Evaluation of Maps using Fixed Shapes: The Fiducial Map Metric

Sören Schwertfeger[*]
National Institute of Standards and Technology
Gaithersburg, MD USA
s.schwertfeger@jacobs-university.de

Adam Jacoff
National Institute of Standards and Technology
Gaithersburg, MD USA
adam.jacoff@nist.gov

Chris Scrapper
The MITRE Corporation
McLean, VA, USA
cscrapper@mitre.org

Johannes Pellenz
Active Vision Group
University of Koblenz-Landau
Koblenz, Germany
johannes.pellenz@uni-koblenz.de

Alexander Kleiner
University of Freiburg
Freiburg, Germany
kleiner@informatik.uni-freiburg.de

ABSTRACT

Mapping is an important task for mobile robots. Assessing the quality of those maps is an open topic. A new approach on map evaluation is presented here. It makes use of artificial objects placed in the environment named "Fiducials". Using the known ground-truth positions and the positions of the fiducials identified in the map, a number of quality attributes can be assigned to that map. Those attributes are weighed to compute a final score depending on the application domain. During the 2010 NIST Response Robot Evaluation Exercise at Disaster City an area was populated with fiducials and different mapping runs were performed. The maps generated there are assessed in this paper demonstrating the Fiducial approach. Finally this map scoring algorithm is compared to other approaches found in literature.

Categories and Subject Descriptors

D.2.8 [**Software Engineering**]: Metrics—*Complexity measures, Performance measures*; I.2.9 [**Artificial Intelligence**]: Robotics—*Miscellaneous*

General Terms

Performance, Algorithms

---

[*]Mr. Schwertfeger is also a PhD Student at the Jacobs University Bremen, Germany

Keywords

Map evaluation, SLAM

1. INTRODUCTION

Mobile robots often generate maps. Those maps are usually used to enable the robot to perform certain tasks like, for example, autonomous navigation using path planning. Maps also assist an operator of a remotely teleoperated robot in locating the robot in the environment. They do this by providing information of features of interest like corners, hallways, rooms, objects, voids, landmarks, etc. Those features are referenced in the map in a global coordinate system defined by the application. This frame of reference can be a geographic coordinate system of the earth or a local one defined by the application (e.g., robot start pose or pose of an operator station).

Maps generated by mobile robots are abstractions of the real world which always contain inaccuracies or errors. Especially on extended missions or in rough terrain maps often contain large errors. But the usefulness of a map not only depends on its quality but also on the application. In some domains certain errors are neglectable or not so important. That is why there is not one measurement for map quality. Different attributes of a map should be measured separately and weighed according to the needs of the application. Those attributes can include:

- Coverage: How much area was traversed or visited.

- Resolution quality: To what level or detail are features visible.

- Global accuracy: Correctness of positions of features in the global reference frame.

- Relative accuracy: Correctness of feature positions after correcting (often the initial error of) the map reference frame.

- Local consistencies: Correctness of positions of different local groups of features relative to each other.

Figure 1: Picture of the maze in which the experiments were performed. Also visible are two barrels that are used as fiducials.

- Topological accuracy: To what level can directions (e.g. "go left at the 2nd crossing") be extracted?

Note that the resolution quality not only depends on the actual size of the gird cells of the map but is also influenced by the quality of the localization. If there are pose errors between scans of the same object its features blur or completely vanish. Depending on the groups chosen there can be multiple local accuracies.

Many factors influence the quality of a map:

- Environment: Sparse environments are difficult for most mapping approaches. These could be wide open areas or hallways with minimal features.

- Robot path: The path that a robot took to gather the sensor data could contain different amounts of loops. Furthermore, the terrain could be even or difficult to traverse, causing the robot to roll and pitch.

- Robotic platform: Features like active sensing or suspended locomotion can increase map quality. Parts of the robot in the field of view of the sensors are disadvantageous for mapping.

- Sensors: The range, field of view, structural errors, accuracy and the position of the sensor on the robot influence mapping algorithms.

- Computational power: Simultaneous Localization And Mapping (SLAM) [5] algorithms can be computationally very intensive. If the map has to be generated online to aid other tasks like path planning, processing time is often sparse on mobile robots. Then less scans can be used, or the number of particles in a particle filter is reduced, or loop closing and graph optimization algorithms are executed less frequently.

- Algorithm: The mapping algorithm itself influences the map quality to a great extent.

2. FIDUCIAL APPROACH

Using fiducials is a new approach to solve the problem of evaluating the quality of maps. Fiducials are artificial features placed in the environment which can be identified in the resulting map. Using the Fiducial approach, several of the above mentioned attributes can be measured to assess the quality of maps. The only information needed to score maps are the (ground-truth) positions of all fiducials in the environment. Upon scoring, each fiducial has to be identified in the map together with its position.

This approach completely abstracts from all other information contained in the map like walls, unexplored and explored areas and other features. However, given a dense enough distribution of fiducials, this method reflects the quality of those features well enough. This is because measuring the localization performance of the SLAM algorithm suffices since applying sensor data to the map given perfect localization is typically easy.

The fiducials used in this paper are cylinders placed in the environment. Those can either be cut in half and (typically) placed on either side of a wall or are separated by a short artificial wall. In the following representations, said cylinders in the actual maps will be refered to as barrels since barrels were used as cylinder approximations in the experiments (see figure 1). Fiducials are then the objects in the actual environment and its model - the ground truth map. As mentioned above there are (usually) two fiducial-parts (A and B) respectively barrel-parts (A and B).

All attributes scored by the Fiducial approach have values between 0 % and 100 % where 0 % means poorest quality while perfect results get a value of 100 %. This allows easy application of dependent weights to the attributes to come to a simple overall score for maps consisting of just one number. The coverage, resolution quality as well as global, relative and local accuracies can be determined using the Fiducial approach. However, the fiducials first have to be identified in the map. Although the Fiducial metric works in principle for 3D maps as well, 2D maps are considered for the rest of the paper.

2.1 Identification of Fiducials in the map

The following steps are performed to find the fiducials in the map and to register their position:

1. **Rasterize**: Render the map to a two-dimensional grid with a sufficiently high resolution (if the map is already present in a raster format this step is obviously already done).

2. **Colorization**: Remove all probabilistic entries in the grid such that there are exactly three color values left:
   - Free (typically white): No obstacle
   - Unknown (typically gray): Unobserved area (e.g. voids, also "content" of barrels)
   - Obstacle (typically black): Obstruction (e.g. walls, barrel)

3. **Identify barrel parts**: Find all obstacles which form parts of circles with the correct radius. The minimum visible angular opening of the circle has to be 2/3 of the actual opening of that part of the barrel.

4. **Assignment**: For each fiducial part, assign one or none of the barrel parts identified in the previous step.

Each of those barrel parts can be assigned to maximum one fiducial part.

5. **Determine Position**: For each barrel part assigned to a fiducial part compute the position of the center point of the circle forming the barrel. This is then the position of the barrel part. Thus the positions of two parts of a cut-in-half-barrel are just separated by the thickness of the wall.

All of the steps above could be computed by applying appropriate algorithms. For example, feature extraction methods like circular Hough transform [2] could be used to find the barrel parts. The problem of assigning barrel parts to fiducial parts could be solved by finding the mapping that resolves in the best global or relative accuracy. Since those algorithms were not implemented yet in the tools used for the experiments presented in this paper, both steps where done by hand.

The following three attributes use distances between two positions to measure the error. For those there are maximum distances $d_{max}^{attribute}$ defined which are considered to be the worst case for the attribute. The values can, but don't have to, be the same for those three attributes. Furthermore the actual distance error d can be discretized to certain values, for example the barrel radius. This can be done in order to avoid differences in scoring which are caused by the inherent error of the ground truth data and to put the resulting scores in bins of similar qualities.

2.2 Global Accuracy

For every barrel-part assigned to a ground-truth-fiducial part calculate the distance d to the (global) position of the corresponding fiducial. Distances d greater than $d_{max}^{accuracy}$ are set to $d_{max}^{accuracy}$. The error e is then calculated as $e = \frac{d}{d_{max}^{accuracy}}$. Average over the errors for all those barrel parts. The value for the global accuracy is then $1 - e$ such that perfect maps get a 100 % number.

2.3 Relative Accuracy

The error of the global accuracy is minimized (or the accuracy value maximized) by rotating, translating or even scaling the map. This can be easily done by just changing the poses of the barrel parts, thus eliminating the identification step for each iteration. Often the value for the transformation is just the error in the start pose. If there was no agreement on a global frame of reference (as in the following experiments) only the relative accuracy can be computed while there can be no score for the global accuracy.

2.4 Local Consistencies

For all groups calculate the distance errors between entries of a group.

In the following experiments the two parts of a fiducial form a group. Those groups are either classified as short range or long range depending on the minimum distance a robot would have to travel in order to see both barrel parts.

For each pair/group where at least one barrel part has been found:

1. Calculate the geometric distance between the positions of the two barrel parts A and B: d_{barrel}.

2. If one of the barrel parts was not identified in the map, set d_{barrel} to a very high value.
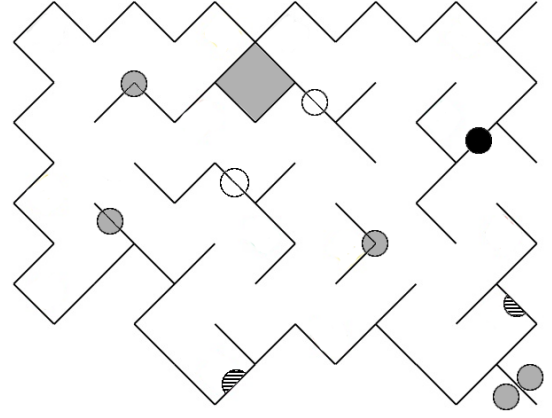

Figure 2: The maze and the fiducials in the ground truth map. The ranges needed to traverse from one fiducial part to the other correspond to the color of the void: gray=four pallets; white=eight pallets; black=twelve pallets; striped=single barrel (no group = no distance).

3. Get the distance between the two corresponding (ground truth) fiducial parts: $d_{fiducial}$.

4. The absolute value of the difference of the distances from step 1) and 2) is the error for this group: $e = min(d_{max}^{consistency}, |d_{barrel} - d_{fiducial}|)/d_{max}^{consistency}$.

The "short range consistency" is thus one minus the average of the error of all short range groups while the "long range consistency" is one minus the average error of the long range groups.

Using barrels or half-barrels on opposite sides of walls has two advantages. First it is very easy to judge the quality of the those pairs by just looking at the map and checking if those barrels are properly aligned and form a good circle without big gaps. This already allows a user to quickly assess a map score without any algorithmic computations.

Second, one can very easily measure the ground truth distance between those fiducial parts. Thus, even when the ground truth positions of the fiducials are unknown or their measurement contains a great error, one can still compute very accurate local consistency scores. For barrels which are simply cut in half and placed on either side of a wall their distance is equal to the thickness of the wall.

Other local consistencies are also possible, for example all fiducials in one room or area.

2.5 Coverage

The ratio of the number of fiducial parts assigned to a barrel part to the total number of fiducial parts. So a value of 100 % means that all fiducials have been mapped while for an error value of 0 no barrels have been found.

2.6 Resolution Quality

The resolution quality can be assessed by measuring the quality of the barrel voids. For each barrel part which is assigned to a fiducial part calculate the area marked as unknown. This can be done by rendering or scaling the map
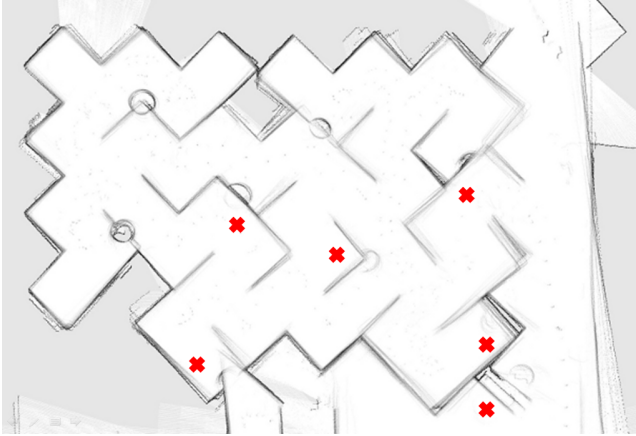
Figure 3: Map 1

and the ground truth to the same scale and simply counting unknown pixels.

The area is summed up for all those barrel parts assigned to a fiducial part. The ratio of the this sum of unknown areas to the actual (unknown) area of all according fiducial parts is the value for this attribute.

This attribute has some problems caused by sensors seeing below or above the actual barrels in the environment and is thus less accurate than the attributes mentioned earlier.

3. EXPERIMENTS

The data presented here has been gathered during the Response Robot Evaluation Exercise Disaster City 2010 [6]. There a maze in a building on an inclined plane has been mapped as well as an adjoining hall and the area in front of this building. Figure 2 shows the ground truth map for the maze. The two different types of fiducials described later have been applied there. Only for the maze the exact poses of the fiducials are known and thus only this part of the maps is used to evaluate their quality in this paper. The software used is a second generation of the Jacobs Map Analysis Toolkit [11].

As fiducials, barrels with a radius of thirty centimeters and a height of one meter are used. They come in two different configurations.

Percent Fiducials consists of two barrels and one piece of square plywood (about 1.2m x 1.2m or 4ft x 4ft). Those are mainly used outdoors where walls are less present.

Wall Fiducials are built by cutting one barrel in half and putting both halves on opposite sides of a wall, forming a nearly exact circle when viewed from the top. They come also in a variation where the barrel is cut into a 1/4th and a 3/4th piece which are placed on corners.

The sensors used are a Hokuyo UTM-30LX laser range finder (LRF) with a field of view of $270°$, an angular resolution of $0.25°$ and a range of above 30 m as well as a Xsens MTi gyro and accelerometer.[1] Those were mounted on a stick and connected to a Laptop. The sensor data was col-

---

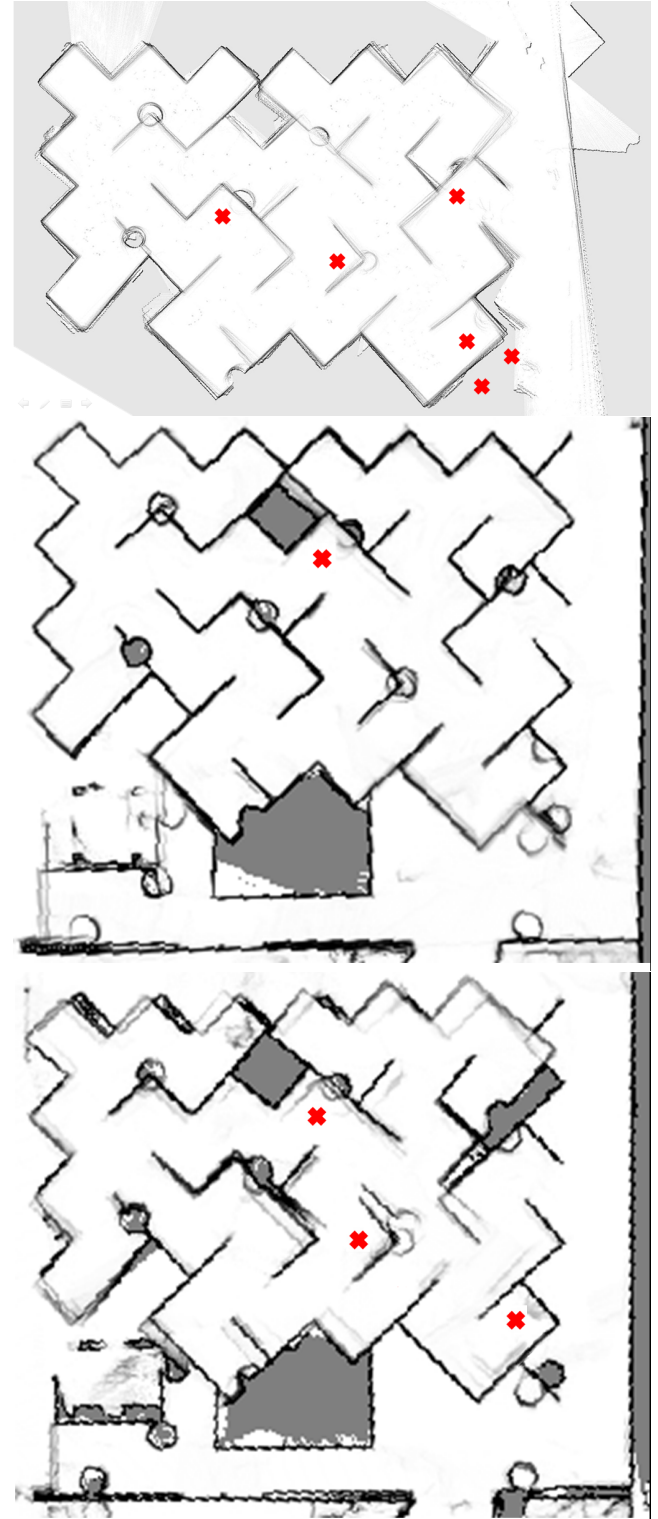[1] Any mention of commercial products is for information only; it does not imply recommendation or endorsement by NIST.
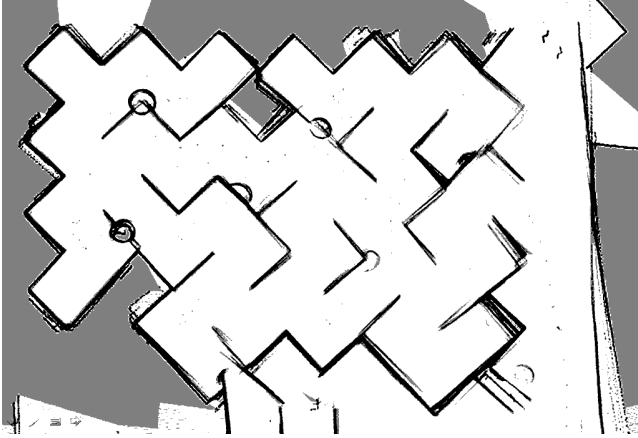

Figure 4: Maps 2, 3 and 4

Figure 5: Map 1 colored to free (white), unknown (gray) and occupied (black) according to section 2.1.

Map	Barrel parts	Short-range pairs	Long-range pairs
1	10	3 of 6	1 of 3
2	10	3 of 6	1 of 2
3	15	6 of 6	2 of 3
4	13	4 of 6	2 (1) of 3

Table 1: Count of identified barrels in the second column. Complete and found short- and long-range groups in the third and fourth column (e.g. 3 of 6 means that for 6 groups at least one barrel was found while for three groups both barrels are identified). One of the long-range groups of Map 4 has a distance d bigger than $d_{max}^{consistency}$.

lected by a person holding the stick with the sensors slowly walking through the maze and the environment.

Two different mapping algorithms were tested. Since the programs did not use a common data format the data was collected repeatedly using different paths and different persons. Two maps were created for each mapping algorithm. Those maps are shown in Figures 3 and 4.

4. RESULTS

This section compares the results of Fiducial map scoring for the four maps gathered. In Figure 2, the ground truth map including the location of the fiducials is shown. There are a total of 16 fiducial parts present. 14 of those form seven groups (pairs) since they are on opposite sides of walls while two don't belong to a complete fiducial.

Each group was assigned a distance, measured in "pallets" (the 1.2 m square area for each element). This distance reflects the minimum number of pallets that has to be traversed to get from on part to the other in the group. The values are 12 for one fiducial, 8 pallets for two more and 4 for the other four fiducials. The 12 and 8 pallet groups are used for the long range consistency while the four four pallet groups are used for the short range consistency.

4.1 Coverage and Local Consistency Results

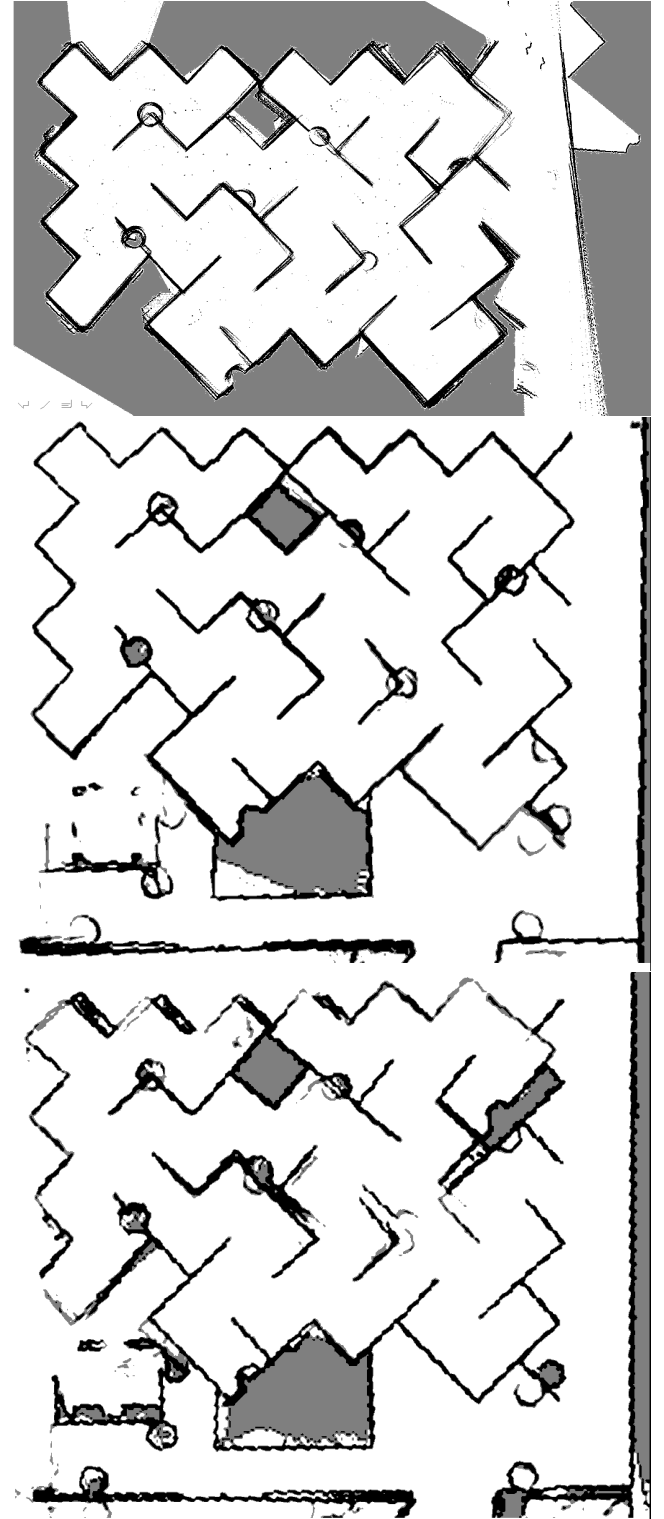Figures 5 and 6 show the maps after the colorization step

Figure 6: Maps 2, 3 and 4 colored.

Map	Barrel parts	Total distance (barrel radii)	Average error (barrel radii)	Average error (cm)
1	10	3.5	0.35	10.5
2	10	1.7	0.17	5.1
3	15	2.2	0.15	4.5
4	13	5.1	0.39	11.8

Table 2: Measured distances/ errors for the identified barrel parts for the relative accuracy attribute.

Map	Barrel Count Unknown > 50 %
1	20 %
2	20 %
3	40 %
4	46 %

Table 3: Unknown void areas. Barrel parts which are filled with at least 50 % unknown area are simply counted.

Map	Coverage	Consistency		Relative Accuracy	Average
		Short	Long		
1	63 %	50 %	33 %	83 %	57 %
2	63 %	50 %	50 %	92 %	64 %
3	94 %	100 %	66 %	93 %	88 %
4	81 %	75 %	33 %	81 %	68 %

Table 4: Results of some attributes. The coverage, short-range and long-range consistencies as well as the relative accuracy are shown.

has been explored in all mapping runs. The reason for this is, that the fiducials do not appear in a good enough quality in the first two maps. So it has to be noted that the coverage attribute of the Fiducial approach only measures the area covered with good enough map quality.

The consistency values reflect the map quality quite well. The best map in the set, Map 3, achieves the highest score. It can also be seen that the short-range consistency is always at least better than the long-range consistency, which is an expected result. The short-range consistency values for the first two maps again suffer from the unrecognizable fiducials. But Map 2 did not see any of the percent-fiducials. Those were also not counted for the long-range consistency such that the average value for the remaining two groups is better. The broken barrel in Map 4 is a long range one and has, in accordance with the algorithm, lowered the long-range consistency score.

The attributes from table 4 were averaged in the last column, giving each attribute the same weight. The average results reflect the subjective map quality quite well such that the Fiducial map scoring algorithm seems to be a viable algorithmic metric for mapping algorithms.

5. COMPARISON WITH PREVIOUS WORK

The analysis of robot generated maps is a relatively new field. Chandran-Ramesh and Newman [4] work in 3D on planar patches. They detect suspicious and plausible arrangements of those planes and classify the map accordingly. A 2D version of their approach could work on lines instead of planes. 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.

Kümmerle et al. [7] proposed a method that does not compare the maps themselves but compares the robot path estimated by the SLAM algorithms with the ground truth path of the robot. This method is an excellent map metric if the preconditions that this algorithm demands are met. Instead of having to have a ground truth map now a ground truth robot path is needed. As discussed in the paper that usually implies human involvement. Especially in an event like RoboCupRescue, where the same environment is explored by different robots (using completely different paths) having a ground truth map representation is more easy to get. The second problem is that, next to the actual map, one needs to obtain the actual pose estimations from the SLAM algorithm. Wulf et al. [13] suggest a similar method. Here the ground truth path is generated using manually supervised Monte Carlo Localization of the 3D scans working on surveyed reference maps. A similarity of both algorithms with the Fiducial metric is the assumption that correct localization is a sufficient enough indicator for a good map

from section 2.1. In Table 1, the counts for identified barrels are provided. The crosses in Figures 3 and 4 mark the missing fiducial parts. Those are used to calculate the coverage as shown in table 4. The first table furthermore contains data for the local consistencies. For the sake of simplicity, the value for $d_{max}^{consistency}$ and for the discretization are chosen to be one barrel radius (30 cm). This way it is easy to just count all those groups/ pairs which are within said distance.

4.2 Relative Accuracy Results

Table 2 contains the data for the relative accuracy. The optimization step was done by hand by overlaying the maps with the ground truth map such that a best possible fit was achieved. The distances between the barrel- and fiducial parts were summed up. The chosen value of the barrel diameter (60 cm) for $d_{max}^{accuracy}$ was in no case exceeded.

Table 4 also contains the values for the relative accuracy calculated after the formula from 2.2. Global accuracy cannot be calculated because there was no global frame of reference.

4.3 Resolution Quality Results

Because of the problem mentioned before, where laser scans can hit the wall below or above the barrel, the results of this attribute are not satisfying. Even in the best map (Map 3), eight of the 16 barrel parts are completely white. For the sake of completeness the results of this calculation are presented in table 3.

4.4 Result Discussion

In this section the results of the Fiducial approach as shown in table 4 are discussed.

The $1/4^{th}$ barrels are more difficult to map and identify than the other sizes. This could be used to measure the Resolution Quality. But those might be too difficult to automatically detect, such that they should be avoided in future experiments.

The coverage values for Maps 1 and 2 are significantly lower than the other two. But the maze, the area of interest,

quality.

Lakaemper and Adluru [8] use a comparison with a ground truth map. They create Virtual Scans out of the target map and measure an alignment energy as map metric. The approach makes use of line- and rectangle-detection which might not be available in unstructured environments. It measures topological correctness and can also quantify the global correctness, just like the Fiducial approach does using the Global Accuracy and the Local Consistencies.

The metric of Wagan, Godil and Li [12] is a feature based approach comparing a map to a ground truth map. A Harris corner detector, the Hough transform and Scale-Invariant Feature Transform (SIFT) are used to extract features from both maps. The features are pairwise matched based on their distance. The quality measure is the number of matched features in the map versus the number of ground truth features. The feature detectors are quite vulnerable to the actual method of rendering the LRF scans. As mentioned in the paper, already changes like noise, jagged lines and distortions pose problems to the feature detectors. This approach can thus only be applied to nearly perfect maps.

Pellenz and Paulus [9] also propose to use feature extraction. Next to using Speeded Up Robust Features (SURF) they also extract rooms as features from the gird map. The extracted rooms are matched with those from the ground truth map, taking into account the topology and the size and shape of the rooms. The average match error per feature is then the quality metric. This work and the Fiducial approach have in common, that they use few, large and easy to detect features (barrel/ room). The rooms have to be fully mapped in order to be used. Distortions could happen within one (bigger) room, potentially hindering the correct mapping to the ground truth room. The use of easy to distribute fiducials seems to be advantageous, since their density can be easily controlled and they can be concentrated in areas of interest. Barrels are small enough that they can be registered with a single scan but still big enough to show in low resolution maps. The room detection will not work in unstructured environments.

Balaguer et al. [1] presents the solution used for the problem of scoring maps in the RoboCup Rescue Virtual Robots competition. There different aspects of the map as well as additional information like features are scored. One of them is the Skeleton Quality which represents the topological structure of the rendered map. The number of false positives and false negatives determine this part of the score, where a false positive is defined as a node which cannot be accessed while a false negative is a clear topological location which is available but has not been included in the skeleton map. The Fiducial approach indirectly supports a similar topological quality measure, if Local Consistency groups with according fiducials are created.

Varsadan, Birk and Pfingsthorn [11] use an image based approach. The Ψ-similarity calculates the average Manhattan distance from the pixels of the ground truth map to the nearest pixel of the same color in the map and vice-versa. This approach only measures the global accuracy and is prone to noise and distortions and can thus only applied to very good maps.

Birk [3] computes one interesting attribute of maps as map metric - the level of brokenness. This structural error is measured by using the Ψ-similarity to find parts of the map which don't fit. This area is then cut-out and registered with the ground truth, thus finding the frame of reference. This is repeated till all of the map is registered. The count of generated sub-maps is the brokenness. This brokeness information can, to some degree, also be found in topological map attributes, for example by comparing different Local Consistencies with each other or the Relative Accuracy.

One big advantage of the Fiducial approach is the low amount of ground truth information needed in oder to compute a score. Kümmerle et al. [7] need a ground truth path while the others use a ground truth map. In the proposed approach just the fiducial positions relative to each other (for Global Accuracy in a global reference frame) have to be provided. If only the Local Consistencies are to be scored, in the proposed wall-barrel-system, the only information needed is the thickness of the walls. The metric can be fully automated while still allowing quick quantitative assessments of the map quality by just looking at the image of the map. The only part of the maps actually evaluated are the fiducials. Thus, as long as the fiducials are detectable in the map, all other mapping errors like noise or broken parts don't effect this algorithm. This metric does not rely on naturally occurring features, although those could be used if they are dense and large enough. This is also the biggest disadvantage of the Fiducial approach, meaning that only environments with such fiducials can be scored.

6. CONCLUSIONS

This paper introduces the problem of map evaluation and proposes a novel approach to it: the Fiducial map metric. The different attributes of this metric are presented and exercised on four maps. Those maps where gathered during the 2010 NIST Response Robot Evaluation Exercise at Disaster City in an environment populated with fiducials in the form of barrels. The resulting numbers for the different attributes and maps support the Fiducial approach. In the previous section the Fiducial metric is compared with other metrics. Demanding low efforts regarding the collection of ground truth data, being robust against many map errors and measuring many different map attributes the Fiducial Map Metric performs well in the task of map scoring, as long as fiducials are present in the environment.

Further work is planned by applying the Fiducial metric to maps from RoboCup 2010, especially in the context of the RoboCupRescue Interleague Mapping Challenge [10]. Another important task is to fully automate all steps of the algorithm. Additional work can also be done by using identified fiducials to form topological paths using ground truth information and scoring the quality of those paths.

Acknowledgments

7. REFERENCES

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

[2] D. H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(3):111–122, 1981.

[3] A. Birk. A Quantitative Assessment of Structural Errors in Grid Maps. *Autonomous Robots*, 28:187–196, 2010.

[4] M. Chandran-Ramesh and P. Newman. Assessing Map Quality using Conditional Random Fields. In C. Laugier and R. Siegwart, editors, *Field and Service Robotics, Springer Tracts in Advanced Robotics*. Springer, 2008.

[5] H. Durrant-Whyte and T. Bailey. Simultaneous localization and mapping: Part I. *IEEE Robotics and Automation Magazine*, pages 99–108, June 2006.

[6] A. Jacoff and E. Messina. DHS/NIST response robot evaluation exercises. In *IEEE International Workshop on Safety, Security and Rescue Robotics, SSRR*, Gaithersburg, MD, USA, 2006.

[7] R. Kümmerle, B. Steder, C. Dornhege, M. Ruhnke, G. Grisetti, C. Stachniss, and A. Kleiner. On measuring the accuracy of slam algorithms, 2009.

[8] R. Lakaemper and N. Adluru. Using virtual scans for improved mapping and evaluation. *Auton. Robots*, 27(4):431–448, 2009.

[9] 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.

[10] S. Schwertfeger. Robocuprescue interleague mapping challenge. http://robotics.jacobs-university.de/mappingChallenge/.

[11] I. Varsadan, A. Birk, and M. Pfingsthorn. Determining Map Quality through an Image Similarity Metric. In L. Iocchi, H. Matsubara, A. Weitzenfeld, and C. Zhou, editors, *RoboCup 2008: Robot WorldCup XII, Lecture Notes in Artificial Intelligence (LNAI)*. Springer, 2009.

[12] A. I. Wagan, A. Godil, and X. Li. Map quality assessment. In *PerMIS '08: Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, pages 278–282, New York, NY, USA, 2008. ACM.

[13] O. Wulf, A. Nuchter, J. Hertzberg, and B. Wagner. Ground truth evaluation of large urban 6d slam. pages 650 –657, oct. 2007.