mean re-projection error in indoor environments could be that the minimum focus distance of the *Ladybug 5* panoramic camera was missed or the higher exposure time in indoors led to motion and image blur. However, shorter object distances in indoors might compensate the higher re-projection error, since the mean precision ranged in all test sites from 0.5 cm to 0.7 cm. The precision achieved by image-based georeferencing was by an order of magnitude higher than the precision achieved by both direct and LiDAR-based georeferencing. Image-based georeferencing thus showed the highest precision of all evaluated georeferencing methods. The mean accuracy of 3D distance observations between neighboring CPs ranged from 2.4 cm to 4.2 cm and the mean absolute coordinate accuracy ranged from 8.5 cm to 13.1 cm. Consequently, subsequent image-based georeferencing improved the mean absolute coordinate accuracy of direct and LiDAR SLAM-based georeferencing by factors 2–7.

Furthermore, image-based georeferencing proved to be highly robust within different environments. In our investigations, all images were successfully oriented, even in the city center with numerous pedestrians and moving vehicles, in the mature forest with low-light conditions as well as in challenging indoor environments with glass façades, texture-less walls and repetitive patterns. Most likely, the constraint rig configuration used for the bundle-adjustment significantly increased the robustness, since it requires a much smaller number of features to georeference a panoramic epoch. However, under really harsh conditions, for example in a mature forest, some outliers with deviations in the dm-range occurred.

Since the portable MMS features a single panoramic camera configuration, leading to single panoramic image sequences, it requires GCPs or accurate image poses from external sources to precisely estimate the scale of the 3D reconstruction. In addition, the processing of the panoramic image sequences results in a drift, which is for example visible in the error vector plot of the indoor test site (see Figure 5.14b). However, the drift could be tackled with additional GCPs. Another limitation of image-based georeferencing is that the processing time significantly exceeds the acquisition time, being by factor 2–4 higher.

Direct Georeferencing Supported with Coordinate and Attitude Updates

Direct georeferencing supported with CUPTs and ATTUPTs that originate form image poses enables the bridging of GNSS-denied areas and results in a highly accurate and dense trajectory, consisting of image poses with 6 DoF. Thus, it is suitable for both the georeferencing of images and LiDAR points.

As already discussed in Subsection 5.3.4, we created the two following differently combined CUPTs and ATTUPTs:

- CUPTs only, whereby we weighted each coordinate component with 5 mm.
- combined CUPTs and ATTUPTs, whereby we weighted each coordinate component with 5 mm, roll and pitch angles with 0.1° and the heading angle with 0.005°.

Combined CUPTs and ATTUPTs showed better internal accuracy than CUPTs only. CUPTs only improved the mean standard deviation of 3D position to the range from 0.5 cm to 7.0 cm, while the mean heading standard deviation remained in the range of direct georeferencing.

Combined CUPTs and ATTUPTs additionally improved the mean heading standard deviation to a range from 0.004° to 0.005°.

However, the mean precision of the CUPT only supported direct georeferencing was by factor 2 lower than the mean precision of the image-based georeferencing, and ranged from 1.3 cm to 1.6 cm. The further addition of ATTUPTs decreased the precision by an order of magnitude so that it came to range between 4.0 cm and 4.7 cm.

The same behavior occurred with the relative accuracy. The mean accuracy of 3D distances between neighboring CPs decreased to a range from 2.9 cm to 5.1 cm and from 4.4 cm to 9.6 cm by supporting the direct georeferencing with CUPTs only and with combined CUPTs and ATTUPTs, respectively. This behavior may be explained by the fact that direct georeferencing smoothed and thereby deteriorated the introduced image pose.

The absolute accuracy of CUPT supported direct georeferencing within a range from 8.8 cm to 12.9 cm is comparable with image-based georeferencing, while a further support of the direct georeferencing with ATTUPT slightly decreased the absolute accuracy to a range from 13.8 cm to 16.8 cm. In total, direct georeferencing with only CUPT support significantly outperformed direct georeferencing with combined CUPT and ATTUPT support. Consequently, direct georeferencing supported only by CUPTs is preferable in the future.

Direct georeferencing supported with CUPTs and ATTUPTs significantly improved the robustness of direct georeferencing. Especially CUPT originating from image-based georeferenced poses, were able to bridge areas with poor GNSS reception and even to bridge indoor environments with complete GNSS absence. Furthermore, CUPT supported direct georeferencing even slightly increased the robustness of image-based georeferencing by smoothing outliers.

However, direct georeferencing supported with CUPTs and ATTUPTs depends on the result of image-based georeferencing, which requires a long processing time. From this point of view, the discussed georeferencing method is not only precise and highly robust, but also resource-intensive and time-consuming.

6.3 Comparison of our Results with Related Work

This section includes a discussion and comparison of our georeferencing methods with related work in terms of the hardware used, the evaluation methods and the achieved results. By general, our results confirm those of our previous work as well as of work from other groups in terms of accuracy, but presents new findings about increasing robustness and versatility.

Lehtola et al. (2017) carried out investigations on point clouds of indoor environments using a comparable backpack MMS. Their deviations reached up to 14 cm and 55 cm in floor heights with the *Leica Pegasus Backpack* and the *Würzburg Backpack*, respectively and are thus in the same order of magnitude. However, their results are not directly comparable to ours, because they only concern the height component and show maximum rather than mean deviations.

By contrast, Corso, Zakhor (2013) investigated the end-to-end system performance of their LiDAR-focused backpack indoor MMS in an indoor test site that is comparable to ours in terms of size and conditions. In contrast to our investigations, they measured CPs by intersecting the manually measured line of sight in the camera image with a plane, that was previously fitted into the mapped LiDAR points closest to the target. Their analysis, using 100 surveyed CPs, showed a mean accuracy of 10 cm, while the mean accuracy of our LiDAR SLAM-based georeferencing showed a mean accuracy of 15 cm using 246 surveyed CPs. However, their offline 3D localization pipeline consists of a 2D LiDAR SLAM that is based on 2D occupancy grids, while they regard the 1D height estimation separately. As the height estimation depends on vertical planes, the method is limited to man-made environments.

While their position part of the pose originates from a 2D LiDAR SLAM and from a 1D height estimation algorithm, they use roll and pitch angles from the IMU and the yaw angle originating from 2D LiDAR SLAM. By contrast, our LiDAR SLAM used fully operates in 3D and uses sub-maps based on 3D occupancy grids. It performs robustly in various environments. However, the limited resolution of the 3D occupancy grid possibly decreased the accuracy of our LiDAR SLAM-based georeferencing. Nonetheless, our mean accuracy of both image-based georeferencing and CUPT supported direct georeferencing amounted to 9 cm and thus slightly outperformed their algorithm.

Tang et al. (2015) performed experiments with an all-terrain-vehicle LiDAR-based MMS in the forest within similar conditions. In mature forest, they reported 2D stem position deviations to the reference with GNSS and IMU as well as with SLAM and IMU in the range of 40 cm to 72 cm and of 4 cm to 45 cm, respectively. Our 3D coordinate differences with direct georeferencing were slightly higher, which might result from the fact that we additionally considered the third dimension and that our acquisition speed was significantly slower. By contrast, their deviations of the SLAM-based georeferencing are comparable to ours. Although, our investigated SLAM-based georeferencing seems to be more robust to environmental changes.

As proven in our previous work (Cavegn et al., 2018), subsequent image-based georeferencing significantly improved the accuracy by a multiple and the precision by an order of magnitude compared to published results with direct or SLAM-based georeferencing in similar environments. Moreover, image-based georeferencing was able to bridge significant deviations and therefore increased the robustness of georeferencing.

Chapter 7

Conclusion and Outlook

This chapter includes a summary of this thesis (see Section 7.1), shows limitations and provides recommendations for further work (see Section 7.2). Finally, Section 7.3 provides an outlook with regard to subsequent research questions as well as current and future technological developments.

7.1 Summary

The overall objective of this thesis was to provide robust and accurate georeferencing of RGB-D images originating from MM in various challenging environments. MM enables rapid and efficient 3D data capturing and its resulting data can be used in cloud-based web services, for example for infrastructure management in combination with GIS or BIM. While Nebiker et al. (2015) indicate the high potential of cloud-based geospatial image spaces using georeferenced RGB-D images, it is meanwhile already commercially deployed and widely used for infrastructure management. However, in outdoor environments, this approach has so far only been used for linear infrastructures, such as roads or railways, that are easily accessible with large platforms, using high-quality sensor technology. Such infrastructures often include areas that are inaccessible to vehicles, such as operational buildings or narrow alleys, for which there is a great demand for equally accurate coverage.

In a first step, we successfully developed a prototypical portable image-focused backpack MMS, which is equipped with a multi-head panoramic camera *Ladybug 5* for omni-directional image coverage, with two multi-profile LiDAR scanners *Velodyne VLP-16* (horizontal and vertical) and with a GNSS and INS-combined navigation unit *SPAN CPT7* with tactical grade performance. Its modular and flexible hardware and software design, which is based on a robust aluminium frame and the ROS, respectively, allowed straightforward adaptions of the sensor configuration. Furthermore, all sensors are precisely synchronized, based on hardware pulses, which is essential for kinematic systems.

We successfully calibrated our backpack MMS in test fields. For camera related parameters (IO, RO and BA), we applied a bundle-adjustment-based calibration approach, which we extended by the equidistant camera model in order to support the wide-angle optics of our panoramic camera heads. Furthermore, we precisely calibrated the RO parameters from both LiDAR scanners to the panoramic camera, using a feature-based calibration approach, which requires no targets but minimizes the LiDAR point cloud deviation by the use of point cloud features. Thereby, the standard deviations of the RO parameters between the panoramic camera heads were in the sub-mm range for the translation and ranged from 0.0044° to 0.0233° for the rotation. Moreover, the standard deviations of the RO parameters between the LiDAR scanners and the panoramic camera were in the same order of magnitude. They were in the

mm range for the translation, whereas they ranged from 0.0001° to 0.0162° for the rotation. Accurate RO parameters between LiDAR scanners and cameras are crucial for LiDAR point cloud and image data fusion.

We considered direct georeferencing and LiDAR SLAM-based georeferencing as two well established methods for outdoor and indoor environments. While direct georeferencing strongly relies on good GNSS reception and thus is only suitable for open field application, LiDAR SLAM depends on environments with distinct geometric structures and with geometric variety. Thus, LiDAR SLAM is also suitable for GNSS-denied environments. Direct georeferencing operates in a global reference frame and has no limitations regarding the expansion of the trajectory. By contrast, most LiDAR SLAM algorithms operate in a local coordinate frame and have a limited maximum map size. Both methods are real-time capable and result in a dense trajectory, which can be used for georeferencing of both image and LiDAR data.

In order to improve the accuracy and robustness of both direct and LiDAR SLAM-based georeferencing, we considered subsequent image-based georeferencing. Cavegn et al. (2018) extended the SfM pipeline *COLMAP* by indirect georeferencing capabilities, which includes the support of prior EO parameters, RO constraints for fixing of the camera rigs, as well as GCP. This showed improvements in accuracy by an order of magnitude compared to direct and LiDAR SLAM-based georeferencing. Consequently, subsequent image-based georeferencing results in a highly accurate but sparse trajectory that can only be used for image georeferencing.

In order to obtain a highly accurate and dense trajectory, we proposed direct georeferencing supported with CUPTs and ATTUPTs originating from image-based georeferenced poses. This method should support and improve the trajectory in areas with poor or no GNSS reception. Finally, we implemented and investigated two variants, one using only CUPTs and the other using combined CUPTs and ATTUPTs. Thereby, we expected that the combined variant would correct drifts in the heading component faster and that it would yield further improvements in the absolute accuracy.

We evaluated and extensively investigated all considered georeferencing methods in three test sites that represent challenging real-world scenarios. The first test site was located in the city center, including narrow alleys with poor GNSS reception, busy with pedestrians and trams. The second test site was located in a forest with mature and dense vegetation and paths that lead through the undergrowth. The third test site represents a public indoor environment with glass façades and repetitive patterns. Beside the internal accuracy of each georeferencing method, we investigated both the precision and the accuracy of image-measured CPs, that we had previously measured with superior accuracy by tachymetry. For this purpose, we performed the image measurements with forward intersection of four consecutive panoramic image epochs. Furthermore, we investigated the relative accuracy by comparing image-measured 3D distances to ground truth.

Subsequent image-based georeferencing showed a median precision that was by an order of magnitude higher than of direct georeferencing and LiDAR SLAM-based georeferencing and ranged from 5 mm to 7 mm in the different test sites. It also showed by factor 2–7 higher absolute accuracies, ranging from 85 mm to 131 mm. Furthermore, subsequent image-based georeferencing showed a higher relative accuracy by factor 2, ranging from 24 mm to 42 mm. The absolute and relative accuracies of direct georeferencing with CUPT support were similar to those of subsequent image-based georeferencing, whereas the mean precision was slightly lower (13 mm to 16 mm). Contrary to our expectations, direct georeferencing supported with combined CUPTs and ATTUPTs showed a lower precision and absolute as well as relative accuracy than direct georeferencing with CUPT support only.

7.2 Recommendations

Our investigations show great potential for complementing 3D geospatial web services of infrastructures, using a high-performance backpack MMS that combines direct georeferencing with CUPTs from image-based georeferencing. As discussed in Section 6.3, our obtained absolute accuracy confirms both the results from previous work, using image-based MMS with fixed stereo bases (Blaser et al., 2018b; Burkhard et al., 2012; Cavegn et al., 2018), and work by other groups, using LiDAR-focused MMS (Tang et al., 2015; Corso, Zakhor, 2013). However, our obtained relative accuracy was slightly lower than the results achieved with fixed stereo bases (Blaser et al., 2018b). Overall, our portable MMS, in combination with the proposed georeferencing methods, has proved itself suitable for an equivalent extension of existing 3D geospatial services for infrastructures in terms of the accuracy and robustness of the georeferencing.

Our prototypical backpack MMS has a high overall weight of 25 kg and the panoramic camera is partially obscured by the top LiDAR scanner as well as by the operator. A new arrangement of the sensors could reduce the occlusion of the panoramic camera, for example by a placement of the panoramic camera directly above the operator's head and the horizontal LiDAR scanner just below the panoramic camera. However, this would require a smaller panoramic camera, in order to avoid collisions with the ceiling or with door frames in indoor environments. Furthermore, the frame structure, as well as the wiring of the sensors, could be massively simplified by a fixed and industrialized configuration, which would bring a significant weight reduction.

The performed overall system calibration included IO, RO and BA parameters of the individual panoramic camera heads and the RO and BA parameters of both LiDAR scanners. However, we had not considered the internal parameters of the LiDAR scanners, as they were not crucial for the LiDAR SLAM-based georeferencing. Nevertheless, Kim et al. (2021) achieved significant improvements of planar misclosure after calibration in the range from 35 % to 81 % by considering the internal parameters. Thus, internal calibration of LiDAR scanners might reduce point cloud noise. This could be particularly relevant for subsequent depth reconstruction in order to obtain georeferenced RGB-D images by the use of georeferenced LiDAR point clouds.

Our proposed multi-stage georeferencing approach for vehicle-denied areas proved to be robust in our test sites with comparable accuracy to direct georeferencing under optimal conditions. However, as Chen et al. (2021b) determine, our LiDAR SLAM algorithm *Cartographer* used tends to wrong loop-closures in indoor environments with uniform structures (e.g. similar-shaped rooms). This is caused by the loop-closure detection algorithm that is based on point cloud similarity. It is therefore advisable to deactivate loop-closure or to use alternative SLAM algorithms for indoor campaigns in large building complexes. For larger combined indoor and outdoor campaigns, we recommend to use direct georeferencing as initial values for the subsequent georeferencing and to use LiDAR SLAM only for local areas with restricted GNSS reception.

The positional accuracy of LiDAR SLAM-based georeferencing could be improved by applying a scale factor, as indicated by the results of our investigations.

Furthermore, we recommend to deactivate the real-time GNSS and INS-combined navigation and only to record raw data for large indoor or underground campaigns with long GNSS absence. GNSS and INS-combined navigation systems have a regulatory velocity limit of 515 m/s and deactivate navigation and all raw data capturing when exceeding the maximum velocity. This limit can be exceeded in case of strong drifts caused by long GNSS outages.

When supporting the direct georeferencing from external sources, for example from image-based georeferenced poses, we recommend to primarily use only CUPTs, since the combined use of CUPTs and ATTUPTs resulted in a significantly lower precision and accuracy. There might be a potential to improve the combined CUPTs and ATTUPTs supported direct georeferencing by optimizing the filter algorithms.

7.3 Outlook

The most obvious research objectives concern dense and accurate 3D reconstruction from both image and LiDAR data to fully meet the requirements for obtaining georeferenced RGB-D imagery. The simplest method would be to georeference the LiDAR data using the dense trajectory from direct georeferencing supported by CUPTs, and afterwards re-project the resulting point clouds into the images. Among other things, the following questions should be addressed:

- Filter and remove points in the point cloud that are invisible in the image.
- Filtering moving objects in the point cloud (e.g. vehicles or pedestrians).
- Handling moving objects visible in the image that were not captured by the LiDAR scanners.

Another possibility would be the 3D reconstruction using georeferenced imagery. However, image-based 3D reconstruction strongly depends on the object texture and the image quality. Furthermore, it would be interesting to develop and investigate hybrid approaches for dense 3D reconstruction, based on both image and LiDAR data balancing the advantages and disadvantages of each other.

With the reconstructed depth information, it would be possible to investigate the end-to-end accuracy by measuring CPs directly in single RGB-D images with mono-plotting.

Other relevant research areas could aim towards high quality SLAM for infrastructure management, by using precise georeferencing for 3D reconstruction on the one hand, and on the other hand to optimize the reconstructed 3D scene, using e.g. machine learning to improve the georeferencing.

Current developments in sensor technology, such as RGB-D cameras equipped with solid-state LiDAR, have great potential to massively simplify and miniaturize the entire MMS. Presently, the achievable accuracy and range of solid-state LiDAR are below the requirements for georeferenced RGB-D imagery for infrastructure management. However, Nebiker et al. (2021) show the great potential of using RGB-D cameras combined with low-cost MM components for applications with accuracy requirements in the m- to sub-m-range (e.g. for on-street parking statistics). With additional maturity and higher accuracy, solid-state LiDAR-based RGB-D cameras have great potential to replace multi-profile LiDAR scanners in future, since it would allow to capture RGB-D images that coincide in time and would provide dense depth information. In addition, this would massively simplify georeferencing indoor environments, as the subsequent direct georeferencing with CUPT support could be dropped when multi-profile LiDAR data is no longer required.

Furthermore, investigations towards the on-site calibration of the overall system would be of great interest. First, this would allow additional consideration of prevailing environmental conditions, such as different temperatures or humidity. Second, great efforts of regular test field calibration could be avoided.

In terms of georeferencing, ongoing developments in visual localization (Sarlin et al., 2019) have great potential to reduce or omit GCPs and CPs within areas with existing accurately georeferenced RGB-D image data, that was previously acquired with a high-performance MMS. Since the visual localization using deep-learned features allows image matching of images with different times and environmental conditions, extensions to existing image data sets or updates in already mapped areas could be performed without any control points. Consequently, the transformation from the local to the global reference frame could be performed using image poses determined with visual localization. As a result, the entire georeferencing process could be fully automated.

In addition, robust and accurate georeferenced RGB-D imagery of urban environments combined with visual localization has an enormous potential for future on-site AR applications with very high practical relevance.

Bibliography

Abraham, S., Förstner, W., 2005. Fish-eye-stereo calibration and epipolar rectification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(5), 278–288. DOI: 10.1016/j.isprsjprs.2005.03.001.

Alismail, H., Baker, L. D., Browning, B., 2012. Automatic Calibration of a Range Sensor and Camera System. *Proceedings of the 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, 286–292. DOI: 10.1109/3DIMPVT.2012.52.

Alsadik, B., Karam, S., 2021. The Simultaneous Localization and Mapping (SLAM)-An Overview. *Surveying and Geospatial Engineering Journal*, 2(01), 01–12. DOI: 10.38094/sgej1027.

Angrisano, A., 2010. GNSS/INS Integration Methods. PhD thesis, Universita' degli studi di Napoli "Parthenope". `https://www.ucalgary.ca/engo_webdocs/other/Angrisano_PhD%20Thesis_ENG2.pdf`, (accessed: 18/5/2022).

Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., Weaver, J., 2010. Google Street View: Capturing the World at Street Level. *Computer*, 43(6), 32–38. DOI: 10.1109/MC.2010.170.

Applanix Corporation, 2017. TIMMS: Fast, Accurate & Cost-Effective Indoor Mapping. `https://www.applanix.com/downloads/products/brochures/2017_Timms2Brochure_r1_feb2017_web.pdf`, (accessed: 18/5/2022).

Aqel, M. O., Marhaban, M. H., Saripan, M. I., Ismail, N. B., 2016. Review of visual odometry: types, approaches, challenges, and applications. *SpringerPlus*, 5(1), 1–26. DOI: 10.1186/s40064-016-3573-7.

Barber, D., Mills, J., Smith-Voysey, S., 2008. Geometric validation of a ground-based mobile laser scanning system. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(1), 128–141. DOI: 10.1016/j.isprsjprs.2007.07.005.

Beers, B. J., 1995. FRANK - the design of a new landsurveying system using panoramic images. PhD thesis, Technische Universiteit Delft. `http://resolver.tudelft.nl/uuid:72a50ee7-332e-43aa-b216-fba148b28d60`, (accessed: 18/5/2022).

Behley, J., Stachniss, C., 2018. Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments. *Robotics: Science and Systems*. DOI: 10.15607/RSS.2018.XIV.016.

Besl, P. J., McKay, N. D., 1992. Method for registration of 3-D shapes. P. S. Schenker (ed.), *Sensor Fusion IV: Control Paradigms and Data Structures*, 1611, International Society for Optics and Photonics, SPIE, 586–606. DOI: 10.1117/12.57955.

Blaser, S., Cavegn, S., Nebiker, S., 2018a. Development of a Portable High Performance Mobile Mapping System Using the Robot Operating System. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1/1, 13–20. DOI: 10.5194/isprs-annals-IV-1-13-2018.

Blaser, S., Meyer, J., Nebiker, S., 2021. Open Urban and Forest Datasets from a High-Performance Mobile Mapping Backpack – a Contribution for Advancing the Creation of Digital City Twins. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B1-2021, 125–131. DOI: 10.5194/isprs-archives-XLIII-B1-2021-125-2021.

Blaser, S., Meyer, J., Nebiker, S., Fricker, L., Weber, D., 2020. Centimetre-Accuracy in Forests and Urban Canyons – Combining a High-Performance Image-Based Mobile Mapping Backpack with new Georeferencing Methods. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-1-2020, 333–341. DOI: 10.5194/isprs-annals-V-1-2020-333-2020.

Blaser, S., Nebiker, S., Cavegn, S., 2017. System Design, Calibration and Performance Analysis of a Novel 360° Stereo Panoramic Mobile Mapping System. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1/W1, 207–213. DOI: 10.5194/isprs-annals-IV-1-W1-207-2017.

Blaser, S., Nebiker, S., Cavegn, S., 2018b. On a Novel 360° Stereo Panoramic Stereo Mobile Mapping System. *Photogrammetric Engineering & Remote Sensing*, 84(6), 347–356. DOI: 10.14358/PERS.84.6.347.

Bosse, M., Zlot, R., Flick, P., 2012. Zebedee: Design of a Spring-Mounted 3-D Range Sensor with Application to Mobile Mapping. *IEEE Transactions on Robotics*, 28(5), 1104–1119. DOI: 10.1109/TRO.2012.2200990.

Brown, D. C., 1971. Close-range camera calibration. *Photogrammetric Engineering*, 37(8), 855–866.

Burkhard, J., Cavegn, S., Barmettler, A., Nebiker, S., 2012. Stereovision Mobile Mapping: System Design and Performance Evaluation. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B5, 453–458. DOI: 10.5194/isprsarchives-XXXIX-B5-453-2012.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J., 2016. Past, Present and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, 32(6), 1309–1332. DOI: 10.1109/TRO.2016.2624754.

Cavegn, S., 2020. Integrated georeferencing for precise depth map generation exploiting multi-camera image sequences from mobile mapping. PhD thesis, Fakultät Luft- und Raumfahrttechnik und Geodäsie, Universität Stuttgart. DOI: 10.18419/opus-11210.

Cavegn, S., Blaser, S., Nebiker, S., Haala, N., 2018. Robust and Accurate Image-Based Georeferencing Exploiting Relative Orientation Constraints. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2, 57–64. DOI: 10.5194/isprs-annals-IV-2-57-2018.

Chan, T. O., Lichti, D. D., Glennie, C. L., 2013. Multi-feature based boresight self-calibration of a terrestrial mobile mapping system. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 112–124. DOI: 10.1016/j.isprsjprs.2013.04.005.

Chen, Q., Li, L., Xu, K., An, X., Wu, Y., 2021a. A New Triple Filtering Algorithm and Its Application for Aerial GNSS/INS-Integrated Direct Georeferencing System. *Journal of Sensors*, 2021, 13. DOI: 10.1155/2021/6527356.

Chen, S., Zhou, B., Jiang, C., Xue, W., Li, Q., 2021b. A LiDAR/Visual SLAM Backend with Loop Closure Detection and Graph Optimization. *Remote Sensing*, 13(14). DOI: 10.3390/rs13142720.

Corso, N., Zakhor, A., 2013. Indoor Localization Algorithms for an Ambulatory Human Operated 3D Mobile Mapping System. *Remote Sensing*, 6611–6646. DOI: 10.3390/rs5126611.

Cramer, M., 2001. Genauigkeitsuntersuchungen zur GPS/INS-Integration in der Aerophotogrammetrie. PhD thesis, Fakultät für Bauingenieur- und Vermessungswesen der Universität Stuttgart.

Durrant-Whyte, H., Bailey, T., 2006. Simultaneous Localization and Mapping: Part I. *IEEE Robotics & Automation Magazine*, 13(2), 99–110. DOI: 10.1109/MRA.2006.1638022.

El-Sheimy, N., Youssef, A., 2020. Inertial sensors technologies for navigation applications: State of the art and future trends. *Satellite Navigation*, 1(2), 1–21. DOI: 10.1186/s43020-019-0001-5.

Elbahnasawy, M., Shamseldin, T., Habib, A., 2018. Image-assisted GNSS/INS navigation for UAV-based mobile mapping systems during GNSS outages. *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 417–425. DOI: 10.1109/PLANS.2018.8373409.

Ellum, C., El-Sheimy, N., 2000. The Development of a Backpack Mobile Mapping System. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIII-B2, 184–191. `https://www.isprs.org/proceedings/XXXIII/congress/part2/184_XXXIII-part2.pdf`, (accessed: 18/5/2022).

Ellum, C., El-Sheimy, N., 2001. A Mobile Mapping System for the Survey Community. *Proceedings of the 3rd International Symposium on Mobile Mapping Technology*.

Ellum, C., El-Sheimy, N., 2002. The Calibration of Image-Based Mobile Mapping Systems. *Proceedings of the 2nd Symposion on Geodesy for Geotechnical and Structural Engineering*, 21–24.

Eugster, H., 2012. Echtzeit-Georegistrierung von Videodaten mit Hilfe von Navigationssensoren geringer Qualität und digitalen 3D-Landschaftsmodellen. PhD thesis, Humboldt-Universität zu Berlin. DOI: 10.13140/RG.2.1.4124.0565.

Eugster, H., Huber, F., Nebiker, S., Gisi, A., 2012. Integrated Georeferencing of Stereo Image Sequences Captured with a Stereovision Mobile Mapping System – Approaches and Practical Results. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B1, 309–314. DOI: 10.5194/isprsarchives-XXXIX-B1-309-2012.

FLIR Inc., 2017. FLIR Ladybug5 USB3, Technical Reference. `https://www.flir.app.boxcn.net/s/owk6vr6x2mflmna0gto9h98uw802t7z1/file/418657863786`, (accessed: 18/5/2022).

Förstner, W., Wrobel, B. P., 2016. *Photogrammetric computer vision*. Springer. DOI: 10.1007/978-3-319-11550-4.

Fraser, C. S., 1997. Digital camera self-calibration. *ISPRS Journal of Photogrammetry and Remote Sensing*, 52(4), 149–159. DOI: 10.1016/S0924-2716(97)00005-1.

Fricker, L., Weber, D., 2019. Portables Mobile Mapping im Aussenraum. Bachelor's thesis, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland. unpublished.

Geo Week News, 2017. Vexcel UltraCam Panther: An Anticipated 3D-Scanning Backpack Arrives. https://www.geoweeknews.com/news/vexcel-ultracam-panther-long-awaited-3d-scanning-backpack-arrives, (accessed: 18/5/2022).

Geoslam, 2022. ZEB REVO RT - Scan, process and track your progress as you go. https://geoslam.com/solutions/zeb-revo-rt/, (accessed: 18/5/2022).

Glennie, C., Kusari, A., Facchin, A., 2016. Calibration and Stability Analysis of the VLP-16 Laser Scanner. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-3/W4, 55–60. DOI: 10.5194/isprs-archives-XL-3-W4-55-2016.

Grejner-Brzezinska, D. A., Toth, C. K., Jwa, Y., Moafipoor, S., Kwon, J. H., 2006. Multi-Sensor Personal Navigator: System Design and Calibration. *International Journal of Urban Sciences*, 10(2), 115–129. DOI: 10.1080/12265934.2006.9693596.

Grejner-Brzezinska, D. A., Yi, Y., Toth, C. K., 2001. Bridging GPS Gaps in Urban Canyons: The Benefits of ZUPTs. *Navigation*, 48(4), 216–226. DOI: 10.1002/j.2161-4296.2001.tb00246.x.

Gu, S., Dai, C., Fang, W., Zheng, F., Wang, Y., Zhang, Q., Lou, Y., Niu, X., 2021. Multi-GNSS PPP/INS tightly coupled integration with atmospheric augmentation and its application in urban vehicle navigation. *Journal of Geodesy*, 95(6), 1–15. DOI: 10.1007/s00190-021-01514-8.

Gunawardena, S., Morton, Y. J., 2020. *Fundamentals and Overview of GNSS Receivers*. John Wiley & Sons, Ltd, chapter 14, 307–338. DOI: 10.1002/9781119458449.ch14.

Haala, N., Peter, M., Kremer, J., Hunter, G., 2008. Mobile LiDAR Mapping for 3D Point Cloud Collection in Urban Areas – A Performance Test. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVII-B5, 1119–1124.

Hamza, V., Stopar, B., Sterle, O., 2021. Testing the Performance of Multi-Frequency Low-Cost GNSS Receivers and Antennas. *Sensors*, 21(6). DOI: 10.3390/s21062029.

Hess, W., Kohler, D., Rapp, H., Andor, D., 2016. Real-Time Loop Closure in 2D LIDAR SLAM. *Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA)*, 1271–1278. DOI: 10.1109/ICRA.2016.7487258.

Hess, W., Kohler, D., Rapp, H., Andor, D., 2021. Cartographer ROS Documentation. https://buildmedia.readthedocs.org/media/pdf/google-cartographer-ros/latest/google-cartographer-ros.pdf, (accessed: 18/5/2022).

Heuvel, F. V. D., Verwaal, R., Beers, B., 2006. Calibration of Fisheye Camera Systems and the Reduction of Chromatic Aberration. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI-5.

Hillemann, M., 2020. Selbstkalibrierung mobiler Multisensorsysteme mittels geometrischer 3D-Merkmale. PhD thesis, KIT-Fakultät für Bauingenieur-, Geo- und Umweltwissenschaften, Karlsruher Institut für Technologie (KIT). DOI: 10.5445/IR/1000125412.

Huitl, R., Schroth, G., Hilsenbeck, S., Schweiger, F., Steinbach, E., 2012. TUMindoor: An Extensive Image and Point Cloud Dataset for Visual Indoor Localization and Mapping. *Proceedings of the 19th IEEE International Conference on Image Processing*, 1773–1776. DOI: 10.1109/ICIP.2012.6467224.

Hübner, P., Weinmann, M., Wursthorn, S., Hinz, S., 2021. Automatic voxel-based 3D indoor reconstruction and room partitioning from triangle meshes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 181, 254–278. DOI: 10.1016/j.isprsjprs.2021.07.002.

Ilci, V., Toth, C., 2020. High Definition 3D Map Creation Using GNSS/IMU/LiDAR Sensor Integration to Support Autonomous Vehicle Navigation. *Sensors*, 20(3), 899. DOI: 10.3390/s20030899.

Kaarta, 2022. Introducing Stencil Pro: The Workhorse of the Kaarta Family. `https://www.kaarta.com/products/stencil-pro/`, (accessed: 18/5/2022).

Kannala, J., Brandt, S., 2006. A Generic Camera Model and Calibration Method for Conventional, Wide-Angle, and Fish-Eye Lenses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8), 1335–1340. DOI: 10.1109/TPAMI.2006.153.

Karam, S., Vosselman, G., Peter, M., Hosseinyalamdary, S., Lehtola, V., 2019. Design, Calibration, and Evaluation of a Backpack Indoor Mobile Mapping System. *Remote Sensing*, 11(8). DOI: 10.3390/rs11080905.

Kersting, A., Habib, A., Rau, J.-Y., 2012. New Method for the Calibration of Multi-Camera Mobile Mapping Systems. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B1, 121–126. DOI: 10.5194/isprsarchives-XXXIX-B1-121-2012.

Khoramshahi, E., Campos, M. B., Tommaselli, A. M. G., Vilijanen, N., Mielonen, T., Kaartinen, H., Kukko, A., Honkavaara, E., 2019. Accurate Calibration Scheme for a Multi-Camera Mobile Mapping System. *Remote Sensing*, 11(23). DOI: 10.3390/rs11232778.

Kim, H. S., Kim, Y., Kim, C., Choi, K. H., 2021. Kinematic In Situ Self-Calibration of a Backpack-Based Multi-Beam LiDAR System. *Applied Sciences*, 11(3). DOI: 10.3390/app11030945.

Kim, N., Bae, J., Kim, C., Park, S., Sohn, H.-G., 2020. Object Distance Estimation Using a Single Image Taken from a Moving Rolling Shutter Camera. *Sensors*, 20(14). DOI: 10.3390/s20143860.

Kruck, E., 1983. Lösung großer Gleichungssysteme für photogrammetrische Blockausgleichungen mit erweitertem funktionalen Modell. PhD thesis, Fachbereich Bauingenieur- und Vermessungswesen der Univeristät Hannover. `https://bingo-atm.de/wp-content/uploads/2015/03/1983_Dissertation-Kruck.pdf`, (accessed: 18/5/2022).

Kumar, A., Kumar, S., Lal, P., Saikia, P., Srivastava, P. K., Petropoulos, G. P., 2021. Chapter 1 - Introduction to GPS/GNSS technology. G. p. Petropoulos, P. K. Srivastava (eds), *GPS and GNSS Technology in Geosciences*, Elsevier, 3–20. DOI: 10.1016/B978-0-12-818617-6.00001-9.

Lauterbach, H. A., Borrmann, D., Heß, R., Eck, D., Schilling, K., Nüchter, A., 2015. Evaluation of a Backpack-Mounted 3D Mobile Scanning System. *Remote Sensing*, 7(10), 13753–13781. DOI: 10.3390/rs71013753.

Lehtola, V., Kaartinen, H., Nüchter, A., Kaijaluoto, R., Kukko, A., Litkey, P., Honkavaara, E., Rosnell, T., Vaaja, M., Virtanen, J.-P., Kurkela, M., Issaoui, A. E., Zhu, L., Jaakkola, A., Hyyppä, J., 2017. Comparison of the Selected State-Of-The-Art 3D Indoor Scanning and Point Cloud Generation Methods. *Remote Sensing*, 9(8), 796. DOI: 10.3390/rs9080796.

Leica Geosystems AG, 2022a. Leica BLK2GO Handheld Imaging Laser Scanner. `https://leica-geosystems.com/en-gb/products/laser-scanners/autonomous-reality-capture/leica-blk2go-handheld-imaging-laser-scanner`, (accessed: 18/5/2022).

Leica Geosystems AG, 2022b. Leica Pegasus Backpack Wearable Mobile Mapping Solution. `https://leica-geosystems.com/en-gb/products/mobile-mapping-systems/capture-platforms/leica-pegasus-backpack`, (accessed: 18/5/2022).

Li, Q., Chen, S., Wang, C., Li, X., Wen, C., Cheng, M., Li, J., 2019. LO-Net: Deep Real-Time Lidar Odometry. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8473–8482. DOI: 10.48550/arXiv.1904.08242.

Liu, T., Carlberg, M., Chen, G., Chen, J., Kua, J., Zakhor, A., 2010. Indoor localization and visualization using a human-operated backpack system. *2010 International Conference on Indoor Positioning and Indoor Navigation*, 1–10. DOI: 10.1109/IPIN.2010.5646820.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91–110. DOI: 10.1023/B:VISI.0000029664.99615.94.

Luber, A., 2015. Ein generisches Abbildungsmodell für Stereokamerasysteme. PhD thesis, Humbold-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II. DOI: 10.18452/17114.

Luhmann, T., Fraser, C., Maas, H.-G., 2016. Sensor modelling and camera calibration for close-range photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 37–46. DOI: 10.1016/j.isprsjprs.2015.10.006.

Luhmann, T., Robson, S., Kyle, S., Boehm, J., 2019. *Close-Range Photogrammetry and 3D Imaging*. De Gruyter. DOI: doi:10.1515/9783110607253.

Macario Barros, A., Michel, M., Moline, Y., Corre, G., Carrel, F., 2022. A Comprehensive Survey of Visual SLAM Algorithms. *Robotics*, 11(1). DOI: 10.3390/robotics11010024.

Meilland, M., Comport, A. I., Rives, P., 2015. Dense Omnidirectional RGB-D Mapping of Large-scale Outdoor Environments for Real-time Localization and Autonomous Navigation. *Journal of Field Robotics*, 32(4), 474–503. DOI: 10.1002/rob.21531.

Nam, D. V., Gon-Woo, K., 2021. Solid-State LiDAR based-SLAM: A Concise Review and Application. *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 302–305. DOI: 10.1109/BigComp51126.2021.00064.

NavVis, 2020. NavVis M6 Indoor Mobile Mapping System. `https://www.navvis.com/m6`, (accessed: 18/5/2022).

NavVis, 2022. NavVis VLX 2.0. `https://www.navvis.com/de/vlx`, (accessed: 18/5/2022).

Nebiker, S., 2017. Amtliches 3D-Bild, Studienauftrag. Technical report, University of Applied Sciences and Arts Northwestern Switzerland. DOI: 10.26041/fhnw-4110.

Nebiker, S., Cavegn, S., Loesch, B., 2015. Cloud-Based Geospatial 3D Image Spaces—A Powerful Urban Model for the Smart City. *International Journal of Geo-Information*, 4(4), 2267–2291. DOI: 10.3390/ijgi4042267.

Nebiker, S., Meyer, J., Blaser, S., Ammann, M., Rhyner, S., 2021. Outdoor Mobile Mapping and AI-Based 3D Object Detection with Low-Cost RGB-D Cameras: The Use Case of On-Street Parking Statistics. *Remote Sensing*, 13(16). DOI: 10.3390/rs13163099.

Neumann, K., Welzenbach, M., Timm, M., 2016. CMOS Imaging Sensor Technology for Aerial Mapping Cameras. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B1, 69–72. DOI: 10.5194/isprs-archives-XLI-B1-69-2016.

Nister, D., Naroditsky, O., Bergen, J., 2004. Visual odometry. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, 1. DOI: 10.1109/CVPR.2004.1315094.

Nocerino, E., Menna, F., Remondino, F., Toschi, I., Rodríguez-Gonzálvez, P., 2017. Investigation of Indoor and Outdoor Performance of Two Portable Mobile Mapping Systems. *Proceedings of SPIE, Range Imaging, and Applications XIV*, 10332. DOI: 10.1117/12.2270761.

Novak, K., 1991. The Ohio State University Highway Mapping System: The Stereo Vision System Component. *Proceedings of the 47th Annual Meeting of The Institute of Navigation*, 121–124.

NovAtel Inc., 2020. SPAN CPT7. `http://a2marinesolution.com/wp-content/uploads/2020/05/SPAN-CPT7-PS.pdf`, (accessed: 18/5/2022).

Nüchter, A., Borrmann, D., Koch, P., Kühn, M., May, S., 2015. A Man-Portable, Imu-Free Mobile Mapping System. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3/W5, 17–23. DOI: 10.5194/isprsannals-II-3-W5-17-2015.

Paparoditis, N., Papelard, J.-P., Cannelle, B., Devaux, A., Soheilian, B., David, N., Houzay, E., 2012. Stereopolis II: A Multi-Purpose and Multi-Sensor 3D Mobile Mapping System for Street Visualisation and 3D Metrology. *Revue française de photogrammétrie et télédétection*, 200(1), 69–79.

Park, C., Moghadam, P., Williams, J. L., Kim, S., Sridharan, S., Fookes, C., 2021. Elasticity Meets Continuous-Time: Map-Centric Dense 3D LiDAR SLAM. *IEEE Transactions on Robotics*, 1-20. DOI: 10.1109/TRO.2021.3096650.

Passaro, V. M. N., Cuccovillo, A., Vaiani, L., De Carlo, M., Campanella, C. E., 2017. Gyroscope Technology and Applications: A Review in the Industrial Perspective. *Sensors*, 17(10). DOI: 10.3390/s17102284.

Prime Computer AG, 2017. Prime Mini Pro Data Sheet. `https://primecomputer.ch/wp-content/uploads/2017/09/Produktdatenblatt_PrimeMiniPro_all_configurations_neu.pdf`, (accessed: 3/4/2018).

Puente, I., González-Jorge, H., Matínez-Sánchez, J., Arias, P., 2013. Review of Mobile Mapping and Surveying Technologies. *Measurement*, 46(7), 2127–2145. DOI: 10.1016/j.measurement.2013.03.006.

Purvis, M., 2020. NovAtel Span Driver, ROS Wiki. `https://wiki.ros.org/novatel_span_driver`, (accessed: 18/5/2022).

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., E., B., Wheeler, R., Ng, A., 2009. ROS: An Open-Source Robot Operating System. *Proceedings of the ICRA Workshop on Open Source Software*, 3, 5. `https://www.willowgarage.com/sites/default/files/icraoss09-ROS.pdf`.

Rao, B. R., McDonald, K., Kunysz, W., 2013. *GPS/GNSS Antennas*. Artech House.

Rau, J., Su, B., Hsiao, K., Jhan, J., 2016. Systematic Calibration for a Backpacked Spherical Photogrammetry Imaging System. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B1, 695–702. DOI: 10.5194/isprs-archives-XLI-B1-695-2016.

Rau, J.-Y., Habib, A. F., Kersting, A. P., Chiang, K.-W., Bang, K.-I., Tseng, Y.-H., Li, Y.-H., 2011. Direct Sensor Orientation of a Land-Based Mobile Mapping System. *Sensors*, 11(7), 7243–7261. DOI: 10.3390/s110707243.

Ravi, R., Lin, Y.-J., Elbahnasawy, M., Shamseldin, T., Habib, A., 2018. Simultaneous System Calibration of a Multi-LiDAR Multicamera Mobile Mapping Platform. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5), 1694–1714. DOI: 10.1109/JSTARS.2018.2812796.

Rechsteiner, F., Wisler, D., 2018. Erstellung eines digitalen Zwillings des Versuchsstollens Hagerbach mittels des BIMAGE Backpacks und anderer 3D Sensoren und Plattformen. Bachelor's thesis, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland. unpublished.

Sarlin, P.-E., Cadena, C., Siegwart, R., Dymczyk, M., 2019. From coarse to fine: Robust hierarchical localization at large scale. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12716–12725. DOI: 10.1109/CVPR.2019.01300.

Scaramuzza, D., Harati, A., Siegwart, R., 2007. Extrinsic Self Calibration of a Camera and a 3D Laser Range Finder From Natural Scenes. *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 4164–4169. DOI: 10.1109/IROS.2007.4399276.

Schönberger, J., Frahm, J.-M., 2016. Structure-from-Motion Revisited. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4104–4113. DOI: 10.1109/CVPR.2016.445.

Schwarz, K., Martell, H., El-Sheimy, N., Li, R., Chapman, M., Cosandier, D., 1993. VIASAT - A Mobile Highway Survey System of High Accuracy. *Proceedings of the Vehicle Navigation and Information Systems Conference, Ottawa*, 476–481. DOI: 10.1109/VNIS.1993.585676.

Shan, J., Toth, C. K., 2018. *Topographic laser ranging and scanning: principles and processing.* CRC press.

Stachniss, C., Leonard, J., Thrun, S., 2016. *Springer Handbook of Robotics.* Springer, Cham, chapter Simultaneous Localization and Mapping, 1153–1176. DOI: 10.1007/978-3-319-32552-1_46.

Tang, J., Chen, Y., Kukko, A., Kaartinen, H., Jaakola, A., Khoramshahi, E., Hakala, T., Hyyppä, J., Holopainen, M., Hyyppä, H., 2015. SLAM-Aided Stem Mapping for Forest Inventory with Small-Footprint Mobile LiDAR. *Forests*, 6(12), 4588–4606. DOI: 10.3390/f6124390.

Thrun, S., 2002. Robotic Mapping: A Survey. *Exploring Artificial Intelligence in the New Millenium*, 1(1), 1–35.

Velodyne, 2016. User's Manual and Programming Guide, VLP-16 Velodyne LiDAR Puck. `http://velodynelidar.com/products/puck/`, (accessed: 18/5/2022).

Verbree, E., Zlatanova, S., Smit, K., 2004. Interactive Navigation Services through Value-Added CycloMedia Panoramic Images. *Proceedings of the 6th International Conference on Electronic Commerce*, ICEC '04, 591–595. DOI: 10.1145/1052220.1052295.

Viametris, 2022a. bMS3D LD5+ Backpack Mobile Scanner. https://viametris.com/backpack-mobile-scanners/, (accessed: 18/5/2022).

Viametris, 2022b. IMS3D Trolley mobile scanner. https://viametris.com/trolley-mobile-scanners/, (accessed: 18/5/2022).

Wanninger, L., Thiemig, M., Frevert, V., 2022. Multi-frequency quadrifilar helix antennas for cm-accurate GNSS positioning. *Journal of Applied Geodesy*, 16(1), 25–35. DOI: 10.1515/jag-2021-0042.

Wen, C., Pan, S., Wang, C., Li, J., 2016. An Indoor Backpack System for 2-D and 3-D Mapping of Building Interiors. *IEEE Geoscience and Remote Sensing Letters*, 13(7), 992–996. DOI: 10.1109/LGRS.2016.2558486.

Withley, J., 2016. Velodyne Driver. ROS Wiki. http://wiki.ros.org/velodyne_driver, (accessed: 18/5/2022).

Wittmer, S., 2017. INS- und bildbasierte Messmethoden für bewaldete Gebiete. Master's thesis, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland. unpublished.

Zhang, J., Grabe, V., Hamner, B., Duggins, D., Singh, S., 2016. Compact, Real-time Localization without Reliance on Infrastructure. *Proceedings of the 3rd Annual Microsoft Indoor Localization Competition*. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/11/Kaarta_RealEarth.pdf, (accessed: 18/5/2022).

Zhang, J., Singh, S., 2017. Low-Drift and Real-Time Lidar Odometry and Mapping. *Autonomous Robots*, 41, 401–416. DOI: 10.1007/s10514-016-9548-2.

Zhang, J., Singh, S., 2018. Laser-Visual-Inertial Odometry and Mapping with High Robustness and Low Drift. *Journal of Field Robotics*, 35(8), 1242–1264. DOI: 10.1002/rob.21809.

Zheng, X., Zhu, J., 2021. Efficient LiDAR Odometry for Autonomous Driving. *IEEE Robotics and Automation Letters*, 6(4), 8458–8465. DOI: 10.1109/LRA.2021.3110372.

Zhou, Y., Daakir, M., Rupnik, E., Pierrot-Deseilligny, M., 2020. A two-step approach for the correction of rolling shutter distortion in UAV photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160, 51–66. DOI: 10.1016/j.isprsjprs.2019.11.020.

Acknowledgements

First of all I express my gratitude to my thesis supervisors Prof. Dr.-Ing. Boris Jutzi and Prof. Dr. Stephan Nebiker for continuous guidance, helpful advices, encouragement and trust. Moreover, I would like to thank Prof. Dr. Corinna Harmening, Prof. Dr.-Ing. Stefan Hinz and Prof. Dr.-Ing. Hansjörg Kutterer for being co-referee.

I would like to thank the University of Applied Sciences and Arts Northwestern Switzerland for providing excellent employment conditions during the preparation of this thesis. Among other thing, this was made possible thanks to the following externally funded projects:

- PReCIS – Portables bildbasiertes Reality Capturing für Infrastruktur- und Smartcity-Management (20200331_04_PReCIS) was supported by Forschungsfonds Kt. Aargau and by our industial partner iNovitas AG.
- cloudIO – Cloud-based Image Orientation for Infrastructure Management (32411.1 IP-ICT) was supported by Innosuisse, the Swiss Innovation Agency and by our industrial partner iNovitas AG.
- BIMAGE – Building Information Management based on Geospatial 3D Imagery (18493.2 PFES-ES) was supported by Innosuisse, the Swiss Innovation Agency and by our industrial partner iNovitas AG.
- infraVIS – Sustainable Infrastructure Management based on Versatile Intelligent 3D Image Spaces (15878.1 PFES-ES) was supported by Innosuisse, the Swiss Innovation Agency and by our industrial partner iNovitas AG.

I am grateful to all my colleagues at the Institute of Geomatics for the friendly and inspiring atmosphere, for interesting discussions, support and the excellent collaboration. Furthermore, I would like to thank iNovitas AG for ongoing interest in our research, for exciting and helpful inputs and discussions, for the use of the indoor calibration field and for the challenging and insightful practical campaigns that we were able to carry out with our backpack MMS.

Last but not least, I would like to thank Manuela for reading through this work and for being understanding during the intense final phase.

Appendices

Appendix A

Coordinate Frames and Transformations

A.1 Coordinate Frames

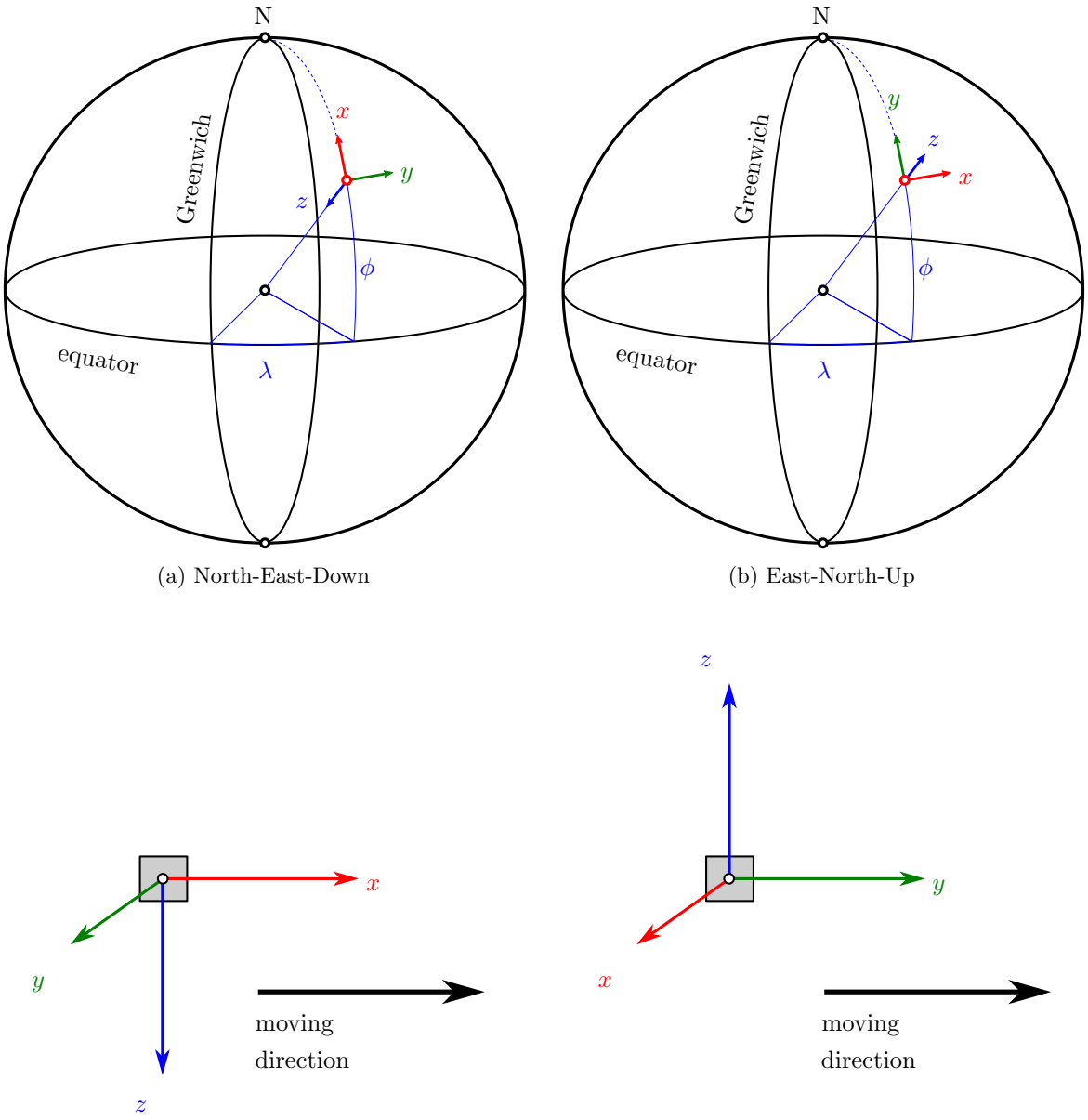
Cramer (2001) provides a comprehensive overview of all coordinate frames involved in platform navigation. This section briefly depicts the different coordinate frame definitions used for both the navigation frame and the body frame. All coordinate frames shown are right-handed systems.

The following coordinate system definitions are commonly used for navigation applications:

- North-East-Down for the navigation frame. Thereby the x-axis points along the meridian to the geodetic north, the y-axis points to the east and the z-axis points downwards along the ellipsoid normal (see Figure A.1a).
- Forward-Right-Down for the body frame. Thereby the x-axis points forward along the moving direction, the y-axis points to the right and the z-axis points downwards (see Figure A.1c).

The following coordinate system definitions are commonly used for photogrammetric and MM applications:

- East-North-Up for the mapping frame. Thereby the x-axis points to the east, the y-axis points along the meridian to the geodetic north and the z-axis points upwards along the ellipsoid normal (see Figure A.1b).
- Right-Forward-Up for the body frame. Thereby the x-axis points to the right, the y-axis points forward along the moving direction and the z-axis points upwards (see Figure A.1d).



(a) North-East-Down

(b) East-North-Up

(c) Forward-Right-Down

(d) Right-Forward-Up

Figure A.1: Different coordinate frame definitions from the navigation frame a and b and from the body frame c and d. The longitude is represented as λ while the latitude is declared as ϕ.

A.2 Pose and 3D Rigid Body Transformation

This section includes the definition of a pose as well as its 3D rigid body transformation, using homogeneous coordinates. A comprehensive theoretical treatment of the transformation with homogeneous coordinates can be found in Förstner, Wrobel (2016). Here we use the mathematical notation proposed in Förstner, Wrobel (2016).

By general, a pose $\mathbf{H}$ mathematically equals with a 3D rigid body transformation. Here, it describes a transformation from coordinate frame a to coordinate frame b:

$$ {}^{b}\mathbf{H}_{a} = \begin{bmatrix} {}^{b}\mathbf{R}_{a} & {}^{b}\mathbf{t}_{a} \\ 0^{T} & 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_{x} \\ r_{21} & r_{22} & r_{23} & t_{y} \\ r_{31} & r_{32} & r_{33} & t_{z} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{A.1} $$

Thereby, we express the 3D rigid body transformation $\mathbf{H}$ with a 4×4 matrix incorporating a 3×3 rotation matrix $\mathbf{R}$ as well as a 3×1 translation vector $\mathbf{t}$. This enables to transform homogeneous point coordinates $\mathbf{p}$ from coordinate frame a to coordinate frame b as following:

$$ {}^{b}\mathbf{p}_{n} = {}^{b}\mathbf{H}_{a}\,{}^{a}\mathbf{p}_{n} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_{x} \\ r_{21} & r_{22} & r_{23} & t_{y} \\ r_{31} & r_{32} & r_{33} & t_{z} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_{x} \\ p_{y} \\ p_{z} \\ p_{w} \end{bmatrix} \tag{A.2} $$

while ${}^{a}\mathbf{p}_{n}$ denotes that the point with point number n refers to the coordinate frame a. The transformation from homogeneous coordinates to cartesian coordinates,

$$ \mathbf{p}_{c} = f(\mathbf{p}_{h}) = \frac{\mathbf{p}_{h}}{p_{w}} = \frac{\begin{bmatrix} p_{x} \\ p_{y} \\ p_{z} \end{bmatrix}}{p_{w}} \tag{A.3} $$

is straightforward as well as the transformation from cartesian to homogeneous coordinates,

$$ \mathbf{p}_{h} = f(\mathbf{p}_{c}) = \begin{bmatrix} \mathbf{p}_{c} \\ 1 \end{bmatrix} = \begin{bmatrix} p_{x} \\ p_{y} \\ p_{z} \\ 1 \end{bmatrix} \tag{A.4} $$

while $\mathbf{p}_{h}$ denotes homogeneous coordinates and $\mathbf{p}_{c}$ shows cartesian coordinates.

Furthermore, 3D rigid body transformations can be concatenated as following:

$$ {}^{c}\mathbf{H}_{a} = {}^{c}\mathbf{H}_{b}\,{}^{b}\mathbf{H}_{a} \tag{A.5} $$

Observe, how the coordinate frame b in the center of the second term diagonally omits.

Finally, 3D rigid body transformations can be inverted by an inversion of the respective 4×4 matrix:

$$ {}^{b}\mathbf{H}_{a} = ({}^{a}\mathbf{H}_{b})^{-1} \tag{A.6} $$

A.3 Rotation Conventions

We use the photogrammetric rotation convention with ω, φ and κ that uses rotated axes (Luhmann et al., 2019), while ω represents a rotation around the x-axis, φ around the y-axis and κ around the z-axis. The rotation matrix can be calculated as following:

$$\mathbf{R} = \mathbf{R}_\omega \mathbf{R}_\varphi \mathbf{R}_\kappa = \begin{bmatrix} \cos\varphi\cos\kappa & -\cos\varphi\sin\kappa & \sin\varphi \\ \cos\omega\sin\kappa + \sin\omega\sin\varphi\cos\kappa & \cos\omega\cos\kappa - \sin\omega\sin\varphi\sin\kappa & -\sin\omega\cos\varphi \\ \sin\omega\sin\kappa - \cos\omega\sin\varphi\cos\kappa & \sin\omega\cos\kappa + \cos\omega\sin\varphi\sin\kappa & \cos\omega\cos\varphi \end{bmatrix} \tag{A.7}$$

Inversely, the angles ω, φ and κ can also be determined from the rotation matrix:

$$\begin{aligned} \varphi &= \arctan\left(\frac{r_{13}}{\sqrt{r_{23}^2 + r_{33}^2}}\right) \\ \omega &= \begin{cases} \arctan\left(-\frac{r_{23}}{r_{33}}\right), & \text{if } \cos\varphi \neq 0 \\ 0, & \text{otherwise} \end{cases} \\ \kappa &= \begin{cases} \arctan\left(\frac{-r_{12}}{r_{11}}\right), & \text{if } \cos\varphi \neq 0 \\ \arctan\left(\frac{r_{22}}{r_{32}}\right), & \text{else if } \varphi = \frac{\pi}{2} \\ \arctan\left(-\frac{r_{22}}{r_{32}}\right), & \text{otherwise} \end{cases} \end{aligned} \tag{A.8}$$

A.4 Transformation of Pose Orientations into the Swiss Reference Frame LV95

This section includes the transformation of resulting pose orientations from direct georeferencing into the projected Swiss reference frame *LV95*. Basically, ${}^{n1_{ned}}\mathbf{R}_{b_{frd}}$ represents the pose orientation resulting from direct georeferencing. It rotates from the body frame b following the forward-right-down (frd) coordinate system definition b_{frd} (see Appendix A.1) to the global navigation frame *WGS84* ($n1$) following the north-east-down (ned) coordinate frame definition $n1_{ned}$. However, for our investigation it is required to transform them to the rotation representation ${}^{p2_{enu}}\mathbf{R}_{b_{rfu}}$ which rotates from the body frame b following the right-forward-up (rfu) coordinate frame definition b_{rfu} to the local projected reference frame *LV95* ($p2$) following the east-north-up (enu) coordinate frame definition $p2_{enu}$.

The following formula represents the entire transformation:

$${}^{p2_{enu}}\mathbf{R}_{b_{rfu}} = {}^{p2_{enu}}\mathbf{R}_{n2_{enu}}\,{}^{n2_{enu}}\mathbf{R}_{n2_{ned}}\,{}^{n2_{ned}}\mathbf{R}_{e2}\,{}^{e2}\mathbf{R}_{e1}\,{}^{e1}\mathbf{R}_{n1_{ned}}\,{}^{n1_{ned}}\mathbf{R}_{b_{frd}}\,{}^{b_{frd}}\mathbf{R}_{b_{rfu}}, \tag{A.9}$$

whereby the rotations from right-forward-up to forward-right-down (${}^{b_{frd}}\mathbf{R}_{b_{rfu}}$) and from east-north-down to east-north-up (${}^{n2_{enu}}\mathbf{R}_{n2_{ned}}$) are the same. Furthermore, ${}^{e1}\mathbf{R}_{n1_{ned}}$ represents the rotation from the navigation frame to the *earth-centered earth-fixed (ECEF)* reference frame, while ${}^{n2_{ned}}\mathbf{R}_{e2}$ represents the inverse rotation. ${}^{e2}\mathbf{R}_{e1}$ represents the rotation of the datum transformation from the global reference ellipsoid *WGS84* ($e1$) to the local reference ellipsoid *CH1903+* ($e2$). Finally, ${}^{p2_{enu}}\mathbf{R}_{n2_{enu}}$ incorporates the rotation from the navigation frame $n2$ to the swiss map projection *Swiss Grid* ($p2$), which basically includes the meridian convergence.

The previously mentioned rotations are represented as following:

- Rotations from the right-forward-up (rfu) to the forward-right-down (frd) coordinate system definition and from the east-north-down (end) to the east-north-up (enu) coordinate system definition respectively and vice versa:

$$ {}^{b_{frd}}\mathbf{R}_{b_{rfu}} = {}^{n2_{enu}}\mathbf{R}_{n2_{ned}} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{A.10} $$

- Rotation from the navigation frame n to the *ECEF* reference frame e:

$$ {}^{e1}\mathbf{R}_{n1_{ned}} = {}^{n2_{ned}}\mathbf{R}_{e2}^{T} = \begin{bmatrix} -\sin\phi\cos\lambda & -\sin\lambda & -\cos\phi\cos\lambda \\ -\sin\phi\sin\lambda & \cos\lambda & -\cos\phi\sin\lambda \\ \cos\phi & 0 & -\sin\phi \end{bmatrix}, \tag{A.11} $$

  whereby ϕ represents the latitude and λ represents the longitude.

- Rotation from the navigation frame n to the Swiss map projection *Swiss Grid* (p):

$$ {}^{p2_{enu}}\mathbf{R}_{n2_{enu}} = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{A.12} $$

  whereby γ represents the meridian convergence.

- Rotation from the global reference frame *WGS84* to the local reference frame *CH1903+*:

$$ {}^{e2}\mathbf{R}_{e1} = \mathbf{I}. \tag{A.13} $$

Since the alignments of the global reference ellipsoid *WGS84* ($e1$) and the local reference ellipsoid *CH1903+* ($e2$) are identical, the formula A.9 is simplified as following:

$$ {}^{p_{enu}}\mathbf{R}_{b_{rfu}} = {}^{p_{enu}}\mathbf{R}_{n_{enu}} {}^{n_{enu}}\mathbf{R}_{n_{ned}} {}^{n_{ned}}\mathbf{R}_{b_{frd}} {}^{b_{frd}}\mathbf{R}_{b_{rfu}} \tag{A.14} $$

A.5 Transformations for Creating Attitude Updates

This section contains transformations of the rotational component necessary to create ATTUPTs for direct georeferencing. It covers the case that the pose refers to the Swiss reference frame *LV95*. For instance, this is given when poses from image-based georeferencing are used. It is assumed that the poses were already transformed from the camera projection center to the body frame, using the calibrated RO and BA parameters.

Basically, ${}^{p2_{enu}}\mathbf{R}_{b_{rfu}}$ represents the pose orientation from image-based georeferencing, whereby it rotates from the body frame (b) following the right-forward-up (rfu) coordinate frame definition to the Swiss reference frame *LV95* ($p2$) following the east-north-up (enu) coordinate frame definition.

However, ATTUPTs require the rotation ${}^{n1_{ned}}\mathbf{R}_{b_{frd}}$ which rotates from the body frame (b) following the forward-right-down (frd) coordinate frame definition to the global navigation frame *WGS84* ($n1$) following the north-east-down (ned) coordinate frame definition.

We rearrange the formula A.9 and obtain

$$ {}^{n1_{ned}}\mathbf{R}_{b_{frd}} = {}^{n1_{ned}}\mathbf{R}_{e1}\,{}^{e1}\mathbf{R}_{e2}\,{}^{e2}\mathbf{R}_{n2_{ned}}\,{}^{n2_{ned}}\mathbf{R}_{n2_{enu}}\,{}^{n2_{enu}}\mathbf{R}_{p2_{enu}}\,{}^{p2_{enu}}\mathbf{R}_{b_{rfu}}\,{}^{b_{rfu}}\mathbf{R}_{b_{frd}}, \tag{A.15} $$

whereby the rotations from forward-right-down to right-forward-up (${}^{b_{rfu}}\mathbf{R}_{b_{frd}}$) and from east-north-up to east-north-down (${}^{n2_{ned}}\mathbf{R}_{n2_{enu}}$) are the same. Furthermore, ${}^{n1_{ned}}\mathbf{R}_{e1}$ represents the rotation from the earth-centered earth-fixed (ECEF) reference frame to the navigation frame, while ${}^{e2}\mathbf{R}_{n2_{ned}}$ represents the inverse rotation. ${}^{e1}\mathbf{R}_{e2}$ represents the rotation of the datum transformation from the local reference ellipsoid *CH1903+* ($e2$) to the global reference ellipsoid *WGS84* ($e1$). Finally, ${}^{n2_{enu}}\mathbf{R}_{p2_{enu}}$ incorporates the rotation from the swiss map projection *Swiss Grid* ($p2$) to the navigation frame ($n2$), which basically includes the meridian convergence.

The previously mentioned rotations are represented as following:

- Rotations from the forward-right-down (frd) to the right-forward-up (rfu) coordinate system definition and from the east-north-up (enu) to the east-north-down (end) coordinate system definition, respectively:

$$ {}^{b_{rfu}}\mathbf{R}_{b_{frd}} = {}^{n2_{ned}}\mathbf{R}_{n2_{enu}} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{A.16} $$

- Rotation from the *ECEF* reference frame (e) to the navigation frame (n):

$$ {}^{n_{ned}}\mathbf{R}_{e} = {}^{e}\mathbf{R}^{T}_{n_{ned}} = \begin{bmatrix} -\sin\phi\cos\lambda & -\sin\phi\sin\lambda & \cos\phi \\ -\sin\lambda & \cos\lambda & 0 \\ -\cos\phi\cos\lambda & -\cos\phi\sin\lambda & -\sin\phi \end{bmatrix}, \tag{A.17} $$

  whereby ϕ represents the latitude and λ represents the longitude.

- Rotation from the Swiss map projection *Swiss Grid* (p) to the navigation frame (n):

$$ {}^{n_{enu}}\mathbf{R}_{p_{enu}} = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{A.18} $$

  whereby γ represents the meridian convergence.

- Rotation from the local reference frame *CH1903+* to the global reference frame *WGS84*:

$$ {}^{e1}\mathbf{R}_{e2} = \mathbf{I}. \tag{A.19} $$

Since the alignments of the global reference ellipsoid *WGS84* ($e1$) and the local reference ellipsoid *CH1903+* ($e2$) are identical, the formula A.15 is simplified as following:

$$ {}^{n_{ned}}\mathbf{R}_{b_{frd}} = {}^{n_{ned}}\mathbf{R}_{n_{enu}}\,{}^{n_{enu}}\mathbf{R}_{p_{enu}}\,{}^{p_{enu}}\mathbf{R}_{b_{rfu}}\,{}^{b_{rfu}}\mathbf{R}_{b_{frd}} \tag{A.20} $$

Appendix B

Camera Models

B.1 Opening Angles of the Equidistant Camera Model

For the equidistant camera model, the radial distance from the principal point corresponds to the angle between the optical ray from the camera to the point and the optical axis. Thus, the diagonal viewing angle α_d can be calculated using the sensor dimensions u, v and the focal length c as following:

$$\alpha_d = \frac{\sqrt{u^2 + v^2}}{c} \tag{B.1}$$

However, the rectangular shaped sensor causes variable horizontal and vertical viewing angles which depend on the y image coordinate component and the x image coordinate component (see Figure 3.1c). The following function shows the relationship between the horizontal viewing angle α_h and the image coordinate component y, using the horizontal sensor dimension u and the focal length c:

$$\alpha_h(y) = 2 \arctan \left(\frac{u \tan \left(\sqrt{\frac{u^2}{4c^2} + \frac{y^2}{c^2}} \right)}{2c \sqrt{\frac{u^2}{4c^2} + \frac{y^2}{c^2}}} \right) \tag{B.2}$$

The following function shows the relationship between the vertical viewing angle α_v and the image coordinate component x, using the vertical sensor dimension v and the focal length c:

$$\alpha_v(x) = 2 \arctan \left(\frac{v \tan \left(\sqrt{\frac{v^2}{4c^2} + \frac{x^2}{c^2}} \right)}{2c \sqrt{\frac{v^2}{4c^2} + \frac{x^2}{c^2}}} \right) \tag{B.3}$$

In order to calculate the minimal image overlap, we specify the minimal horizontal viewing angle as well as the minimal vertical viewing angle. They extend along the axes of the image coordinate system with the origin in the principal point (see Figure 3.1a). As we put zero for x and y into $\alpha_h(x)$ and $\alpha_v(y)$, the equations can be simplified as following:

$$\alpha_{h_{min}} = \frac{u}{2c} \tag{B.4}$$

$$\alpha_{v_{min}} = \frac{v}{2c} \tag{B.5}$$

B.2 Collinearity Equations

The following collinearity equations are taken from Luhmann et al. (2019) and were slightly adopted to the formal logic of this thesis:

$$x' = h_x + z'\frac{k_x}{k_z} + \Delta x' = h_x + z'\frac{r_{11}(p_x - t_x) + r_{21}(p_y - t_y) + r_{31}(p_z - t_z)}{r_{13}(p_x - t_x) + r_{23}(p_y - t_y) + r_{33}(p_z - t_z)} + \Delta x', \tag{B.6}$$

$$y' = h_y + z'\frac{k_y}{k_z} + \Delta y' = h_x + z'\frac{r_{12}(p_x - t_x) + r_{22}(p_y - t_y) + r_{32}(p_z - t_z)}{r_{13}(p_x - t_x) + r_{23}(p_y - t_y) + r_{33}(p_z - t_z)} + \Delta y', \tag{B.7}$$

where,

x', y' = Image coordinates of the object point $\mathbf{p}$
h_x, h_y = Image coordinates of the principal point
z' = Principal distance ($z' = -c$)
c = Calibrated focal length
p_x, p_y, p_z = Coordinates of the object point $\mathbf{p}$
t_x, t_y, t_z = Translational components of the image pose ${}^{w}\mathbf{H}_{cam}$
$r_{11}, ..., r_{33}$ = Rotational components of the image pose ${}^{w}\mathbf{H}_{cam}$
$\Delta x', \Delta y'$ = Distortions of the image