

EVALUATION OF A MOBILE MULTI-SENSOR SYSTEM FOR SEAMLESS OUTDOOR AND INDOOR MAPPING

D. Iwaszczuk^{1,2*}, Z. Koppányi², J. Pfrang², C. Toth²

¹ Remote Sensing & Image Analysis, Technische Universität Darmstadt, Darmstadt, Germany

² Dept. of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, USA
iwaszczuk@geod.tu-darmstadt.de; janpfrang@hotmail.com; zoltan.koppanyi@gmail.com; toth.2@osu.edu

Commission I, WG I/10

KEY WORDS: Mapping, Indoor, Calibration, LiDAR, Coregistration, Sensor system

ABSTRACT:

Indoor mapping has been gaining importance recently. One of the main applications of indoor maps is personal navigation. For this application, the connection to the outdoor map is very important, as users typically enter the building from outside and navigate to their destination inside. Obtaining this connection, however, is challenging, as the georeferencing of indoor maps is difficult due to the weak or total lack of GPS signal which makes positioning impossible in general. One solution for this problem could be matching indoor and outdoor datasets. Unfortunately, this is difficult due to the very low or non-existing overlap between the indoor and outdoor datasets as well as the differences in different. To overcome this problem, we propose a mobile mapping system, which can seamlessly capture the outdoor and indoor scene. Our prototype system contains three laser scanners, six RGB cameras, two GPS receivers and one IMU. In this paper, we propose an approach to seamlessly map a building and define the requirements for the mapping system. We primarily describe the construction phase of this system. Finally, we evaluate the performance of our mapping system with regard to the defined requirements.

1. MOTIVATION

Due to urbanization and expected growth of the population in the cities, modern solutions for urban planning and data management are required. In response to this need, the Smart City concept has been developed over the past years with a main focus on mobility, including autonomous driving, pedestrian navigation in buildings, and smart building infrastructure. All these objectives require accurate models of urban areas, especially buildings, resulting in increased interest in mapping large urban areas, typically using mobile mapping systems. Data delivered from these systems include detailed geometries of an urban environment, including road networks, vegetation, city furniture, and building façades together with detailed structures, such as windows and doors. Mapping building interiors, however, requires other measurement strategies. Various approaches employing photogrammetric methods, laser scanners or their combination have been proposed to map building interiors. One of the main challenges while mapping indoor environments is the sensor georeferencing and alignment with outdoor geometries.

The difficulty of this georeferencing is related to the fact, that GPS signal is unreliable or non-existent in indoor environments and IMU suffers from drift which leads to a measurement error increasing over time. An automatic alignment of indoor scene with the outdoor scene could be a solution. These two scenes, however, even if showing the same object, practically do not overlap. Only transition areas, where both datasets can find common measurements, such as windows and doors. In the last years, promising algorithms based on these transition areas have been developed (Cohen et al., 2016; Koch et al., 2016; Speciale, 2018). In practice, however, many of these transitions are not visible in the data, due to occlusions (particularly vegetation in

outdoor) and inaccessible rooms. To cope with this, a mapping system which can seamlessly map transitions between outdoor an indoor environment would be advantageous. In this paper, we describe the development phase of such a system as well as evaluate this system in terms of its suitability for seamless outdoor-indoor mapping.

2. SEAMLESS OUTDOOR-INDOOR MAPPING

Our approach for seamless outdoor-indoor mapping relies on construction of a mapping system, which can operate efficiently in outdoor as well as in indoor environment. This system must be also flexible enough to continue measurement while passing the indoor-outdoor transition.

In this approach, we aim at obtaining geometry and semantic information of building interior, including its connection to the outdoor scene. For mapping, the following sensors have been considered:

- LiDAR
- Camera
- RGB-D Sensor

LiDAR-bases systems have the advantage that they deliver a point cloud, measured directly by light pulses. LiDAR can deliver high point density metric coordinates on all specularly reflecting surfaces, also on homogeneous surfaces, such as white walls, which are very common in building interiors. LiDAR point clouds are, however, more challenging for semantic labelling. Semantic information can be usually better derived from RGB information, which can be achieved with cameras. Using a single camera and Structure from Motion (SfM) technique also 3D point clouds can be derived. This point

* Corresponding author

clouds, however, are reconstructed in an internal, arbitrary coordinate system and have to be scaled to the real-world coordinates as well as localized. RGB-D sensors, such as Kinect, fuse depth images derived, using structured-light or time-of-flight technology with RGB images. These sensors operate in close range and have difficulties with measurements under daylight conditions.

Considering those properties of the above-mentioned sensors, a combination of LiDAR sensor and RGB camera seems to be the most suitable choice. RGB-D camera can be considered as a supplementary device used in addition to those sensors.

The main idea for the data collection is to map building interiors using a laser-based system for obtaining 3D geometries with additional RGB cameras, which should be primarily used for semantic labelling. In order to connect them to outdoor geometries by:

1. seamlessly mapping the outdoor-indoor transitions including entrances, followed by
2. registration refinement using transitions, such as doors and windows.

As LiDAR delivers already metric point clouds, the main challenge for georeferencing is to find the correspondence between the LiDAR frames and transform the point cloud to the world coordinate system. The outdoor georeferencing is performed using both, GPS and IMU. As soon as the system crosses the outdoor-indoor transition, only IMU can be used. IMU has a high short-term precision, but on longer distances suffers from a drift. Therefore, in addition to using IMU, we want to perform ICP (iterative closes point) algorithm between the neighbouring laser frames and calculate the trajectory based on the IMU data and ICP output using the Kalman filter approach.

In order to apply this approach, these transitions must be detected. These transitions can be detected directly from the navigation data by analysing GPS uncertainty (Pipelidis et al., 2017). In this project, we want to investigate a detection method based on visual data. These transitions should be detected by first conducting pixel-level classification on images, and then, by detecting objects, such as entrances. For the pixel-wise classification, we use an encoder-decoder Convolutional Neural Network (CNN). Object labels being output of the CNN-based classification are then transferred to the point clouds, measured by the laser scanners. Knowing the relative orientation between the laser scanners and the camera, the RGB information and the labels can be assigned to the 3D points as shown in Figure 1. Transferring labels from images to the 3D points. In this way, indoor-outdoor transitions are localization in the 3D point cloud of the indoor datasets.

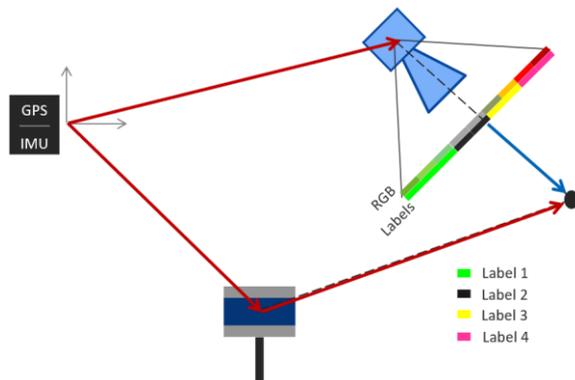


Figure 1. Transferring labels from images to the 3D points.

Knowing the position of the transitions in outdoor data, we can search for corresponding transitions in the indoor data by creating correspondence hypotheses, which can be verified using e.g. RANSAC approach (Fischler & Bolles, 1981). After finding the best correspondence hypothesis, we can estimate the similarity transformation between them, which leads to the alignment of the two datasets.

3. SYSTEM DESIGN

From our project requirements, we defined the requirements for the mapping system:

- Flexibility regarding the operation in indoor and outdoor. More specifically, in outdoor scene facades must be well covered, while indoors the entire scene, including stairways and narrow corridors should be well covered.
- Good point cloud coverage of objects in outdoor and in indoor area.
- Flexibility to pass through outdoor-indoor transitions including stairways and elevators.
- Acquisition geometry delivering point clouds which can help orient the system (using ICP).
- Good image coverage of entire indoor scenes, adaptable field of view.
- Simultaneous images with known stereo basis.

These requirements made up the main guidelines for the design of the system and its construction.

In order to ensure the flexibility, our prototype mobile mapping system was conceived in a backpack configuration. We decided to use three laser scanners from Velodyne:

- VELO_1: mapping horizontally, placed at the top of the mapping system
- VELO_2: mapping at angle of about 45°
- VELO_3: mapping vertically

VELO_1 is mostly useful in indoor scenes, where we expect to use it for orientation support. VELO_2, which is mapping the scene perpendicularly to the operator's moving direction is useful in indoor and in outdoor. In outdoor scene VELO_2 and VELO_3 ensure that also higher facades can be mapped. In addition to the Velodyne laser scanners, six different cameras were foreseen for the system.

Prior to construction of the mapping system, it was first drafted in CAD software. The rack was modelled by support of a human model to make it possibly ergonomic (Figure 3. System design: project created using CAD software (left) and ready prototype (right). This system should integrate multiple sensors for navigation and mapping. All sensors including battery and recording devices were placed in a specially designed stiff lightweight aluminium rack, which was carried as backpack.

The sensor platform was assembled and adjusted to ensure overlap of the neighbouring sections of different sensors, but at the same time to keep this overlap small in order to fully use the image and scan sections. Another goal during the planning phase was to minimize shadows in the field of view of each sensor, by parts of the rack, by other sensors or by parts of the operator's body while moving.

In order to achieve these goals, all the mapping sensors, including their field of view, were modelled in the CAD

software as well. Main assumption was to map an indoor cross section of approximately 3x3 m, which represents, for example, a hallway (Figure 2.).

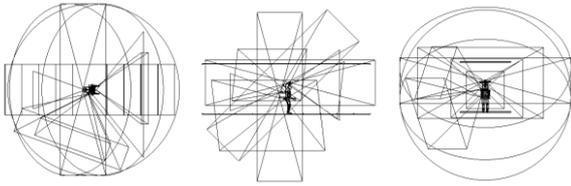


Figure 2. CAD simulation, including the sensors and their field of view; three views: top, side, back. As the cameras were placed on the backpack symmetrically, only one side is presented here, in order to improve the readability of the image

During the planning phase, different configurations of the sensors were simulated in CAD software and we realized that the visibility conditions can be achieved, when the sensors are directed mostly to the back (opposite direction of the operator’s moving direction). After the general configuration was known, we refined the position and orientation of every sensor, so that the occlusions were minimal and the system was still small enough to be operational, particularly while passing through narrow passages and doors. Based on this conceptual design, the prototype was constructed (Fig. 3, right).

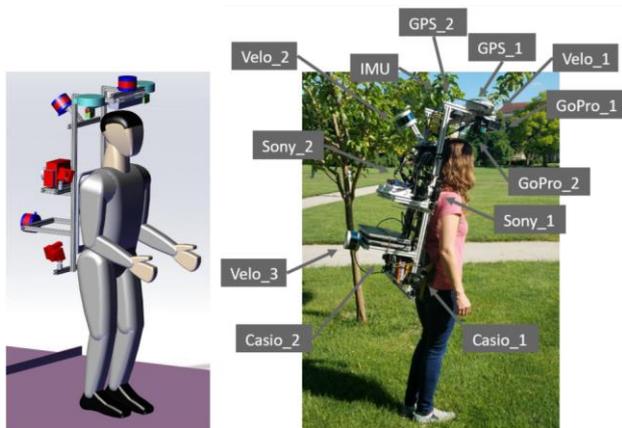


Figure 3. System design: project created using CAD software (left) and ready prototype (right)

The implemented system consists of three LiDAR sensors, one in horizontal position, one in vertical position and one at approximately 45° to the others. In addition, six RGB cameras, two looking forward (GoPro_1, GoPro_2) and four looking backwards (Sony_1, Sony_2, Casio_1, Casio_2) were mounted. For positioning two GPS receivers, and one inertial measurement unit (IMU). The components of the mapping system are listed in Tab. 1. GoPro cameras were mounted directly on the rack in the up-side-down manner without any flexibility to set the viewing direction. The other four cameras were mounted using a coupling, which enables flexible adjustment of the viewing direction prior to every measurement, or even during the operation. This feature improves the flexibility of the system, but at the same time, it requires a calibration every time when the orientation of the camera is changed.

Abbrev.	Description
Velo_1	Velodyne VLP-16 (LiDAR)
Velo_2	Velodyne VLP-16 (LiDAR)
Velo_3	Velodyne VLP-16 (LiDAR)
Sony_1	Sony Alpha 6000 (high resolution still images)
Sony_2	Sony Alpha 6000 (high resolution still images)
Casio_1	Casio EX- H20G (P&S, video)
Casio_2	Casio EX- H20G (P&S, video)
GoPro_1	GoPro HERO5 (video)
GoPro_2	GoPro HERO5 (video)
GPS_1	GPS Antenna Novatel 502
GPS_2	GPS Antenna Novatel 502
IMU	Inertial Measurement Unit STIM

Table 1. Components of the mapping system

Our system is similar to the system presented in Blaser et al. (2018). The main difference of our system is usage of the third laser scanner at 45°.

4. SYSTEM EVALUATION

In this study, the focus is on the evaluation of the designed mapping system with regard to the requirements defined in Section 3.

The lever-arm and boresight calibration between the laser scanners and IMU were conducted in two steps:

1. The nominal angles and distances between the sensors were taken from the CAD sketch.
2. These parameters were refined based on captured point clouds.

This refinement was conducted in static mode. The mapping system was placed on a table in an office. All three laser scanners were recording this scene for some time. Then random, synchronized frames were selected from each scanner and they were visually aligned to each other, which led to refined lever-arm and boresight parameters. The result of this alignment is presented in Figure 9. Exemplary results of labelling the indoor scene.

Figure 9. Exemplary results of labelling the indoor scene. shows also the positions of the laser scanners (cyan dots) and of the IMU (red dot). It can be seen that around the position of the VELO_2 scanner, there is a small set of green points. These points are registered by VELO_3 and are measured on the surface of VELO_2. This means that VELO_2 is partially occluding the view of VELO_3. This could not be avoided, because a completely occlusion-free acquisition would require very long offset of VELO_3 from the operators back. This would make the usage of the backpack more uncomfortable for the operator and would increase the size of the device, so that it would not be possible to pass through most of the doors.

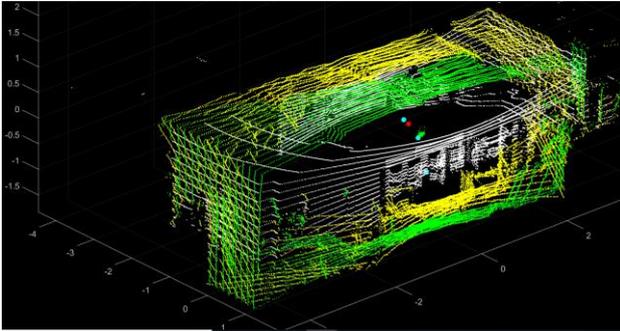


Figure 4. Lever-arm and boresight calibration between the laser scanners. White: points coming from horizontal scanner VELO_1; green: vertical VELO_3; yellow: 45° scanner VELO_2; cyan: positions of laser scanners; red: position of IMU

Our system can operate in outdoor scene as well, what is shown in Figure 5. Point cloud captured in outdoor environment (single frames from VeloView): horizontal scanner VELO_1 (top), vertical scanner VELO_3 (middle), 45° scanner VELO_2 (bottom). As we can see, the best view/observation of the scene can be delivered by the horizontal scanner. The other two scanners are mainly useful for achieving higher point density and capturing higher facades. In Figure 6. Outside view from Sony_1 also a camera view in an outdoor scene is presented.

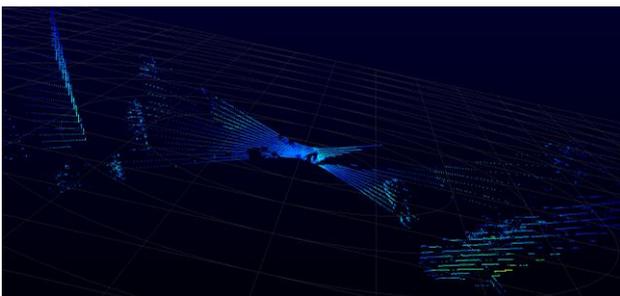
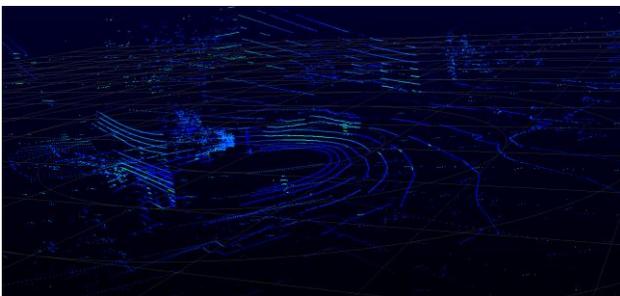


Figure 5. Point cloud captured in outdoor environment (single frames from VeloView): horizontal scanner VELO_1 (top), vertical scanner VELO_3 (middle), 45° scanner VELO_2 (bottom).



Figure 6. Outside view from Sony_1

Our prototype system proved to be suitable for mapping entrances and stairways. To pass through a door, however, the operator needs support to hold the door open. Also, manoeuvrability of the system is limited. Figure 9. Exemplary results of labelling the indoor scene shows the moment of passing through a building entrance.



Figure 7. View from Sony_2 while passing through a building entrance

In general, occlusions could be avoided for the cameras. Only GoPro cameras suffer from very small coverage by operator's head (Figure 8).



Figure 8. A frame from GoPro video from the calibration room.

5. FIRST RESULTS ON SEMANTIC LABELLING

One of the goals of our approach is to extract semantic information from our data, particularly detect doors and windows. This is done by pixel-wise semantic labelling. As we showed in Koppányi et al. (2019), an encoder-decoder CNN

architecture can be successfully used for pixel-wise labelling of indoor scenes. We achieved about 76% accuracy while training the network, using one building, and then testing on another one. These results show that it is not necessary to obtain extensive ground truth data in the test area. This opens the possibility to use public dataset to high extent.

In order to pixel-wise label our images, we use a CNN, trained on a dataset consisting of 90% public datasets (Armeni et al., 2017) and 10% manually labelled data originating from our data collection. An example labelling result is presented in Figure 9.



Figure 9. Exemplary results of labelling the indoor scene

6. CONCLUSION AND FUTURE WORK

In this paper, we presented the conceptual design and implementation of a mobile mapping system in a backpack configuration for seamless outdoor-indoor mapping. We evaluated this system in terms of the requirements defined in this paper based on the planned approach.

The mapping system has been shown to be flexible enough and able to deliver data suitable for our application.

Our next steps will focus on processing the data delivered by the system. Also, some minor improvements to the system are planned. First of all, we plan to include Ultrawide Band (UWB) transmitter and use it to obtain reference trajectory data.

While the lever-arm and boresight calibration of the laser scanners is valid for every measurement, the calibration of the cameras must be conducted every time the system is put into operation. We want to test another solution, which is applying Structure from Motion (SfM) algorithm for all available cameras, followed by matching the LiDAR point cloud and the SfM point cloud, in order to calculate the final orientation of all cameras. This approach would also solve the problem of synchronization of all sensors, which is not effectively working for the cameras for now.

REFERENCES

Armeni, I., Sax, A., Zamir, A. R. and Savarese, S., 2017. Joint 2D-3D-Semantic Data for Indoor Scene Understanding. *ArXiv e-prints*.

Blaser, S., Cavegn, S., Nebiker, S. (2018) Development of a Portable High Performance Mobile Mapping System Using The Robot Operating System. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume IV-1, 2018. ISPRS TC I Mid-term Symposium "Innovative Sensing –*

From Sensors to Methods and Applications", 10–12 October 2018, Karlsruhe, Germany

Cohen A, Schönberger JL, Speciale P, Sattler T, Frahm JM, Pollefeys M., 2016. Indoor-Outdoor 3D Reconstruction Alignment. *European Conference on Computer Vision 2016*. Springer International Publishing.

Fischler, M., Bolles, R. (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24 (6): 381–395.

Koch, T., Korner, M., Fraundorfer, F., 2016. Automatic Alignment of Indoor and Outdoor Building Models using 3D line segments. *CVPR Workshops (2016)*

Koppanyi, Z., Iwaszczuk, D., Zha B, Saul, C.J., Toth, C., Yilmaz, A. (2019) Multi-Modal Semantic Segmentation: Fusion of RGB and Depth Data in Convolutional Neural Networks. In: *Multi Modal Scene Understanding*. Elsevier

Pipelidis, G., Moslehi Rad, O.R., Prehofer, Ch., Iwaszczuk, D., Hugentobler, U. (2017) A Novel Approach for Dynamic Vertical Indoor Mapping through Crowd-sourced Smartphone Sensor Data. *Proceedings of Indoor Positioning and Indoor Navigation - IPIN 2017, September 18-21, 2017*

Speciale, P., Paudel, D.P., Oswald, M.R., Riemenschneider, H., V. Gool, L., Pollefeys, M. 2018. Consensus Maximization with Linear Matrix Inequality Constraints. *Conference on Computer Vision and Pattern Recognition (2018)*