Using a Fiducial Map Metric for Assessing Map Quality in the context of RoboCup Rescue

Sören Schwertfeger, Adam Jacoff, Johannes Pellenz and Andreas Birk

Abstract— **Mapping is an important task for mobile robots in general and for Safety, Security, and Rescue Robotics (SSRR) in particular. It is hence one core aspect which is evaluated in the RoboCup Rescue league. But assessing the quality of maps in a simple and efficient way is not trivial, especially if no detailed, complete ground truth data of the environment is available. 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. Depending on the application domain those attributes can weighed to compute a final score. Results are presented that are based on using this method during the RoboCup Rescue competition 2010 in Singapore where maps were generated by different teams in an maze populated with fiducials. Those maps are evaluated here and compared to a human judgment, showing the effectiveness of the fiducial approach.**

I. Introduction

Mapping is an important task for mobile robots, because maps are often essential to enable the robot to perform its tasks [16]. One prominent example for such a task is autonomous navigation using path planning. Maps also assist an operator of a remotely teleoperated robot in locating the robot in the environment 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). Furthermore, maps are in the context of Safety, Security, and Rescue Robotics (SSRR) not only an important mean for robot control but they are often also a crucial mission deliverable [3]. It is hence of interest to be able to assess the quality of maps, respectively mapping approaches, to identify working solutions as well as open problems.

Maps generated by mobile robots are abstractions of the real world, which always contain inaccuracies or errors. There has been great progress in mapping in the last two decades, especially with respect to Simultaneous Localization and Mapping (SLAM) techniques. But especially in scenarios that are of high interest for SSRR, namely on extended missions, respectively in unstructured environments, maps still often contain large errors. Furthermore, the usefulness of a map not only depends on its quality but also on the application [9]. In some domains certain errors are negligible 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 [8]. Those attributes can include:

- Coverage: How much area was traversed/visited.
- Resolution quality: To what level/detail are features visible.
- Global accuracy: Correctness of positions of features in the global reference frame.
- Relative accuracy: Correctness of feature positions after correcting (the initial error of) the map reference frame.
- Local consistencies: Correctness of positions of different local groups of features relative to each other
- Topological accuracy: To what level can directions (e.g. "go left at the 2nd crossing") be extracted?

Note that the resolution quality is not only depending 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 consistencies.

Many factors influence the quality of a map:

- Environment: Sparse environments are difficult for most mapping approaches. Examples include wide open areas or hallways with minimal features.
- Robot path: The path that a robot took to gather the sensor data could contain different numbers of loops. Furthermore could the terrain 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 while parts of the robot in the field of view of the sensors are disadvantageous.
- Sensors: The range, field of view, structural errors, accuracy and the position of the sensor on the robot are factors influencing mapping algorithms.
- Computation power: SLAM algorithms can be computational very intensive. If the map has to be generated on-line to aid other tasks like path planning processing time is often scarce 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.

Sören Schwertfeger and Andreas Birk are with the Dept. of Computer Science, Jacobs University Bremen, 28759 Bremen, Germany [s.schwertfeger, a.birk]@jacobs-university.de

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

Johannes Pellenz is with the Bundeswehr Technical Center for Engineer and General Field Equipment (WTD 51), 56070 Koblenz, Germany johannespellenz@bwb.org

- Algorithm: Of course the mapping algorithm itself influences the map quality to a great extend.

Within RoboCup Rescue [5], there are different competitions to evaluate the performance of robotic systems for search and rescue mission, including their mapping capabilities. While comparing mapping approaches one can try to control some of the above mentioned aspects; the different RoboCup Rescue competitions have different focuses in this respect:

- VirtualRoboCupRescue [4]: Here in a virtual, simulated 3D environment a group of agents (robot models) have to (mainly autonomously) explore and map an area. The organizers control the environment and to a great extend the robot platform, sensors and computation power. Hence, the robot path (autonomy algorithms) and mapping algorithms determine the mapping result.
- RealRescue [15]: Physical robots build by the teams have to search a simulated disaster site (see Figure 1). There is a yellow area for autonomous robots and orange and red areas with greater mobility challenges for teleoperated robots. Here the organizers only control the environment. All other factors vary strongly from robot to robot.
- Interleague Mapping Challenge [11]: Real sensor data (laser range finders (LRF), gyro, potentially odometry) of a run through the arena is collected in the RealRescue maze. Teams are asked to provide their mapping algorithms. The organizers then run their mapping algorithm and build maps using this data. Thus the organizers control everything except the algorithm that each participant is using for mapping.

This paper approaches the task of map evaluation using artificial features called fiducials [12]. The ground truth information about where the fiducials are located, also with respect to each other, is used to calculate the scoring attributes. The significant advantage of this method is that it does not require a complete, detailed ground truth representation (i.e. a perfect maps) of the environment.

This map metric works with maps depicting free space and obstacles in an arbitrary environment (outdoor as well as indoor). The sensors used to create the map are not important to this algorithm, although maps are typically created using 2D or 3D LRF. Nevertheless, any sensors giving information about free space and obstacles such as ultrasound sensors, echo sounders (underwater), stereo-imaging, etc. can be used. As an additional bonus the approach of systematically placing fiducials also enables humans to quickly asses the quality of maps by just looking at them.

The experiments used to demonstrate the usefulness of this metric are performed on maps generated by various teams during the RoboCup Rescue competition 2010 in Singapore with real robots as well as maps from the RoboCup Rescue Interleague Mapping Challenge that used data collected in Singapore.

The analysis of robot generated maps is a relatively new field, nevertheless it has already received some attention. Using ground-truth robot paths [19] and [6] compare those paths with those estimated by SLAM algorithms. But obtaining the actual robot paths is a difficult problem and can only be done in very controlled environments. Most other approaches use accurate maps of the environment as ground truth, which is particularly easy if simulations are used [13], [14]. One can, for example, measure the alignment error of virtual scans in the ground truth map [7] or use image-based techniques. Image similarity methods [17] have their limitations due to the common errors in maps, because maps often have structural errors like noise, structures appearing more than once due to localization errors and the like. Nevertheless, certain attributes like the level of brokenness [2], can still be obtained. The advantage of the fiducial approach compared to the image based ones is, that errors described above do not effect the score as long as the fiducials can be identified in the map.

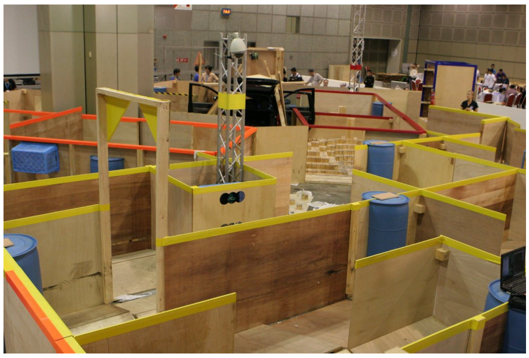

Fig. 1. One part of the RoboCupRescue 2010 arena. Some fiducials (blue barrels) are visible.

Quite often features are detected in the images of the maps. Harris corner detector, the Hough transform and SIFT are used in [18] while SURF and a room detection approach are applied in [10]. Those approaches have in common, that they use the pose of the detected features to determine the map quality. They also require detailed ground truth representations of the environment to be usable. A combination of methods is used in the RoboCup Rescue Virtual Robots competition [1], especially the topological structure of the map is also taken into consideration.

II. Fiducial Approach

Using the Fiducial map evaluation approach, several of the above mentioned attributes can be measured to asses 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. From a practical viewpoint, especially in the case of competitions like RoboCup Rescue where a multitude of maps has to be evaluated in an environment that is changed between runs to ensure variable conditions, this has the great advantage that no detailed complete ground truth representation is required.

This approach completely abstracts from all other information contained in the map like walls, unexplored and explored areas and other aspects because it just takes the fiducials and their position into account. However, given a reasonable distribution of fiducials, this method reflects the quality of those aspects well enough. This is because measuring the localization performance of the SLAM algorithm suffices since incorporating 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 two cylinders are separated by a short artificial wall. In order to ease the understanding, in the following representations of said cylinders in the actual maps will be refereed 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 so to say ground truth map. As mentioned above there are (usually) two fiducial-parts (A and B) respectively barrel-parts (A and B). Circular fiducials are chosen because they stand out nicely in the environment.

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 us to easily apply application 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 and relative accuracies and local consistencies can be determined using the Fiducial approach. However, first the fiducials have to be identified in the map. The Fiducial metric works in principle for 3D as well as 2D maps, if the 3D positions of the barrels can easily be extracted from the 3D maps. For the rest of the paper and the experiments the algorithm is described using 2D maps.

A. *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 can obviously be skipped).
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): Unknown, unobserved area (e.g. voids, also “content” of barrels).
   - Obstacle (typically black): Obstacle (e.g. walls, barrel).
3) **Identify barrel parts**: Find all obstacles which form parts of circles with the right radius. The minimum visible angular opening of the part-circle has to be 2/3rd 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 coordinates of the center point of the circle (-part) visible in the map. 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.

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 maximums 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.

B. *Determining Attributes*

1) 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. So d can be seen as the error in the position between the ground-truth barrel and the one in the map. 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) 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.

3) 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.
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. Firstly one can very easily measure

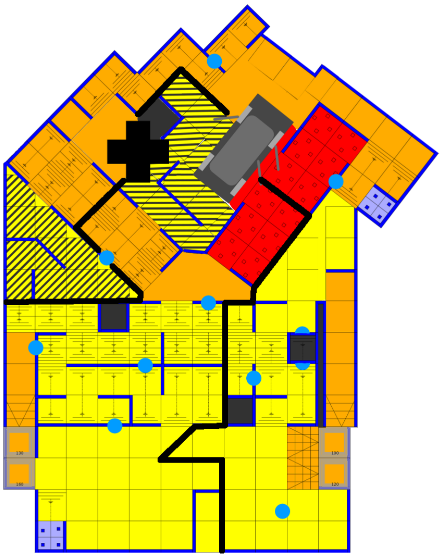

Fig. 2. A ground truth map of the RoboCupRescue arena with iconic representation of the different terrain types. The arena is split into two parts to accommodate two parallel runs during the preliminary rounds. The bold black line shows the border between the two arenas. The blue circles depict the locations of the fiducials.

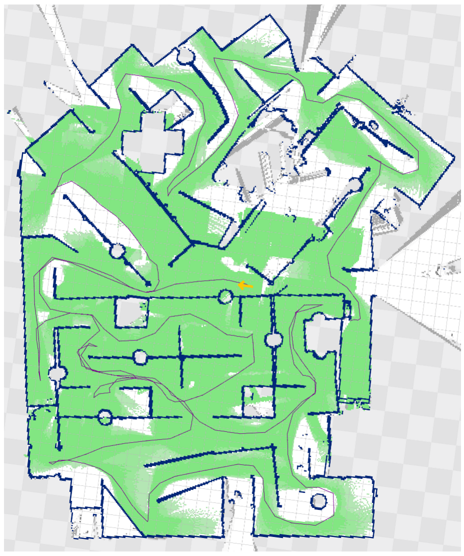

Fig. 3. A very good map created by a human carrying a lightweight robot around. In contrast to Figure 2, the arena is now configured as one big maze for the autonomy challenge. The map is colored after the RoboCup standard, marking white areas as observed but additionally marking areas that were extensively searched for simulated disaster-victims as green. The robot-path, start pose and a 1m grid in the background are also visible.

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.

Secondly 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.

4) 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.

III. Experiments

In order to demonstrate the usefulness of the fiducial map metric several maps are evaluated in the following using this approach. The results are compared to a ranking provided by human judgment, which used to be the method of choice for judging maps in previous years in RoboCup Rescue. The maps were created during the RoboCupResuce World-cup 2010 competition in Singapore as well as the RoboCup Rescue Interleague Mapping Challenge 2010 that used sensor data collected in Singapore. The maze used in the competition is depicted in Figure 1 and a schematic ground-truth map is shown in Figure 2. This ground truth map does not reflect the exact arrangement of the arena. Furthermore, the arena has been reconfigured throughout the competition and the positions of some of the fiducials were also changed. The software used is a second generation of the Jacobs Map Analysis Toolkit [17].

Barrels with a radius of thirty centimeters and a height of one meter are used as fiducials. They are build by cutting one barrel in half and putting the halves on both sides of a wall, forming a nearly exact circle when viewed from the top. One pair is put on both sides of a void, which could be interpreted as a very thick wall (1.2 meter).

Figure 3 shows a map generated using nearly perfect sensor data (a lightweight robot was carried through the maze).

The maps created during the rescue competition were generated by different robotic platforms using different sensors, navigation techniques (autonomous as well as teleoperated) and mapping algorithms. In contrast, for the interleague challenge maps sensor data was collected in the maze and later-on fed to the algorithms.

Due to different arena configurations three groups of maps can be distinguished. For the preliminary rounds in the competition the arena was parted into two parts (A and B) to allow simultaneous runs, while later-on one big maze was available. For part A ten maps were scored while there are 7 maps for part B. For the whole arena 24 maps were tested, of which 16 were generated during the interleague challenge.

The sensors used for the data collection of the interleague challenge are a Hokuyo UTM-30LX LRF with a field of view of 270°, an angular resolution of 0.25° and a range of above 30m as well as a Xsens MTi gyro and accelerometer. Those were mounted on a stick and connected to a laptop. The sensor data was collected by a person holding the stick with the sensors slowly walking through the maze and the environment.

For illustration of the fiducial metric two local consistencies are defined: short-range consistency and long-range consistency. For those the two barrel parts on opposite sides of a wall form one group. Each group was assigned a distance, measured in “pallets” (the 1.2m square area for each element - a 4x4 foot plywood sheet used to build the arena). 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 long-range consistency is thus the average value for all groups which are more than six pallets away from each other, while the short-range consistency is calculated for fiducials with six or less pallets distance.

The value for $d_{max}^{attribute}$ is set to 4 times the radius for the evaluations while the actual distance value d is discretized to the radius (30 cm).

For the dataset used in this experiment, the global coordinates of the fiducials were unknown, and therefore the computation of the global accuracy was skipped. One approach to estimate the positions of the fiducials is to use the well-structured design of the RoboCup Rescue arena and derive the positions from the floor plan as depicted in Figure 2. However, the quality of the maps nowadays sometimes exceeds the accuracy of the so called ”ground truth map”, which does not take into account smaller movements of the wall elements or of the fiducials itself. Another way to determine the ground truth positions of the fiducials is to ”fly” through the arena with a mapping device under very controlled conditions (as done for the map in Figure 3). The resulting maps have a higher quality of the maps produced by the robots, but still depend on the mapping algorithm and the limited accuracy of the laser range finder. As a solution to generate a highly accurate position estimation for the fiducials, we are using the following approach:

1) On each fiducial, attach a bright marker (e.g. a nail with a large white head) with a unique number (see Figure 4).
2) Set up a tachymeter (a precise measurement device used on construction sites) at a location from which all fiducials can be seen.
3) Measure the Cartesian coordinates of the markers attached to the fiducials.

Since modern tachymeters provide the feature to define a reference line first (e.g. a fixed line outside the arena), the positions can be measured in a global coordinate frame from multiple locations. This approach yields very precise results for the reference coordinates for several reasons:

- The accuracy of tachymeters is much higher than the accuracy of laser range finders used on mobile robots.
- The location of the fiducials is inspected from a single or only a few different locations.
- Since it takes only a short time to capture the data, the reference coordinates can measured right before each individual run. This way, also small modifications in the environment can be captured.

Fig. 4. Left: The middle of a fiducial with marker. Right: A tachymeter set up next to a RoboCupRescue arena to measure the position of the fiducials.

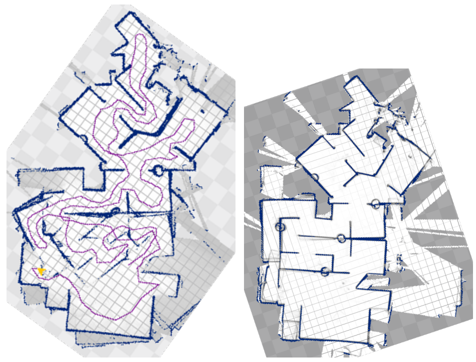

Fig. 5. Maps 1 and 2 from the RoboCupRescue Interleague Mapping Challenge. As a service to the reader, these and all following maps are rotated by the authors to be roughly parallel with the ground truth map.

The use of a tachymeter hence further increases the efficiency - in terms of time it takes to establish a reference base - and the accuracy of the reference; an according device is hence used at RoboCup World Cup 2011 in Istanbul.

IV. Results

In this paper only the results for one group of maps (part A) consisting of ten maps are presented in more detail, while the experiments performed on the other two groups in general further support these results. The maps 1 and 2 (Figure 5) are from the RoboCupRescue Interleague Mapping Challenge. Since the sensor data collected there contained all of this part of the maze a 100% score was possible for this arena. Unfortunately, the algorithm generating map 1 suffered from a severe localization error during its estimations, hence overwriting already correct parts of the map with wrong data.

The other maps (Figures 6 and 7) are from the RoboCupRescue competition. Since no global coordinate system was defined the global accuracy cannot be calculated. The robots took different paths exploring different amounts of the arena, thus corresponding barrels on the other side of walls have often not been mapped (properly). This leads to low scores for the consistency attributes.

Table I show the results of the Fiducial Map Metric. For the average first the two consistency values are averaged and then this value is used to calculate an average score also using the coverage and the relative accuracy. The coverage value is sometimes surprisingly low (e.g. maps 1 and 9) compared to the area actually visible. This is due to the fact that the fiducials are not properly visible in the map due

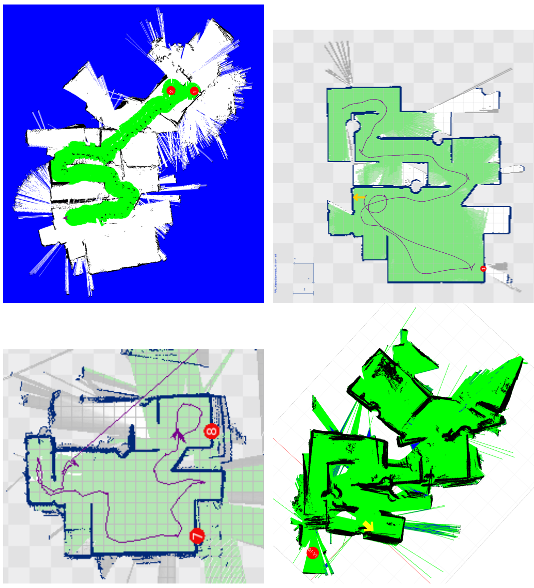

Fig. 6. Maps from the RoboCup Rescue competition: maps 3 (top-left), 4 (top-right), 5 (bottom-left) and 6 (bottom-right).

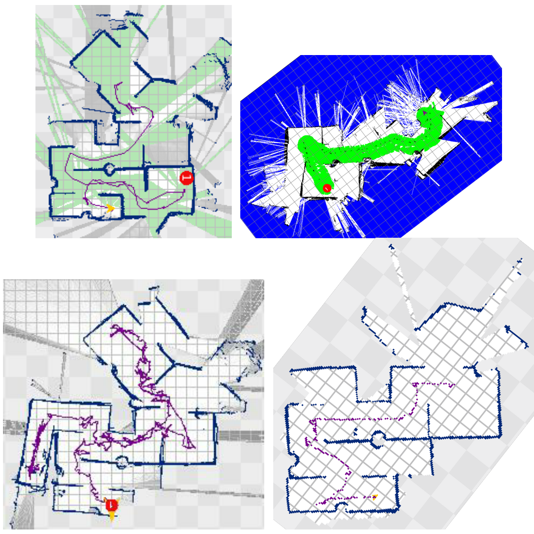

Fig. 7. Maps from the RoboCup Rescue competition: maps 7, 8, 9 and 10.

to mapping errors. A low coverage value also means that the consistency calculation will often miss the corresponding barrel of a fiducial pair, thus generating a low score. If there is, on the other hand, just one pair with a good consistency, the score will be very good. So for maps with a good coverage it is more challenging to get a good consistency value. This is less of a problem with the bigger maps with more fiducials (or a higher fiducial density).

Table II shows the results of the human judgment of the maps. This judgment was not done using the fiducials but using the mapped area (for coverage) and locations and consistencies of walls. As mentioned before, this used to be the method of choice for judging maps in RoboCup Rescue in the years before due to a lack of better alternatives, especially as the generation of ground data for comparison would require a lot of effort. This human judgment of the map quality is of course quite subjective, but the ranking generated out of it should reflect the actual map attributes quite well.

TABLE I

SCORES FOR THE MAP ATTRIBUTES FROM THE FIDUCIAL MAP METRIC.

Map	Coverage	Relative Accuracy	Consistency		Average
			Short-range	Long-range	
1	56%	75%	100%	0%	60%
2	100%	97%	100%	100%	99%
3	44%	94%	100%	0%	63%
4	67%	100%	100%	100%	89%
5	33%	100%	0%	100%	61%
6	67%	71%	38%	25%	57%
7	56%	95%	100%	0%	67%
8	33%	67%	75%	0%	46%
9	44%	81%	100%	50%	67%
10	44%	100%	100%	0%	65%

TABLE II

ATTRIBUTES FOR THE MAPS APPROXIMATED BY A HUMAN JUDGE.

Map	Coverage	Relative Accuracy	Consistency (Topology)	Average
1	80%	50%	70%	67%
2	100%	90%	90%	93%
3	50%	50%	70%	57%
4	60%	90%	90%	80%
5	40%	50%	90%	60%
6	50%	40%	50%	47%
7	50%	90%	90%	77%
8	30%	60%	80%	57%
9	60%	70%	80%	70%
10	40%	90%	90%	73%

The results of the fiducial algorithms and the human scoring are compared in Table III. The first rank in each cell is the one gathered with the Fiducial Map Metric (e.g. tie between places 4 and 5 for coverage of map 1) while the number after the slash is the judges rank (e.g. place 2 for coverage of map 1).

The big differences (two ranks) for coverage for the maps 1, 6 and 9 are again due to the fact that for those maps fiducials are often missing due to mapping inaccuracies. In so far the fiducial metric is even advantageous to the

TABLE III

RANK COMPARISON. THE FIRST RANK IS THE ONE CALCULATED WITH THE FIDUCIAL MAP METRIC WHILE THE RANK AFTER THE SLASH IS THE ONE FROM THE HUMAN JUDGMENT.

Map	Coverage	Relative Accuracy	Consistency (Topology)	Average
1	4-5 / 2	8 / 7-9	4-6 / 8-9	8 / 6
2	1 / 1	4 / 1-4	1-2 / 1-5	1 / 1
3	6-8 / 5-7	6 / 7-9	4-6 / 8-9	6 / 8-9
4	2-3 / 3-4	1-3 / 1-4	1-2 / 1-5	2 / 2
5	9-10 / 8-9	1-3 / 7-9	4-6 / 1-5	7 / 7
6	2-3 / 5-7	9 / 10	9 / 10	9 / 10
7	4-5 / 5-6	5 / 1-4	7-8 / 1-5	3-4 / 3
8	9-10 / 10	10 / 6	10 / 6-7	10 / 8-9
9	6-8 / 3-4	7 / 5	3 / 6-7	3-4 / 5
10	6-8 / 8-9	1-3 / 1-4	7-8 / 1-5	5 / 4

human judgment since only area that has been properly mapped without errors should be counted for the coverage calculation.

For the relative accuracy the values for maps 5 and 8 differ significantly. This is due to the fact that those maps are really small. The few (3) fiducials for map 5 are actually at the right places while the subjective appearance of the map is not so good. The walls for map 8 overlap quite well with the ground truth, but some of the actual fiducial positions are off.

The consistency values differ to up to three ranks. This is due to the relatively low amount of fiducial pairs (two for each short and long range) which yield to extreme results depending on whether the pair is complete or now. For the maps with greater coverage in the of the full sized arena this is effect is less prominent and the results are thus better.

Nevertheless, the average of the results delivers a fairly decent result. As can be seen in Table III, the actual rankings for the averages correspond quite nicely, having a rank difference of one in four cases, of two in two cases and the same rank for four maps.

V. Conclusions

This paper proposes a novel approach for map evaluation: the Fiducial map metric. The different attributes of this metric are presented and exercised on maps gathered during and after the RoboCupRescue World-Cup 2010 in Singapore. The resulting numbers for the different attributes and maps support the Fiducial approach.

One big advantage of the Fiducial approach is the low amount of ground truth information needed in order to compute a score, especially there is no need for a detailed complete ground truth representation of the environment. In the proposed approach, just the fiducial positions relative to each other have to be provided. If the Global Accuracy is to be calculated the fiducial positions also have to be known in a global reference frame. 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 do not effect this algorithm. This metric does not rely on naturally occurring features, although those could be used if they are dense and large enough. A disadvantage of using natural features, such as rectangular corners is, that there are often ambiguities which corner of the environment is shown if the maps contain big localization errors. That is something which can be avoided with not so dense and skillfully placed fiducials. Natural features also do not allow for the very easy measurement of the local consistencies as do the cut-in-half barrels on either sides of a wall.

But this is also the biggest disadvantage of the Fiducial approach, meaning that only environments with such fiducials can be scored. On the other hand, this avoids that the method has to rely on special assumptions about certain environment properties that would have to serve as natural features.

The work presented here works in mainly planar scenarios. For more difficult terrain like piles of rubble a 3D map representation is needed. If fiducials are detectable in those 3D maps this map scoring method can be easily applied to 3D. This remains future work for now.

VI. Acknowledgements

Part of this work has been supported by the German Academic Exchange Service (DAAD) for the cooperation project "Response Robots Performance Evaluation" by the National Institute of Standards and Technology (NIST) and Jacobs University Bremen.

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] A. Birk. A quantitative assessment of structural errors in grid maps. *Autonomous Robots*, 28:187–196, 2010.
[3] A. Birk and S. Carpin. Rescue robotics - a crucial milestone on the road to autonomous systems. *Advanced Robotics Journal*, 20(5):595–695, 2006.
[4] S. Carpin, A. Birk, M. Lewis, and A. Jacoff. *High fidelity tools for rescue robotics: results and perspectives*. Lecture Notes in Artificial Intelligence (LNAI). Springer, 2006.
[5] H. Kitano and S. Tadokoro. Robocup rescue. a grand challenge for multiagent and intelligent systems. *AI Magazine*, 22(1):39–52, 2001.
[6] R. Kümmerle, B. Steder, C. Dornhege, M. Ruhnke, G. Grisetti, C. Stachniss, and A. Kleiner. On measuring the accuracy of slam algorithms, 2009.
[7] R. Lakaemper and N. Adluru. Using virtual scans for improved mapping and evaluation. *Auton. Robots*, 27(4):431–448, 2009.
[8] D. C. Lee. *The Map-Building and Exploration Strategies of a Simple Sonar-Equipped Mobile Robot: An Experimental, Quantitative Evaluation*. Distinguished Dissertations in Computer Science. Cambridge University Press, 2003.
[9] J. J. Leonard and H. F. Durrant-Whyte. *Directed Sonar Sensing for Mobile Robot Navigation*. Springer, 1992.
[10] 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.
[11] S. Schwertfeger. Robocuprescue interleague mapping challenge, 2010. `http://robotics.jacobs-university.de/mappingChallenge/`.
[12] 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.
[13] 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)*, pages 1–6, 2007.
[14] C. Scrapper, R. Madhavan, and S. Balakirsky. Performance analysis for stable mobile robot navigation solutions. In *Proceedings of SPIE*. International Society for Optical Engineering, 2008.
[15] S. Tadokoro, H. Kitano, T. Takahashi, I. Noda, H. Matsubara, A. Shinjoh, T. Koto, I. Takeuchi, H. Takahashi, F. Matsuno, M. Hatayama, J. Nobe, and S. Shimada. The robocup-rescue project: A robotic approach to the disaster mitigation problem. *Proceeding of the 2000 IEEE International Conference on Robotics and Automation*, pages 4089–4094, 2000.
[16] S. Thrun. *Robotic Mapping: A Survey*. Morgan Kaufmann, 2002.
[17] I. Varsadan, A. Birk, and M. Pfingsthorn. *Determining Map Quality through an Image Similarity Metric*, pages 355–365. Springer, 2009.
[18] 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.
[19] O. Wulf, A. Nuchter, J. Hertzberg, and B. Wagner. Ground truth evaluation of large urban 6d slam. pages 650 –657, oct. 2007.

Schwertfeger, S., A. Jacoff, J. Pellenz, and A. Birk, "Using a Fiducial Map Metric for Assessing Map Quality in the context of RoboCup Rescue", IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR): IEEE Press, pp. 1-6, 2011.

http://dx.doi.org/10.1109/SSRR.2011.6106762