DEVELOPMENT OF A PORTABLE HIGH PERFORMANCE MOBILE MAPPING SYSTEM USING THE ROBOT OPERATING SYSTEM

S. Blaser [1], S. Cavegn [1,2], S. Nebiker [1]

[1] Institute of Geomatics, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland
- (stefan.blaser, stefan.cavegn, stephan.nebiker)@fhnw.ch
[2] Institute for Photogrammetry, University of Stuttgart, Germany

Commission I, WG I/7

KEY WORDS: Mobile Mapping, Indoor, Robot Operating System, ROS, SLAM, Sensor Orientation, Infrastructure

ABSTRACT:

The rapid progression in digitalization in the construction industry and in facility management creates an enormous demand for the efficient and accurate reality capturing of indoor spaces. Cloud-based services based on georeferenced metric 3D imagery are already extensively used for infrastructure management in outdoor environments. The goal of our research is to enable such services for indoor applications as well. For this purpose, we designed a portable mobile mapping research platform with a strong focus on acquiring accurate 3D imagery. Our system consists of a multi-head panorama camera in combination with two multi-profile LiDAR scanners and a MEMS-based industrial grade IMU for LiDAR-based online and offline SLAM. Our modular implementation based on the Robot Operating System enables rapid adaptations of the sensor configuration and the acquisition software. The developed workflow provides for completely GNSS-independent data acquisition and camera pose estimation using LiDAR-based SLAM. Furthermore, we apply a novel image-based georeferencing approach for further improving camera poses. First performance evaluations show an improvement from LiDAR-based SLAM to image-based georeferencing by an order of magnitude: from 10-13 cm to 1.3-1.8 cm in absolute 3D point accuracy and from 8-12 cm to sub-centimeter in relative 3D point accuracy.

1. INTRODUCTION

Digitalization trends on a broad scale lead to a rapid and massive transformation of the construction and real estate industries. New methods and tools such as BIM (Building Information Modeling) and VDC (Virtual Design and Construction) in combination with new technologies for geospatial data capture and exploitation enable a paradigm shift in the way buildings are designed, tested, built, maintained and refurbished. The establishment of fully three-dimensional collaborative processes and workflows with stakeholders from multiple domains require accurate, detailed and up-to-date 3D geodata. New image-based mobile reality capturing techniques in combination with cloud technologies, such as presented by Nebiker et al. (2015), hold the potential to provide such data and services in a rapid, cost-efficient and user-friendly manner. As shown by Puente et al. (2013) current commercial mobile mapping systems (MMS) are dominated by LiDAR as primary sensors and with cameras as complimentary sensors. However, first road- and railway-based mobile mapping experiments were based on stereo camera systems and date back to the early 1990ies (Novak, 1991; Schwarz et al., 1993). Since then image-based outdoor mobile mapping systems have evolved into multi-stereo systems (Cavegn & Haala, 2016) and into high-performance 360° stereo systems (Blaser et al., 2017; Meilland et al., 2015) with an unparalleled information richness and density.

In terms of positioning technologies, the vast majority of outdoor mobile mapping solutions currently rely on direct georeferencing using GNSS and INS. However, alternative approaches are required for positioning and pose estimation in indoor environments and in environments with poor GNSS coverage such as forests or urban canyons. Recent developments in indoor localization and mapping benefit from and build on methods and

techniques, namely SLAM (simultaneous localization and mapping), from the robotics and more recently from the computer vision communities (Stachniss et al., 2016; Thrun, 2002). Indoor mobile mapping not only requires new georeferencing strategies, but also new platforms. Solutions using carts, like the Viametris iMMS (Thomson et al., 2013), work well in large unobstructed spaces but are not suitable for typical indoor environments with stairs, closed doors, obstructed floors due to ongoing construction etc. Thus, the focus in research and commercial development has shifted towards portable or 'personal' mobile mapping systems (Lehtola et al., 2017; Nüchter et al., 2015). But with very few exceptions, such as the image-based UltraCam Panther (Vexcel, 2018), most developments focus on LiDAR-based systems and point clouds.

In this paper, we introduce the BIMAGE backpack, a portable high-performance mobile mapping research platform and system built on top of the Robot Operating System (ROS). The system is designed for the creation of 3D image data spaces and for the image-based building and infrastructure management (Nebiker et al., 2015). It features LiDAR SLAM-based real-time 3D mapping with a subsequent novel image-based georeferencing approach using relative orientation constraints leading to significant improvements in relative and absolute accuracies.

First, we discuss the modular mechanical, electronic and software design of the BIMAGE backpack, allowing the integration of multiple medium- to low-cost LiDAR sensors and panoramic camera sensors. In particular, we introduce a low-cost architecture for the mission-critical accurate triggering and synchronization of all types of sensors involved. We also provide details on the calibration of the multi-sensor system. The acquisition system provides a number of novel features, such as real-time 3D positioning and mapping capabilities and geometry-driven camera triggering for optimal data coverage. Our highly

automated processing workflow offers both LiDAR SLAM- and image-based georeferencing. In a performance evaluation, we demonstrate that our novel image-based georeferencing approach offers an improvement in relative and absolute accuracies in the order of magnitude under real-world conditions.

2. SYSTEM REQUIREMENTS

In order to create image-based services for indoor infrastructure management with accurate 3D measurement functionality, we need an image-based indoor mobile mapping platform for efficient data acquisition.

As the sensor technology improves rapidly, a prototypic indoor mobile mapping system for research purposes should provide the flexibility to easily integrate new components. Furthermore, it should also be possible to implement and investigate different sensor configurations. In order to achieve this flexibility, a modular design is indispensable. This affects both hardware and software. A trolley system like NavVis (Szwarc, 2017) would offer more flexibility and fewer limitations in terms of weight and size than a portable system. However, the applications of a trolley system are limited. For example, a trolley system cannot be used for capturing staircases or construction sites. For our research, it is important that the system is able to cover as many use cases as possible.

As in our earlier mobile mapping solution (Burkhard et al., 2012), it should be possible to derive 3D imagery through dense image matching, ideally assigning a depth value to each image pixel. These 3D images are subsequently aggregated into 3D image data spaces, which can easily be navigated and used for simple 3D measurements in single images (Nebiker et al., 2015). Ideally, a camera configuration should cover the indoor environment with multiple views and as completely as possible. For measuring purposes, both precise camera calibration and precise image orientation are essential. For indoor use, the camera has to perform well in low light condition and should be able to adapt rapidly to changing lighting conditions. Since the platform is mobile, the cameras must be triggered synchronously, ideally with an electronic trigger signal.

Precise georeferencing is essential in mobile mapping. Inertial navigation systems (INS) combining GNSS and IMU are widely used for direct georeferencing in outdoor mobile mapping applications. The absence of GNSS in indoor environments precludes this sensor combination from being applied to indoor mobile mapping. A possible way to overcome the lack of GNSS is to replace INS with LiDAR-based simultaneous localization and mapping (SLAM) (Hess et al., 2016).

An approach to increase the accuracy of direct georeferencing is to apply an additional image-based georeferencing. With this approach Cavegn et al. (2016) achieved an accuracy improvement by an order of magnitude in urban street environments. However, the success of this approach depends on the number and distribution of detected and matched features in the outdoor environment. We assume that image-based georeferencing will provide a significant accuracy improvement in structurally rich indoor environments.

In contrast to most commercial indoor mobile mapping systems (Leica Geosystems, 2018; Vexcel, 2018), our system should support the complete system initialization in an indoor environment. For reasons of efficiency, this is essential for the mapping of large building complexes. In addition, a status indicator should provide real-time information about completeness and accuracy of data acquisition.

3. SYSTEM COMPONENTS

Our portable mobile mapping system consists of different components for navigation, mapping, data registration and data pre-processing. A panorama camera captures images in order to map the entire indoor environment. Multi-profile laser scanner data and IMU data fused in a LiDAR SLAM algorithm deliver the system position and attitude. A computer with high RAM and storage capacities and low energy consumption processes the SLAM algorithm and registers the captured data.

3.1 Panorama Camera

For image-based indoor mapping, the system uses a FLIR Ladybug5 panorama camera. The Ladybug5 panorama camera consists of six different single camera heads. Five camera heads are pointing sideways, one camera head is pointing upwards. Each camera head has a resolution of 5MP and features ultra-wide-angle optics (see Table 1). Previous investigations had shown that the equidistant camera model (fisheye) optimally fits the Ladybug5 camera heads (Blaser et al., 2017). A general-purpose input and output interface (GPIO) features several triggering modes with an electronic signal (FLIR Inc., 2017).

Sensor Type	Sony ICX655 CCD
Shutter Type	Global shutter
Resolution [px]	2448 x 2048
Pixel size [μm]	3.45
Focal length [mm]	4.3
Field-of-view [deg]	113 x 94

Table 1. Specifications of the panorama camera. Technical data of an individual Ladybug5 camera head (FLIR Inc., 2017).

3.2 Laser Scanner

The system uses two multi-profile Velodyne VLP-16 laser scanners for navigation as well as for mapping. Velodyne designed the VLP-16 as a rugged medium- to low-cost laser scanner primarily for collision avoidance applications in the automotive industry. The multi-profile laser scanner consists of 16 different laser modules mounted in a compact housing. The housing spins with an adjustable rate in-between 5 and 20 rotations per second (Velodyne, 2016). Therefore, the laser scanner delivers a horizontal full field-of-view (see Table 2). The horizontal resolution is limited by the speed of the laser modules and the rotation rate. In contrast, the 16 fixed laser modules limit the vertical field-of-view as well as the vertical resolution. Glennie et al. (2016) provide a calibration and a stability analysis of the Velodyne VLP-16 laser scanner. Their evaluations reveal a long-term walk in individual lasers. Thus, a geometric calibration does not remain temporally stable. The VLP-16 supports time synchronization with a pulse-per-second (PPS) signal in conjunction with a one-per-second NMEA sentence (Velodyne, 2016).

Max. range [m]	100
Typical accuracy [cm]	3.0
Number of channels	16
Angular resolution (H x V) [deg]	0.1-0.4 x 2.0
Field-of-view (H x V) [deg]	360 x 30
Rotation rates [Hz]	5-20
Max. points per second	300'000

Table 2. Specifications of the Velodyne VLP-16 Puck LiDAR scanner (Velodyne, 2016)

3.3 Inertial Measurement Unit

The system contains an XSens MTI-300 IMU for navigation. The MTI-300 is a MEMS-based attitude heading reference system (AHRS), which returns drift-free roll and pitch as well as magnetic referenced yaw from an extended Kalman filter fusing raw data from gyroscopes, accelerometers and magnetometers (Kooi, 2014). Further processing algorithms of our system exclusively use raw accelerometer and gyroscope data, which are accessible as well. Based on its specifications (see Table 3), the MTI-300 fits the classification of an industrial grade IMU. The IMU features a clock synchronization with a pulse per second (PPS) signal.

	Gyroscope	Accelerometer
Bias repeatability (1 yr)	0.5 deg/s	0.05 m/s^2
In-run bias stability	10 deg/h	40 µg
Noise density	0.015 deg/s/√Hz	150 µg/√Hz

Table 3. Specifications of the MEMS-based XSens MTI-300 AHRS IMU (Kooi, 2014)

3.4 Computer

The mobile mapping system is equipped with a minicomputer Prime Mini Pro with high RAM and data storage capacities and with a processor for mobile computers (see Table 4). The computer has low energy consumption and no mechanically moving parts.

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

Table 4. Specifications of the on-board computer Prime Mini Pro (Prime Computer AG, 2017)

3.5 Flashlights

Self-designed, strip-shaped LED elements provide a flashlight for the artificial lighting of low-light indoor environments. We glued commercial low-cost LED strips onto an aluminum rail with a matt cover, which diffuses the light. A strip-shaped LED element with a length of 50 centimeter generates a luminous flux between 1200 and 1500 lm.

4. SYSTEM CONFIGURATION AND CAMERA CALIBRATION

When creating our prototypic mobile mapping system, we generally focused on a flexible design for the hardware and software components. We equipped our system with a 360° panorama camera primarily for the use as a mapping sensor. Our aim is to orient single panorama images in indoor environments with the aid of two LiDAR sensors and an IMU as navigation sensors as precisely as possible. We implemented a time synchronization based on electric signals between all sensors to avoid systematic errors caused by software-induced time shifts. To further increase the accuracy, we performed a complete panorama camera calibration.

4.1 Mechanical Design

We assembled all sensors and components on a robust aluminum frame. The modular design allows rapid changes to the sensor configuration.

We fixed the multi-head panorama camera on the frame so that the first camera head (cam0) faces backwards. The other four sidewards facing camera heads (cam1-cam4) are arranged in clockwise order. The sixth camera head (cam5) faces to the top. Moreover, we tilted the panorama camera by a few degrees to achieve a roughly horizontal camera plane when a person carries the backpack (see Figure 1, no. 2). The overlapping of images allows stitching them together to a panorama image, so that the surrounding frame is barely visible.

One of the two LiDAR scanners sits on top of the frame and is tilted by 30° to cover mainly the walls as well as some parts of the floor and the ceiling (see Figure 1, no. 1). In contrast, the second vertically mounted LiDAR scanner (see Figure 1, no. 5) covers mainly the floor and the ceiling. The second LiDAR is complementary to the first one and gives additional geometric stability to the resulting point cloud. We fixed the IMU stably in the lower part of the frame (see Figure 1, no. 7).
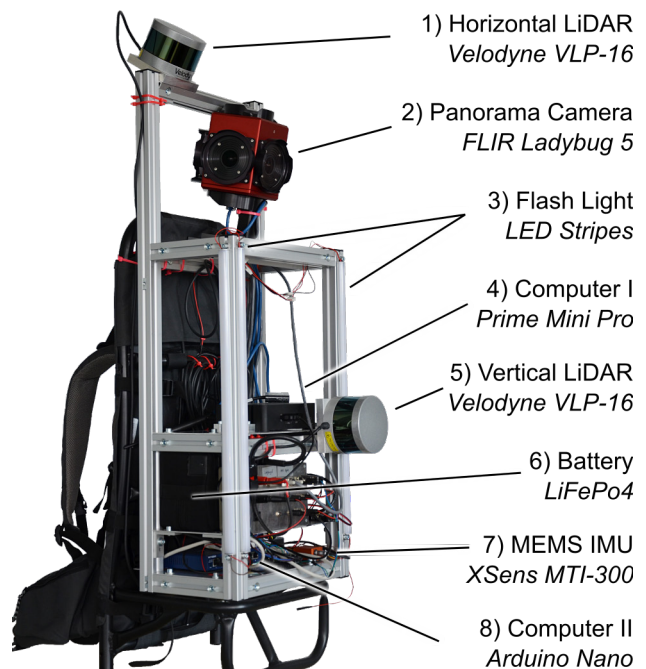

1) Horizontal LiDAR
Velodyne VLP-16

2) Panorama Camera
FLIR Ladybug 5

3) Flash Light
LED Stripes

4) Computer I
Prime Mini Pro

5) Vertical LiDAR
Velodyne VLP-16

6) Battery
LiFePo4

7) MEMS IMU
XSens MTI-300

8) Computer II
Arduino Nano

Figure 1. System overview demonstrating our backpack indoor mobile mapping system

4.2 Electronic Design

A 12V lithium ion battery with a capacity of 20Ah (see Figure 1, no. 6) supplies power for the system. The single board computer II (Arduino Nano) synchronizes the time between the sensors and triggers the camera and the flashlights with an electronic trigger signal. The computer II contains a GPS simulator and produces a pulse per second (PPS) signal and an NMEA message per second to synchronize both LiDAR scanners (see Figure 2). A PPS signal also synchronizes the internal IMU clock. A post-processing step solves the ambiguity of the internal IMU clock, using the NMEA message, which the computer II sends directly to computer I. For triggering the camera, computer I (Prime Mini Pro) sends a trigger command to computer II. Then computer II triggers the camera and the flashlights with an electronic signal and sends a message with the trigger timestamp back to computer I. All

sensors send the data to computer I for registration and pre-processing.
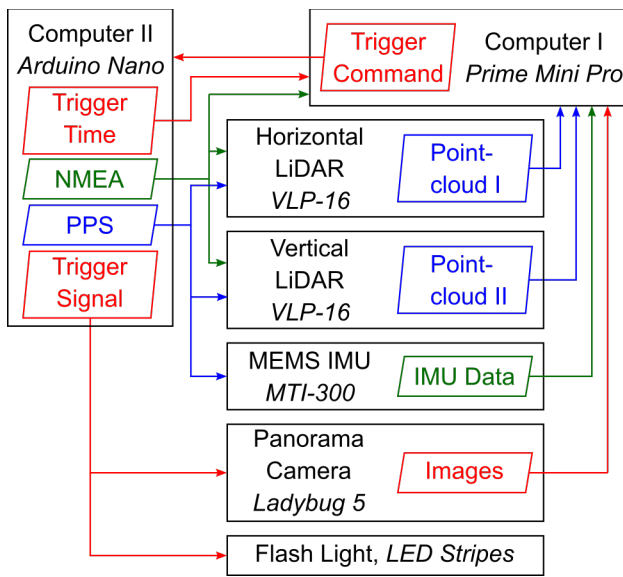

Figure 2. Schema with sensor synchronization and data flow

4.3 Software Design

Our acquisition software runs on computer I with a Linux Ubuntu 16.04 installation. Each external device with Virtual Network Computing (VNC) support is able to control computer I via remote desktop over a WLAN connection. We implemented our entire acquisition software with the Robot Operating System (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). 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). Either, a node can publish messages to one or more topics (red arrows) or 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. We represent services and topics as rectangles in our software schema (see Figure 3) but present the communication with services with green arrows.

For both LiDAR scanners, we use the Velodyne driver with minor modifications by Withley (2016). The driver converts the sensor raw data into ROS `pointcloud2` messages and publishes them. For the IMU, we use the XSens driver by Colas (2016) which converts the raw data into ROS `Imu` messages. For post-processing, the ROS node "Bag" stores the IMU messages as well as the messages from both LiDAR scanners in a so-called bag file. Furthermore, we fuse the LiDAR and IMU data with the 3D LiDAR SLAM Cartographer. Hess et al. (2016) introduce the functionality of Google's Cartographer exemplarily with their 2D LiDAR SLAM. Our new Conditional Trigger Node supports spatial trigger criteria, such as distance and orientation differences, as well as non-spatial criteria such as time difference and triggers a command if necessary. The Conditional Trigger Node uses the system pose coming from the Cartographer to

calculate spatial differences. The node of Computer II advertises a trigger command service, publishes the trigger time and logs all incoming messages from Computer II. As described in chapter 4.2, Computer II triggers the camera and the flashlights hardware-based.

Furthermore, we use the driver by Rockey & Purvis (2015) for the Panorama Camera, which returns all six images from the camera heads strung together in one image. The Image Slicer node re-separates the images and the different Image Writer node instances write the single images.
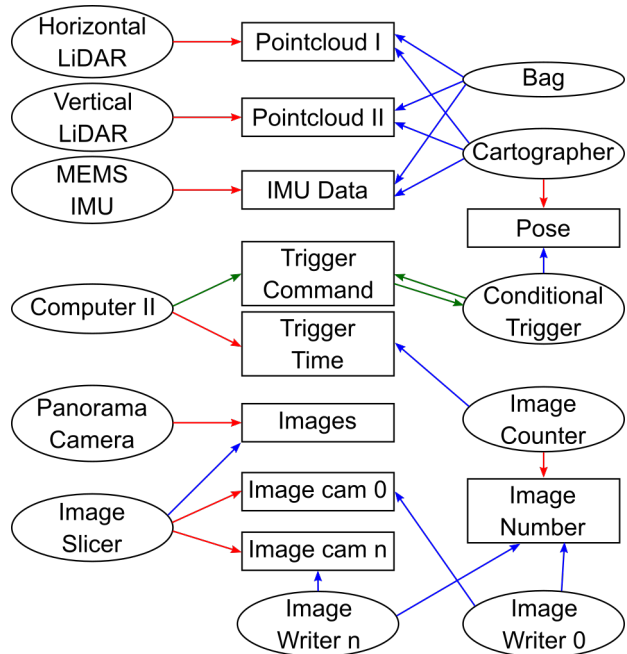

Figure 3. Schema with the implemented ROS software structure. ROS nodes are represented as circles and ROS topics as rectangles. Red arrows depict ROS publishers, blue arrows ROS subscribers and green arrows ROS services.

4.4 Camera Calibration

The calibration of the multi-head panorama camera consists of the interior orientation parameters (IOPs) of each camera head as well as the relative orientation parameters (ROPs) of each camera head to the first camera head (cam0). We used the constrained multi-system calibration software implemented by Burkhard et al. (2012) that is based on the bundle adjustment approach by Kersting et al. (2012). Blaser et al. (2018) extended this software with fisheye camera support using the equidistant camera projection model (Abraham & Förstner, 2005).

We calibrated the panorama camera in our indoor calibration field, which features 188 well-distributed target points on all four walls as well as on the ceiling (see Figure 4). We measured the target points with a contactless high-precision industrial measurement system, thus the standard deviation of a target point is below 0.3 mm.

Schmeing et al. (2011) modelled Ladybug3 panorama camera images as virtual spherical images. Rau et al. (2016) applied the same model to the Ladybug5 panorama camera as described in the Ladybug5 technical reference (FLIR Inc., 2017). Since we model each single camera head separately, no stitching error will occur. Our camera calibration investigations showed that the equidistant projection model (Abraham & Förstner, 2005) fits best for the Ladybug5 camera heads; therefore we apply the equidistant projection model to all individual camera heads. We

estimated the following IOPs: focal length, principal point, two radial and two tangential distortion parameters. We performed the calibration with five image epochs at different locations and with 1010 image observations. The mean reprojection error amounts to 0.23 pixel. Table 5 shows the standard deviations of the ROPs between the first camera head and each other camera head.

Figure 4. Indoor calibration field for multi-camera configurations of mobile mapping systems

Camera head	X [mm]	Y [mm]	Z [mm]	ω [deg]	φ [deg]	κ [deg]
cam1	0.27	0.18	0.34	0.033	0.023	0.030
cam2	0.20	0.17	0.38	0.016	0.022	0.007
cam3	0.24	0.17	0.40	0.017	0.023	0.009
cam4	0.29	0.18	0.34	0.033	0.022	0.029
cam5	0.34	0.62	0.46	0.016	0.010	0.016
Mean	0.27	0.27	0.39	0.023	0.020	0.018

Table 5. Standard deviations of the calibrated ROPs between the first camera head (cam0) and each other camera head (cam1-cam5)

5. DATA ACQUISITION AND PROCESSING

Our prototypic MMS initializes completely indoors. The starting position of the LiDAR SLAM defines the origin of a local coordinate frame. The LiDAR SLAM algorithm computes real time sensor poses, which enable geometrically constrained camera triggering. We have implemented a time-based, a distance-based and a horizontal angle-based camera trigger constraint. In addition, the LiDAR SLAM provides a map overview containing the trajectory as well as the captured environment (see Figure 5). When data acquisition is completed, the LiDAR SLAM exports the so-called Cartographer state, which inherits the optimized sensor trajectory. Further data acquisition results are the ROS bag file with LiDAR and IMU messages (see section 4.3), the raw images from the panorama camera as well as a list containing the camera trigger events with their corresponding timestamps.

In order to perform image measurements, we require undistorted images and their respective exterior orientations. Figure 6 gives an overview of our entire post-processing workflow.

First, we undistort the raw images using the calibrated IOPs. Applying the calibrated IOPs will correct the distortions to the equidistant camera model. Abraham & Förstner (2005) provide the formulas for the conversion from the equidistant to the perspective camera model. We apply this conversion in a further step directly to the image measurements.

Second, in order to obtain the corresponding exterior image orientations (EOPs) based on LiDAR SLAM, we extract the trajectory from the Cartographer state. For this, we extended the Google Cartographer with a batch function, which writes the final trajectory as a text file, by performing Google Cartographer's assets writer. In a next step, we perform time-based interpolation of the image events between the trajectory events. We conduct a linear interpolation for the positions and a spherical linear interpolation based on quaternions for the orientations. Finally, we transform the interpolated positions and orientations with the calibrated ROPs and the manually measured boresight alignment parameters from the body frame to the frame of each camera head.
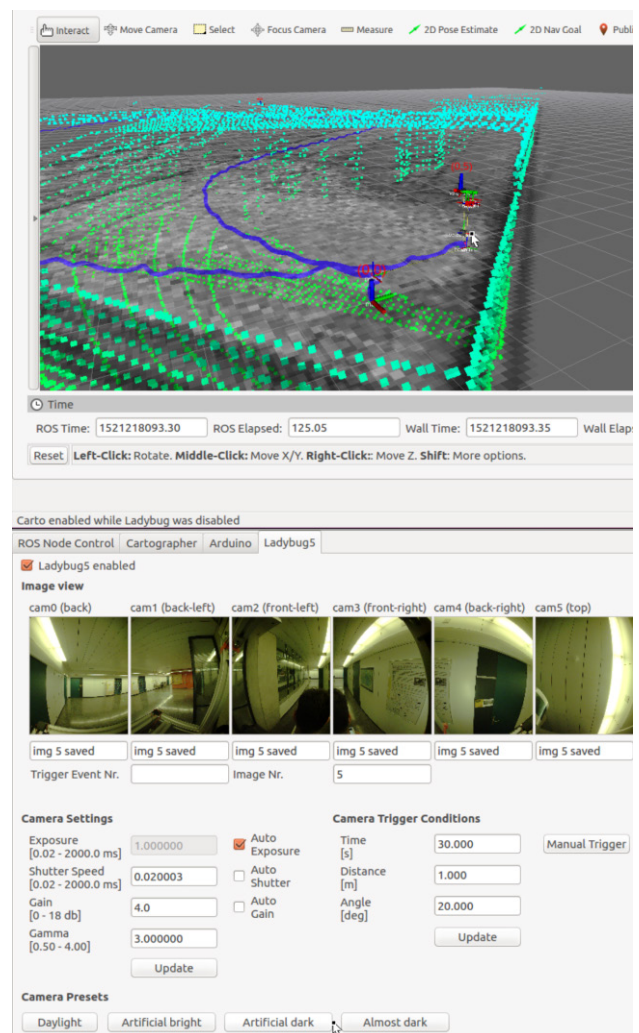

Figure 5. Screenshots of the data acquisition software. Top: Cartographer SLAM preview, Bottom: Image preview and conditional trigger settings

In application scenarios with high accuracy requirements, we aim to improve the LiDAR SLAM-based EOPs with a subsequent image-based georeferencing. For this purpose, we introduce the undistorted images into the incremental structure-from-motion (SfM) tool COLMAP (Schönberger & Frahm, 2016), which

Cavegn et al. (2018) extended with georeferencing capabilities by integrating prior EOPs and exploiting ROP constraints. Hence, we use LiDAR SLAM-based EOPs of the first camera (cam0) as initial values, and fix the ROPs of the other camera heads with the pre-calibrated values. The complete process of integrated georeferencing based on COLMAP is described in detail by Cavegn et al. (2018). Good lighting conditions and a structurally rich environment for proper and well-distributed feature detection are essential.
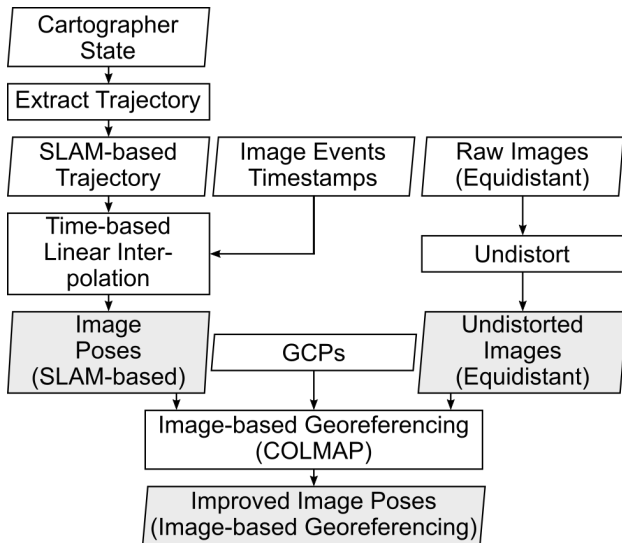

Figure 6. Flow chart indicating our data post-processing workflow

6. PERFORMANCE EVALUATION

In order to compare the performance of LiDAR SLAM-based image poses and subsequently improved poses using image-based georeferencing, we carried out investigations in an indoor environment.

6.1 Test Site and Data Acquisition

Our indoor study area is located on the sixth floor of the main campus building of the University of Applied Sciences and Arts Northwestern Switzerland (FHNW) in Muttenz close to Basel. It features a hallway with a dimension of ca. 27 m x 24 m, that leads to several offices, computer rooms, lecture rooms, staircases and elevators. The typical corridor width is approximately 3 m, with a minimum of about 2 m (left part in Figure 7) and a maximum of 4 m (right part, no. 3 in Figure 7). We determined 3D coordinates of many reference points representing natural markings e.g. on doorframes, elevator and room corners by tachymetry. These reference points have an accuracy of approximately 5 mm. We used several of them for our investigations as ground control points (GCPs) or check points (CPs).

We started the data acquisition at the origin of the local geodetic coordinate frame marked with a green diamond on 21.03.2018 10:50. The acquisition required 24 minutes. The camera trigger constraints were set to a distance interval of 1 m and an azimuth change of 15 degrees. In total, we captured a set of 1518 images with the panorama camera corresponding to six images at 253 different locations, indicated as blue filled circles in Figure 7.

6.2 Image-Based Georeferencing

We defined the ROP configuration as well as its corresponding calibrated values in a JSON file, where the backward facing camera head cam0 serves as master, and fixed these ROP for the subsequent processing. Since the images from the upward facing camera head cam5 predominantly contain homogeneous surfaces leading to few feature correspondences, we only processed images from the horizontal pointing camera heads cam0-cam4 captured at 253 locations. COLMAP exploited previously computed EOP from online LiDAR SLAM, used the simple radial fisheye camera model and performed no refinement of the interior orientation parameters.

COLMAP extracted DSP-SIFT features and carried out spatial feature matching with a maximum distance of 10 m, 200 maximum neighbours, a maximum angle constraint of 100 degrees and a maximum ratio of 0.9. This resulted in 1115 registered images, 92'597 points, 779'782 observations, a mean track length of 8.4, 699.4 mean observations per image and a mean reprojection error of 0.79 pixel. These values are similar to the results obtained from a different mapping at the same location on 27.11.2017 (Cavegn et al., 2018).
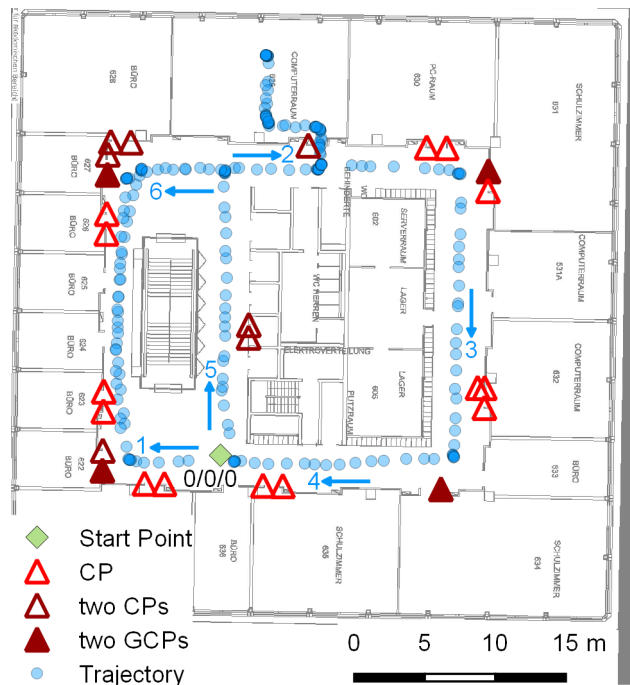

Figure 7. Floor base map with overlaid projection centers of camera head cam0 (Trajectory), ground control points (GCPs) and check points (CPs) in the local geodetic coordinate system. At dark red point symbols, we measured both points at the top and the bottom of a doorframe.

6.3 Precision Analysis

In order to perform a precision analysis, we determined 36 points in total and used 28 of them as CPs and 8 as GCPs (see Figure 7). We measured each point in a minimum of four different images, performed a bundle adjustment-based forward intersection with a Python tool and estimated the 3D point coordinates. In a first run, we used the LiDAR SLAM-based image poses and in the second run, we utilized the improved image poses from image-based georeferencing to perform the forward intersection. The mean precision of the forward intersection with SLAM-based image poses amounts to 11.6 cm. In contrast, with image poses improved by image-based georeferencing, the mean precision of

the forward intersection improves to 0.3 cm. Table 6 shows the mean precisions for GCPs and CPs respectively. This precision is a good indicator for the relative accuracy of typical 3D distance and area measurement tasks, which could be carried out irrespective of the presence of ground control points.

Furthermore, we transformed the resulting 3D points with eight GCPs (see Figure 7) into the ground truth coordinate frame in order to perform the accuracy evaluation. Thus, we estimated three translation parameters as well as three rotation parameters. We compared the transformed 3D points of both sensor orientation approaches separately with the ground truth from tachymetry. The RMSE of 3D point residuals measured with LiDAR SLAM-based image poses amounts to 13.3 cm. In contrast, the RMSE of 3D point residuals measured with image-based georeferenced image poses is 1.8 cm. The results of image-based georeferencing representing the absolute 3D measurement accuracy are in the same order as in Cavegn et al. (2018).

	LiDAR SLAM-based poses		Improved poses by image-based georeferencing	
	GCPs	CPs	GCPs	CPs
Precision [cm]	8.2	12.3	0.2	0.3
Accuracy [cm]	10.6	13.3	1.3	1.8

Table 6. Precision and accuracy values for ground control points and check points from both LiDAR SLAM-based image poses and image-based georeferencing. Precision indicates the mean RMSE of forward intersection of single point measurements. Accuracy shows the RMSE of residuals to tachymetry.

The precision as well as the accuracy values of the 3D point measurements based on the LiDAR SLAM image poses are in the same order. Thus, the LiDAR SLAM image poses are suitable e.g. for point object digitalization in the decimeter range. By performing a subsequent image-based georeferencing, we improved the accuracies by an order of magnitude. The results of measured 3D points with the improved image poses show that relative measurements in the sub-centimeter range and absolute measurements in the centimeter range are achievable. It should be noted that image-based measurements are far more intuitive than measurements in 3D point clouds. They also require very little training – in contrast to measurements in 3D point clouds, which typically require expert skills.

7. CONCLUSIONS AND OUTLOOK

We successfully developed a portable mobile mapping system for efficient image-based infrastructure management. Our research platform and working prototype features a multi-head panorama camera for the actual image acquisition as well as two multi-profile LiDAR scanners (horizontal and vertical) and a MEMS-based industrial grade IMU for LiDAR-based SLAM. Our modular design allows straightforward adaptations of the sensor configuration and of the acquisition software based on the Robot Operating System (ROS). We also provided details on the calibration of the multi-head panorama camera Ladybug5 using the equidistant fisheye camera model. We furthermore implemented and presented a workflow for indoor data acquisition without a need for outdoor initialization. The workflow utilizes LiDAR-based SLAM for online mapping and progress monitoring as well as for post-processing purposes. Furthermore, we improved the camera poses with subsequent image-based georeferencing using relative orientation constraints.

Our accuracy investigation delivered absolute 3D point accuracies in the range of 10.6 to 13.3 cm using image poses directly from LiDAR SLAM. Thus, using image poses from direct georeferencing based on LiDAR SLAM, 3D point coordinates can be determined with an accuracy at the decimeter level. By performing our image-based georeferencing, an accuracy improvement by an order of magnitude can be expected. Our investigations show absolute 3D point accuracies in the range of 1.3 to 1.8 cm. The excellent 3D point measurement precision of 0.2-0.3 cm, obtained after image-based georeferencing, indicates that the final 3D services will provide relative accuracies for typical measurement tasks well within the sub-centimeter level.

In future work we will tackle the overall system calibration, i.e. the precise alignment of the laser scanners and the panorama camera to the IMU. Moreover, we plan to assemble additional cameras that will allow investigations with fixed stereo bases, presumably leading to more accurate image poses especially in narrow passages. Furthermore, we will fuse the IMU data as well as the trajectory points from LiDAR SLAM within an Extended Kalman Filter (EKF). Since we expect an accuracy increase of EKF-based camera poses, this method could be an alternative to the image-based georeferencing. We expect that this approach should perform well also in indoor environments with poor structure and different lighting conditions.

REFERENCES

Abraham, S., & Förstner, W., 2005. Fish-eye-stereo calibration and epipolar rectification. *ISPRS J. Photogramm. Remote Sens.*, *59*(5), pp. 278–288.

Blaser, S., Nebiker, S., & Cavegn, S., 2017. System Design, Calibration and Performance Analysis of a Novel 360° Stereo Panoramic Mobile Mapping System. In: *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, Hannover, Germany, Vol. IV-1/W1, pp. 207–213.

Blaser, S., Nebiker, S., & Cavegn, S., 2018. On a Novel 360° Panoramic Stereo Mobile Mapping System. *Photogramm. Eng. & Remote Sens.*, 84(6), pp. 347–356.

Burkhard, J., Cavegn, S., Barmettler, A., & Nebiker, S., 2012. Stereovision Mobile Mapping : System Design and Performance Evaluation. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Melbourne, Australia, Vol. XXXIX, Part B5, pp 453–458.

Cavegn, S., Nebiker, S., & Haala, N., 2016. A systematic comparison of direct and image-based georeferencing in challenging urban areas. In: *Int. Arch. of the Photogramm. Remote Sens. Spatial Inf. Sci.*, Prague, Czech Republic, Vol. XLI, Part B1, pp. 529–536.

Cavegn, S., & Haala, N., 2016. Image-Based Mobile Mapping for 3D Urban Data Capture. *Photogramm. Eng. & Remote Sens.*, 82(12), pp. 925–933.

Cavegn, S., Blaser, S., Nebiker, S., & Haala, N., 2018. Robust and Accurate Image-Based Georeferencing Exploiting Relative Orientation Constraints. In: *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, Riva del Garda, Italy, Vol. IV-2, pp 57–64.

Colas, F., 2016. XSens Driver, ROS Wiki http://wiki.ros.org/xsens_driver (3 April 2018).

FLIR Inc., 2017. FLIR Ladybug5 USB3, Technical Reference. Richmond, BC. https://www.ptgrey.com/support/downloads/10128 (3 April 2018).

Glennie, C. L., Kusari, A., & Facchin, A., 2016. Calibration and stability analysis of the VLP-16 laser scanner. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Lausanne, Switzerland, Vol. XL-3/W4, pp. 55–60.

Hess, W., Kohler, D., Rapp, H., & Andor, D., 2016. Real-Time Loop Closure in 2D LIDAR SLAM. In: *IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden, pp. 1271–1278.

Kersting, A. P., Habib, A., & Rau, J.-Y., 2012. New Method for the Calibration of Multi-Camera Mobile Mapping Systems. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Melbourne, Australia, Vol. XXXIX, Part B1, pp. 121–126.

Kooi, B., 2014. MTi User Manual, MTi 10-series and MTi 100-series. http://www.farnell.com/datasheets/1935846.pdf (3 April 2018).

Lehtola, V.V., Kaartinen, H., Nüchter, A., Kaijaluoto, R., Kukko, A., Litkey, P., Honkavaara, E., Rosnell, T., Vaaja, M.T., Virtanen, J.-P., Kurkela, M., El Issaoui, A., Zhu, L., Jaakkola, A. & Hyyppä, J., 2017. Comparison of the Selected State-Of-The-Art 3D Indoor Scanning and Point Cloud Generation Methods. *Remote Sens.* 9(8), https://doi.org/10.3390/rs9080796

Leica Geosystems, 2018. Leica Pegasus Backpack Wearable Mobile Mapping Solution. https://leica-geosystems.com/products/mobile-sensor-platforms/capture-platforms/leica-pegasus-backpack (3 April 2018).

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. *J. F. Robot.*, 32(4), pp. 474–503.

Nebiker, S., Cavegn, S., & Loesch, B., 2015. Cloud-Based Geospatial 3D Image Spaces—A Powerful Urban Model for the Smart City. *ISPRS Int. J. Geo-Information*, 4(4), pp. 2267–2291.

Novak, K., 1991. The Ohio State University Highway Mapping System: The Stereo Vision System Component. In: *Proc. of the 47th Annual Meeting of the Institute of Navigation*, Williamsburg, VA, pp. 121–124.

Nüchter, A., Borrmann, D., Koch, P., Kühn, M., & May, S., 2015. A Man-Portable, Imu-Free Mobile Mapping System. In: *Int. Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, La Grande Motte, France, Vol. II-3/W5, pp. 17–23.

Prime Computer AG, 2017. PrimeMini Pro Data sheet. https://primecomputer.ch/wp-content/uploads/2017/09/Produktdatenblatt_PrimeMiniPro_all_configurations_neu.pdf (3 April 2018).

Puente, I., González-Jorge, H., Martínez-Sánchez, J., & Arias, P., 2013. Review of mobile mapping and surveying technologies. *Measurement*, 46(7), pp. 2127–2145.

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Berger, E., Wheeler, R., Ng, A., 2009. ROS: an open-source Robot Operating System. *ICRA workshop on open source software*, 3(3.2), p. 5.

Rau, J. Y., Su, B. W., Hsiao, K. W., & Jhan, J. P., 2016. Systematic calibration for a backpacked spherical photogrammetry imaging system. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Prague, Czech Republic, Vol. XLI, Part B1, pp. 695–702.

Rockey, C., & Purvis, M., 2015. Pointgrey Camera Driver. ROS Wiki http://wiki.ros.org/pointgrey_camera_driver (3 April 2018).

Schmeing, B., Laebe, T., & Förstner, W., 2011. Trajectory Reconstruction Using Long Sequences of Digital Images from an Omnidirectional Camera. In: *DGPF Tagungsband 20/2011*, Mainz, Germany, pp. 443–452.

Schönberger, J. L., & Frahm, J.-M., 2016. Structure-from-Motion Revisited. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 4104–4113.

Schwarz, K. P., Martell, H. E., El-Sheimy, N., Li, R., Chapman, M. A., & Cosandier, D., 1993. VIASAT - A mobile highway survey system of high accuracy. In: *Proceedings of the Vehicle Navigation and Information Systems Conference*, Ottawa, Canada, pp. 476–481.

Stachniss, C., Leonard, J. J., & Thrun, S., 2016. Simultaneous Localization and Mapping. *Springer Handbook of Robotics*, pp. 1153–1176.

Szwarc, H., 2017. Increasing information in building models. *GIM International*, 31(5), pp. 31–33.

Thomson, C., Apostolopoulos, G., Backes, D., & Boehm, J., 2013. Mobile Laser Scanning for Indoor Modelling. In: *Int. Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, Antalya, Turkey, Vol. II-5/W2, pp. 289–293.

Thrun, S., 2002. Robotic mapping: A survey. *Exploring Artificial Intelligence in the New Millennium*, 1, pp. 1–35.

Velodyne, 2016. User's Manual and Programming Guide, VLP-16 Velodyne LiDAR Puck. http://velodynelidar.com/docs/manuals/VLP-16 User Manual and Programming Guide 63-9243 Rev A.pdf (3 April 2018).

Vexcel, 2018. UltraCam Panther. http://www.vexcel-imaging.com/products__trashed/ultracam-panther/ (3 April 2018).

Withley, J., 2016. Velodyne Driver. ROS Wiki http://wiki.ros.org/velodyne_driver (3 April 2018).