

CENTIMETRE-ACCURACY IN FORESTS AND URBAN CANYONS – COMBINING A HIGH-PERFORMANCE IMAGE-BASED MOBILE MAPPING BACKPACK WITH NEW GEOREFERENCING METHODS

S. Blaser¹, J. Meyer¹, S. Nebiker¹, L. Fricker¹, D. Weber¹

¹ Institute of Geomatics, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland - (stefan.blaser, jonas.meyer, stephan.nebiker)@fhnw.ch, lea.delia.fricker@gmail.com, tapi@gmx.ch

Commission I, ICWG I/IV

KEY WORDS: Backpack Mobile Mapping, Performance Investigation, Georeferencing, SLAM, Bundle Adjustment, Urban, Forest

ABSTRACT:

Advances in digitalization technologies lead to rapid and massive changes in infrastructure management. New collaborative processes and workflows require detailed, accurate and up-to-date 3D geodata. Image-based web services with 3D measurement functionality, for example, transfer dangerous and costly inspection and measurement tasks from the field to the office workplace. In this contribution, we introduced an image-based backpack mobile mapping system and new georeferencing methods for capture previously inaccessible outdoor locations. We carried out large-scale performance investigations at two different test sites located in a city centre and in a forest area. We compared the performance of direct, SLAM-based and image-based georeferencing under demanding real-world conditions. Both test sites include areas with restricted GNSS reception, poor illumination, and uniform or ambiguous geometry, which create major challenges for reliable and accurate georeferencing. In our comparison of georeferencing methods, image-based georeferencing improved the median precision of coordinate measurement over direct georeferencing by a factor of 10-15 to 3 mm. Image-based georeferencing also showed a superior performance in terms of absolute accuracies with results in the range from 4.3 cm to 13.2 cm. Our investigations showed a great potential for complementing 3D image-based geospatial web-services of cities as well as for creating such web services for forest applications. In addition, such accurately georeferenced 3D imagery has an enormous potential for future visual localization and augmented reality applications.

1. INTRODUCTION

Ongoing progress in digitalization leads to rapid and massive changes in infrastructure management. The establishment of three-dimensional collaborative processes and workflows with stakeholders from multiple domains require detailed accurate and up-to-date 3D geodata. Image-based mobile reality capturing techniques in combination with cloud technologies, such as presented by Nebiker et al. (2017), hold the potential to provide such data and services in a rapid, cost-efficient and user-friendly manner. First image-based outdoor mobile mapping systems (MMS) date back to the early 1990ies (Novak, 1991; Schwarz et al., 1993). Burkhard et al. (2012) present a stereo image-based MMS and performed accuracy investigations using different types of industrial cameras. In order to capture urban environments with a maximal coverage, image-based MMS have evolved into systems with (multi-) panorama camera configuration (Meilland et al., 2015). Blaser et al. (2017) present a MMS configuration with two tilted panorama cameras, which constitute multiple stereo systems to the sides in order to capture entire façades of the buildings.

However, most MMS use LiDAR as primary sensors and cameras as complimentary sensors in order to generate textured point clouds (Heuvel et al., 2006; Puente et al., 2013). Nebiker et al. (2015) discuss some the advantages of image-based over LiDAR-based MM data in terms of temporal coherence in the acquisition and density of information. Current developments in the field of MM are moving towards updating already existing 3D databases in indoor spaces (Hasler et al., 2019; Kostoeva et al., 2019; Saran et al., 2019) as well as in outdoor environments

(Hasler et al., 2020) using consumer devices (smartphones and tablets).

Other work has focused on accuracy improvement within GNSS-denied areas. Jende (2019), for example, improved the trajectory in urban areas in a largely automated process using aerial imagery. While his investigations yield accuracies at the decimetre level, Cavegn et al. (2019) obtained accuracies at the centimetre level by applying image-based georeferencing using constrained bundle-adjustment and ground control points.

In recent years, some portable as well as indoor MMS came onto the market. Lehtola et al. (2017) provide a comparison of numerous state-of-the-art LiDAR-based indoor MMS. By contrast, Tang et al. (2015) conducted performance investigations in forests using SLAM positioning and show an improvement of the accuracy by 38 % compared to direct georeferencing. Our very first investigations 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 such as image-based georeferencing with constrained bundle-adjustment (Wittmer, 2017).

In conjunction with the development of new georeferencing methods, Blaser et al., (2018) present the development of a portable image-based MMS and provide accuracy analysis in indoor environments with promising results within the centimetre range. Blaser et al. (2019) confirmed the high accuracy potential in a challenging underground environment.

In this contribution we first expand our earlier image-based backpack MMS with direct georeferencing. Thus, with direct georeferencing, SLAM-based georeferencing (Blaser et al., 2018) and image-based georeferencing (Cavegn et al., 2018) three independent georeferencing approaches are now available, which we evaluate at two large-scale outdoor test sites. Thereby, we even focus on areas not accessible to vehicles, which often have restricted GNSS reception, poor illumination and difficult geometric conditions (e.g. narrow streets in the city centre or small footpaths in the forest). Finally, we discuss the results of the different georeferencing methods and the suitability of our high-performance backpack MMS for outdoor use.

2. MOBILE MAPPING SYSTEMS AND METHODS

2.1 State-of-the-art portable and indoor systems

Several portable MMS already exist for data acquisition in inaccessible and often GNSS restricted environments. Most of them focus on capturing a 3D LiDAR point cloud. Tucci et al. (2018) examined three commercially available LiDAR-based portable MMS. Both *Zeb-Revo* and *Kaarta Stencil* are handheld MMS with only SLAM-based navigation. By contrast, the *Leica Pegasus Backpack* is a multi-sensor backpack MMS using GNSS and IMU as well as SLAM for navigation and georeferencing. Maboudi et al. (2017) present a point cloud-based comparison of the *Zeb-Revo* with the trolley-based indoor MMS *Viametris iMS3D*. Lehtola et al. (2017) examined different types of commercially available indoor MMS as well as prototypical research indoor MMS. Among them are a) handheld and exclusively LiDAR SLAM-based MMS, such as the *Zebedee*, the *Kaarta Stencil*, and *Aalto VILMA*, b) terrestrial indoor mapping systems such as the *Matterport*, c) trolley- and cart-based systems like *NavVis* and *FGI Slammer* as well as d) backpack MMS such as the *Leica Pegasus Backpack* and the *Würzburg Backpack* (Nüchter et al., 2015). Further commercially available and more image-focussed backpack MMS are the *Vexcel Panther* (Vexcel, 2020) and the *Viametris Backpack bMS3D LD5+* (Viametris, 2020). Blaser et al. (2018) discuss different MMS platform types in terms of flexibility and acquisition efficiency and define system requirements in order to acquire and create image-based services for infrastructure management with accurate 3D measurement functionality.

2.2 Georeferencing methods for mobile mapping data

Current georeferencing methods for data acquired with mobile mapping systems can be divided into direct georeferencing, SLAM-based georeferencing, and additionally image-based georeferencing – in case of image-based MMS.

In general, **direct georeferencing** refers to platform and sensor pose estimation using on-board sensors only. Since the early days of MMS, the sensor combination of GNSS and IMU is widely used for direct georeferencing. Schwarz et al. (1993) describe the advantages of GNSS and IMU sensor integration in detail. However, the accuracy of direct georeferencing strongly depends on GNSS reception (see Table 1). In case of partial or total signal shading, the accuracy can decrease from the centimetre range to the decimetre or even metre range – even with high-end equipment (Cavegn et al., 2018). Thus, direct georeferencing works well in outdoor environments with good GNSS coverage. Direct georeferencing is also real-time capable and – in case of additional post-processing – requires a comparatively

small computational effort. Direct georeferencing provides the trajectory directly within a global reference frame.

Simultaneous localization and mapping (SLAM) has been widely used in robotics and was first used for the navigation of robots in indoor environments (Durrant-Whyte, Bailey, 2006). Cadena et al. (2016) give an overview of different SLAM approaches based on various sensors. Blaser et al. (2018) used 3D LiDAR SLAM with loop closure support for georeferencing purposes in indoor environments and proved an accuracy potential within the decimetre range using entry-level LiDAR sensors. In case of 3D LiDAR SLAM, the accuracy and robustness depend on the geometric properties of the environment as well as on the trajectory shape. High geometric variability with clear corners, edges and surfaces as well as a loop shaped trajectory improve the robustness and increase the accuracy. 3D LiDAR SLAM is real-time capable but requires high CPU resources. Usually, SLAM operates in a local coordinate frame. Blaser et al. (2019) show that post-processing with optimized parameters improves the result, whereby post processing time is typically in excess of the acquisition time.

Subsequent **image-based georeferencing** can significantly improve image poses from direct georeferencing or from SLAM-based georeferencing. Cavegn et al. (2018) significantly improved direct georeferencing by processing images of all stereo systems within a SfM pipeline. Furthermore, fixed pre-calibrated relative orientation parameters stabilize the bundle block adjustment with a reduction in the unknown parameters. Resulting accuracies were in the centimetre range. However, the processing time significantly exceeds the acquisition time. Moreover, the accuracy as well as the robustness strongly depend on lighting conditions and the radiometric texture. In principle, image-based georeferencing can be used in outdoor as well as in indoor and underground environments.

Some of the main characteristics of the georeferencing methods discussed above are summarized in Table 1.

	Direct	SLAM-based (LiDAR)	Image-based
Accuracy range	dm to cm	dm	cm
Operation areas	Outdoor	Outdoor, indoor, underground	Outdoor, indoor, underground
Environmental requirements	GNSS reception	Geometric texture	Radiometric texture
Reference frame	Global	Local	Local
Post processing time vs. acquisition time	- (real-time capable)	+ (real-time capable)	++

Table 1. Comparison of the discussed georeferencing methods.

3. BACKPACK MOBILE MAPPING SYSTEM

Commercially available backpack MMS are often closed systems. Thus, investigations and improvements in the sensor configuration as well as in georeferencing and data processing are very limited or not possible at all. Therefore, we developed the prototypical and modular image-based portable indoor MMS *BIMAGE Backpack*. In previous contributions, we already described the system development in detail. In this chapter, we focus on the new system extension with direct georeferencing. This includes hardware and software enhancements as well as

innovations in the post processing workflow. A major challenge was that the changes should not affect previous key features, such as the indoor initialization and the independent SLAM-based system navigation and mission progress monitoring.

3.1 System configuration

Blaser et al. (2018) describe the original indoor configuration of the *BIMAGE Backpack* in detail. A horizontal and a vertical LiDAR scanner *Velodyne VLP-16* as well as an industrial grade IMU *XSens MTI-300* formed the navigation sensors. The multi-head panorama camera *Ladybug 5* was used as the main mapping sensor. In order to enable outdoor applications, we modified and extended the system configuration as follows (see Figure 1):

- Replacement of the industrial grade IMU *XSens MTI-300* by the IMU *NovAtel SPAN CPT7* with tactical grade performance
- Extension by the helix GNSS antenna *HX-CHX600A*
- Removal of the flashlights and the single board computer *Arduino Nano* since the *NovAtel SPAN CPT7* synchronizes both LiDAR scanners and provides a more precise time reference for camera timestamps than the *Arduino Nano*

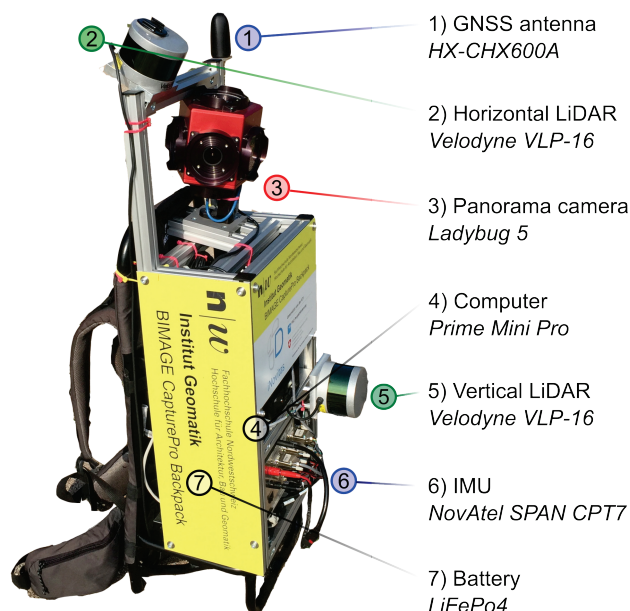


Figure 1. *BIMAGE Backpack* sensor configuration for indoor and outdoor applications. Sensors marked with blue lines are mainly used for direct georeferencing, green lines show sensors used for LiDAR-SLAM-based georeferencing and red lines components for image-based georeferencing. Black lines represent supporting components.

The specifications of the tactical grade MEMS-based IMU *NovAtel SPAN CPT7* state a position accuracy of horizontally 10 mm and vertically 20 mm under good GNSS coverage and after post-processing. Accuracies of the attitude angles roll and pitch are specified as 0.005° and of the heading angle as 0.010° . A GNSS outage of 60 s degrades the horizontal accuracy to 150 mm, the vertical accuracy to 50 mm, the roll and pitch attitude accuracies to 0.007° and the heading accuracy to 0.012° (NovAtel Inc., 2020). The specifications are in the same range

as those of the tactical grade IMU *UIMU-LCI* with fibre-optic gyros used on the MMS vehicle (Cavegn et al., 2018).

Our self-developed acquisition and system control software is based on the Robot Operation System (ROS) (Quigley et al., 2009). We used the ROS *novatel_span_driver* for *SPAN CPT7* support (Purvis et al., 2019). We subsequently extended the driver to record raw GNSS and INS data as well as camera timestamps for post processing.

In outdoor campaigns, we used the real-time backpack pose provided by the *SPAN CPT7* for navigation and for the geometrically constrained camera triggering. In contrast, LiDAR SLAM-based real-time navigation is available for indoor campaigns, but it requires more CPU power.

3.2 Post-processing workflow

The *BIMAGE Backpack* records GNSS and INS raw data from the *SPAN CPT7*, LiDAR raw data from both *Velodyne VLP-16* and raw single images from each *Ladybug 5* camera head (see Figure 2, top line). GNSS raw data from a reference station have to be obtained externally.

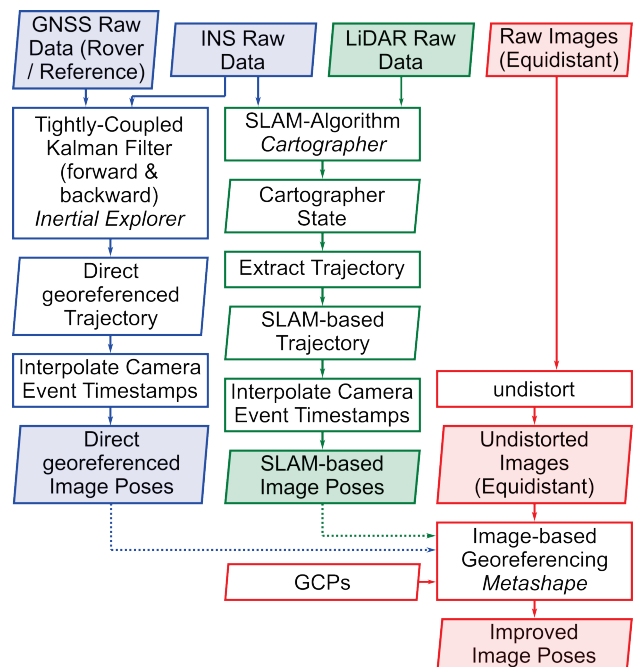


Figure 2. Extended flow chart indicating our data post-processing workflow (Blaser et al., 2018). The blue elements belong to the direct georeferencing, while the green elements represent the SLAM-based georeferencing and the red elements concern to the image-based georeferencing.

The *SPAN CPT7* extension enables direct georeferencing using tightly coupled GNSS and INS sensor data fusion with a Kalman filter. For this, we used the *Waypoint Inertial Explorer* software, which supports forward and backward trajectory processing as well as camera event interpolation based on timestamps (see Figure 2, left). By considering pre-calibrated lever arms and misalignments between the camera heads and the navigation centre, georeferenced image poses are obtained directly in the global reference frame.

Furthermore, LiDAR SLAM-based georeferencing is available independently from direct georeferencing. Based on IMU and

LiDAR raw data, the 3D SLAM algorithm *Google Cartographer* (Hess et al., 2016) continuously estimates the trajectory and the Voxel map. Then, a self-developed exporter extracts the trajectory from the so-called *Cartographer State* and interpolates the camera events based on recorded timestamps. By considering lever arms and misalignments between the camera heads and the navigation centre, georeferenced image poses are obtained in a local coordinate frame with the origin at the start of the campaign (see Figure 2, centre).

Finally, subsequent image-based georeferencing can significantly improve image poses originating from direct georeferencing or SLAM-based georeferencing (Cavegn et al., 2018). For this purpose, we introduced the undistorted images into the structure-from-motion software *Agisoft Metashape*. Further, we fixed the pre-calibrated relative orientation parameters between the individual camera heads and introduced SLAM-based or directly georeferenced image poses as initial values. Poses in a local coordinate frame can be transformed to the reference coordinate frame using ground control points (GCPs). The results of the bundle-adjustment are improved image poses in the desired reference frame (see Figure 2, bottom).

4. TEST SITES

We used two different test sites to carry out extensive performance investigations with the *BIMAGE Backpack* in outdoor environments. One test site is located in a city centre and the other in a forest. Both test sites have sections with restricted GNSS reception and paths, which are not accessible to a MMS vehicle. They also represent real-world scenarios for acquisition campaigns. Therefore, they are suitable to investigate the potential and the limitations and to evaluate different georeferencing approaches.

4.1 City centre

The first test site is located in the city centre of Basel (Switzerland) and covers an area of 150 x 200 m. It includes different road and path widths including a place with good GNSS reception for system initialization (see Figure 3, Image 1). By contrast, it also includes narrow alleys only accessible to pedestrians with steps and slopes up to 16 % (see Figure 3, Image 2). Such narrow alleys are challenging for all three georeferencing methods because of restricted GNSS reception, geometric homogeneity and poor illumination. Wide pedestrian promenades with shops on both sides dominate other parts of the test site (see Figure 3, Image 3). Image 3 in Figure 4 shows the main traffic axis through the city centre with busy tram and bicycle traffic. The dense street network in the city centre allows for acquisition patterns with multiple loops. Furthermore, the test site comprises 79 reference points. Most of them are well-defined natural reference points and some were marked with photogrammetric targets. Fricker, Weber (2019) provide a detailed description of the reference point measurements by tachymetry and show a 3D standard deviation below 5 mm.

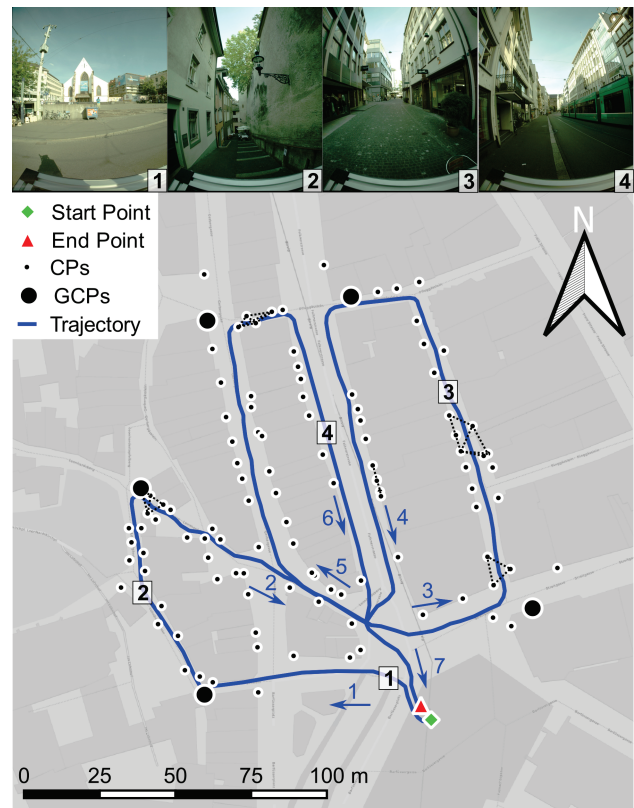


Figure 3. Map from the first test site located in the city centre of Basel with images showing environmental conditions.

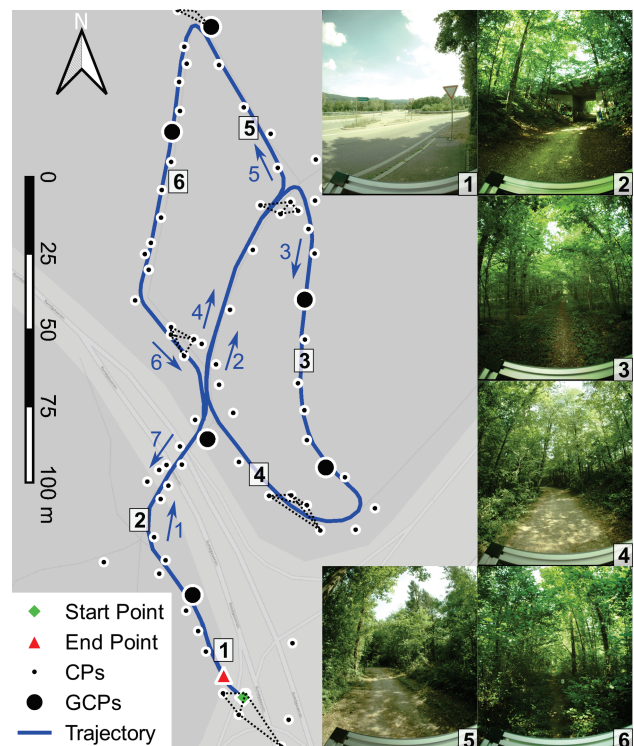


Figure 4. Map from the second test site located in the forest with images showing environmental conditions.

4.2 Forest

The second test site is situated in a forest in Münchenstein near Basel. The extent of the test site is approx. 100 x 200 m. This test site also incorporates an area with good GNSS reception at a highway exit for system initialization (see Figure 4, Image 1). Furthermore, the forest path leads through a road underpass (see Figure 4, Image 2). Narrow paths only accessible to pedestrians with dense vegetation at ground level dominate the scenery in images 3 and 6 of Figure 4. Because of restricted GNSS reception, poor and variable illumination as well as ambiguous, repetitive geometry, such forest scenarios represent a major challenge for all georeferencing methods. In addition, the test site also includes driveable forest roads with less dense vegetation (see Figure 4, Images 4 and 5). The test site includes 89 reference points, which were marked with photogrammetric targets and fixed either on trees or on driven-in pillars. Fricker, Weber (2019) describe the reference point measurements by tachymetry with closed polygons as well as the geodetic evaluation, which shows a 3D standard deviation of 5 mm.

5. PERFORMANCE INVESTIGATIONS

Firstly, we aimed at investigating the suitability of our backpack MMS in outdoor environments under real-world conditions. Secondly, we compared the different georeferencing methods in terms of accuracy and reliability in different environments. For this, we carried out data acquisition campaigns in the two different large-scale test sites. Furthermore, we aimed at providing meaningful statements on the accuracy potential by the comparison of image-based coordinate measurements with reference point coordinates. While the standard deviations of coordinate measurements indicate their precision, coordinate differences to the ground truth show the potential of absolute accuracy. Furthermore, we defined 3D distances with different lengths in order to evaluate the relative accuracy.

5.1 Data acquisition

In both test sites, we initialized our system in a location with good GNSS coverage at the beginning as well as at the end of a data acquisition campaign. In order to align the body frame with the local-level frame, we ran a few laps under avoidance of pivoting movements.

Test site	City centre	Forest
Targets (CPs, GCPs) [n]	79 (74, 5)	89 (83, 6)
Panorama images [n]	721	843
Single images [n]	4326	5052
Trajectory length [m]	800	740
Acquisition time [min]	24	25
Data volume [GB]	14	16

Table 2. Key figures relating to data acquisition for the performance investigations.

Since image storage requires somewhat less than one second, we aimed at an acquisition speed of one meter per second in order to get an image capturing interval of about one meter. In the city centre, the trajectory length was about 800 m and the acquisition took 24 minutes (see Table 2). The trajectory length in the forest was about 740 m and the data acquisition required 25 minutes. In total, we captured 4326 single images at 721 locations in the city and 5052 single images at 843 locations in the forest respectively. The LiDAR data formed the majority of

the resulting data volume of about 14 GB in the city centre and 16 GB in the forest.

5.2 Data processing and datum transformation

Using the recorded raw data from GNSS, IMU, LiDAR scanners and panorama camera, we determined the image poses with direct, SLAM-based, and image-based georeferencing as already introduced in chapter 3.2 (see Figure 2). For image-based georeferencing, we used the SLAM-based image poses as initial values for the bundle-adjustment.

In order to compare all coordinates within the same reference frame, we transformed the directly georeferenced and the SLAM-based image poses with a 6 DoF transformation using well distributed 5 GCPs in the city centre (see Figure 3) and 6 GCPs in the forest (see Figure 4). As described in Blaser et al. (2018 & 2019), we carried out the 6 DoF transformation using a self-developed Python program. Since the directly georeferenced image poses were already in a global reference frame, the transformation parameters only comprised mean offsets and rotations to the GCPs. By contrast, the transformation parameters of the SLAM-based image poses additionally contained the transition from the local to the global reference frame. For image-based georeferencing, we introduced the transformed SLAM-based image poses as initial values and measured the same GCPs as used for the 6 DoF transformation in four consecutive images for the subsequent bundle adjustment. This ensured that the image poses from image-based georeferencing were in the same reference frame as the transformed directly georeferenced image poses and the transformed SLAM-based image poses.

5.3 Coordinate and distance measurements

In our previous investigations in indoor and in underground environments (Blaser et al., 2018 & 2019), we estimated the 3D point coordinates using a forward intersection with image measurements in four consecutive images. Our self-developed bundle adjustment-based forward intersection Python program supports using the same image measurements for different image pose sources. In addition, the program also provides the standard deviation of the forward intersection, which represents the precision of a 3D point measurement. Furthermore, we defined 3D distances between known reference points with lengths from 0.03 to 21.08 m. They are distributed across different locations in both test areas (see dotted lines in Figure 3 and 4). We measured both the start and end of the 3D distance in different image sets.

6. RESULTS AND DISCUSSION

First, we evaluated the standard deviations of 3D coordinate measurements from the forward intersection. They represent the precisions of coordinate observations. Further, the precision is a good measure for the relative measuring accuracy within the same images. The distribution of precisions shows a significant number of outliers across all test sites and georeferencing methods (see Figure 5). In order to evaluate the potential in accuracy, we used the median, which is more robust against outliers than the mean or the RMSE. The median precision using directly georeferenced image poses amounted to 3.4 cm in the city centre and 2.3 cm in the forest. This is slightly lower than the median precision using SLAM-based image poses, which amounts to 5.2 cm in the city centre and 4.0 cm in the forest

(see Figure 5 and Table 3). Image-based georeferencing improves the median precision by a factor of 10-15 to 3 mm within both test sites. This confirms the major improvements that we also achieved in previous work (Blaser et al., 2018 & 2019).

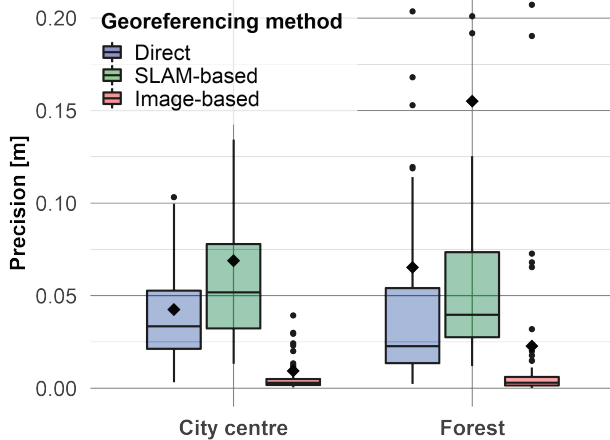


Figure 5. Boxplot with the precision of the 3D coordinate observations for the test sites “City centre” (left) and “Forest” (right). The blue boxplot on the left represents direct georeferencing, the green boxplot in the middle SLAM-based georeferencing, and the red boxplot on the right image-based georeferencing. The black diamond symbol indicates the mean precision.

In order to evaluate the absolute accuracy, we compared the 3D coordinates of measured points with the ground truth. The distribution of coordinate differences shows outliers across all test fields and georeferencing methods (see Figure 6). The median 3D coordinate deviation of direct georeferencing amounts to 45.2 cm in the city centre and 100.7 cm in the forest. The much poorer GNSS reception in the forest might cause the significant difference to the city centre. By contrast, the median 3D coordinate deviations of SLAM-based georeferencing with 36.6 cm in the city centre and 21.0 cm in the forest are in the same order of magnitude. However, the higher number of outliers show that SLAM-based georeferencing in the forest is less reliable than in the city centre. Image-based georeferencing significantly reduced the median 3D coordinate deviations in both test sites to 4.3 cm in the city centre and 13.4 cm in the forest (see Table 3).

Finally, we analysed the length deviations of 3D distances to the ground truth in order to evaluate the relative accuracy. Thereby, the predefined 3D distances varied in length from 3 cm to 23.1 m and we measured the start and the end of the distances in different image sets. Figure 7 depicts the length deviations of 3D distances related to the measured distances. By contrast, Figure 8 shows the length deviations using different georeferencing methods for both test sites.

All median values of length deviations using directly georeferenced poses as well as SLAM-based poses did not significantly differ. They are between 7.1 and 9.4 cm (see Table 4). However, a larger number of outliers occurred when using directly georeferenced poses. Poses from image-based georeferencing significantly improved the relative accuracy, so that the median of length deviations decreased to 1.9 and 2.0 cm.

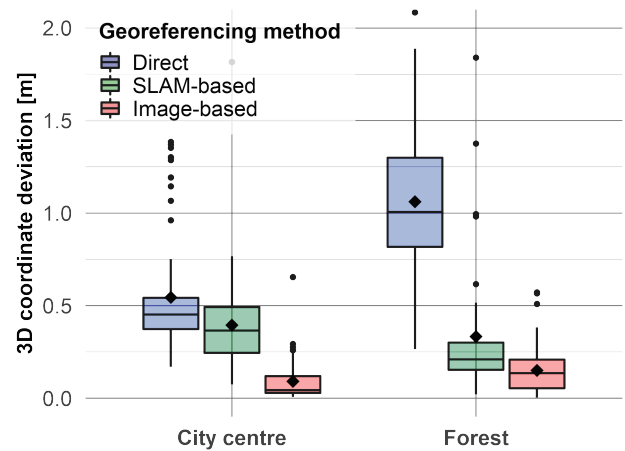


Figure 6. Boxplot with 3D coordinate deviations to the ground truth that represent the absolute accuracy. All samples are listed from left to right by test area (“City centre” and “Forest”) and georeferencing method (“Direct” (blue), “SLAM-based” (green) and “Image-based” (red)). The black diamond symbol indicates the mean 3D coordinate deviation.

Method	Test site	Pts. [n]	Precision [mm]		Accuracy [mm]	
			mean	median	mean	median
Direct	1)	79	42	34	545	452
	2)	89	65	23	1061	1007
SLAM-based	1)	79	69	52	395	366
	2)	89	155	40	333	210
Image-based	1)	79	9	3	91	43
	2)	89	23	3	150	134

Table 3. Summary of precisions and accuracies of 3D coordinate observations using different georeferencing methods (Direct, SLAM-based and Image-based) for both test sites 1) “City centre” and 2) “Forest”. The table contains both the mean and median precision and accuracy values. Precision represents the RMSE of forward intersection of a single point measurement and accuracy shows the 3D coordinate deviation to the ground truth.

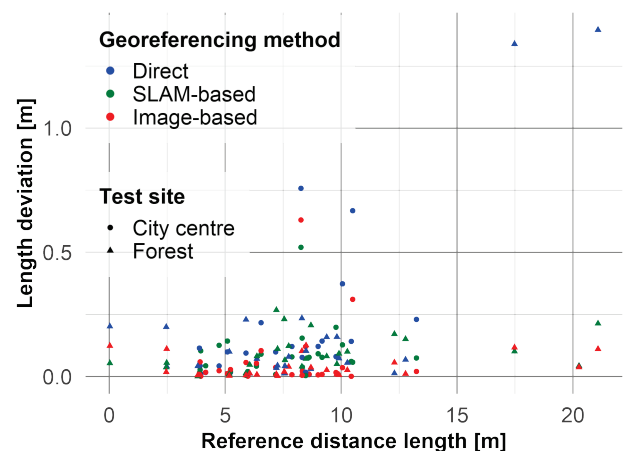


Figure 7. Scatter plot with length deviations between measured 3D distances and ground truth. The colours show the different georeferencing methods, while the point shape represents the test site.

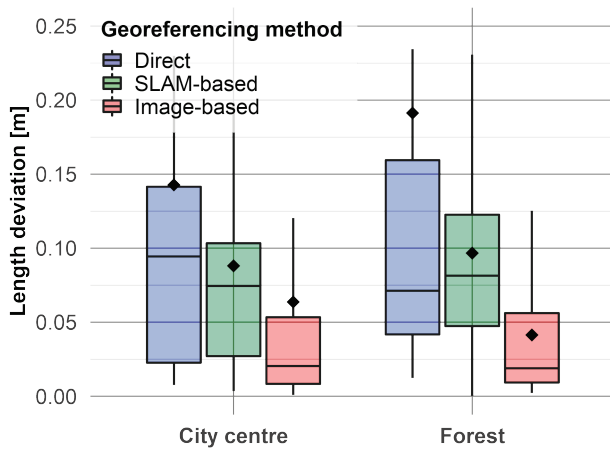


Figure 8. Boxplot with length deviations between measured 3D distances and ground truth representing the relative accuracy.

All samples are listed from left to right by test area (“City centre” and “Forest”) and georeferencing method (“Direct” (blue), “SLAM-based” (green) and “Image-based” (red)). The black diamond symbol indicates the mean 3D distance deviation.

Method	Test site	Distances [n]	Accuracy [mm]	
			mean	median
Direct	1)	25	143	94
	2)	25	191	71
SLAM-based	1)	25	88	74
	2)	25	97	81
Image-based	1)	25	64	20
	2)	25	41	19

Table 4. Summary of accuracies of 3D distance observations using different georeferencing methods (Direct, SLAM-based and Image-based) for both test sites: 1) “City centre” and 2) “Forest”. The table contains both the mean and median accuracy values. Accuracy represents the length deviation to the ground truth.

Our results confirm those of our previous work as well as of work from other groups. Lehtola et al. (2017) carried out investigations on pointclouds of indoor environments using comparable backpack MMS. Their deviations up to 14 cm and 55 cm in floor heights with the *Leica Pegasus Backpack* and the *Würzburg Backpack* respectively are in the same order of magnitude. However, their results are not directly comparable because of the different environmental conditions. Furthermore, they only concern the height component. By contrast, 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 40-72 cm and 4-45 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, the deviations of the SLAM-based georeferencing are comparable to ours. Although, our investigated SLAM-based georeferencing seems to be more robust to environmental changes. In addition, as proven in our previous work, subsequent image-based georeferencing significantly improves the accuracies by a multiple and the precisions by an order of

magnitude over published results with direct or SLAM-based georeferencing in similar environments. However, a closer fusion of all three georeferencing methods has great potential to further improve accuracy and robustness.

7. CONCLUSION AND OUTLOOK

In this contribution, we extended our image-based backpack MMS by direct georeferencing capabilities and we carried out performance investigations within two large-scale test sites. One test site is situated in a city centre and the other in a forest. Both test sites consist of areas with restricted GNSS reception, poor illumination and uniform or ambiguous geometry, which are challenging for any georeferencing method. With our investigations, we demonstrated the suitability of our research backpack MMS under real conditions. Further, we empirically compared the performance of direct, SLAM-based and image-based georeferencing in both test sites. We obtained median precisions of 3D coordinate measurements of 3 mm using image poses from image-based georeferencing. Thus, we achieved an improvement in precision from direct or SLAM-based georeferencing to image-based georeferencing by a factor of 10 to 15. These precision improvements offered by image-based georeferencing by about an order of magnitude confirm the findings of our previous investigations in indoor and underground environments.

The median coordinate differences of direct georeferencing were 45.2 cm in the city centre and 100.2 cm in the forest due to restricted GNSS reception. SLAM-based georeferencing with median coordinate differences of 36.6 cm and 21.0 cm performed slightly better. With image-based georeferencing, we achieved median coordinate differences of 4.3 cm in the city centre and 13.4 cm in the forest. This also corresponds to an accuracy improvement by a factor between 2 to 10. Finally, we investigated the relative accuracy by comparing measured 3D distances with the ground truth. The median deviations were between 1.9 cm and 2.0 cm using image poses from image-based georeferencing. The absolute accuracies as well as the relative accuracies are comparable to previous work and to image-based MMS with fixed stereo bases.

Our investigations show a great potential for complementing 3D image-based geospatial web-services of cities as well as for creating such web services for forest applications. In addition, accurately georeferenced imagery of urban environments has an enormous potential for future visual localization and augmented reality applications.

Nevertheless, we intend to improve further the robustness as well as the accuracy by combining different georeferencing methods. In addition, we will complete the overall system calibration of our backpack MMS and analytically determine pending lever arms and misalignment from LiDAR scanners to the panorama camera. Ongoing work also focuses on the robust generation of accurate 3D depth information, which will benefit from highly accurate image-based georeferencing. The additional depth layer will enable 3D measurement directly in the image with just one mouse click and will thus significantly enhance the user-friendliness of the cloud-based 3D services.

ACKNOWLEDGEMENTS

The research project cloudIO (32411.1 IP-ICT) was supported by Innosuisse, the Swiss Innovation Agency and by our industrial partner iNovitas AG.

REFERENCES

- Abraham, S., Förstner, W., 2005. Fish-eye-stereo calibration and epipolar rectification. *ISPRS J. Photogramm. Remote Sens.* 59(5), 278-288. doi.org/10.1016/j.isprsjprs.2005.03.001
- Blaser, S., Cavegn, S., Nebiker, S., 2018. Development of a Portable High Performance Mobile Mapping System Using the Robot Operating System. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-1, 13-20. doi.org/10.5194/isprs-annals-IV-1-13-2018
- Blaser, S., Nebiker, S., Cavegn, S., 2017. System Design, Calibration and Performance Analysis of a Novel 360° Stereo Panoramic Mobile Mapping System. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-1/W1, 207-213. doi.org/10.5194/isprs-annals-IV-1-W1-207-2017
- Blaser, S., Nebiker, S., Wisler, D., 2019. Portable Image-based High Performance Mobile Mapping System in Underground Environments – System Configuration and Performance Evaluation. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-2/W5, 255-262. doi.org/10.5194/isprs-annals-IV-2-W5-255-2019
- Burkhard, J., Cavegn, S., Barmettler, A., Nebiker, S., 2012. Stereovision Mobile Mapping: System Design and Performance Evaluation. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XXXIX-B5, 453-458. doi.org/10.5194/isprsarchives-XXXIX-B5-453-2012
- Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J.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.org/10.1109/TRO.2016.2624754
- Cavegn, S., Blaser, S., Nebiker, S., Haala, N., 2018. Robust and Accurate Image-based Georeferencing Exploiting Relative Orientation Constraints. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-2, 57-64. doi.org/10.5194/isprs-annals-IV-2-57-2018
- Durrant-Whyte, H., Bailey, T., 2006. Simultaneous Localization and Mapping: Part I. *IEEE Robotics & Automation Magazine*, 13(2), 99-110. doi.org/10.1109/MRA.2006.1638022
- Fricker, L., Weber, D., 2019. Portables Mobile Mapping im Aussenraum. Bachelor Thesis. FHNW University of Applied Sciences and Arts Northwestern Switzerland, Switzerland (unpublished).
- Hasler, O., Blaser, S., Nebiker, S., 2019. Implementation and First Evaluation of an Indoor Mapping Application Using Smartphones and AR Frameworks. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W17, 135-141. doi.org/10.5194/isprs-archives-XLII-2-W17-135-2019
- Hasler, O., Blaser, S., Nebiker, S., 2020. Performance Evaluation of a Mobile Mapping Application Using Smartphones and Augmented Reality Frameworks. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, Nice, France.
- Hess, W., Kohler, D., Rapp, H., Andor, D., 2016. Real-Time Loop Closure in 2D LiDAR SLAM. *IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden, 1271-1278. doi.org/10.1109/ICRA.2016.7487258
- Heuvel, F. A. Van Den, Verwaal, R., Beers, B., 2006. Calibration of Fisheye Camera Systems and the Reduction of Chromatic Aberration. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XXXVI Part 5, unpaginated CD-ROM.
- Jende, P. L. H., 2019. Position estimation of mobile mapping imaging sensors using aerial imagery. Doctor of Philosophy, Faculty of Geo-Information Science and Earth Observation, Enschede. doi.org/10.3990/1.9789036548847
- Kostoeva, R., Upadhyay, R., Sapar, Y., Zakhor, A., 2019. Indoor 3D Interactive Asset Detection Using a Smartphone. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W13, 811-817. doi.org/10.5194/isprs-archives-XLII-2-W13-811-2019
- Lehtola, V. V., Kaartinen, H., Nüchter, A., Kaijaluo, R., Kukko, A., Litkey, P., Honkavaara, E., Rosnell, T., Vaaja, M. T., Virtanen, J-P., Kurkela, M., Issaoui, A. El, Zhu, L., Jaakkola, A., Hyypä, J., 2017. Comparison of the Selected State-Of-The-Art 3D Indoor Scanning and Point Cloud Generation Methods, *Remote Sens.*, 9(8), 796. doi.org/10.3390/rs9080796
- Maboudi, M., Bánhid, D., Gerke, M., 2017. Evaluation of indoor mobile mapping systems. *Proceedings of the GFal Workshop 3D North East*, Berlin, 125-134.
- 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.org/10.1002/rob.21531
- Nebiker, S., 2017. 3D Imagery for Infrastructure Management – Mobile Mapping meets the Cloud. *The 56th Photogrammetric Week*, University of Stuttgart, Institute for Photogrammetry, Stuttgart, Germany.
- 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), 2267–2291. doi.org/10.3390/ijgi4042267
- Novak, K., 1991. The Ohio State University Highway Mapping System: The Stereo Vision System Component. *Proc. of the 47th Annual Meeting of the Institute of Navigation*, Williamsburg, VA, 121–124.
- Novatel Inc., 2020. SPAN CPT7. www.novatel.com/assets/Documents/Papers/SPAN-CPT7-PS.pdf (5 February 2020).

Nüchter, A., Borrmann, D., Koch, P., Kühn, M., May, S., 2015. A Man-Portable, Imu-Free Mobile Mapping System. *Int. Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, II-3/W5, 17–23. doi.org/10.5194/isprsannals-II-3-W5-17-2015

Puente, I., González-Jorge, H., Martínez-Sánchez, J., Arias, P., 2013. Review of mobile mapping and surveying technologies. *Measurement*, 46(7), 2127–2145. doi.org/10.1016/j.measurement.2013.03.006

Purvis, M., 2020. NovAtel Span Driver. wiki.ros.org/novatel_span_driver (5 February 2020).

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), 5.

Saran, V., Lin, J., Zakhori, A., 2019. Augmented Annotations: Indoor Dataset Generation with Augmented Reality. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W13, 873–879. doi.org/10.5194/isprs-archives-XLII-2-W13-873-2019

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. *Proceedings of the Vehicle Navigation and Information Systems Conference*, Ottawa, Canada, 476–481.

Tang, J., Chen, Y., Kukko, A., Kaartinen, H., Jaakola, A., Khoramshahi, E., Hakala, T., Hyypä, J., Holopainen, M., Hyypä, H., 2015. SLAM-Aided Stem Mapping for Forest Inventory with Small-Footprint Mobile LiDAR. *Forests*, 6(12), 4588–4606. doi.org/10.3390/f6124390

Tucci, G., Visintini, D., Bonora, V., Parisi, E.I., 2017. Examination of Indoor Mobile Mapping Systems in a Diversified Internal/External Test Field. *Appl. Sci.*, 8(3). doi.org/10.3390/app8030401

Vexcel Imaging, 2020. UltraCam Panther. www.vexcel-imaging.com/ultracam-panther/ (5 February 2020).

Viametris, 2020. bMS3D LD5+ Backpack Mobile Scanner. www.viametris.com/backpackmobilescannerbms3d (5 February 2020).

Wittmer, S., 2017. INS- und bildbasierte Messmethoden für bewaldete Gebiete. Master Thesis, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Switzerland (unpublished).