Using Fiducials in 3D Map Evaluation

Sören Schwertfeger[1] and Andreas Birk[2]
[1]ShanghaiTech University, 319 Yueyang Road, Shanghai 200031, China
soerensch@shanghaitech.edu.cn, http://robotics.shanghaitech.edu.cn
[2]Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany
a.birk@jacobs-university.de, http://robotics.jacobs-university.de

***Abstract*— For missions in Safety, Security, and Rescue Robotics (SSRR) maps are often a core deliverable. Hence it is of high interest to assess the quality of maps in a simple and efficient way. Since SSRR is mostly taking place in unstructured environments, 3D mapping has become more and more important. Here a method to evaluate the quality of 3D maps is presented that extends the previously developed 2D Fiducial approach to the third dimension. The artificial features are identified and located in 2D cross-sections of the map as well as in the 3D maps. It is then attempted to proof this concept using a ground truth map and robot generated maps from the RoboCup Rescue competition 2013 in Eindhoven.**

I. Introduction

Mapping is an important task in many applications of mobile robots. But those maps are not perfect. The amount and type of error has to be measured in order to make statements of the quality of the maps, the mapping system and thus the performance of the whole robot. Although there has been great progress in mapping in the last two decades, especially with respect to Simultaneous Localization and Mapping (SLAM) techniques, it has to be noted that especially in scenarios that are of high interest for SSRR, namely on extended missions and in unstructured environments, maps still often contain large errors.

There are different approaches to asses the quality of a mapping system. Ground-truth robot paths are used in [1] and [2] and compared with the paths estimated by the SLAM algorithms. But obtaining the actual robot paths is a difficult problem and can only be done in very controlled environments or in simulations. A metric for measuring the error of the manually corrected trajectory of datasets is available to the public in [3].

Often ground truth information in form of maps or feature locations are used. This is easy when working with simulations [4]. Image similarity methods [5] and pixel-level feature detectors [6], [7] have their limitations due to the common errors in maps, because maps often have structural errors like noise or structures appearing more than once due to localization errors.

Capturing the topology of the environment and comparing it against the ground truth gives evaluation results that strongly correlate with the usefulness of the maps for human navigation, for example in a scenario where a first responder has to reach a victim by navigation the robot generated map. In [8] this is applied in the RoboCup Rescue Virtual Robots competition. The brokenness is one such measure that was

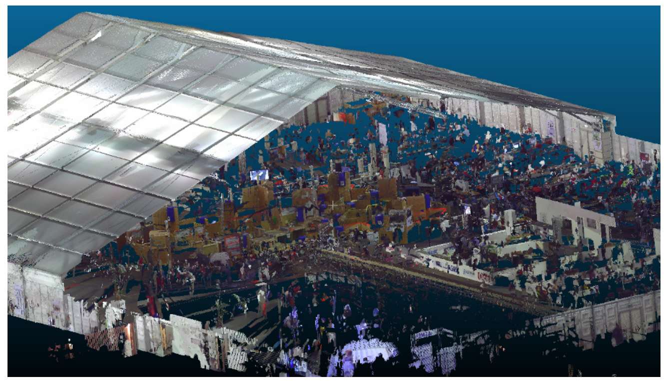


Fig. 1. The arena in the original scans with a partially removed tent (75 million colored points). The RoboCup at Home arena is on the right side.

introduced in [9] and calculated in a more topological way in [10] and [11]. Please refer to [12], [11] for a more detailed survey on map evaluation.

All of the above mentioned methods have in common that they work on 2D grid maps. But the use of 3D mapping is essential in SSRR scenarios and becoming more common [13], [14], [15], [16]. Very little work has been done on 3D map evaluation. 3D planar patches are used in [17]. They detect suspicious and plausible arrangements of those planes and classify the map accordingly. So assumptions about the environment, e.g., the presence of planar walls, are made, and the local consistency of this assumption is quantified. The algorithm does not make use of any ground truth information, thus a map which looks nice might get a good score even if it seriously broken at some point. Also, compared to human-made environments, there are fewer planes in unstructured environments, so it is less applicable in the search and rescue domain.

Datasets of 3D sensor measurements and ground truth poses for benchmarking the performance of SLAM algorithms are provided in [18] and [19]. The ground truth information has been obtained using a tracking system and by creating the data in a simulation, respectively.

There is no single measurement for map quality, but different attributes of a maps should be measured separately and weighed according to the needs of the application [20]. Those attributes can include the coverage, the accuracy (Correctness of positions of features in the global reference frame) and topological measurements like the consistency

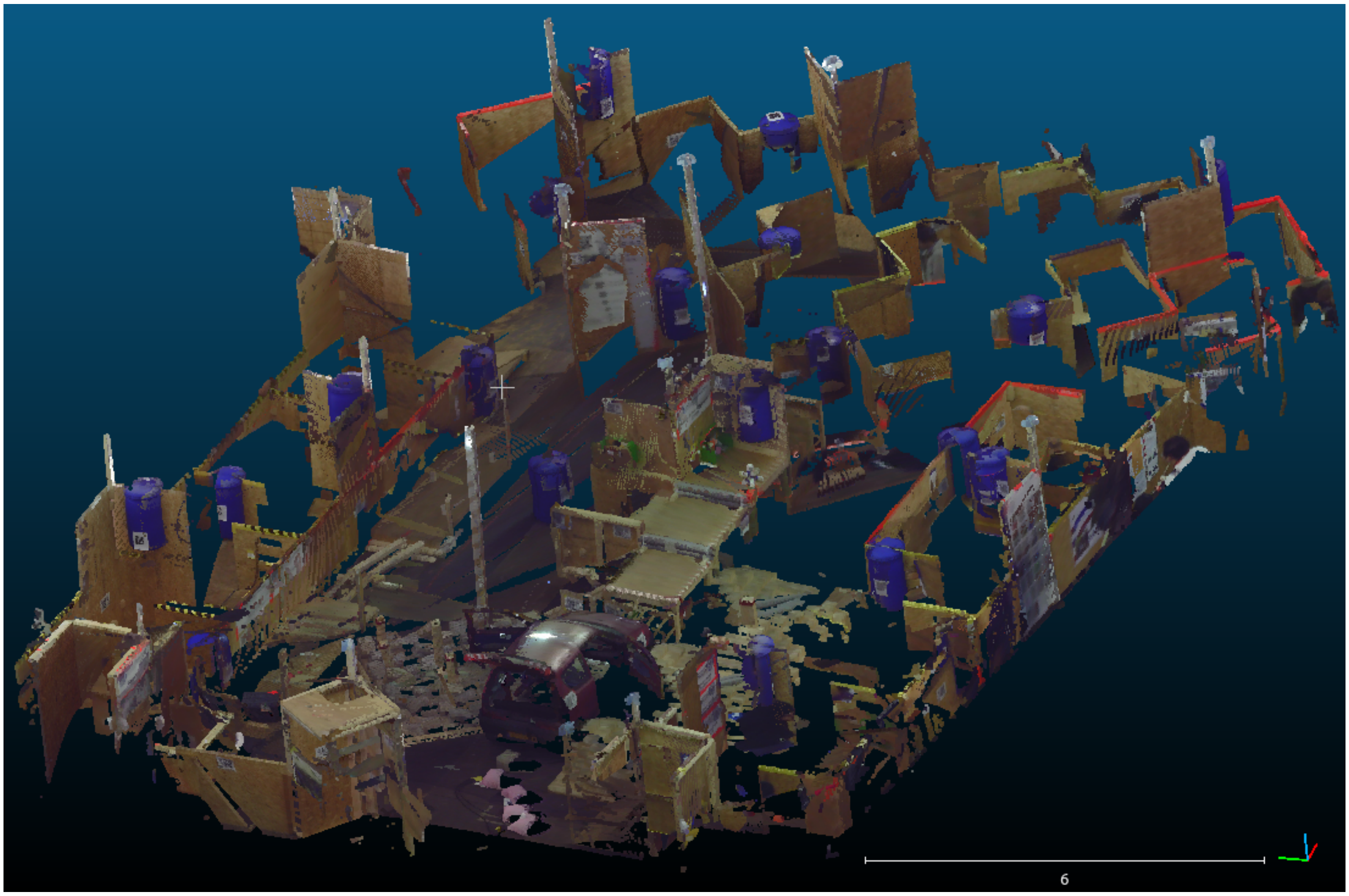


Fig. 2. 3D ground truth map, composed out of four Faro scans, one from each corner of the arena. (12.6 million colored points)

of local groups of features. An approach using artificial Fiducials (barrels) that can calculate those attributes has been proposed in [21], [22] and is used as a tool for grading 2D maps at RoboCup Rescue competitions. This paper describes how this concept can be extended to 3D and performs some proof of concept experiments.

In the next section this paper presents the environment that is represented in the maps evaluated in this paper, namely the RoboCup Rescue arena. The gathering and the properties of the ground truth map are also described there. Section III gives information about the maps. The evaluation of the maps and the generated results are discussed in Section IV. The paper concludes in Section V.

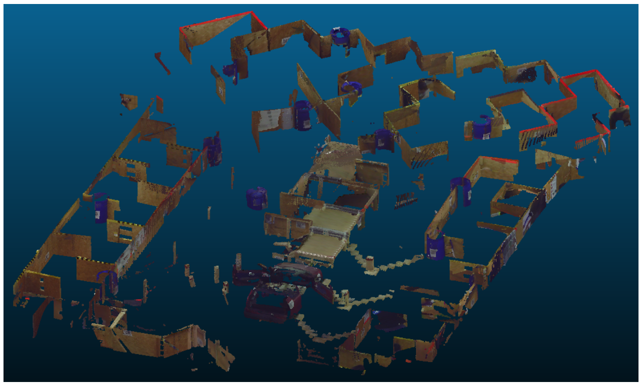
II. RoboCup Rescue and Ground Truth Map

In this paper, maps created during the 2013 RoboCup Rescue competition in Eindhoven, Netherlands are used. This competition is an excellent tool to guide the scientific development of robots towards the needs of the SSRR domain [23], [24]. On the other hand, standard test methods for response robots can be thoroughly evaluated on a number of different robots there [25]. The arena (see Figure 2) consists of maze-like structures (top, bottom and right side in Figure 3(c)), two types of difficult terrain divided by a car (left) and an elevated area (center). The wall panels are typically 1.2m high, with the exception of several walls with double that height.

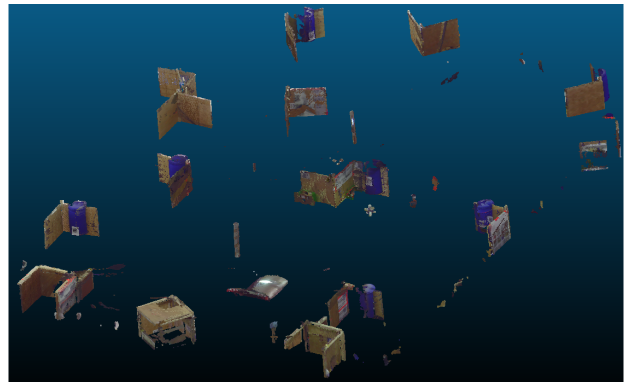
The ground truth map was created using a FARO scanner. Four scans were made from every corner of the arena. The horizontal field of view (fov) was 360 degrees while it was from -60 to 90 degrees vertically. The resolution was 8192 x 3414 beams, resulting in 28 million points. That corresponds to a point to point distance of 7.7 mm at 10 m distance. Each scan took about 4 minutes 15 seconds. Restricting the field of view would result in fewer data to process and shorter scan times, but it is more error prone (selecting a sufficient fov and aligning the sensor accordingly) such that a full scan was selected.

Registering the four scans by hand was easy using the software Cloud Compare [26]. After a rough manual alignment common distinctive points were selected. The software then computes a best fit. Since the sensor has a range of 120 m the whole tent the competition was held in was mapped (See Figure 1). All non-arena parts were then removed. Also persons that were captured in the scans and points belonging to the net that was spanned over the arena for the aerial vehicles were removed. The resulting point cloud has 12.6 million colored points (see Figure 2).

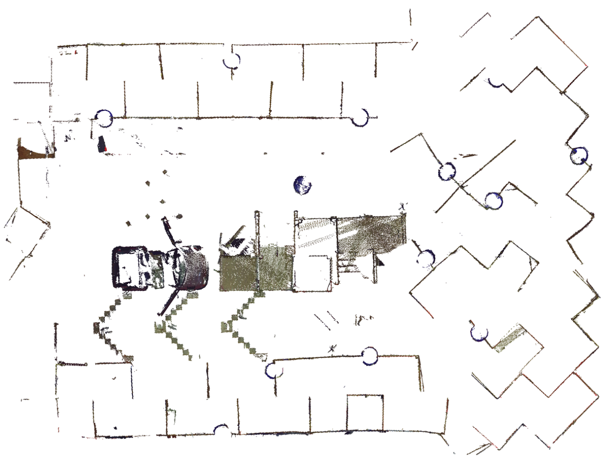
The use of the FARO scanner gives very accurate results and also the registration of the four scans works very reliably.



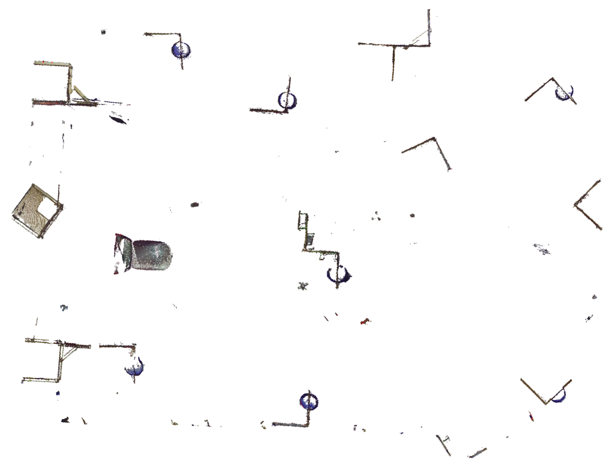
(a) The ground truth 3D map cut between heights of 0.4m and 1.2m. (4 million points)



(b) The ground truth 3D map cut between heights of 1.3m and 2.4m. (3.7 million points)



(c) 1st level 2D ground truth map generated from (a). (86.127 occupied cells)



(d) 2nd level 2D ground truth map generated from (b). (33.618 occupied cells)

Fig. 3. 2D ground truth maps.

Nevertheless it has to be noted that the resulting map has some flaws. The most important is, that not all the walls have been captured, especially on the upper right corner of Figure 3(c). This is because those walls were not seen by the scan taken near that corner (fov restriction) and they were occluded by the high walls for all three other scans. Also quite often the lower parts of the walls and the ground are occluded. In the future this should be remedied by taking more scans, also from inside the arena.

For the experiments later on also 2D ground truth maps are required. They can be easily and reliably extracted from the 3D point cloud by cutting out cross sections of the 3D ground truth point cloud at a certain height range. Figure 3(a) shows the cross-section between 0.4 m and 1.2 m while Figure 3(b) shows a cross section between 1.3 m and 2.4 m. The latter one thus does not contain the lower, ground level walls and only represents the high walls and the elevated obstacles (car, ramp, staircase, etc.). The resulting point clouds are then projected on the x, y plane in a 2D grid with a resolution of 16 mm per pixel (same resolution as the 2D maps provided in the next section). The resulting 2D ground truth maps can be seen in Figure 3(c) and Figure 3(d), respectively.

Run	Map	Abbr.	Num	2D or 3D
Preliminary run 1	normal 2D	Pre1	1	2D
	crossection 2 m	Pre1_2m	2	2D
	3D octomap	Pre1_3D	3	3D
Preliminary run 2	normal 2D	Pre2	4	2D
	crossection 2 m	Pre2_2m	5	2D
	3D octomap	Pre2_3D	6	3D
Preliminary run 4	normal 2D	Pre4	7	2D
	crossection 2 m	Pre4_2m	8	2D
	3D octomap	Pre4_3D	9	3D
Best in Class Autonomy	normal 2D	BiC	10	2D
	crossection 2 m	BiC_2m	11	2D
	3D octomap	BiC_3D	12	3D

TABLE I

THE MAPS USED IN THE EXPERIMENTS.

All maps can be downloaded at http://robotics.shanghaitech.edu.cn/research/SSRR2015_dataset .

III. Evaluation Maps

During the RoboCup Rescue competition only one team created and submitted 3D maps: Team Hector Darmstadt

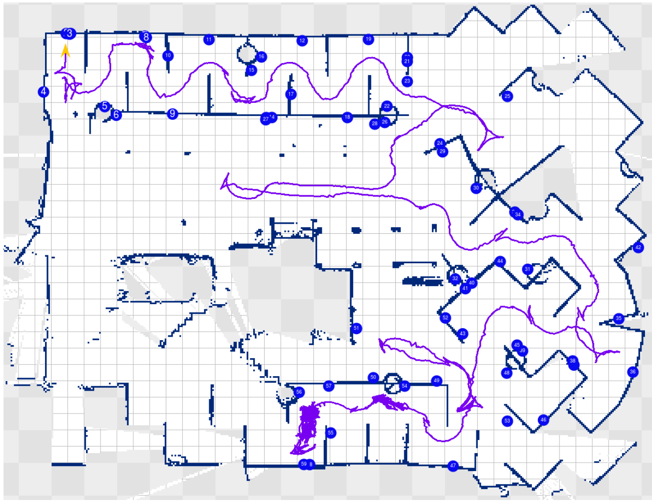


Fig. 4. Map BiC: RoboCup Rescue 2013 ”Best in Class” 2D map from Hector with robot start pose (yellow), robot path (purple) and locations of QR codes found (blue dots).

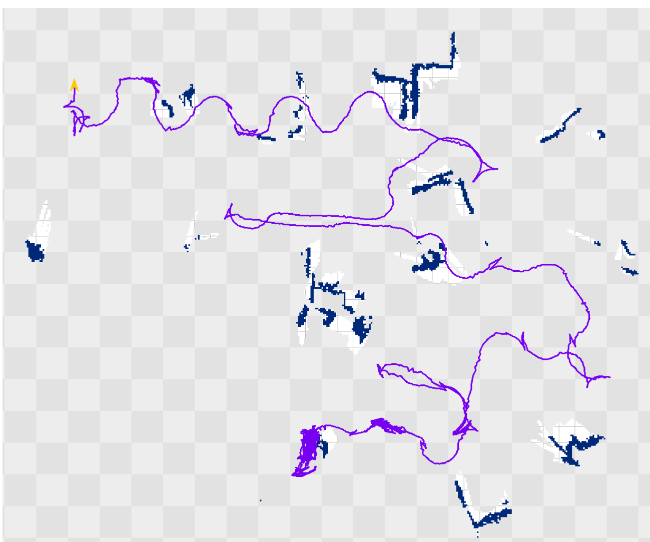


Fig. 5. Map BiC_2m: The cut at 2m from the data from Figure 4.

from the Technische Universität Darmstadt. This team is very active in the RoboCup Rescue community and develops and provides software modules for search and rescue robots [27]. Especially the ROS mapping software ”Hector SLAM” [28] is used by most of the teams participating. The 3D maps used here were created using the 6D localization of Hector SLAM and an actuated RGBD camera filling an octomap [29]. The octomap was developed by the University of Freiburg.

There are three unique sources of data that can be used for the experiments in this paper: The 3D (ground truth) Faro data, the 2D maps produced by Hector and the 3D octomaps from Hector. The Hector maps were generated during the four preliminary rounds of RoboCup Rescue 2013 (although there was no map generated in the 3rd round due to technical difficulties) and the ”Best in Class in Autonomy” competition. Each of those runs produced a 2D map on

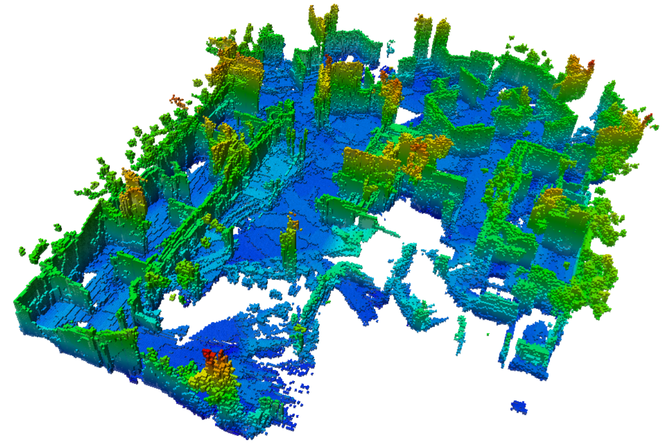


Fig. 6. Map BiC_3D: RoboCup Rescue 2013 ”Best in Class” 3D octomap from Hector (leaf size 0.05 m; 2.3 million nodes - pruned 0.9 million). The corresponding point cloud has 417.000 points. The points are colored after the value of their z coordinate.

ground level, a 2D map in 2 m height and a 3D map. See Table I for an overview of the maps and abbreviations used when regarding to each map. Figure 4 shows BiC, from the run covering the most area. Figure 5 shows BiC_2m from the same run, which is basically the 3D octomap BiC_3D (Figure 6) sampled at a certain height. The 2D maps of the other runs are shown in Figure 7. Please also see the accompanying video for a visualization of the 3D data which is available at `http://ieeexplore.ieee.org`.

It has to be noted that the ground truth map was created during the final runs of the competition. The arena configuration thus differs slightly from the preliminary runs (where the arena was actually separated in two parts) and also the best in class autonomy run. The BiC configuration was very close to the one from the ground truth map - just the position of one stair was changed and the shape of one wall changed from zig-zag to straight (on the right side). All maps can be downloaded at `http://robotics.shanghaitech.edu.cn/research/SSRR2015_dataset` .

IV. Evaluation and Discussion

The suggested map evaluation approach closely follows the Fiducial approach presented in [21] and [22]. For the ground level exactly the same method is used: The coverage is calculated by dividing the number of barrels identified by the number of barrels available at that height. The accuracy is a measure for the error of the location of the found Fiducials. The consistency measures how well the two parts of one barrel are mapped together. Long-range consistency is measured for two barrel parts where the shortest robot-drivable path between places of observation of the two barrel parts is very long (over 8 pallets = 9.6 m) and short range consistency is measured for barrels with a shorter path between them. Please refer to [22] for more details.

In this paper, additionally the cross-section maps around the 2m height (see the previous section) are applied to the Fiducial algorithm. For the Fiducial algorithm the locations

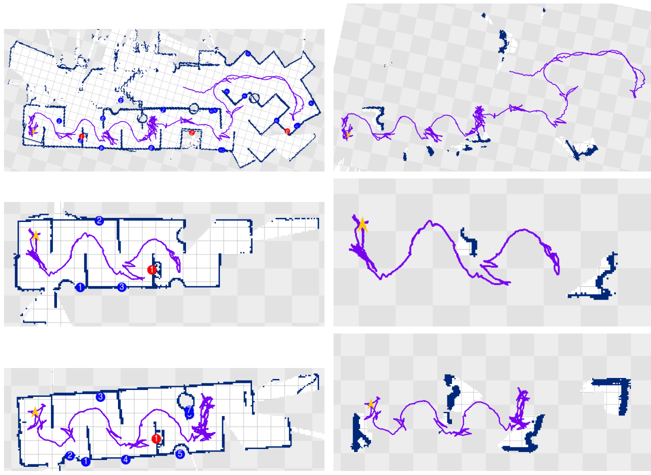


Fig. 7. Maps from the Preliminary Rounds 1, 2 and 4 from Team Hector Darmstadt. Ground level maps on the left and elevated maps (at 2m) on the right. Maps 2 and 4 share the start point with Figure 4 while map 1 starts on the lower left corridor.

TABLE II
SCORES FOR THE MAP ATTRIBUTES FROM THE PROPOSED ALGORITHMS.

Map	Coverage	Relative Accuracy	Consistency Short-range	Consistency Long-range
Pre1	33%	88%	98%	96%
Pre2	13%	95%	–	–
Pre4	17%	94%	95%	–
BiC	83%	97%	95%	89%
Pre1_2m	14%	92%	–	–
Pre2_2m	14%	87%	–	–
Pre4_2m	14%	95%	–	–
BiC_2m	36%	97%	–	–

of the different Fiducials are carefully carefully extracted by hand by the authors.

The ground truth positions of the Fiducials were extracted from the 2D ground truth maps.

This approach is then compared to method where the Fiducials on all levels are directly detected in the 3D map provided by the team. Since there is currently no software support for that, the accuracy could not be measured. The results of those calculations can be found in Table III.

Additionally the 3D maps were inspected by overlaying them with the ground truth map.

With just four maps these experiments cannot evaluate the quality of this map evaluation approach. But it is sufficient as a proof of concept. The numbers of the ground and the elevated level in Table II correlate, even though there are just 14 elevated half-barrels compared to 24 half-barrels on the ground level. Comparing the 2D approaches with the 3D search in Table III one can see that for most of the maps more high Fiducials can be found in 3D. This is because the cross-section sometimes misses a barrel for which then more identifiable points exists at a greater height.

For BiC_3D fewer ground Fiducials were found than in BiC. This is because the ground maps were created not using a cross-section of the 3D map but using the 2D LRF mounted on the robot. This device and the 2D map have a higher

TABLE III
PERCENTAGE OF IDENTIFIABLE FIDUCIALS IN THE 3D MAP.

Map	Coverage Ground	Coverage High	Total Coverage
Pre1_3D	33%	21%	29%
Pre2_3D	13%	14%	13%
Pre4_3D	17%	21%	18%
BiC_3D	63%	57%	61%

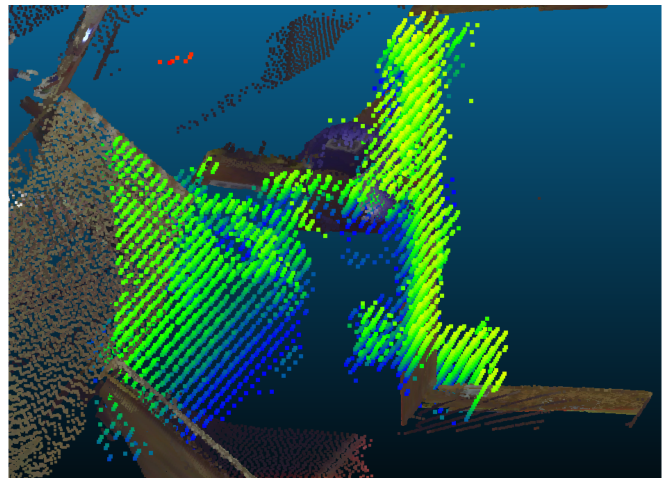


Fig. 8. Attempt to visualize where a Fiducial that was found in BiC could not be identified in BiC_3D. The map point cloud (blue near the ground to green near the top) was cut around the Fiducial. The ground truth map is visible as photo-colored point cloud.

resolution than the 3D map (on the x, y plane): 16 mm per pixel vs. 50 mm per voxel. Also a messy 2D scan of a Fiducial might still be more identifiable than a messy 3D scan. See Figure 8 for an example of a Fiducial that can be identified in the 2D map (Figure 4 in the center, right of the elevated floor, below the blue QR-code dot number 32) but not in 3D. Nevertheless all numbers clearly show that the ”Best in Class” run was very successful, as can also be seen when overlaying the map with the ground truth (Figure 9).

In Figure 10 one can see that the localization of Hector SLAM apparently failed at some point during the preliminary 1 run, since there is a considerable amount of points below the ground level.

It was also observed that a number of times the back side (with respect of the driving direction of the robot) of higher Fiducials have not been mapped. This is since the pan tilt head with the mounted RGBD camera is not looking backwards when exploring the arena.

A. Application to SSRR

Mapping and maps can be useful in many different ways for Safety, Security, and Rescue Robotics (SSRR). Depending on the particular application mapping systems have to be evaluated by weighing the attributes differently. For two examples proposed weights are shortly outlined:

- *Search in a large building*: The goal is to find a victim or hazard in a maze like structure. On success human first responders will navigate to the found location using

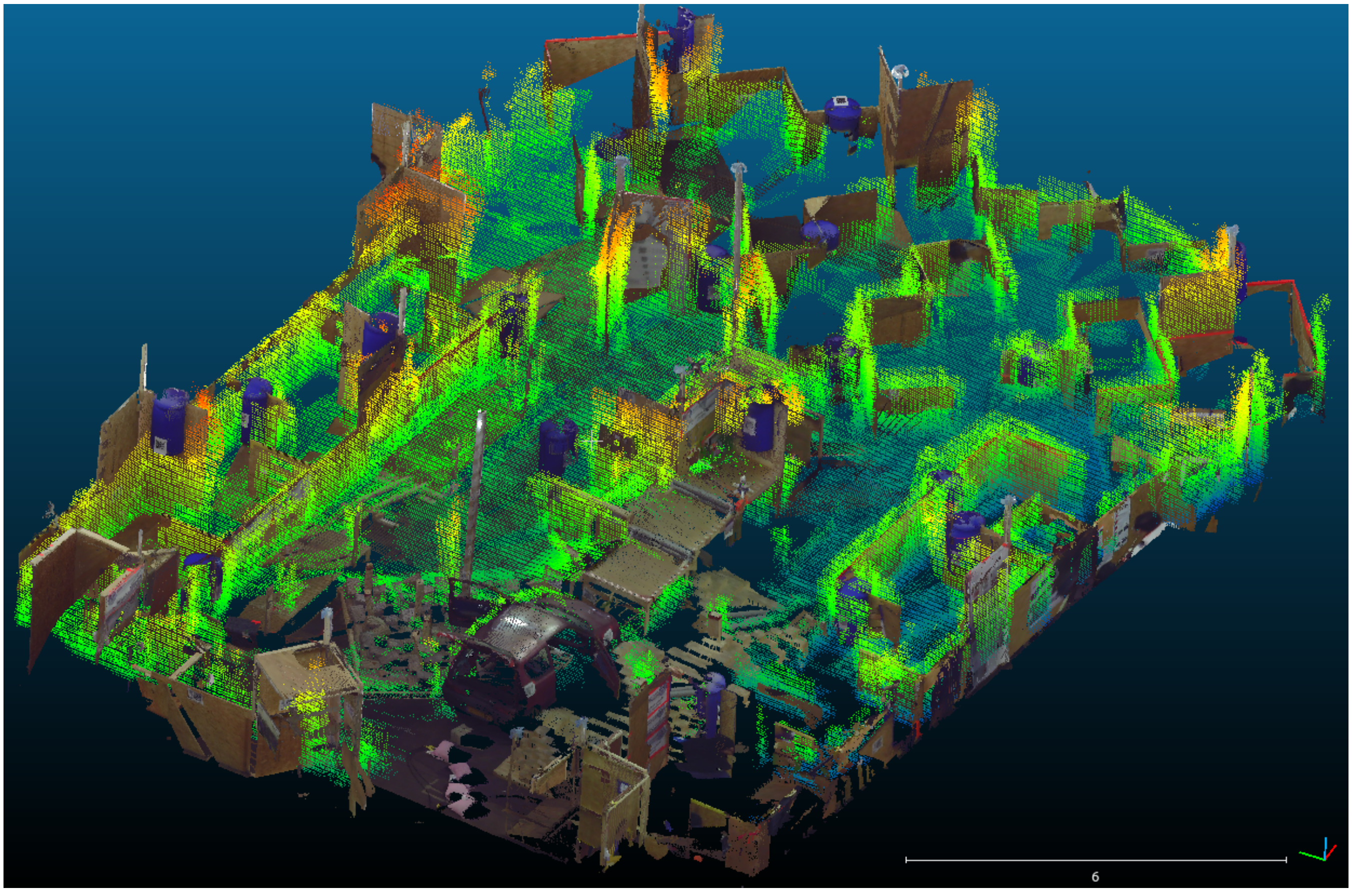


Fig. 9. The "Best in Class" point cloud (from Figure 6) overlaid with the ground truth map.

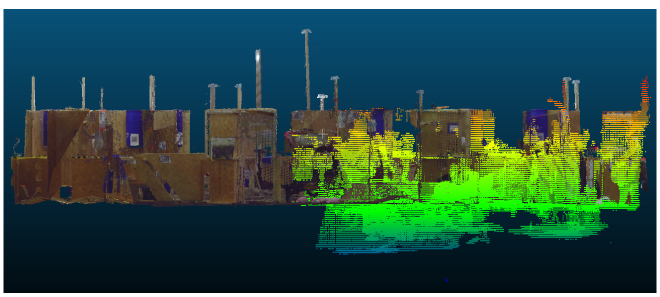


Fig. 10. Prel_3D with considerable amounts of erroneous data below the ground level.

the map. Clearly the coverage (of the map and thus the search) should be one important map attribute. Humans can navigate quite successfully even when presented slanted or bend maps, thus we do not need a good global or relative accuracy - local consistency (short range and to a certain degree long range) are more important.

- *Search in rubble pile*: Here the goal is to find a victim or hazard in a rubble pile using small robots. On success the rescue personnel will dig a hole from the surface straight down to the location. Clearly the most important map attributes are the global (if available) and relative accuracy: If we found a victim we have to dig the access hole at the right (x, y) location and to the right depth (z-axis) - any big error in the global position would lead to a mission failure.

V. Conclusions

This paper shows a proof of concept of a 3D Fiducial approach for evaluating 3D maps. We demonstrated the collection of a 3D ground truth map. Although it is not completely modeling the whole arena, it is more then enough to extract the ground truth Fiducial locations and for visual comparison with the 3D maps. The four runs that created 3D maps were evaluated with the 2D approach, on the elevated level and with the 3D approach. All methods delivered convincing results. But it could also be seen that the Fiducial approach does not capture all errors which are becoming more likely in 3D maps. Also the inherent need for lower resolutions in 3D maps makes the identification of reasonably sized Fiducials more difficult. Cases like in Figure 10 where map data exists below the ground level cannot be measured with this approach.

Acknowledgements

The authors would like to thank Team Hector Darmstadt for generating and providing the robot maps.

REFERENCES

[1] O. Wulf, A. Nüchter, J. Hertzberg, and B. Wagner, "Ground truth evaluation of large urban 6d slam," oct. 2007, pp. 650 –657.

[2] R. Kümmerle, B. Steder, C. Dornhege, M. Ruhnke, G. Grisetti, C. Stachniss, and A. Kleiner, "On measuring the accuracy of slam algorithms," *Autonomous Robots*, vol. 27, no. 4, pp. 387–407, 2009.

[3] W. Burgard, C. Stachniss, G. Grisetti, B. Steder, R. Kummerle, C. Dornhege, M. Ruhnke, A. Kleiner, and J. Tardos, "A comparison of slam algorithms based on a graph of relations," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, 2009, pp. 2089–2095.

[4] C. Scrapper, R. Madhavan, and S. Balakirsky, "Stable navigation solutions for robots in complex environments," in *IEEE International Workshop on Safety, Security and Rescue Robotics (SSRR)*, 2007, pp. 1–6.

[5] I. Varsadan, A. Birk, and M. Pfingsthorn, "Determining map quality through an image similarity metric," in *RoboCup 2008: Robot World-Cup XII, Lecture Notes in Artificial Intelligence (LNAI)*, L. Iocchi, H. Matsubara, A. Weitzenfeld, and C. Zhou, Eds. Springer, 2009, pp. 355–365.

[6] J. Pellenz and D. Paulus, "Mapping and Map Scoring at the RoboCupRescue Competition," Quantitative Performance Evaluation of Navigation Solutions for Mobile Robots (RSS 2008, Workshop CD), 2008.

[7] R. Lakaemper and N. Adluru, "Using virtual scans for improved mapping and evaluation," *Auton. Robots*, vol. 27, no. 4, pp. 431–448, 2009.

[8] B. Balaguer, S. Balakirsky, S. Carpin, and A. Visser, "Evaluating maps produced by urban search and rescue robots: lessons learned from robocup," *Autonomous Robots*, vol. 27, pp. 449–464, 2009.

[9] A. Birk, "A quantitative assessment of structural errors in grid maps," *Autonomous Robots*, vol. 28, pp. 187–196, 2010.

[10] S. Schwertfeger and A. Birk, "Evaluation of map quality by matching and scoring high-level, topological map structures," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE Press, 2013.

[11] ——, "Map evaluation using matched topology graphs," *Autonomous Robots*, pp. 1–27, 2015. [Online]. Available: http://dx.doi.org/10.1007/s10514-015-9493-5

[12] S. Schwertfeger, "Robotic mapping in the real world: Performance evaluation and system integration." Ph.D. dissertation, Department of Electrical Engeneering and Computer Science, Jacobs University Bremen, 2012. [Online]. Available: http://www.jacobs-university.de/phd/files/phd20120706_Schwertfeger.pdf

[13] A. Nuechter, O. Wulf, K. Lingemann, J. Hertzberg, B. Wagner, and H. Surmann, "3d mapping with semantic knowledge," in *RoboCup-2005: Robot Soccer World Cup IX*, ser. Lecture Notes in Artificial Intelligence (LNAI), A. Bredenfeld, A. Jacoff, I. Noda, and Y. Takahashi, Eds. Springer Verlag, 2006, vol. 4020.

[14] K. Pathak, A. Birk, N. Vaskevicius, M. Pfingsthorn, S. Schwertfeger, and J. Poppinga, "Online 3d slam by registration of large planar surface segments and closed form pose-graph relaxation," *Journal of Field Robotics, Special Issue on 3D Mapping*, vol. 27, no. 1, pp. 52–84, 2010.

[15] J. Pellenz, D. Lang, F. Neuhaus, and D. Paulus, "Real-time 3d mapping of rough terrain: a field report from disaster city," in *Safety Security and Rescue Robotics (SSRR), 2010 IEEE International Workshop on*. IEEE, 2010, pp. 1–6.

[16] S. Schwertfeger and A. Birk, "A short overview of recent advances in map evaluation," in *IEEE International Symposium on Safety, Security, Rescue Robotics (SSRR)*. IEEE Press, 2012.

[17] M. Chandran-Ramesh and P. Newman, "Assessing map quality using conditional random fields," in *Field and Service Robotics, Springer Tracts in Advanced Robotics*, C. Laugier and R. Siegwart, Eds. Springer, 2008.

[18] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, Oct 2012, pp. 573–580.

[19] A. Handa, T. Whelan, J. McDonald, and A. Davison, "A benchmark for rgb-d visual odometry, 3d reconstruction and slam," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, May 2014, pp. 1524–1531.

[20] D. C. Lee, *The Map-Building and Exploration Strategies of a Simple Sonar-Equipped Mobile Robot: An Experimental, Quantitative Evaluation*, ser. Distinguished Dissertations in Computer Science. Cambridge University Press, 2003.

[21] S. Schwertfeger, A. Jacoff, C. Scrapper, J. Pellenz, and A. Kleiner, "Evaluation of maps using fixed shapes: The fiducial map metric," in *Proceedings of PerMIS*, 2010.

[22] S. Schwertfeger, A. Jacoff, J. Pellenz, and A. Birk, "Using a fiducial map metric for assessing map quality in the context of robocup rescue," in *International Workshop on Safety, Security, and Rescue Robotics (SSRR)*. IEEE Press, 2011.

[23] R. Sheh, T. Kimura, E. Mihankhah, J. Pellenz, S. Schwertfeger, and J. Suthakorn, "The robocuprescue robot league: Guiding robots towards fieldable capabilities," in *Advanced Robotics and its Social Impacts (ARSO), 2011 IEEE Workshop on*. IEEE, 2011, pp. 31–34.

[24] R. Sheh, A. Jacoff, A.-M. Virts, T. Kimura, J. Pellenz, S. Schwertfeger, and J. Suthakorn, "Advancing the state of urban search and rescue robotics through the robocuprescue robot league competition," *8th International Conference on Field and Service Robotics*, 2012.

[25] A. Jacoff, R. Sheh, A.-M. Virts, T. Kimura, J. Pellenz, S. Schwertfeger, and J. Suthakorn, "Using competitions to advance the development of standard test methods for response robots," in *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*. ACM, 2012, pp. 182–189.

[26] D. Girardeau-Montaut, "Détection de changement sur des données géométriques tridimensionnelles," Ph.D. dissertation, Télécom Paris-Tech, 2006.

[27] S. Kohlbrecher, K. Petersen, G. Steinbauer, J. Maurer, P. Lepej, S. Uran, R. Ventura, C. Dornhege, A. Hertle, R. Sheh, , and J. Pellenz, "Community-driven development of standard software modules for search and rescue robots," in *Proceedings of the 10th IEEE International Symposium on Safety Security and Rescue Robotics (SSRR 2012)*, 2012.

[28] S. Kohlbrecher, J. Meyer, O. von Stryk, and U. Klingauf, "A flexible and scalable slam system with full 3d motion estimation," in *Proc. IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR)*. IEEE, November 2011.

[29] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An efficient probabilistic 3D mapping framework based on octrees," *Autonomous Robots*, 2013, http://octomap.github.com.