

ROBUST AND ACCURATE IMAGE-BASED GEOREFERENCING EXPLOITING RELATIVE ORIENTATION CONSTRAINTS

S. Cavegn^{1,2}, S. Blaser¹, S. Nebiker¹, N. Haala²

¹ Institute of Geomatics, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Muttensz, Switzerland
- (stefan.cavegn, stefan.blaser, stephan.nebiker)@fhnw.ch

² Institute for Photogrammetry, University of Stuttgart, Germany - norbert.haala@ifp.uni-stuttgart.de

Commission II, WG II/1

KEY WORDS: Sensor Orientation, Georeferencing, Bundle Adjustment, Relative Orientation, Constraints, Mobile Mapping, Indoor

ABSTRACT:

Urban environments with extended areas of poor GNSS coverage as well as indoor spaces that often rely on real-time SLAM algorithms for camera pose estimation require sophisticated georeferencing in order to fulfill our high requirements of a few centimeters for absolute 3D point measurement accuracies. Since we focus on image-based mobile mapping, we extended the structure-from-motion pipeline COLMAP with georeferencing capabilities by integrating exterior orientation parameters from direct sensor orientation or SLAM as well as ground control points into bundle adjustment. Furthermore, we exploit constraints for relative orientation parameters among all cameras in bundle adjustment, which leads to a significant robustness and accuracy increase especially by incorporating highly redundant multi-view image sequences. We evaluated our integrated georeferencing approach on two data sets, one captured outdoors by a vehicle-based multi-stereo mobile mapping system and the other captured indoors by a portable panoramic mobile mapping system. We obtained mean RMSE values for check point residuals between image-based georeferencing and tachymetry of 2 cm in an indoor area, and 3 cm in an urban environment where the measurement distances are a multiple compared to indoors. Moreover, in comparison to a solely image-based procedure, our integrated georeferencing approach showed a consistent accuracy increase by a factor of 2-3 at our outdoor test site. Due to pre-calibrated relative orientation parameters, images of all camera heads were oriented correctly in our challenging indoor environment. By performing self-calibration of relative orientation parameters among respective cameras of our vehicle-based mobile mapping system, remaining inaccuracies from suboptimal test field calibration were successfully compensated.

1. INTRODUCTION

Image-based mobile mapping systems featuring multi-view stereo camera configurations enable efficient data acquisition, for both outdoor (Cavegn & Haala, 2016; Blaser et al., 2017) and indoor environments (Holdener et al., 2017). In order to obtain accurate geospatial 3D image spaces consisting of collections of georeferenced multi-view RGB-D imagery (Nebiker et al., 2015; Nebiker, 2017), which can be exploited for 3D mono-plotting applications as well as for 3D point cloud and mesh generation, depth maps of high quality need to be computed. These depth maps are preferably generated by performing multi-view stereo matching using imagery captured at different epochs (Cavegn et al., 2015). In order to efficiently apply coplanarity constraints during dense stereo matching, sub-pixel accurate relative orientations of the image sequences are required. Since Cavegn et al. (2016) revealed trajectory discontinuities from direct georeferencing of up to 15 cm in urban environments, this can only be achieved by image-based georeferencing. This furthermore allows the elimination of trajectory offsets in the range of several decimeters leading to consistent image sequences, which might be captured at different days and daytimes.

Eugster et al. (2012) feed stereovision-based position updates back into a Kalman filter and thus achieve a consistent improvement of absolute 3D measurement accuracies from several decimeters to 5-10 cm for land-based mobile mapping. Since airborne surveys are much less affected by GNSS degradations experienced by ground-based mobile mapping systems (Haala et al., 2008; Puente et al., 2013), Nebiker et al. (2012) proposed the fusion of ground-based imagery from mobile

mapping systems with aerial imagery. Similarly, Jende et al. (2017) perform feature-based registration of mobile mapping and aerial nadir images in order to adjust the mobile platform locations in GNSS-denied urban environments. Integrated georeferencing is also necessary indoors due to the limited accuracy of online SLAM, which is the predominant pose estimation procedure for indoor mapping systems (Lehtola et al., 2017; Nocerino et al., 2017).

Rumpler et al. (2017) developed a complete processing pipeline from image capturing to mesh generation. They consider weighted ground control points (GCP) and GNSS positions in the final bundle adjustment. Nonetheless, they do not provide support for constraining relative orientation parameters (ROP) between cameras in a multi-camera setup, which is also the case for the approach of Toldo et al. (2015) as well as for the automated image orientation packages evaluated in Remondino et al. (2012). Moreover, oblique images from the individual camera heads are usually treated independently in aerial triangulation (Cavegn et al., 2014; Rupnik et al., 2015; Karel et al., 2016). However, Sun et al. (2016) parameterize oblique camera poses with nadir camera poses as well as constant relative poses between oblique and nadir cameras, so that the number of unknown parameters and the dimension of the normal equations decrease, which dramatically reduces the computational complexity and memory cost. Relative observations that relate the position and attitude parameters of two consecutive epochs are exploited by Rehak (2017) as well as by Schönberger et al. (2014) in the UAV case. Klingner et al. (2013) optimize trajectories from Google Street View cars by leveraging precise relative poses from IMU. Kersting et al. (2012) introduce a single-step calibration method for multi-camera mobile mapping

systems, which has the ability of estimating two sets of ROP, namely the lever arm offsets and the boresight angles relating the cameras and the IMU body frame as well as the ROP among the cameras. Pix4D is the first commercial photogrammetric software enabling multi-camera rig processing including ROP self-calibration. Exploiting ROP constraints coupled with the incorporation of GCP as well as prior exterior orientation parameters (EOP) for highly redundant multi-view image sequences into bundle adjustment should lead to a significant increase of the georeferencing quality in terms of robustness and accuracy.

Our main contributions in this paper are as follows:

- Extension of the structure-from-motion pipeline COLMAP with georeferencing capabilities by integrating exterior orientation parameters from direct sensor orientation or SLAM as well as ground control points into bundle adjustment in order to meet our high accuracy requirements of a few centimeters
- Exploitation of constraints for relative orientation parameters among all cameras in bundle adjustment
- Evaluation of our integrated georeferencing approach on two data sets, one captured outdoors by a vehicle-based multi-stereo mobile mapping system and the other captured by a portable panoramic mobile mapping system in an indoor environment

2. INTEGRATED GEOREFERENCING APPROACH BASED ON COLMAP

COLMAP is a powerful incremental structure-from-motion (SfM) tool, which is mainly being developed by Johannes Schönberger (Schönberger & Frahm, 2016). It starts with feature extraction and matching, followed by geometric verification. The resulting scene graph serves as the foundation for the reconstruction stage, which seeds the model with a carefully selected two-view reconstruction, before incrementally registering new images, triangulating scene points, filtering outliers, and refining the reconstruction using bundle adjustment. We extended COLMAP for the purpose of integrated georeferencing (see Figure 1), thus assuming initial values for exterior orientation parameters with a reasonable accuracy from direct sensor orientation or Simultaneous Localization and Mapping (SLAM) (Cadena et al., 2016). Furthermore, our data usually features a limited number of ground control points in a global geodetic reference system.

Using our Python framework, which allows automated processing of all COLMAP modules, we first reduce prior projection centers as well as GCP coordinates to the EOP centroid, i.e. 3D coordinate translation into a local geodetic reference system. Since rotations are often given as Euler angles, we then transform these to quaternions, and store the computed local poses in a COLMAP database. Correspondence search and

global reconstruction result in refined local poses, which can be transformed back to the initial global geodetic reference system.

2.1 Correspondence Search

Schönberger et al. (2017) show that DSP-SIFT performs better than SIFT. As more features are extracted at the cost of longer computation times, DSP-SIFT is especially beneficial in environments with poor texture.

Since we rely on prior EOP, we use the spatial feature matcher implemented in COLMAP, which only considers camera positions closer than a given maximum radius from the current image for search space reduction. Moreover, we added a maximum angle constraint in order to further speed up the process as feature matching is the most time consuming step in the COLMAP procedure.

Geometric verification of potentially overlapping image pairs is performed as described in Schönberger & Frahm (2016). SfM verifies the matches by trying to estimate a transformation that maps feature points between images using projective geometry. If a valid transformation maps a sufficient number of features between the images, they are considered as geometrically verified. Since the correspondences from matching are often outlier-contaminated, robust estimation techniques, such as RANSAC, are then required. The output is a scene graph with images as nodes and verified pairs of images as edges.

2.2 Global Reconstruction

First of all, we triangulate 2D image features to natural 3D points incorporating all available images based on prior EOP, followed by bundle adjustment (Triggs et al., 2000) which is the joint refinement of camera parameters and point parameters in a non-linear optimization. In our case, bundle adjustment (BA) is carried out based on Google's Ceres Solver library for non-linear least squares problems (Agarwal et al., 2018). As in Rumpler et al. (2017), our BA procedure (see Equation 1) not only minimizes the reprojection errors between the projected natural 3D points as well as ground control points and its corresponding 2D measurements in image space (see Equation 2), but it also minimizes differences of 3D projection center coordinates from direct georeferencing and photogrammetric reconstruction (see Equation 3). Furthermore, we use the robust Cauchy function to potentially down-weight outliers.

Optimization problem:

$$f^* = \min \sum E(P) + \sum E^{gcp}(R) + \sum E^{dg}(S) \quad (1)$$

where E = error function
 P = natural 3D points
 R = reference / ground control points
 S = projection centers (from direct georeferencing)

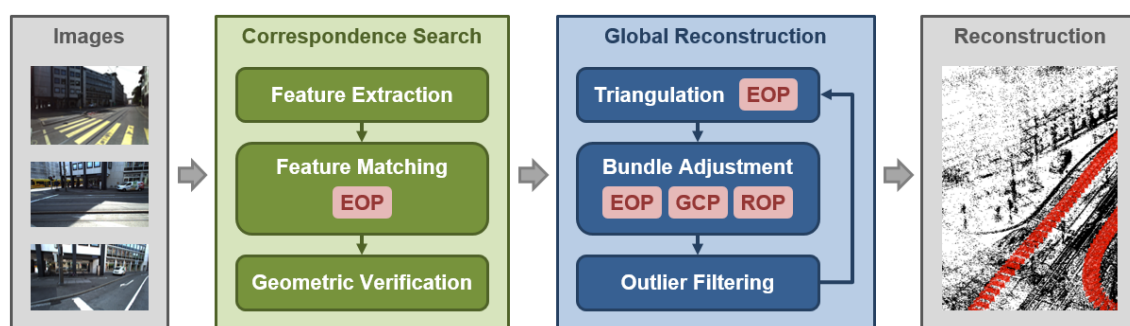


Figure 1. Adapted processing pipeline of COLMAP based on prior exterior orientation parameters (EOP)

Error function for 3D points (natural and GCP):

$$E(X) = \sum_{x^i \in X_P} \rho(C^P(\Gamma_i(X), x^i)) \quad (2)$$

where ρ = loss function (robust Cauchy function)
 C^P = 2D reprojection error / Euclidean distance in 2D
 $\Gamma(X)$ = projected 3D point into image
 x = observed 2D measurement

Error function for projection centers:

$$E(C) = \sum_{M \in S} \rho(C^{dg}(M, C)) \quad (3)$$

where ρ = loss function (robust Cauchy function)
 C^{dg} = Euclidean distance in 3D
 M = projection center from direct georeferencing
 C = reconstructed projection center

Moreover, we constrain calibrated relative orientation parameters or define constraints for ROP among cameras (e.g. constant ROP for base over all image sequences if self-calibrated) in bundle adjustment.

Subsequently, COLMAP completes 3D point tracks, merges 3D points that are very close to each other, filters and re-triangulates observations before performing a new bundle adjustment computation. This iterative process is continued until convergence is reached.

3. INVESTIGATIONS IN AN OUTDOOR ENVIRONMENT

In order to compare the performance of our integrated georeferencing approach based on COLMAP with previous results, we carried out some investigations at our test site in Basel.

3.1 Vehicle-Based Mobile Mapping System

All data used for our investigations in an urban environment was captured by the vehicle-based multi-sensor stereovision mobile mapping system of the Institute of Geomatics (IGEO), University of Applied Sciences and Arts Northwestern Switzerland (FHNW). The multi-sensor system features several industrial stereo cameras with CCD sensors as well as a GNSS/IMU positioning system. They are mounted on a rigid platform to ensure stability of relative orientations and offsets between all sensors, which are synchronized by hardware trigger signals from a custom-built trigger box.

Two campaigns with different sensor configurations were performed, the first in July 2014 and the second in August 2015. In both cases, the main stereovision system facing forward consists of two 11 MP cameras and a calibrated stereo base of 905 mm. These stereo cameras have a resolution of 4008 x 2672 pixels at a pixel size of 9 μ m, a focal length of 21 mm and a resulting field-of-view of 81° in horizontal and 60° in vertical direction. For the campaign in July 2014, the multi-camera configuration is completed by two additional stereovision systems comprising HD cameras with a resolution of 1920 x 1080 pixels, a pixel size of 7.4 μ m, a focal length of 8 mm and a field of view of 83° x 53° (see Figure 2). While the stereo cameras pointing back-right have a base of 779 mm, the stereo cameras facing left are separated by 949 mm.

We used a NovAtel SPAN inertial navigation system for direct georeferencing of the imagery acquired at typically 5 fps. The navigation system consists of a tactical grade inertial measurement unit featuring fiber-optics gyros of the type UIMU-LCI and a L1/L2 GNSS kinematic antenna. In case of good

GNSS coverage, these sensors provide an accuracy of horizontally 10 mm and vertically 15 mm during post-processing. Accuracies of the attitude angles roll and pitch are specified with 0.005° and heading with 0.008°. A GNSS outage of 60 seconds degrades the horizontal accuracy to 110 mm and the vertical to 30 mm.



Figure 2. Sensor configuration of our vehicle-based mobile mapping system for the campaign in Basel in July 2014 which captured image sequences 1.0, 2.0 and 3.0

3.2 Test Site and Data

The study area depicted in Figure 3 is located at a very busy junction of five roads in the city center of Basel, Switzerland. It includes large and rather tall commercial properties, which create a very challenging environment for GNSS positioning. Three street sections of this test site were mapped three times, once in July 2014 and twice during a day in August 2015 (Cavegn et al., 2016). In all nine cases data acquisition was performed shortly before noon and at good weather conditions. For our investigations we used 85 up to 191 stereo image pairs per sequence from the forward facing stereovision system on a sequence length between 108 m and 217 m as well as images from the back-right and left stereovision systems for the campaign in July 2014 (see Figure 4). An along-track distance between successive image exposures of 1 m was targeted, but larger distances occurred at velocities higher than 18 km/h since the maximum frame rate was 5 fps. While the campaign in July 2014 was part of a complete survey of the city-state of Basel, the campaign in August 2015 was specifically performed for the investigations at our study area.

We determined 3D coordinates of more than 50 points mainly on corners of road markings in a global geodetic reference system by tachymetry. They served either as ground control or check points and have an absolute 3D accuracy of better than 1 cm.

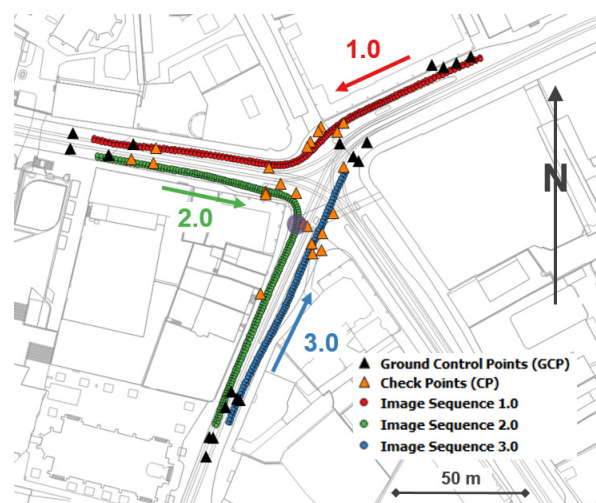


Figure 3. Base map of the study area with overlaid projection centers of selected stereo image sequences, ground control points, check points and vehicle location at the time of capturing the images of Figure 4 (purple filled circle) (Source: Geodaten Kanton Basel-Stadt)

We calibrated all sensors mounted on the rigid frame of the vehicle-based mobile mapping system in an extensive and rigorous process. First, we determined interior as well as relative orientation parameters between all cameras by constrained bundle adjustment exploiting imagery taken in different indoor calibration fields for the two campaigns. The precisely calibrated interior orientation parameters of all cameras served for distortion and principal point correction of the stereo imagery (see Figure 4). Second, we computed lever arm and misalignment to the left camera of the forward looking stereo system using imagery which was captured on our outdoor calibration field (Burkhard et al., 2012).

We processed navigation data in tightly coupled mode using the GNSS and inertial post-processing software Inertial Explorer from NovAtel. Furthermore, we performed processing in multi-pass directions and additionally smoothed trajectories. By incorporating the previously computed boresight alignment as well as the relative orientation parameters, we calculated directly georeferenced sensor orientations for all images.



Figure 4. Forward stereo (top), back-right stereo (middle) and left stereo images (bottom) captured at the same location (see Figure 3) during the campaign in Basel in July 2014

3.3 Integrated Georeferencing

Same as for the PhotoScan projects documented in Cavegn et al. (2016), we selected two, three or four GCP at both ends of each image sequence (see Figure 3). Subsequently, we measured sensor coordinates for each GCP in several stereo image pairs of the forward facing cameras using a Python tool. Furthermore, we defined the ROP configuration and its corresponding calibrated values in a JSON file as expected from COLMAP, where the left camera of the forward pointing stereovision system serves as master camera. Due to a moderately precise calibration, bundle adjustment with fixed ROP resulted in large check point residuals. Therefore, we estimated the ROP among respective cameras over all stereo images for the final computation, which corresponds to a self-calibration of stereo bases as well as offsets and rotations between the stereovision systems.

First, we processed images from the forward facing stereo system (single stereo) incorporating previously computed EOP from direct sensor orientation for the seven image sequences listed in Table 2. We mainly used the COLMAP standard parameters so that SIFT features were extracted. However, we did not refine interior orientation parameters. Moreover, we defined a

maximum distance of 10 m and a value of 50 for maximum neighbors but no maximum angle constraint in spatial feature matching.

Second, we performed integrated georeferencing exploiting all images from the three stereovision systems (multi-stereo) captured in July 2014 in the same process (see Figure 5) resulting in a mean reprojection error of 0.51 pixel (see Table 1). Differently to single stereo, we defined a value of 200 for maximum neighbors and a maximum angle constraint of 100 degrees in spatial feature matching, which would even allow matching of images from the forward and left pointing systems.

	Single stereo (mean of 7 proc.)	Multi-stereo (1 processing)
Registered images	298	2190
3D points	118'026	1'038'228
Observations	929'986	6'972'788
Mean track length	7.9	6.7
Mean obs. per image	3152	3184
Mean reproj. error [px]	0.75	0.51

Table 1. COLMAP processing statistics for single stereo and multi-stereo image sequences

We performed all experiments on a Linux laptop with an Intel Xeon E3-1535M 8-Core processor (2.9 GHz), 32 GB RAM and a Nvidia Quadro M2000M graphics card. In order to give an idea, the total processing time of the seven single stereo sequences was 9.9 minutes for feature extraction, 116.1 minutes for feature matching and 26.2 minutes for bundle adjustment.

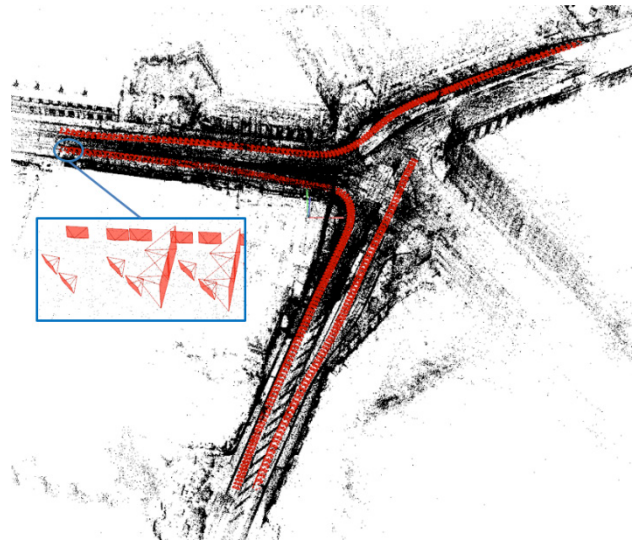


Figure 5. Georeferenced mobile mapping images (red) and 3D tie points (black) at our outdoor test site using our modified COLMAP processing pipeline

3.4 Check Point Investigations

We computed check point residuals to tachymetry in order to compare our integrated georeferencing approach with the solely image-based procedure of PhotoScan (Cavegn et al., 2016) and to verify the assumed advantages of using multi-stereo camera systems. For that, we determined 3D coordinates of check points by image measurements in single stereo pairs of the forward looking cameras and by incorporating EOP from the previous single stereo as well as multi-stereo processing. For all sequences but one, check points with significant deviations needed to be classified as outliers (see Table 2). Reasons are measurements

close to image borders as well as large distances between image positions and check points, difficult check point identification and 3D coordinate computation based on just two image measurements. Probably since measured in multiple images, none of the deviations to check point coordinates determined by PhotoScan was significant.

Over all forward image sequences, a mean RMSE value of 182 mm was computed for PhotoScan check point residuals, a value of 91 mm for all COLMAP residuals and a value of 58 mm for COLMAP residuals without outliers, which is an improvement by a factor of 2-3. Table 2 reveals that all non-outlier contaminated RMSE values of COLMAP are smaller than the PhotoScan values. Sequences 2.1 and 2.2 show for both PhotoScan and COLMAP the largest RMSE values.

For the campaign performed in July 2014 (image sequences 1.0, 2.0 and 3.0), we computed a mean RMSE value of 39 mm for single stereo and a value of 28 mm for multi-stereo (forward stereo, back-right stereo and left stereo). There is no significant degradation for multi-stereo values compared to single stereo and sequence 2.0, which has a sharp curve, shows an accuracy increase by a factor of more than two.

Se- quence	# CP no outliers (all)	Photo- Scan single stereo $\Delta 3D$ [mm]	COL- MAP single stereo $\Delta 3D$ [mm]	COL- MAP multi- stereo $\Delta 3D$ [mm]
1.0	14 (15)	137	29 (54)	30 (66)
1.1	11 (11)	42	30 (30)	
1.2	10 (11)	121	41 (78)	
2.0	10 (11)	76	51 (58)	22 (46)
2.1	11 (12)	432	90 (152)	
2.2	10 (11)	425	131 (148)	
3.0	6 (8)	42	36 (119)	31 (117)

Table 2. RMSE values for check point (CP) residuals between image-based georeferencing and tachymetry (results from all observations, including outliers are shown in brackets)

The obtained mean RMSE values for check point residuals of 58 mm for all forward image sequences and 28 mm for multi-stereo sequences lie in the range of the GCP residuals. When processing image sequences of trajectories with sharp curves, 3D points cannot be tracked in so many images as in nearly straight segments, hence deteriorating the accuracy. Especially in such cases, multi-stereo images from opposite driving directions are beneficial since even images from the back-right and left stereo systems can be matched leading to a more precise co-registration. Furthermore, this reduces the required number of GCP to a minimum.

4. INVESTIGATIONS IN AN INDOOR ENVIRONMENT

In the following sections, we show that our integrated georeferencing approach based on COLMAP is also able to successfully process a challenging data set captured by our portable mobile mapping system in an indoor environment.

4.1 Portable Mobile Mapping System

The portable panoramic mobile mapping system of the Institute of Geomatics (IGEO), University of Applied Sciences and Arts Northwestern Switzerland (FHNW) consists of multiple sensors mounted on a rigid frame, which is attached to a backpack (see Figure 6). A multi-head 360° panorama camera of the type FLIR

Ladybug5 serves for image capturing. Each of the six camera heads of the Ladybug5 camera has a resolution of 2448 x 2048 pixels (5 MP) at a pixel size of 3.45 μm , a focal length of 4.3 mm and a field-of-view of about 113° x 94°. The labeling of the five horizontally arranged camera heads is depicted in Figure 7, starting with cam0 facing backward and increasing in clockwise direction. The panorama camera is tilted by a few degrees in Figure 6, but the arrangement of the five consecutive camera heads turns almost horizontal in the case the backpack is carried by a person. The Velodyne LiDAR PUCK (VLP-16) mounted on top of the frame has a 360° horizontal field-of-view and a 30° vertical field-of-view, with $\pm 15^\circ$ up and down. It supports 16 channels, captures ca. 300'000 points per second, has a measurement range of up to 100 m and a typical accuracy of ± 3 cm. It is tilted by approx. 30 degrees in order to map points on floors as well as on ceilings, which improves the performance of LiDAR SLAM. The MEMS based Xsens IMU of the type MTi-300 further supports 3D LiDAR SLAM. For dynamic use, the accuracy of the attitude angle roll is specified with 0.3°, pitch and heading with 1.0°. In addition, there is an on-board computer, a synchronization box and a battery for power supply as well as four LED strips for illumination on the backpack. Further information about our portable mobile mapping system is given in Blaser et al. (2018).

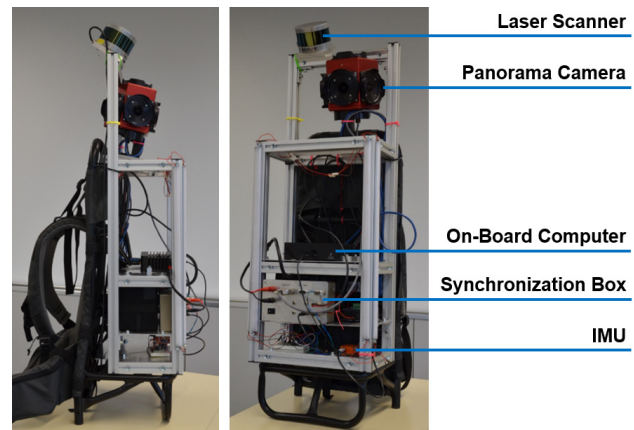


Figure 6. Sensor configuration of our portable panoramic mobile mapping system

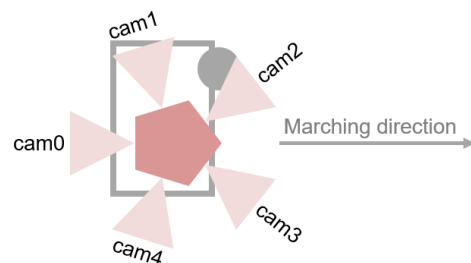


Figure 7. Top view of our backpack with camera head naming

4.2 Test Site and Data

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 Muttens close to Basel. As depicted in Figure 8, it features a hallway that has a dimension of ca. 27 m x 24 m, and leads to several offices, computer rooms, lecture rooms, further rooms, two staircases as well as five elevators. The typical corridor width amounts to approx. 3 m, but is also 2 m (left part, no. 6 in Figure 8) or ca. 4 m (right part, no. 3 in Figure 8).

We started the mapping at the origin of the local geodetic coordinate system marked with a dark gray diamond on 27.11.2017 17:17 and it took 20 minutes. The panorama camera captured images whenever reaching an along-track spacing of 1 m or an azimuth change of 15 degrees (see trajectory of cam0 in Figure 8). During the complete mapping process, online 3D LiDAR SLAM using the Google Cartographer (Hess et al., 2016) based on laserscanner and IMU data was performed. The real-time computation served for user guidance as well as for image triggering. The resulting SLAM EOP were used in the subsequent integrated georeferencing step as priors.

We determined the 3D coordinates of many 3D points representing e.g. natural markings on door frames or elevator corners (see Figure 9) by tachymetry. They have an accuracy of approximately 5 mm and we used several of them for our indoor investigations as ground control points or check points.

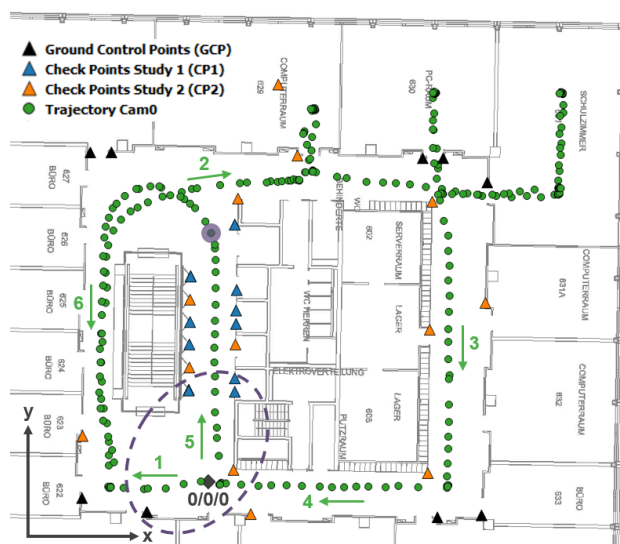


Figure 8. Floor base map with overlaid projection centers of camera head cam0, ground control points, check points, local geodetic coordinate system, area covered by Figure 9 (purple dashed line) and MMS location at the time of capturing the images of Figure 10 (purple filled circle)

We previously calibrated the interior orientation parameters (IOP) of all camera heads as well as the relative orientation parameters among these in an indoor calibration field. Hence, we utilized these IOP in order to correct our captured fisheye images for principal point and distortions to the equidistant camera model (Blaser et al., 2017). Figure 10 not only shows the mapping area at a specific camera position, but also gives an impression of the difficult lighting conditions and poor texture.

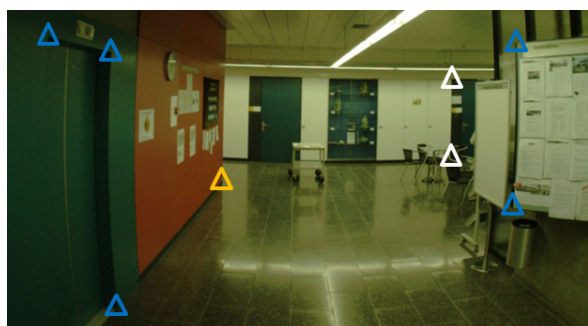


Figure 9. Image section from camera head cam0 with marked ground control points (white) as well as check points (blue and orange)



Figure 10. Images captured at the same location (see Figure 8) from the upward facing camera head cam5 (bottom right), backward pointing camera head cam0 (top left) and the consecutive camera heads cam1 (top middle), cam2 (top right), cam3 (bottom left), cam4 (bottom middle)

4.3 Integrated Georeferencing

We chose three GCP in each corner and measured sensor coordinates for each of these 12 GCP in three consecutive images of camera head cam2 using a Python tool. Where only two GCP per group are visible in Figure 8, two of them lie at the same 2D position but on two different height levels (see Figure 9). In addition, 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 270 locations (see Figure 11). COLMAP exploited previously computed EOP from online LiDAR SLAM, used the simple radial fisheye camera model instead of the simple pinhole camera model as for the outdoor data and performed no refinement of the interior orientation parameters.



Figure 11. Georeferenced mobile mapping images (red) and 3D tie points (black) at our indoor test site using our modified COLMAP processing pipeline

COLMAP extracted DSP-SIFT features and carried out spatial feature matching with a maximum distance of 10 m, 200 maximum neighbors, a maximum angle constraint of 100 degrees and a maximum ratio of 0.9. This resulted in 1350 registered images, 88'955 points, 876'096 observations, a mean track length of 9.8, 649 mean observations per image and a mean reprojection error of 0.80 pixel. Our Linux laptop needed 40.2 minutes for feature extraction, 131.5 minutes for feature matching and 29.6 minutes for bundle adjustment. The considerably longer feature extraction time compared to outdoors is due to DSP-SIFT instead of SIFT, which led to a larger number of features. However, these features and the resulting feature correspondences are not necessarily better distributed all-over the images.

4.4 Check Point Investigations

We performed two studies, both evaluating the absolute 3D measurement accuracy of our portable mobile mapping system. Study 1 aimed at assessing the calibrated relative orientation parameters among all camera heads. Hence, we selected six blue check points close to the elevators at three locations on two different height levels for each camera head (see Figure 8 and Figure 9). Then, we determined 3D check point coordinates by image measurements in four consecutive images using a self-developed Python tool supporting the equidistant camera model and computed residuals to tachymetry per camera head (see Table 3). The RMSE values for 3D check point residuals vary from 15 to 20 mm with a mean value of 17 mm. Thus, measurements for 3D point determination can be performed in arbitrary camera heads without accuracy degradation that is of high practical relevance.

Camera head	# CP	Δx [mm]	Δy [mm]	Δz [mm]	$\Delta 3D$ [mm]
cam0	6	10	11	8	17
cam1	6	12	7	9	17
cam2	6	10	6	8	15
cam3	6	8	9	12	16
cam4	6	10	13	11	20

Table 3. RMSE values of study 1 for check point residuals between image-based georeferencing and tachymetry

For study 2, we determined the 3D coordinates of the 13 orange check points distributed all-over the hallway by image measurements in four consecutive images of camera head cam2 or cam3 using our Python tool (see Figure 8). The resulting RMSE value for 3D check point residuals between image-based georeferencing and tachymetry amounts to 22 mm (see Table 4), which is not significantly larger than the values achieved in study 1. To sum up, based on arbitrary images captured by our portable mobile mapping system that are processed using our integrated georeferencing approach, absolute 3D point coordinates can be computed with an accuracy of approx. 2 cm.

Camera head	# CP	Δx [mm]	Δy [mm]	Δz [mm]	$\Delta 3D$ [mm]
2 and 3	13	10	14	12	22

Table 4. RMSE values of study 2 for check point residuals between image-based georeferencing and tachymetry

5. CONCLUSIONS AND OUTLOOK

Our integrated georeferencing approach based on COLMAP incorporating priors from SLAM or direct sensor orientation as well as a limited number of ground control points and exploiting

relative orientation constraints among cameras enables absolute 3D natural point accuracies in the order of 2 cm indoors and 3 cm outdoors. However, distances between the respective mobile mapping system and check points are considerably smaller indoors with ca. 2 m whereas about 7 m outdoors. Compared to the solely image-based procedure of PhotoScan, our integrated georeferencing approach led to an accuracy increase by a factor of 2-3 at the outdoor test site in Basel. Due to pre-calibrated ROP values, images of all camera heads were oriented correctly in our indoor environment, even though they hardly overlap as well as they mainly contain homogenous surfaces or repetitive patterns. By performing self-calibration of ROP among respective cameras of our vehicle-based mobile mapping system, remaining inaccuracies from suboptimal test field calibration were successfully compensated.

Future work will comprise the modification of our ROP self-calibration procedure so that just individual offsets or rotations can be estimated and the other ROP components can be fixed with pre-calibrated values (e.g. only adjustment of rotations between stereovision systems). Furthermore, a workflow for processing multiple large image sequences, e.g. survey of a complete city, needs to be developed. By incorporation of multi-view stereo images from opposite driving directions, all images will be co-registered precisely, thus reducing the required number of GCP to a minimum.

ACKNOWLEDGEMENTS

Thanks are due to iNovitas AG (Baden-Dättwil, Switzerland) and to the City of Basel (Bau- und Verkehrsdepartement Kanton Basel-Stadt) for providing the vehicle-based mobile mapping data of the campaign performed in July 2014. This work was co-funded by the Swiss Innovation Agency (Innosuisse, formerly CTI) as part of the BIMAGE project (No. 18493.2 PFES-ES).

REFERENCES

- Agarwal, S., Mierle, K. & Others, 2018. Ceres Solver. <http://ceres-solver.org> (4 April 2018)
- 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., Cavegn, S. & Nebiker, S., 2018. Development of a Portable High Performance Mobile Mapping System Using the Robot Operating System. In: *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, Karlsruhe, Germany (submitted).
- 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.
- Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I.D. & Leonard, J.J., 2016. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, 32(6), pp. 1309-1332.
- Cavegn, S., Haala, N., Nebiker, S., Rothermel, M. & Tutzauer, P., 2014. Benchmarking High Density Image Matching for Oblique Airborne Imagery. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Zurich, Switzerland, Vol. XL-3, pp. 45-52.

- Cavegn, S., Haala, N., Nebiker, S., Rothermel, M. & Zwölfer, T., 2015. Evaluation of Matching Strategies for Image-Based Mobile Mapping. In: *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, La Grande Motte, France, Vol. II-3/W5, pp. 361-368.
- Cavegn, S., Nebiker, S. & Haala, N., 2016. A Systematic Comparison of Direct and Image-Based Georeferencing in Challenging Urban Areas. In: *Int. Arch. 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. *Photogrammetric Engineering & Remote Sensing*, 82(12), pp. 925-933.
- 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. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Melbourne, Australia, Vol. XXXIX, Part B1, pp. 309-314.
- Haala, N., Peter, M., Kremer, J. & Hunter, G., 2008. Mobile LiDAR Mapping for 3D Point Cloud Collection in Urban Areas - A Performance Test. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Beijing, China, Vol. XXXVII, Part B5, pp. 1119-1124.
- 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, pp. 1271-1278.
- Holdener, D., Nebiker, S. & Blaser, S., 2017. Design and Implementation of a Novel Portable 360° Stereo Camera System with Low-Cost Action Cameras. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Hamburg, Germany, Vol. XLII-2/W8, pp. 105-110.
- Jende, P., Nex, F., Gerke, M. & Vosselman, G., 2017. Fully Automatic Feature-Based Registration of Mobile Mapping and Aerial Nadir Images for Enabling the Adjustment of Mobile Platform Locations in GNSS-Denied Urban Environments. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Hannover, Germany, Vol. XLII-1/W1, pp. 317-323.
- Karel, W., Ressel, C. & Pfeifer, N., 2016. Efficient Orientation and Calibration of Large Aerial Blocks of Multi-Camera Platforms. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Prague, Czech Republic, Vol. XLI, Part B1, pp. 199-204.
- 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.
- Klingner, B., Martin, D. & Roseborough, J., 2013. Street View Motion-from-Structure-from-Motion. In: *International Conference on Computer Vision (ICCV)*, Sydney, Australia, pp. 953-960.
- 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., 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 Sensing*, 9(8), doi:10.3390/rs9080796
- Nebiker, S., Cavegn, S., Eugster, H., Laemmer, K., Markram, J. & Wagner, R., 2012. Fusion of Airborne and Terrestrial Image-Based 3D Modelling for Road Infrastructure Management – Vision and First Experiments. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Melbourne, Australia, Vol. XXXIX, Part B4, pp. 79-84.
- Nebiker, S., Cavegn, S. & Loesch, B., 2015. Cloud-Based Geospatial 3D Image Spaces—A Powerful Urban Model for the Smart City. *ISPRS International Journal of Geo-Information*, 4(4), pp. 2267-2291.
- Nebiker, S., 2017. 3D Imagery for Infrastructure Management – Mobile Mapping meets the Cloud. In: *Proceedings of 56th Photogrammetric Week*, Stuttgart, Germany.
- Nocerino, E., Menna, F., Remondino, F., Toschi, I. & Rodríguez-González, P., 2017. Investigation of indoor and outdoor performance of two portable mobile mapping systems. In: *Proc. of SPIE, Videometrics, Range Imaging, and Applications XIV*, Munich, Germany, Vol. 10332, doi: 10.1117/12.2270761
- Puente, I., González-Jorge, H., Martínez-Sánchez, J. & Arias, P., 2013. Review of Mobile Mapping and Surveying Technologies. *Measurement*, 46(7), pp. 2127-2145.
- Rehak, M., 2017. Integrated Sensor Orientation on Micro Aerial Vehicles. PhD thesis, École Polytechnique Fédérale de Lausanne, Switzerland.
- Remondino, F., Del Pizzo, S., Kersten, T. & Troisi, S., 2012. Low-Cost and Open-Source Solutions for Automated Image Orientation – A Critical Overview. In: *EuroMed 2012*, Limassol, Cyprus, LNCS 7616. Springer, pp. 40-54.
- Rumpler, M., Tscharf, A., Mostegel, C., Dafttry, S., Hoppe, C., Pretenthaler, R., Fraundorfer, F., Mayer, G. & Bischof, H., 2017. Evaluations on multi-scale camera networks for precise and geo-accurate reconstructions from aerial and terrestrial images with user guidance. *Computer Vision and Image Understanding*, 157, pp. 255-273.
- Rupnik, E., Nex, F., Toschi, I. & Remondino, F., 2015. Aerial multi-camera systems: Accuracy and block triangulation issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101(60), pp. 233-246.
- Schönberger, J.L., Fraundorfer, F. & Frahm, J.-M., 2014. Structure-from-Motion for MAV Image Sequence Analysis with Photogrammetric Applications. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Zurich, Switzerland, Vol. XL-3, pp. 305-312.
- 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.
- Schönberger, J.L., Hardmeier, H., Sattler, T. & Pollefeys, M., 2017. Comparative Evaluation of Hand-Crafted and Learned Local Features. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, USA, pp. 1482-1491.
- Sun, Y., Sun, H., Yan, L., Fan, S. & Chen, R., 2016. RBA: Reduced Bundle Adjustment for oblique aerial photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 121, pp. 128-142.
- Toldo, R., Gherardi, R., Farenzena, M. & Fusiello, A., 2015. Hierarchical structure-and-motion recovery from uncalibrated images. *Computer Vision and Image Understanding*, 140, pp. 127-143.
- Triggs, B., McLauchlan, P.F., Hartley, R.I. & Fitzgibbon, A.W., 2000. Bundle Adjustment — A Modern Synthesis. In: *Vision Algorithms'99*, LNCS 1883. Springer, pp. 298-372.