

RoboCupRescue Interleague Challenge 2009: Bridging the Gap between Simulation and Reality

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

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

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

ABSTRACT

The RoboCupRescue initiative, represented by real-robot and simulation league, is designed to foster the research and development of innovative technologies and assistive capabilities to help responders mitigate an emergency response situation. This competition model employed by the RoboCupRescue community has proven to be a propitious model, not only for fostering the development of innovative technologies but in the development of test methods used to quantitatively evaluate the performance of these technologies. The Interleague Challenge has been initiated to evaluate real-world performance of algorithms developed in simulation, as well as to drive the development of a common interface to simplify the entry of newcomer teams to the robot league. This paper will discuss 1) the development of emerging test methods used to evaluate robotic-mapping, 2) the development of a common robotic platform, and 3) the development of a novel map evaluation methodology deployed during the RoboCupRescue competition 2009.

Keywords

SLAM, Performance Metric, Simulation, Mapping, Rescue

1. INTRODUCTION

Response robots refer to a broad class of mobile robots intended to assist emergency response personnel in a variety of application domains; such as Urban Search and Rescue (USAR), Explosive Ordnance Disposal (EOD), and Intelligence, Surveillance, and Reconnaissance (ISR). These platforms serve as an extension of the operator to improve remote situational awareness and to provide assistive capabilities that minimizes the risk to responders and maximize the effectiveness and efficiency of a response in a tactical environment. Although recent advancements in the technical capabilities of these robots have improved the flexibility, utility, and survivability of overall system, in large these systems have failed to achieve a technology readiness level

suitable for fielded systems deployed in their respective operational domains.

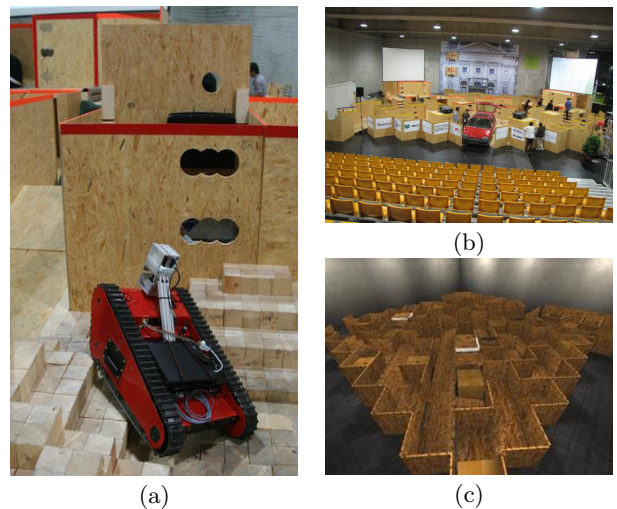


Figure 1: (a) Common robotic platform. (b) 2009 RoboCupRescue Maze in Graz. (c) Virtual maze in USARSim.

Test methods establish a confident connection between developers and consumers regarding the expectations and performance objectives of robotic technologies. This is a cardinal step in fostering innovation and assessing the maturity of evolving technologies. They consist of well-defined testing apparatuses, procedures, and objective evaluation methodologies that isolate particular aspects of a system in known testing conditions [ASTM, 2007]. The development of test methods start with a comprehensive analysis of the application domain to identify requirements with associated metrics and the range of performance, starting from a baseline threshold to the objective “best-case” performance. This analysis provides the basis for developing test methods and testing scenarios that are intentionally abstract so as to be repeatable across a statistically significant set of trials and reproducible by other interest parties. This also provides developers with a basis for understanding the objective performance of a system and allows consumers to confidently select systems that will meet their requirements.

Robotic competitions have also proven to be a propitious model for fostering innovation and assess the maturity emerging robotic technologies. Commonly, test methods provide

(c) 2009 Association for Computing Machinery. ACM acknowledges that this contribution was authored or co-authored by a contractor or affiliate of the U.S. Government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only. PerMIS'09, September 21-23, 2009, Gaithersburg, MD, USA. Copyright 2009 ACM 978-1-60558-747-9/09/09...10.00

the basis for evaluating the performance of robots in the competitions; setup as either elemental tests or embedded in operational environments to produce testing scenarios. For instance, Defense Advanced Research Project Agency (DARPA) and the European Space Agency have employed this model to assess autonomous ground vehicle in urban environments [Darpa, 2007] and in extraterrestrial environments, such as lunar landscapes or Mars exploration [ESA, 2008]. Not only do these competitions provide a means to assess the performance of emerging technologies, they also provide feedback as validity of the tests themselves.

The RoboCupRescue initiative also leverages the competition model to foster the research and development of the key capabilities to assist in the mitigation of an emergency response situation. This initiative partitions the emergency management problem into three competitions; the RoboCupRescue Robot, the RoboCupRescue Virtual Robot Simulation, and RoboCupRescue Agent competitions. Each of the competitions explore the partitioned problem space at different levels of abstraction, ranging from search and rescue of a single building to the development of an Incident Command System. The relevance of this initiative can be gauged according to two aspects: 1) the ability of the competitions to communally develop comprehensive systems capable of achieving an appropriate technology readiness level, and 2) the development of quantitative benchmarks and test methods capable to assess emerging technologies and assisting responders making deployment and purchasing decisions.

Therefore, the 2009 RoboCupRescue competitions sponsored an Interleague Challenge between the Robot competition and the Virtual Robot Simulation competition to demonstrate how well robotic algorithms developed in simulation can perform on a real robot. The challenge utilizes a common robotic platform and emerging standard test methods to explore the stability and accuracy of online mapping technologies, emphasizing the impact on an operator’s ability to efficiently and completely map an unknown environment. This paper will discuss the development of emerging test methods used to evaluate robotic-mapping, the development of a common robotic platform, and the development of a novel map evaluation methodology deployed during the 2009 RoboCupRescue Interleague Challenge.

The remainder of this paper provides an overview of the Interleague Challenge and is structured as follows. Section 2 provides an overview of the test methods and robotic platform employed at the challenge. Section 3 will detail the 2D map evaluation framework used to evaluate the competing maps at the challenge. This map evaluation framework consists of a ground truth generation process and a map assessment tool. The evaluation results from the challenge are presented in Section 4, followed by the conclusion in Section 5.

2. THE INTERLEAGUE CHALLENGE

The evaluation of robot-generated maps is often based on qualitative approaches that do not take into account how specific environmental conditions, differing sensor configurations, and *in situ* decisions impact the performance of the system. While this type of analysis provides some indication

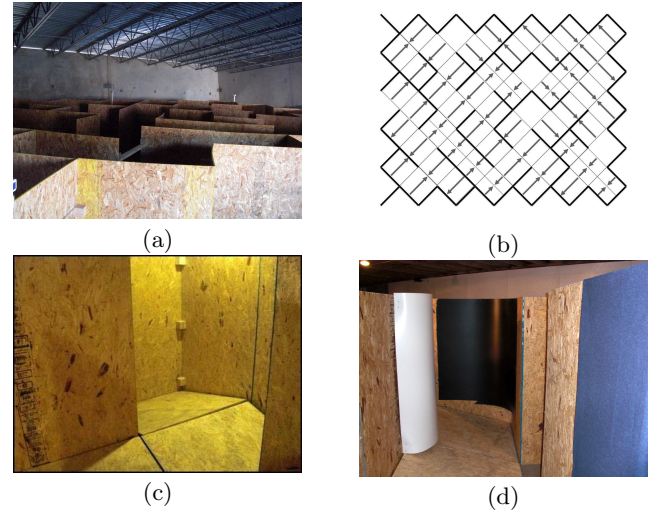


Figure 2: (a) The Maze is a testing apparatus that limits complexities in the environment in order to evaluate the objective performance of particular mapping systems. (b) The configuration of the maze, whose dimension is 15 meters by 10 meters, with black lines representing the maze layout and the gray arrows showing the configuration of 15° pitch and roll ramp flooring, shown in (c). (d) The maze utilizes a variety of materials and shapes to produce additional mapping features, such as concave and convex surfaces.

of the overall performance, it does not allow researchers to identify problematic situations, analyze how errors propagate throughout the system, or compare the performance of one system with competing approaches.

To address these issues the Interleague Challenge leverages emerging test methods for robotic mapping and the development of a common robotic platform to enable repeatable and reproducible testing scenarios. This constrains the variability in the test to abet in the evaluation of mapping algorithms as they transition from a purely virtual world to the real world. The remainder of this section is dedicated to describing the testing method and common robotic platform used for this challenge.

2.1 A Test Method for Robotic Mapping

Associated Metrics

Arguably, the most common mapping paradigm employed for robotic navigation is the metric mapping paradigm. This intuitive mapping paradigm provides a representation where the spatial relationship between any two objects in the map is proportional to the spatial relationship of the corresponding objects in the actual environment. Therefore, assessing the quality of metric maps is based on the spatial consistency of features, such as walls and hallways, between the map produced by the robot and the ground truth map of the actual environment. Error propagation and sensitivity to performance singularities [Scrapper *et al.*, 2008] idiosyncratic to most robotic mapping systems suggests the associated metric needs to quantify the local (or regional) consistency of

areas within the map, as well as the global consistency of the overall map.

Apparatus

The Maze testing apparatus, seen in Figure 2, is part of an emerging suite of test methods for characterizing the performance of the robotic mapping for response robots [ASTM E54.08, 2009]. The apparatus is essentially a random maze, whose overall dimension is 15 meters by 10 meters, and is constructed from oriented strand board (OSB) to form a series of hallways that are approximately 1.2 meters wide. The non-flat flooring, comprising of 15° ramps, makes the vehicle pitch and roll as it traverses the maze. The modularity of the apparatus enables the randomization of the maze configurations for repetitive testing. The use of OSB to construct this apparatus provides a surface friction similar to dust covered concrete and a cost-effective testing apparatus that is easy to replicate.

This apparatus was chosen for the Interleague Challenge because it provides a feature-rich scenario that limits environmental complexities. This should provide the best-case scenario, where mapping systems should perform optimally. The configuration of this apparatus generates a closed set of distinct mapping features and vertical walls that provide a mapping system with distinct observations throughout the apparatus. The presence of distinct features and the lack of occlusions found in this apparatus reduce the uncertainty in the mapping system when associating features, which increases the likelihood of determining valid correspondences. Additionally, surfaces perpendicular to the motion of the vehicle are present in almost every scan. This increases the ability of the system to make accurate measurements of the surrounding environment and accentuates the displacement of features between observations of the external world. Limiting the environmental complexities allows developers to tune their systems and establishes a baseline for comparison.

2.2 A Common Robotic Platform

A common robot with a fixed sensor configuration is used to challenge researchers and ensure compatibility. As mentioned above, this platform also support the entry to robotics of newcomer teams.

Robotic Platform

The base robotic platform used for the Interleague is designed to provide a cost-effective robotic platform capable maneuvering in the complex terrain. The dimensions of the common robotic platform is .533m x .304m x .762m (21 x 12 x 30 inches), it weighs approximately 18 kg (40 lb), and it is equipped with 4 rechargeable 12 volt batteries allowing it to operate in for up to 10 hours. The platform is capable of operating in all weather conditions and a high degree of mobility allows it to navigate on rough terrain and to climb stairs up to 50° inclination. The motors are equipped with encoders for controlling set-velocities and computing the wheel odometry.

Sensor Configuration

The sensor configuration is defined by two statically mounted laser scanners (one horizontal scan and one vertical scan),

and an inertial measurement unit (IMU). Both laser scanners deliver range readings with a field of view of 240° with an angular resolution of 0.36° and a range of approximately 4 meters. The IMU provides measurements of the three Euler angles; yaw, roll, and pitch. Although the IMU supports compass-based yaw measurements, i.e., orientation measurements relative to magnetic North, the current implementation does not utilize this information. Therefore, the orientation angles obtained from the IMU are based solely on measurements take from the gyroscopes and accelerometers. This is motivated by the fact that, particularly in harsh environments, magnetometer readings can strongly be perturbed.

Robotic Interface

The platform provides two robotic interfaces: a low-level serial interface and a high-level robotic interface. The serial interface provides developers with direct access to the on-board controller, giving them direct access to motor commands and supporting the integration of both laser scanners and the IMU via an USB interface implementing a vendor specific protocol. In order to create a more user-friendly interface, the common robotic platform has implemented a high-level robotic interface. This interface is built on a wrapper application that automatically collects data from all the sensors and provides them via a TCP/IP server executed on-board the robot. The messaging protocol used by this process server is chosen according to the protocol of the USARSim simulation application [Wang and Balakirsky, 2009]. Thereby users can control the robot and access sensor data in the same way as if they would connect to the USARSim simulator used in the RoboCupRescue Virtual Robot Simulation competition. Hence, the migration of software developed in simulation towards execution on the real platform is simplified.

3. A 2D MAP EVALUATION FRAMEWORK

As mentioned in Section 2.1, the development of an evaluation framework for quantifying the performance of robotic mapping algorithms in the metric paradigm requires two key components. First is the ability to obtain or generate an accurate ground truth of the test method. Second is the ability to quantify the local consistency, incremental errors arising in bounded regions within the map, as well as the globally consistency, accumulation of errors throughout the entire map. The remainder of this section will describe a 2D map evaluation framework consisting of a generalized ground truth generation methodology (Section 3.1), and a methodology for quantifying the performance of the mapping algorithms. (Section 3.2).

3.1 A Ground Truth Generation Tool

The process developed for generating ground truth of the test methods at the Interleague Challenge, as described in this section, is an attempt to automate the production of ground truth with minimal *a priori* information about the environment. However, this ground truth generation process requires human intervention. Depending on the length of the runs and sampling frequencies employed, this process can be laborious. It is the belief of the authors that some level of human intervention is necessary when developing ground

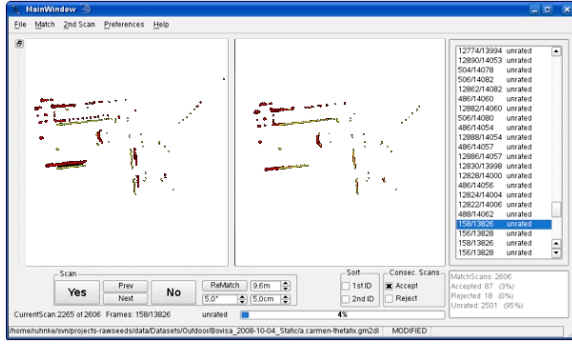


Figure 3: Manual verification of automated laser scan alignment for ground truth generation.

truth of test environments, especially environments where *a priori* information is not readily available.

The process used for generating ground truth during the Interleague Challenge is a two-step process. The first step uses a *Simultaneous Localization and Mapping* (SLAM) algorithm, supervised by a *subject-matter expert* (SME), that directly processes raw sensor streams to estimate a globally consistent trajectory and to derive initial displacement candidates $\delta_{i,j}$, referred to in this paper as *constraints*. During the second step the SME is required to inspect every constraint and verify the displacement. For example, the SME inspects two consecutive observations obtained from the laser range finder and the corresponding pose estimates at time step i and j belonging to the two poses of the constraint, respectively (see Figure 3). During the verification process, the SME can either *accept* or *reject* the constraints previously estimated by the SLAM algorithm. In the case of an *accept*, the SME verifies the final constraint, $\delta_{i,j}^*$, and adjusts the displacement as necessary.

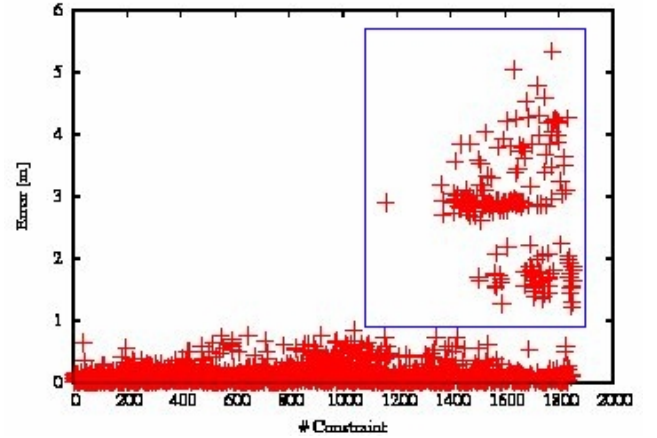
3.2 A Map Assessment Tool

Many of the methodologies being developed for quantifying metric maps employ feature extraction and image registration techniques on an occupancy grid (or image) of the global map. Commonly, these approaches do not take into account the resolution and scale of the grid and do not consider the local consistency within the global map. Therefore, the approach employed at the Interleague Challenge attempt to quantify the errors in a map using the *relative* displacements between poses, i.e. relative motion, as the evaluation criteria. This motivation for using this evaluation criterion is based on two anecdotal factors. First, to consider the pose instead of the resulting occupancy grid based map is advantageous because features in grid-based maps can become obscured beyond recognition, especially after closing loops, although the actual pose error is within centimeters. Second, using the relative displacements of the poses minimizes error propagation during the evaluation, providing a more accurate metric that isolates regional errors when considering the absolute pose error. For example, pose estimates towards the end of the run are not impacted by the pose errors that occurred at the beginning of the run.

Essentially this map assessment tool quantifies errors in a



(a)



(b)

Figure 4: An example illustrating the utilized metric for map assessment. (a) A visualization of the ground truth constraints (red lines) superimposed on the map generated by the SLAM algorithm. (b) The error plot showing the relative displacement error at each constraint plotted over time. As depicted by the blue rectangle the error increases drastically in the last third where the robot returned after driving a long hallway without features.

robot-generated map by comparing the constraints from the ground truth generation process and the pose estimates being produced by a given mapping system being considered. An associative relationship is built between the constraints and the pose estimates based on time¹. It then uses this associative relationship to compute the error in the relative displacements for each of the sets of corresponding data.

The map evaluation process can be expressed formally as follows: Given the estimated trajectory of a robot, $x_{1:T}$, consisting of a series of the pose estimates, x_i , produced by the mapping system at timestep i from 1 to T . Let $x_{1:T}^*$ be

¹While correlating data between two systems based on time is suitable in some situations, it is not a good assumption and it will be addressed in later versions of the tool.

the set of the corresponding reference poses generated during the ground truth process. The *relative* displacement can then be defined as $\delta_{i,j} = x_i \ominus x_j$, where \oplus is the standard motion composition operator and \ominus its inverse. Instead of comparing x to x^* (in the global reference frame), the operation is based on δ and δ^* as

$$\varepsilon(\delta) = \sum_{i,j} (\delta_{i,j} \ominus \delta_{i,j}^*)^2. \quad (1)$$

The major advantage of this tool is the ability to adjust the resolution of the evaluation criteria, which is defined by the constraints produced during the ground truth generation process. This suggests that this tool can be dynamically adjusted to evaluate the global or local consistency of a map. It can also be used to identify and test performance singularities in mapping systems by evaluating how different environmental conditions impact a particular mapping system.

A demonstration of the map assessment used to evaluate a SLAM algorithm in a difficult environment with long hallways and glass walls is shown by Figure 4(a). Ground truth constraints are depicted by red lines in the map. Figure 4(b) shows the temporal analysis of errors in the map for each constraint over the course of the entire run. As can be seen, the error increases drastically in the last third (depicted by a rectangle), which corresponds to the real situation where the robot returned after driving a long hallway without features. This cluster clearly identifies a weak point in the estimated algorithm, which did not manage to close the loop in this particular situation, shown by the shearing effect in the corridor. A more detailed description can be found in previous work [Burgard *et al.*, 2009].

4. THE EVALUATION AND RESULTS

At the RoboCup competition in Graz 2009 the Interleague Challenge provided a testing scenario to facilitate the inter-comparison of the robotic mapping algorithms developed in simulation and the applicability of these algorithms to real data captured from a real robot while operating complex terrain. This section will provide an overview of the challenge and present the results of the three teams demonstrating the most proficient performance. In the remainder of this section, these teams will be referred to as *Team 1*, *Team 2*, and *Team 5*.

Prior to actual the challenge the common robotic platform (described in Section 2.2) and a variation of the *The Maze* test method (described in Section 2.1) was modeled in USARSim (see Figure 1c) and provided to the teams. These models allowed teams to tune their mapping systems in this particular environment.

For the actual challenge, the evaluation was based on a sensor data set logged while teleoperating the real common robotic platform through the maze built in support of the RoboCupRescue Robot competition, shown in see Figure 1b. The sensor data set logged the data streams of the two laser range finders, mounted vertically and horizontally, and the

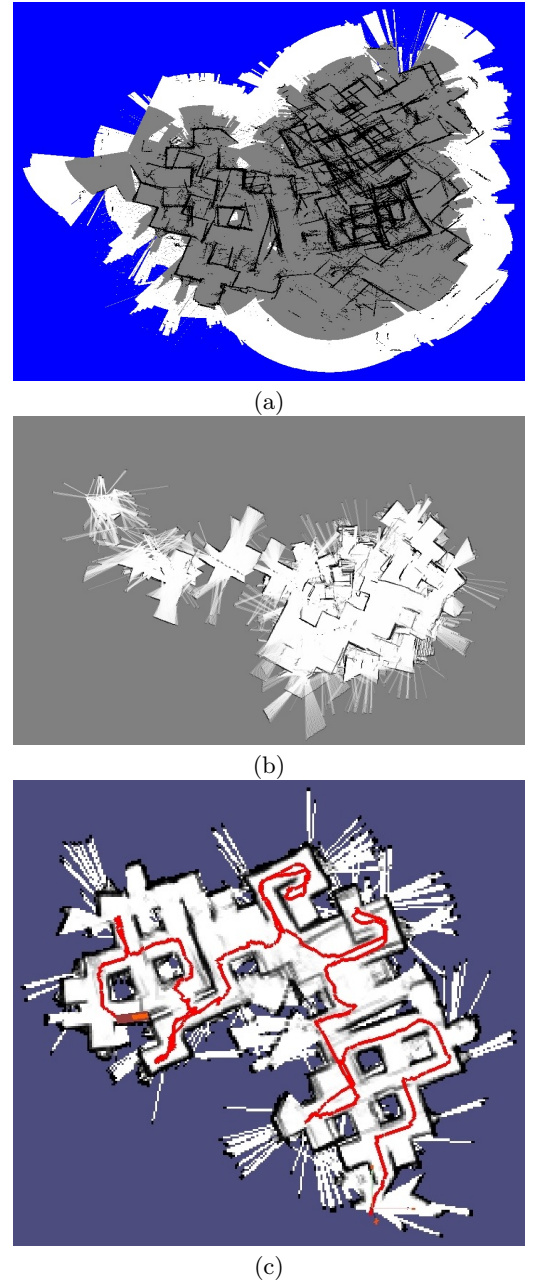


Figure 5: (a-c) Visual mapping results from team 1, team 2, and team 5, respectively.

IMU sensor using the USARSim message protocol. A simple server application, simulating the USARSim interface to the actual robot, publishes the contents of the sensor data set over a TCP/IP socket at the same data rates found on the actual robot. Simulating a USARSim connection to a real robot not only reduces the integration required for teams to deploy their mapping algorithms but provides a repeatable testing scenario that eliminates the impact of *in situ* decisions on the mapping process.

Teams participating in the mapping challenge ran their respective mapping algorithms against the data from the real robot using the server application mimicking the USARSim

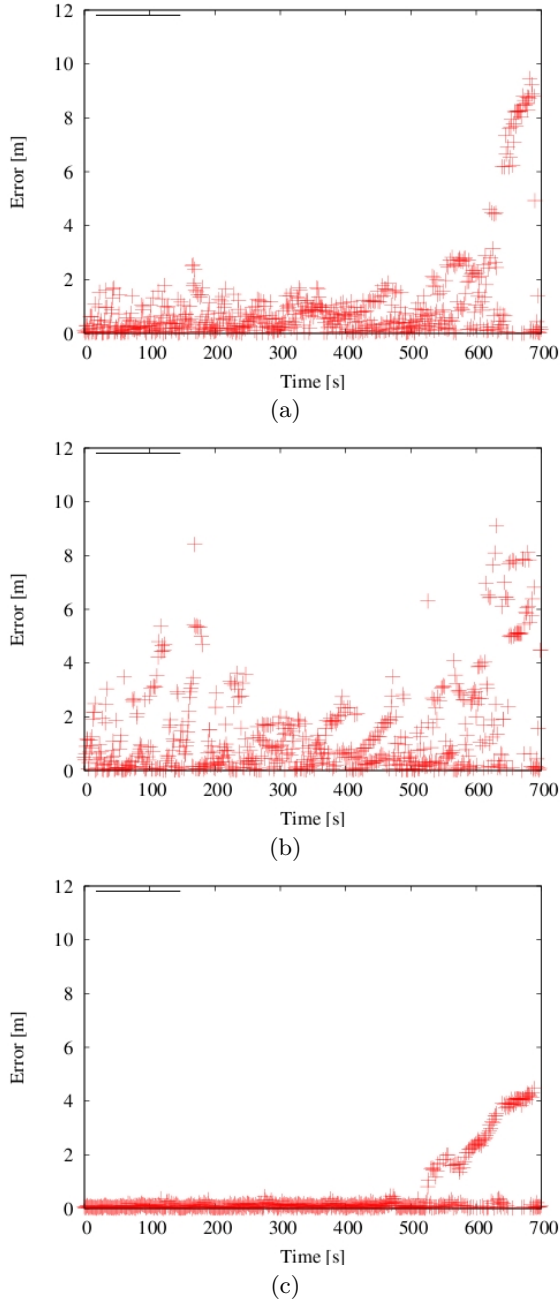


Figure 6: (a-c) Visualization of position errors at each constraint for team 1, team 2, and team 5, respectively. As can be seen, the resulting metric corresponds to the visual impression given by the according maps.

interface to the real robot. At the end of the run, each of the teams were required to submit a log file reporting their results as a list of pose estimates, x_t, y_t, θ_t , computed for each observation at time t . Each of the tuples in the log file, denoted as a $\langle x, y, \theta, t \rangle$, corresponds to observations used to construct ground truth constraints used in the map assessment tool, described in Section 3.2.

Table 1: Quantitative results (avg. position error)

Team	Abs. Err. [m]	Sqr. Err. [m^2]	Max. Err. [m]
1	1.08 ± 1.81	4.44 ± 14.84	9.46
2	1.40 ± 1.77	5.08 ± 11.74	9.11
5	0.48 ± 1.04	1.32 ± 4.00	4.46

The resulting metric maps from each of the teams are shown in Figure 5. Qualitative analysis suggests that divergent behavior has occurred in each of the approaches but does not provide any empirical evidence on which team was able to more accurately map the environment. For instance, the map produced by Team 5 appears to be more consistent but this does not provide a basis for distinguishing differences in the maps produced by Team 1 and Team 2. Additionally, the convolution of features in the maps makes it hard to assess if the divergent behavior seen in the map is due to regional errors that have propagated through the system or a catastrophic error.

The application of the map assessment tool quantifies the errors as they arise during the run and provides the means to temporally assess the errors that occurred in each of the respective maps. The error plots shown in Figure 6 show the relative displacement as compared to the constraints formulated during the ground truth generation process. This temporal analysis shows the stability of the mapping system and helps classify the impact of the errors that have occurred. This shows the map produced by Team 5 appears to be more stable than the competing approaches but is plagued with some regional errors arising during the course of the run, probably due to problems arising from loop closure at the end of the run. It also indicates that the mapping system deployed by Team 1 is more stable than Team 2 and could be suffering from a configuration or data association problem. This information could be used by developers to improve their approach.

The overall performance of the competing mapping systems is summarized in Table 1 by averaging the relative displacements measured by the map assessment tool over the whole run. Again this shows that the approach deployed by Team 5 outperformed the other two approaches.

5. DISCUSSION

The Interleague Challenge made its debut at the 2009 RoboCup Competition in Graz, in the first attempt to literally bridge the gap between simulated and real robot platforms. The challenge, built on a common robotic platform and a common testing apparatus, laid the foundation for not only assessing the applicability of algorithms developed in simulation on real data but also the need to develop quantitative metrics and test methods capable of evaluating the local and global consistency of robot generated maps. This challenge also exemplifies there is still work to be done.

The undulating terrain found in the maze presented a particularly problematic environment for the fixed sensor configuration. As the robot traversed the pitch and roll ramps, the horizontally aligned laser range finder periodically scanned the ceiling and floor of the maze. This produces artifacts in

the data that further complicate the data association problem inherent to many techniques used for mapping, such as incremental scan matching. While there are mechanical techniques used to address this issue, i.e. actuating laser scanners to continuously level the orientation of the sensor with respect to world reference frame, the sensor configuration of the common robotic platform was intended to realize a survivable mapping solution that limits the number of moving parts. The surface friction of the OSB exacerbates the non-systematic errors in skid-steered vehicles rendering the encoder-based odometry almost useless.

In order to facilitate the transfer of technologies from simulated world to actual implementations on real robotic platforms, there needs to be a tighter coupling between simulated systems and their real world counterparts. For example, the simulated IMU or INS sensor in USARSim reports location and orientation of the robot assuming a Gaussian noise model, which is not consistent with available inertial measurement systems.

The evaluation of robot-generated maps is often based on qualitative approaches that do not take into account how specific environmental conditions or *in situ* decisions impact the performance of the system. While this type of analysis provides some indication of the overall performance, it does not allow researchers to understand what errors a specific system is prone to, how these errors impact the overall performance of that system, and how performance of that system compares with competing approaches. Developing testing scenarios for evaluating robot-generated maps can greatly benefit from the development repeatable and reproducible testing scenarios that isolate potential failure conditions in a controlled environment.

The development of a common robotic interface that enables developers to seamlessly transition robotic algorithms from simulation to the real world can help foster innovation and expedite the transfer of the technologies from the lab to fielded systems. The continuing development of this interface will lower the entrance barrier to robotics for newcomers and help improve the development cycle.

In the future, we plan to extend the Interleague Challenge to address other performance benchmarks for robotics; for example, an autonomous behavior challenge that will focus on the robot's ability to autonomously navigate a vaguely defined testing apparatus in a known way. In this challenge, teams will have to directly control the real robot via the USARSim interface according to a given task description.

6. ACKNOWLEDGMENTS

We would like to thank the teams during RoboCup 2009 in Graz, Austria for contributing their mapping results.

7. DISCLAIMER

Commercial equipment and materials are identified in this paper in order to adequately specify certain procedures. Such identification does not imply recommendation or endorsement by affiliated institutions, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

8. REFERENCES

- [ASTM E54.08, 2009] ASTM E54.08. ASTM Committee E54.08.01 on Homeland Security Operations; Operational Equipment; Robots, 2009. <http://www.astm.org/COMMIT/SUBCOMMIT/E5408.htm>.
- [ASTM, 2007] ASTM. In *Form and style for ASTM Standards*. ASTM International, March 2007.
- [Burgard *et al.*, 2009] W. Burgard, C. Stachniss, G. Grisetti, B. Steder, R. Kümmerle, C. Dornhege, M. Ruhnke, A. Kleiner, and J. D. Tardós. Trajectory-based comparison of slam algorithms. In *Proc. of the Int. Conf. on Intelligent Robots and Systems (IROS)*, 2009. To appear.
- [Darpa, 2007] Darpa. Darpa Urban Challenge, 2007. <http://www.darpa.mil/grandchallenge/>.
- [ESA, 2008] ESA. Lunar robotics challenge, 2008. http://www.esa.int/esaCP/SEM4GKRTKMF_index_0.html.
- [Scraper *et al.*, 2008] C. Scrapper, R. Madhavan, and S. Balakirsky. Performance analysis for stable mobile robot navigation solutions. volume 6962, page 696206. SPIE, 2008.
- [Wang and Balakirsky, 2009] J. Wang and S. Balakirsky. *USARSim V3.1.3 - A Game-based Simulation of mobile robots*. <http://sourceforge.net/projects/usarsim/>, 2009.