Best in Class: Leveraging Robot Performance Standards in Academic Competitions to Encourage Development and Dissemination

Raymond K Sheh[1,2] and Adam S Jacoff[2]

ABSTRACT

Standard test methods and academic competitions share much in common. We detail how we use standard test methods to promote education, research, development, and dissemination among the academic community. Since 2014, we have used competitions and open source robot designs to focus students, and particularly high school students, on the challenges of emergency response and public safety robotics. Our two main initiatives are the Rapidly Manufactured Robot Challenge (RMRC), which forms part of the RoboCup Rescue Robot League (RRL), and the Open Academic Robot Kit (OARKit). The RRL and RMRC leverage Standard Test Methods for Response Robots, developed under ASTM International Subcommittee E54.09 on Homeland Security Applications: Response Robots. The standards developed under ASTM E54.09 are significantly supported by the pre-normative research collaboration between the US Department of Homeland Security (DHS) and the US National Institute of Standards and Technology (NIST). These standard test methods are an effective way of communicating the challenges of the domain and focusing research and development on open problems. By measuring the performance of prototypical implementations in a consistent, comparable manner, standard test methods also allow students to compare performance with each other, as well as with commercial, deployed robots. The OARKit aims to capitalize on the ease of comparison and collaboration that comes with the use of standard

[1] Department of Computing, Curtin University, Bentley, WA, 6102, Australia; 0000-0002-9693-3184

[2] Intelligent Systems Division, National Institute of Standards and Technology, Gaithersburg, MD, 20899, USA

test methods by lowering the resource and expertise barriers for entry into response robotics research. This family of robot designs form ideal starting points for new and existing teams to enter the RMRC. Teams build these robots by following basic instructions and then improve them in their area of expertise. The results are then rigorously measured in the competition and disseminated to other teams. Thus the community of teams share each other's developments and push the state-of-the-art forward.

Keywords

Robot Competitions, Response Robotics, Education

Introduction

Standard test methods share many similar characteristics with competitions, particularly competitions that focus on individuals or teams achieving a score that is then compared with others. In both standard test methods and competitions, performance is measured in a way that is intended to be fair, repeatable, reproducible, and significant in some way.

The RoboCup Rescue Robot League (RRL) has been running since 2000. This research competition, originally for university research students, aims to foster the development of technologies to address gaps in the deployed capabilities of response robots[3], such as those deployed for addressing natural disasters, explosive ordnance disposal, or nuclear incident response. This competition forms an integral part of the development process for the Standard Test Methods for Response Robots, developed under ASTM International Subcommittee E54.09 on Homeland Security Applications: Response Robots. The RRL is an incubator for conceiving,

[3] Certain commercial products are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products identified are necessarily the best available for the purpose.

developing, refining, and validating new test methods, in the presence of prototypical robots that embody new capabilities and ideas. It also disseminates the test methods among the academic and developer communities and provides an opportunity to perform repeatability and reproducibility testing.

The RRL Rapidly Manufactured Robot Challenge (RMRC) is a research competition aimed at high school and early undergraduate students. It also leverages these test methods to expose these younger students to open research challenges in the field of response robotics. Since 2014, we have been developing the RMRC alongside the Open Academic Robot Kit (OARKit), an initiative that aims to bring interesting research-level robots into high school and undergraduate classrooms. It lowers the barrier of entry in terms of cost and resources by leveraging open source principles and low cost, rapid manufacturing and prototyping equipment such as 3D printers and laser cutters. In this paper, we present the history, motivation, and latest developments of these two initiatives and their close relationship to the ASTM E54.09 standard test method development efforts.

Robotics competitions bring significant benefits to younger students and research communities alike. For example, they provide inspiration to students, both by way of the application as well as through observing their peers from the other side of town or the other side of the world. They are a conduit for disseminating information about the challenges and best-in-class solutions to those challenges. They guide the thinking of students towards gaps in current capabilities and provide a way to measure their progress towards addressing them. They also bring to the attention of industry and government the capabilities being developed in classrooms and institutions that may otherwise not be apparent, and present them in a way that is directly comparable to the deployed state-of-the-science.

Standard test methods can play an important role in amplifying these benefits by ensuring that the challenges that guide and inspire the students direct them towards measurable deficiencies in the current state-of-the-art. The Standard Test Methods for Response Robots are designed to be elemental, abstracted tasks that are directly relevant to the real world tasks that response robots must perform in applications such as search and rescue, explosive ordnance disposal, hazardous and nuclear waste cleanup, disaster survey, military reconnaissance, and the like.

They do not replace testing of robots in operational scenarios; rather they provide an indication of the strengths and weaknesses of the robotic system and are a filter to determine which systems are ready for operational testing. To use a sporting analogy, they are like the basic tests of running, jumping, catching, throwing, and so-on that a coach might use to determine how their team is performing and where individual players may need to improve. These tests would also be used by the coach to filter new players before they play their first test game.

THE HISTORY OF ROBOCUP

Since 1997, the RoboCup Federation has been holding robotics competitions. These first focused on the Artificial Intelligence (AI) research community. In the 20 years since, it has branched out across both the wide array of robotics-related topics as well as the various age groups, from junior school students through to early career researchers.

The objective of RoboCup is that "By the middle of the 21st century, a team of fully autonomous humanoid robot soccer players shall win a soccer game, complying with the official rules of Fédération Internationale de Football Association (FIFA, the world soccer governing organization), against the winner of the most recent World Cup." [1]. Conceived as a "Grand Challenge", RoboCup is a vehicle to promote robotics and AI research. It is also a type of "standard problem" that can be replicated around the world, with rules that are familiar to many around the world and are well-documented and understood.

The RoboCup Rescue Robot League

In 2000-2001, the RoboCup Rescue Robot League was introduced. This competition, first piloted at the Association for the Advancement of Artificial Intelligence (AAAI) 1999-2000 conferences, brought the challenge of the Urban Search and Rescue task to university students, who were tasked with building a robot that could survey a disaster site, find victims, determine their condition, and build maps of the environment [2]. In these early years, the arena consisted of various random items of furniture, debris, and other household items, arranged into three regions that represented different mobility challenges [3] [4].

The competition was run in three rounds. First, a preliminary round is run with all teams being given multiple opportunities to search and survey the arena. The arena was often split in half, allowing two teams to run at the same time before switching sides. Teams scored points by finding mannequin body parts as simulated “victims”, and other objects of interest, overcoming debris and other obstacles strewn around the arena. The points were added up and a clear break in the scores was sought that separated out the top 5-10 teams, thus avoiding the situation where a team “just” missed out. These teams progressed to the finals. Teams were given the whole arena to search and survey, with the team that found the most victims and brought back the most information was declared the winner.

During these two rounds, teams were prevented from observing the arena until after their runs, thus the competition was truly a search task. While realistic, this competition was less reliable as a standard test due to the highly random nature of the behavior of the robots, as well as the element of luck involved in searching the arena.

A third round, called Best-in-Class, allowed teams that specialized in particular aspects of the competition to demonstrate their capabilities, even if they did not have the broad base of expertise that would allow them to win the overall competition. Initially, Best-in-Class awards were given for Mobility and Autonomy. These focused on the robot's ability to traverse terrain and to navigate autonomously, respectively. In later years, this was joined by Manipulation, which focused on robots with the ability to manipulate objects.

Starting in 2005, in conjunction with the launching of the ASTM standardization effort supported by DHS, prototypical standard test method elements began being introduced into the arena. This started with the Random Stepfields [5] shown in Figure 1, a repeatable, reproducible analog for unstructured terrain. The random mannequin body parts that were the simulated "victims" in previous years were also replaced by standard artefacts, designed to test the abilities of the robots' various sensors, such as thermal and visible light cameras, microphones, carbon dioxide sensors, and so-on. In subsequent years, the arena shifted to being based on standard test method apparatuses, primarily for terrain traversal such as crossing ramps, symmetric stepfields, and stairs. Thus, as teams ran their robots through the arena, they were also, in effect, performing informal tests within the standard test methods.

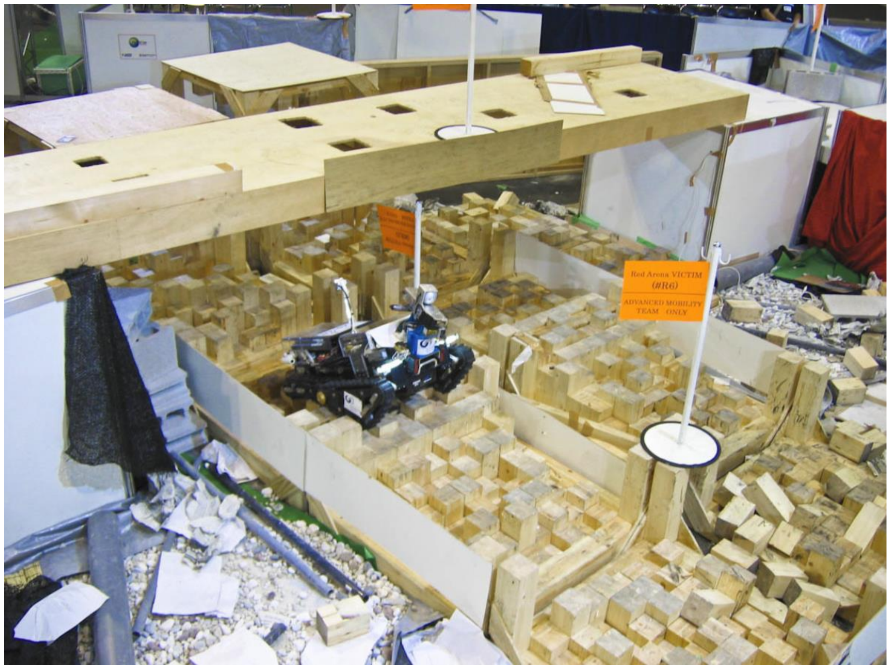

Figure 1: The Random Stepfields at the RoboCup Rescue Robot League 2005.

To further reduce the element of chance in the competition, in the late 2000's the secrecy component of the competition was also dropped, with team members being allowed to enter the arena prior to the run and verify that it had been re-set correctly after the previous team's run. This had the effect of switching the competition from a search task to one of focusing on the test methods integrated into the arena. The more test methods teams could overcome, the more "victim" sensor test boards they could reach and thus more points they could obtain. At the same time, the Best-in-Class competitions were introduced as separate rounds within the standard test method apparatuses. Most obvious of these was the Best-in-Class Mobility competition, which was a direct application of ASTM E2828-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Terrains: Symmetric Stepfields [6].

Around 2015, the competition progressed further towards standardization. Instead of running one or a small number of standard test method rounds as a Best-in-Class round, the preliminary rounds were replaced by standard test method runs within separate lanes. The goal became that of statistically significant testing in the preliminary rounds with as many teams as possible running tests in parallel. The scores for each test were normalized according to the best performing team in that test and then combined to determine which teams progressed to the final rounds.

Outreach Activities

The RoboCup Rescue Robot League goes beyond just a competition. The overall goal of the League is to advance the state-of-the-science in response robotics, in part by fostering the sharing of capabilities between teams. The competition itself goes some way towards this by allowing teams to observe each other's capabilities, as tested in comparable test methods. However, it can be quite difficult for teams to share information during competition, if only because time and development effort pressures mean that many teams, especially those who are performing particularly well, would rather focus on their next competition run.

As a result, the RoboCup Rescue Robot League also hosts regular teaching camps [7] or summer schools. The League community is also active in academic conferences, particularly the RoboCup Symposium and the International Symposium on Safety, Security and Rescue Robotics (SSRR), run by the Institute of Electrical and Electronic Engineers (IEEE) Robotics and Automation Society (RAS). These venues provide a more relaxed environment in which those who developed the Best-in-Class capabilities of the previous competition can present their work.

The teaching camps in particular are structured to be highly practical. Usually half of the time is set aside for traditional lectures and presentations, while half of the time is spent in multi-tracked practical sessions where smaller groups are led in the implementation, or re-implementation, of a capability that they can then take and build upon or use as a point of comparison. This provides opportunities for other teams to bring their own robots and learn, in a hands-on fashion, how to implement these capabilities. Examples of capabilities that have been spread through the league through these events include various 2D and 3D Simultaneous Localization and Mapping (SLAM) implementations [8], inverse kinematics, aerial robots, and user interfaces.

These events also form a valuable opportunity to form collaborations with stakeholders outside of traditional university academia. The teaching camps and summer schools are usually hosted at a site where responders train and always have a contingent of responders in attendance, such as firefighters and police officers. These personnel attend alongside the students, with the goal of sharing their knowledge of the application, grounding the students' understanding of where their technologies may be useful, as well as to learn about new capabilities that may be of use to them but may not be well known outside the laboratory. Other academic sectors have also attended these events to learn about potential opportunities to expand their activities. Indeed, the RMRC was launched at the 2014 Safety, Security and Rescue Robotics Summer School and Workshop [9], which was jointly hosted by the Perth Artifactory makerspace and Curtin University. For the first time, a contingent of high school teachers joined the event and formed the core teams that participated in the pilot competitions in the following years.

Finally, these events provide an opportunity to analyze the outcomes of the prototypical test methods, experiment with them in a more informal setting, and document them in preparation for potential standardization. They also provide an avenue to disseminate in more detail the test methods among the broader academic community.

COMPETITION TRADE-OFFS

In discussing robot and AI competitions and their ability to advance research goals, we find it useful to consider competitions according to two different trade-offs. The first is the task, be it abstract or real-world application based. The second is expected novelty, be it known or research.

Competition Task Trade-Off

An example of an abstract task is robotic chess. This is a game, developed for a particular purpose. On the one hand, while the research challenges are still very real and applicable, the game is heavily abstracted from open real-world challenges. On the other hand, the problem is also very well defined with a well understood set of semantics, problem definition, and evaluation metric.

In contrast, a real-world application based task might be the US Defense Advanced Research Projects Agency (DARPA) Urban Challenge [10] which involves autonomous robotic vehicles driving through an urban area, complete with other autonomous and human-operated vehicles. The top performers in such a competition have produced a working solution that solves an open real-world application problem. Of course there is usually still significant work to be done to make their entry viable in the real world from safety, economic, production, and regulatory perspectives (among others). However, administering and competing in such competitions tend to be very resource intensive. This limits both the ability to run such competitions as well as the ability of teams to compete.

All competitions trade off between these two characteristics depending on their goals, a compromise that is generally mutually exclusive. Our goal with the RRL and RMRC is to balance being abstract enough to make it easy to administer and compete in these competitions, while also being close enough to the application that they can foster and evaluate the development of technologies that are applicable to real-world problems in the short to medium term.

We leverage the Standard Test Methods for Response Robots as a way to achieve a balance between these two characteristics. Test methods, by their nature, must be both abstract enough to be reproducible and repeatable, scientifically rigorous, and economically viable to reproduce and disseminate, and yet the resulting metrics must clearly be an accurate reflection of a real-world need. By building a competition based around standard test methods, we take advantage of all of the development effort, industry and end-user consultations, and international experience behind these test methods; resources that are generally unavailable to most competitions.

Competition Novelty Trade-Off

Competitions also vary in terms of the novelty of the challenges and expected solutions. On the one hand are competitions where the underlying problems are relatively well understood. Teams are differentiated by how well they can answer this problem, which may still require some degree of novelty, but in general what is required of teams is well understood. Examples include track racing or flat-floor mazes. Such competitions encourage refinement and optimization of solutions.

On the other hand are competitions that pose problems for which there is no known satisfactory solution. Such competitions encourage a much wider variation in solutions, but these solutions also tend to be less well optimized and more experimental. The RoboCup Rescue Robot League is an example of an application that tends towards this end of the spectrum.

RELATED WORK

Research competitions and open source hardware initiatives have been gaining significant traction over the last decade as their value to both teaching and research have become recognized. In this section we will discuss some related competitions and initiatives.

Related Competitions

Robotics competitions have been a feature of high schools for many decades as teachers have leveraged their ability to consolidate the many aspects of what is now called STEAM – Science, Technology, Engineering, Art, and Mathematics – into an event that inspires creativity and problem solving. As a result there is currently a plethora of different competitions, many of which are surveyed in [11].

RoboCup itself is a family of competitions for participants from 14 years old (some regional competitions have lower limits) up to early career researchers (and beyond). In the Junior category, which allows students up to 19 years old, there are three main competitions: OnStage (performance), Soccer, and Rescue. There are also additional challenges such as the CoSpace challenge [12] as well as the RMRC itself. Apart from the RMRC, these competitions are based on games and abstract tasks.

Beyond RoboCup, perhaps one of the best known families of robotics competitions in the present day are the "For Inspiration and Recognition of Science and Technology", or FIRST, Robotics Competitions [13]. Catering to students from kindergarten through high school, the various competitions that make up FIRST aim to not only inspire students to partake in the study of science and technology, but to also teach students about other life skills such as "self-confidence, communication, and leadership".

To this end, RoboCup Junior and FIRST quite deliberately pick somewhat abstract games, made up of combinations of tasks such as moving balls into goals, navigating mazes, pushing levers, and delivering objects on a flat playing field. The point of the competition isn't the solution to the problem itself, but rather the journey that students take in solving the problem. Competitions such as RoboCup Junior OnStage also focus on artistic aspects, particularly as it relates to performance art. CoSpace augments the real robotics competition with a parallel virtual robotics environment to enable more advanced programming and algorithm development. FIRST augments these problem solving tasks with those that encourage teamwork, professionalism, and community service, such as having to form ad-hoc teams during the competition and giving presentations. Indeed, having a more concrete, real-world problem could arguably get in the way of these goals.

Traditionally, competitions where the goal is the solution, rather than the journey, has been the realm of more senior students. These competitions pose open research problems and teams that do well will have developed a novel contribution to a real-world problem. Examples of such competitions include the DARPA Grand and Robotics Challenges, the World Robot Summit 2020 [14], and the RoboCup Rescue Robot League [15]. These competitions tend to require more resources to administer and compete in, both in terms of materials as well as expertise.

In a sense, developing these competitions is very much like developing standard test methods. The goal is to strike a balance in developing a test that is sufficiently abstract that it can be reproduced, understood, and disseminated reliably and fairly, while also being relevant to the real world task, with all the variability that this entails.

Related Open Source Robotics Initiatives

One way to lower the material and expertise cost of participation for new and existing teams into any of these competitions is to allow the teams to build on others' work. Different initiatives have been proposed to assist in this task.

Some competitions make use of specified or mostly specified parts. Examples include competitions that make use of construction kits, such as the FIRST Lego League and several of the RoboCup Junior competitions [16]. Other competitions, such as the FIRST Robotics Challenge [13], provide a specific, common kit of parts along with documentation and examples.

Construction kits that combine physical components with sensors, actuators, and computation, such as Lego Mindstorms, have been pivotal in bridging the gap between computer science and physical robots in the high school because they remove a lot of the complexity involved in actually building the robots that are necessary to embody code and algorithms. The use of such kits is also significantly safer for younger students than building robots in a more traditional manner out of metal, wood, or plastic. However, it is limited in its ability to build larger, more complex, more durable robots due to the need to limit designs and structures to those that are compatible with the particular build system. It is also difficult to teach more advanced concepts in design, particularly as it relates to design for manufacture. Dissemination and re-use of other teams' work can also be challenging with these kits due to the high level of design coupling between different parts of the robot. Thus, while such kits are a tremendously valuable tool to teach introductory robotics, it is difficult to build larger, more complex robots and to leverage other students' work.

In recent years, 3D printing and laser cutting have become much more widespread among high schools, particularly as open source equipment and software have lowered the financial and expertise cost. Practical, low cost 3D printers capable of printing robot components out-of-the-box, with minimal expertise, can be purchased for less than $300 USD, while larger printers, capable of printing whole robot parts without needing to glue them together afterwards, can be purchased for less than $1,000 USD. In a classroom environment, laser cutters have also become more commonplace and can often be more practical as they can produce larger (albeit flat) components much more quickly than 3D printers. Here too, prices have fallen with easy to use units of a practical size starting from $2,500 USD.

Recent years have also seen rapid advancements in the power, flexibility, availability, and ease of use of open source electronics such as Arduino [17] and Raspberry Pi [18] development boards. These advancements are in large part due to the fact that, with open source projects such as these, anyone can make improvements based on their area of expertise and share them with the broader community. As a result, it has become even easier to build embedded computation, advanced sensing, and reliable communications into classroom projects.

The widespread availability of these systems has resulted in a plethora of open source robotics initiatives, including our Open Academic Robot Kit [19]. The goal is to enable ad-hoc collaboration between students in different parts of the world and across year levels. For example, we have developed a basic platform that is intended to be easy to build upon, complete with basic electronics, sensors, and code [20]. A PhD student from Australia might implement a new vision algorithm that allows the robot to avoid obstacles using its camera. A high school student in Thailand might add a better arm design, while a makerspace group in the United States might contribute better wheels. These groups – and others – can benefit from each other's work.

Other open source robotics designs of interest to high school students include the Robotis TurtleBot 3 [21], Poppy [22], Vorpal Robots [23], and Yale OpenHand [24]. All of these designs share these common characteristics of providing students with a basic design with all components either 3D printed or easily available off-the-shelf, plus instructions and source code that allows them to "close the loop" on building an interesting robot that they can then extend. Some of them even have "standard" tasks that can be used as the basis of student competitions.

The Rapidly Manufactured Robot Challenge

The RMRC, as a sub-competition within the RRL, is designed to foster research and development among young students, including contributions to the open source community. The design of the competition and arena is therefore focused on providing students with opportunities to experiment, specialize in particular areas of interest, demonstrate their capabilities, and share these capabilities back to the competition and research community.

COMPETITION STAGES

The RMRC starts with teams qualifying based on their Team Description Materials (TDMs). Teams that are selected based on these materials come to the competition and proceed through two rounds. During the initial, preliminary round, the Standard Test Methods for Response Robots are laid out individually. Teams select test methods in which to evaluate their robots with the aim of achieving the highest metric possible within a prescribed 10 minute period. This is followed by one or more final rounds where the test methods are arranged in a maze or course with teams scoring points according to how many test methods they overcome in sequence.

Qualification

Teams are qualified to compete in the World Championships of the RMRC via three avenues. First, teams may qualify by submitting TDMs. These correspond to the Team Description Papers that are requested of the Major RoboCup Rescue Robot League teams (who are mostly undergraduate or graduate students), but in a more general form that makes it conducive to integration into high school media curriculum such as a website, blog post series, and/or videos. Teams may also be selected if they reached the finals in the previous year or if they place in regional competitions.

The TDM is required to cover the following topics, regardless of the format in which it is presented.

- Logistical info: Team Name, Organization, Country, Contact details, Website.

- Introduction summarizing:
    - The team.
    - The technical aspects that it focuses on.
- System description, describing:
    - Hardware.
    - Software.
    - Communications.
    - Human-robot interface.
- Application, describing:
    - Setup and packup of the robot and operator station.
    - Mission strategy.
    - Experiments and testing that have been done or will be done.
    - How the particular strengths of the team are relevant to applications in the field.
- Conclusion, summarizing:
    - What the team has learned so far.
    - What the team plans on doing between now and the competition.
- Appendix containing:
    - One table per robot outlining the components and estimated cost of the robot.
    - At least one picture, 3D rendering, or technical drawing of the robot.
    - A list of software packages, hardware, and electronic components that have been used, or plan to use, particularly those from the Open Source community, through the Open Academic Robot Kit or otherwise.
    - A list of software packages, hardware, and electronic components and designs that have

been, or plan to be, contributed to the Open Source community, through the Open Academic Robot Kit or otherwise.

- References (to other work that the team has made use of).

These materials are evaluated by a panel of high school and university academics. Teams are qualified not just for the potential to perform, but also for the potential to learn and for excellence in specific capabilities.

Preliminary Round

Teams that come to the World Championship first enter a preliminary round that runs over two full days. The goal of this round is to give teams the most number of opportunities to demonstrate their capabilities on the test methods that they are able to complete. This includes giving them time to experiment, to fail if necessary, and re-attempt.

The test methods are divided into two categories: those that relate to the ability of the robot to reach its destination and those that relate to its ability to perform a task once at the destination. We will refer to the former as the traversal test methods and the latter as the sensing and manipulation test methods, both of which are critical in measuring the performance of these systems as a whole, as they relate to their ability to perform in the response robotics application. The total scores achieved in each of these categories are multiplied together to yield the teams' final score.

Traversal test methods

The traversal test methods represent the ability of the robots to reach their destination in order to perform a task. These are generally maneuvering or terrain tasks such as centering between obstacles, climbing stairs (ASTM E2804-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Obstacles: Stairs/Landings [25]), or overcoming rough terrain (e.g., ASTM E2826-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Terrains: Continuous Pitch/Roll Ramps [26], ASTM E2827-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Terrains: Crossing Pitch/Roll Ramps [27], and ASTM E2828-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Terrains: Symmetric Stepfields [6]). These are all presented in the form of terrain pallets as shown in Figure 2. Each has a start and an end with the traversal challenge in the middle and points are scored for reaching each end of the pallet. One exception is the nodal manipulation ("Pipestar") apparatus, which also exists in the arena as a specific test for manipulation-focused teams as shown in Figure 3. Robots may perform one or more of five tasks at each pipe for one point each: Precision touch (ASTM WK54272 Evaluating Ground Response Robot Dexterity: Touch or Aim), Rotation of the object (ASTM WK54273 Evaluating Ground Response Robot Dexterity: Rotate), Extraction of the object, Insertion of the object (ASTM WK54274 Evaluating Ground Response Robot Dexterity: Extract and Place), and Inspection of the interior of the pipe (ASTM WK54271 Evaluating Ground Response Robot Dexterity: Inspect). Approved standards are indicated by a label that begins with an "E" whereas draft test methods are considered work items and indicated by a "WK". These draft test methods are being refined in preparation for balloting as standards [28].

Figure 2: Terrain test method apparatuses arranged as individual lanes for preliminary trials.

Figure 3: The Nodal Manipulation ("Pipestar") apparatus.

The tests are performed according to the ASTM International standard test method procedure [29] with the metric being the number of half-laps of the terrain pallet that the robot performs within the prescribed time, usually 10 minutes. These pallets are arranged individually as shown in Figure 2 so that they can be run in parallel. Points are assigned for each test method that a team successfully completes a measurement run in. Teams have the opportunity to try each traversal test method several times with their maximum metric for each test method counting towards their final total. Each test method's scores are normalized such that the top metric in that test method is equal to 100. Thus if a team is the only one to successfully achieve a measurement in a given test method, perhaps because it is particularly specialized or difficult, that team automatically gains 100 points for that test method. In contrast, a team must achieve excellence in an easier or more popular test to gain the same number of points.

A scheduling matrix, an example of which is shown in Figure 4, is drawn up. Teams are invited, in random order, to claim a slot representing a particular test method, at a particular time (in the figure, teams are denoted by a two-letter code appearing in the upper left corner). Once each team has claimed a slot, the cycle repeats with their second slot and so-on. The placement is constrained by the inability of teams to place two of their own slots on the same line (as they can only do one test at once), some test methods cannot run in parallel (because they share equipment), and the total number of taken slots on each line cannot exceed the number of referees available to administer each test. The order is then inverted for the following half-day. Once each team has claimed all the slots that they wish to, remaining spots may be used by any team to practice. Teams may also move slot claims to empty (future) slots or negotiate with other teams to swap. As tests are performed, the results are recorded in each slot, along with the name of the administrating referee.

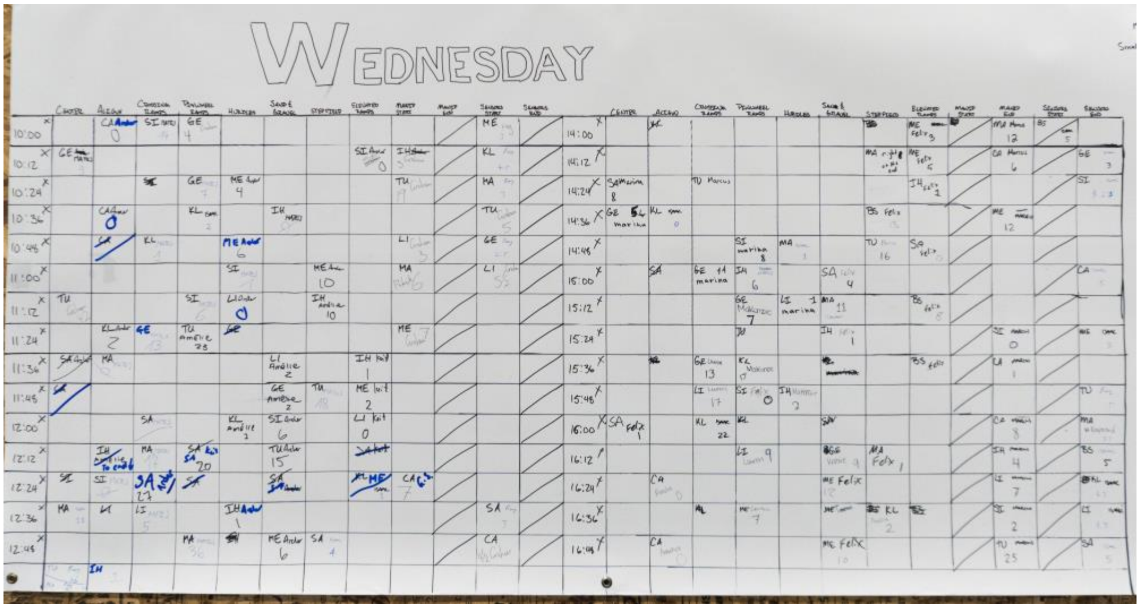

Figure 4: A scheduling matrix for the preliminary (standard testing) rounds.

To the maximum extent possible, teams are provided with the ability to run as many times as their own scheduling and reliability allows. This encourages teams to come prepared and to develop reliable systems. Runs are scheduled every half-day during team leader meetings in the morning and at lunch time. Each run is 10 minutes long with 5 minutes for changeover, yielding four runs per hour. Each half-day runs for around 3-4 hours and there are generally 4-5 referees to administer tests; thus there are generally around 96-160 test opportunities per day or 192-320 test opportunities across the qualification round. In the 2017 World Championships, being the first year that the RMRC was run as an open competition, six viable teams competed yielding over 100 standardized test results. With around 10-13 teams as anticipated going forward, this allows each team between 15 and 30 testing opportunities across the qualification round. There are generally around 10 test methods available, thus teams will generally not be schedule-limited in their ability to try test methods multiple times to achieve their best score.

Sensing and Manipulation test methods

The sensing and manipulation test methods represent the ability of the robot to perform its tasks once it arrives at its destination. These include tests of vision, directed perception, audio acuity, retrieving objects, pushing buttons, and turning valves. These tests are repeated every half-day, during one of the time slots selected in the matrix, to ensure that the capabilities are maintained through the competition. The result is used as a multiplier and applied to the scores of the team's runs for that half-day. This means that teams need to exercise care when driving their robots so that delicate sensors and manipulators (such as robot arms) are not damaged.

Finals

The scores from the Preliminary round are tabulated and around five teams selected. The specific cutoff score is decided by looking for a distinct gap in the preliminary scores. It should be undeniable that the worst performing team that goes through to the finals has performed significantly better than the best team that did not make the cut.

For the final rounds, the traversal test methods are rearranged into a maze as shown in Figure 5, with the arrangement decided in consultation with the teams that qualified. Teams each have between 10 and 15 minutes per run with 5 minute change-overs, depending on how many teams are admitted to the finals. Following a half-day break from the Preliminary rounds for arena reconfiguration and team practice, the Final rounds are run across two half-days. The aim is to allow each team to run at least five times through the arena over the course of the two half-days.

Figure 5: The terrain pallets arranged into a maze for finals.

During each finals run, teams are invited to start their robot anywhere in the arena and the goal is to accumulate as many points through the arena maze as possible. Each terrain pallet yields up to two points, one in each direction, with each point awarded when the robot either touches the end of the pallet or exits onto a connected pallet. As a result, teams need to be strategic in choosing their starting point and their path so as to maximize the number of points while minimizing excessive terrain traverses or the need for excessive risk early in their run. Teams are allowed to drop their worst-performing score, allowing teams to experiment with the different format. This score is then multiplied by their score in the Sensing and Manipulation test methods, which are performed once during the day.

ARENA AND TEST METHODS

The arena is made up of standard test method pallets, arranged first as individual test lanes for the preliminary rounds and then as a maze for the finals. It is important to note that these test method pallets are not simplified versions of the test methods. They are full test methods, built at the 30 cm scale to represent smaller, confined environments such as collapsed buildings, air conditioning ducts, and industrial plants. Currently the following test methods are implemented. Examples of these are shown in Figure 6, in order:

- Center Between Obstacles (no ASTM number assigned yet).
- ASTM WK53649 Evaluating Ground Response Robot Maneuvering: Align Edges.
- ASTM E2827-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Terrains: Crossing Pitch/Roll Ramps [27].
- ASTM E2802-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Obstacles: Hurdles [30].
- ASTM E2992/E2992M-17 Standard Test Method for Evaluating Response Robot Mobility: Traverse Sand Terrain [31].
- ASTM E2991/E2991M-17 Standard Test Method for Evaluating Response Robot Mobility: Traverse Gravel Terrain [32].
- ASTM E2828-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Terrains: Symmetric Stepfields [6].
- Elevated ramps (no ASTM number assigned yet).
- Manipulation pipestar (supporting ASTM WK54272, WK54273, WK54274, and WK54271 as described previously).

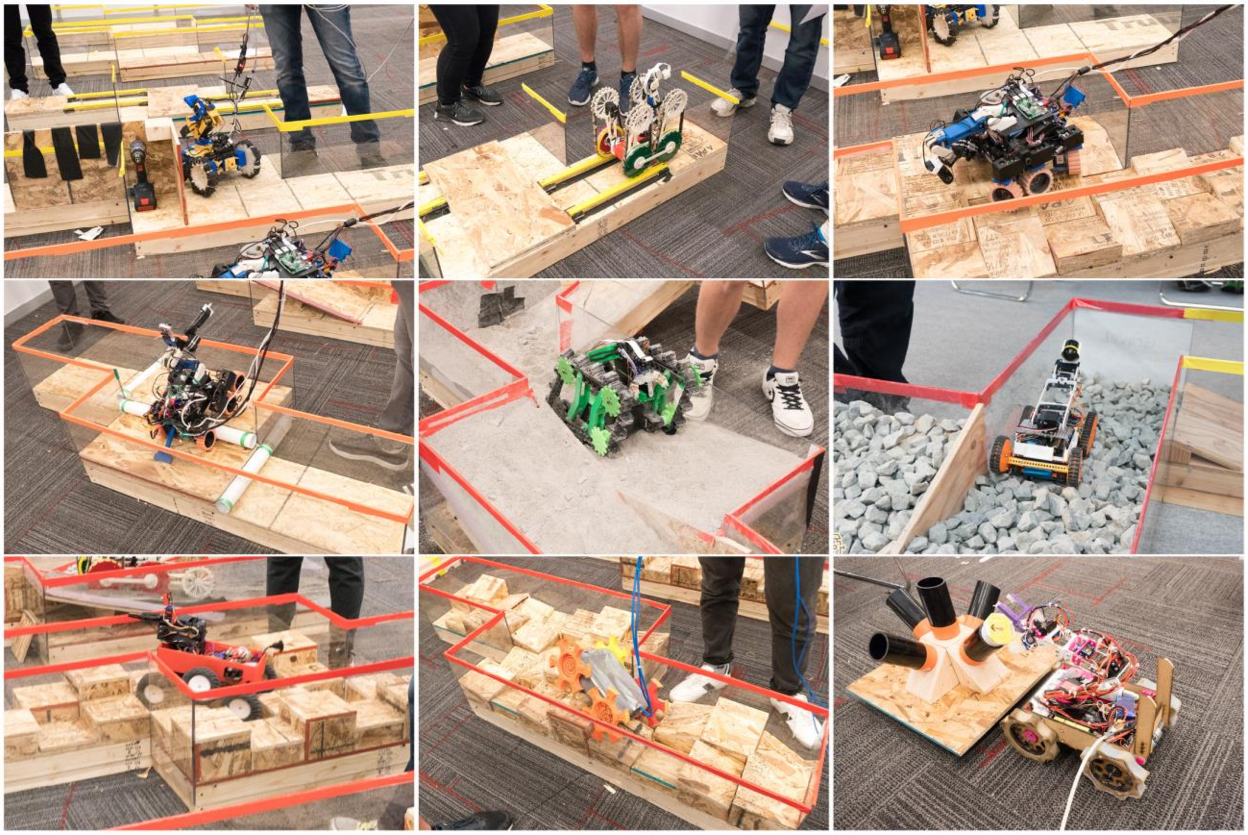

Figure 6: Examples of the different terrains represented in the RMRC, along with examples of rapidly manufactured (3D printed or lasercut) robots.

Sensor tests were embodied by the “Readiness Board”, as shown in Figure 7. This consisted of embedded versions of test methods for ASTM E2566-17a Standard Test Method for Evaluating Response Robot Sensing: Visual Acuity [33], ASTM WK57967 Evaluating Ground Response Robot Sensing: Thermal Image Acuity, ASTM WK60783 Evaluating Ground Response Robot Sensing: Audio Speech Intelligibility, and prototypical test methods for Survey Acuity, Motion Detection, CO_2 detection, and Hazardous Material Label Recognition. These are performed at a standard near-field distance of 40 cm (16 in) and each test is thresholded to represent a point each.

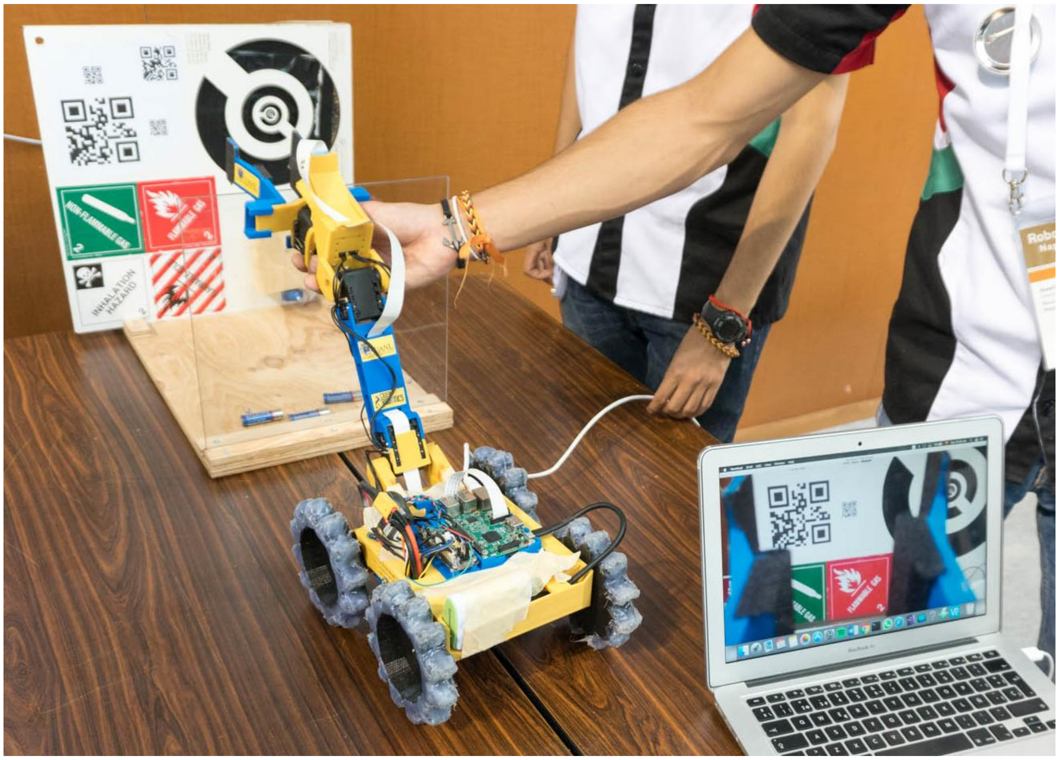

Figure 7: A robot performing the sensor "readiness" test.

Taken together, the terrains, manipulation test, and readiness board represent challenges that robots face getting to the site where they must work, manipulating what they must work on, and then observing it. Thus they each contribute to the scoring. During the preliminaries, there are 12 tests offered.

Four represent mobility and maneuvering challenges that all robots can attempt:

- Center Between Obstacles (no ASTM number assigned yet).
- ASTM WK53649 Evaluating Ground Response Robot Maneuvering: Align Edges.
- ASTM E2827-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Terrains: Crossing Pitch/Roll Ramps [27].
- A variant on the above called Pinwheel Ramps.

Four represent advanced mobility tasks:

- ASTM E2802-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Obstacles: Hurdles [30].
- ASTM E2992/E2992M-17 Standard Test Method for Evaluating Response Robot Mobility: Traverse Sand Terrain [31], and ASTM E2991/E2991M-17 Standard Test Method for Evaluating Response Robot Mobility: Traverse Gravel Terrain [32]. (combined for the preliminaries),
- ASTM E2828-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Terrains: Symmetric Stepfields [6].
- Elevated Ramps (no ASTM number assigned yet).

Finally, two represent "payload" tasks:

- Manipulation pipestar (supporting ASTM WK54272, WK54273, WK54274, and WK54271 as described previously).
- Readiness board (supporting embedded variants of ASTM E2566-17a Standard Test Method for Evaluating Response Robot Sensing: Visual Acuity [33], ASTM WK60783 Evaluating Ground Response Robot Sensing: Audio Speech Intelligibility, ASTM WK57967 Evaluating Ground Response Robot Sensing: Thermal Image Acuity, and ASTM WK54755 Evaluating Ground Response Robot Sensing: Match Colors).

The latter two were repeated twice, once at the start of the preliminary day to evaluate the capabilities that robots started with and again at the end of the day to determine if any capabilities had been lost due to damage/wear during the rest of the testing.

The final runs are a combination of tests that together represent a mission. To ease the logistics of the competition, the readiness board is performed at the start and end of the day, representing the capabilities of the robot at the start and after any wear or damage during the day. This score is then multiplied by the score that the robot achieves during its run through the maze, including both the various terrains as well as the manipulation task.

The Open Academic Robot Kit

The aim of the Open Academic Robot Kit (OARKit) is to make it as easy as possible for a new team to enter the RMRC. The initial OARKit designs have all mechanical parts 3D printable on a low cost printer. All other parts are drawn from a relatively small set of components that are easily available by mail order and all designs, instructions, and source code are available online under an open source license. This way anyone, anywhere in the world, can follow the downloadable instructions and build a robot that can be used as a starting point for their competition entry.

The use of 3D printing for the mechanical structure, rather than a standard kit of parts, allows for much greater flexibility in the robot designs and the opportunity for students to learn proper structural design under manufacturing constraints. It also allows designs to be shared more easily as teams seeking to replicate a design need only own a suitable 3D printer and a cache of standard parts.

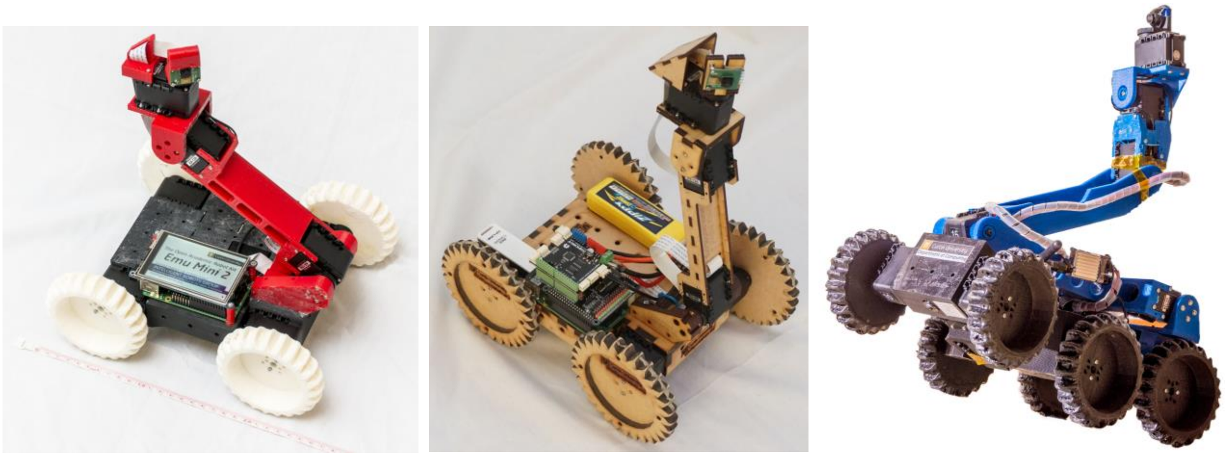

Figure 8: The initial robot designs from the Open Academic Robot Kit.

The first two designs in the OARKit are the “Emu Mini 2” (3D printed and lasercut) and the “Excessively Complex 6-Wheeled Robot”, both shown in Figure 8. These robots are not intended to be the best, or even particularly good, robots for tackling all of the challenges in the RMRC. Rather, they are intended to be good for building additional capabilities from. Over the four years since the OARKit project started, many teams have built variants. Most of these variants share the same parts as the original OARKit: Raspberry Pi for the computation, Arduino for embedded interfacing, and Dynamixel AX-12 servos for motion.

First published in 2014, the OARKit has already spawned a plethora of successors across high schools in Australia, the US, and Europe. Examples of these have appeared at RMRC competitions in the intervening years and some are shown in Figure 9.

Figure 9: Examples of robots that were based (at least in part) on the original Open Academic Robot Kit designs.

Results

The RMRC has been run as an open competition during both the 2017 and 2018 RoboCup Rescue Robot League World Championships, held in Nagoya, Japan and Montreal, Canada, respectively. For the 2017 competition, 13 teams applied of which 10 were qualified and participated in the competition. Preliminary rounds were conducted over one and a half days, resulting in a total of 139 standard tests and five teams qualifying for the finals.

In 2018, 19 teams applied of which 13 were qualified either on the basis of team description materials or performance in the 2017 competition. 11 of those participated in the competition, completing a total of 297 standard tests over two and a half days and again qualifying five teams for the finals. The results of the 2018 preliminary competition are shown in Figure 10. To provide some anonymization, robots have been denoted by two-letter codes. Each robot's performance in each test was normalized such that 100 represented the (possibly equal) best performance in that test. The radar charts are ordered in descending overall performance and show how the performance of the robots change, both in terms of overall performance as well as in specific areas.

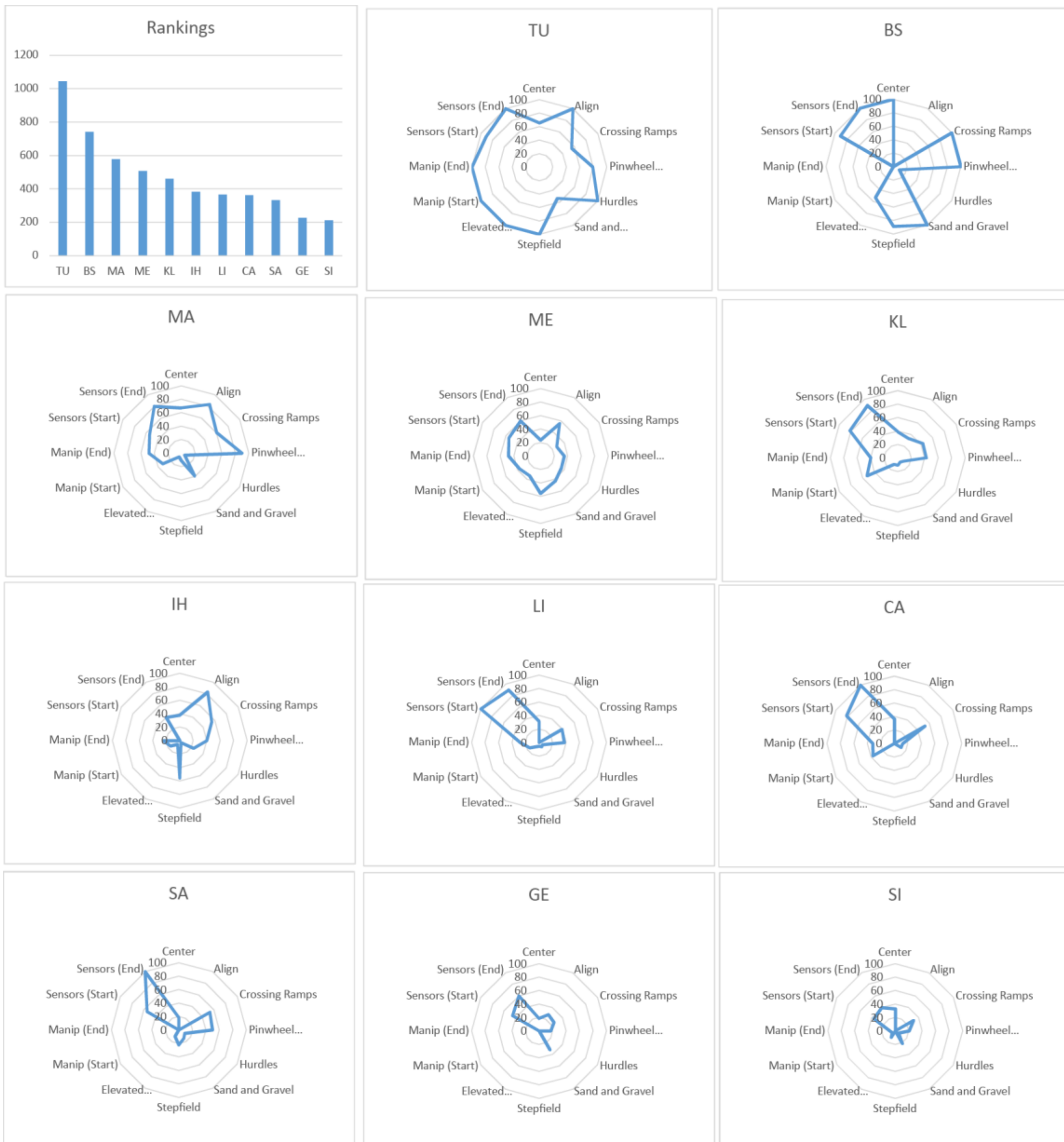

Figure 10: Results of each robot after the preliminary rounds of the 2018 competition. Radar plots for each test are normalized such that 100 represents the best performance in that test. Robots anonymized to two-letter codes.

CONCLUSIONS AND FUTURE WORK

The RMRC is still very much in development as a competition, after two years of open competitions and a preceding two years of trials. Yet we have already observed tremendous growth and maturity in the international community of teams that have developed around it over the last few years. In 2019, we will once again hold the RMRC, in conjunction with the broader RRL, in Sydney, Australia. There we will be refining the rules of the competition as well as the procedures of the test methods, in collaboration with our ASTM E54.09 colleagues. In particular, we will be addressing some issues relating to robot size, whereby several of the robots in the competition barely fit in the mobility test methods and thus the walls became a significant influence on their behavior. This would suggest that these effects either need to be included in the robot performance or otherwise some accommodation made to ensure that their influence is minimized.

Beyond the test methods, we will be providing opportunities for the RMRC and RRL to become more closely integrated again. When the confined space arena, the precursor to the RMRC arena, was first conceived, it was intended to be a "shortcut" into the more difficult areas of the RRL arena for smaller robots. This represented the access options that smaller robots might have in a real response situation that larger robots may not be able to take advantage of. Our plan in 2019 will be to introduce an intermediate arena of test methods at the 60 cm (24 in) scale that will be connected to both the RMRC and RRL arenas for the finals. This will allow robots from both competitions to attempt an intermediate scale before possibly transitioning into the others' arena. The introduction of the intermediate sized arena may also address some concerns about the robots being too big for the arena, albeit with the disadvantage of making the arena larger and thus less cost effective for high school teams to reproduce.

Over the past several years, we have seen the original Open Academic Robot Kit robots proliferate and further develop as teams improve and share the designs. To further encourage contributions to the open source community, we will be adding a multiplier onto teams' preliminary scores, based on a report, document, or other resource that teams submit two weeks prior to the competition. This resource will be scored based on how useful it is for another team to replicate a particular feature or innovation that the team has developed.

The introduction of standard test methods into the RRL and associated RMRC has accelerated the development of robotics for emergency response and public safety. They have helped to communicate the challenges of the application to students and researchers. In turn, they have helped to communicate the prototypical capabilities within academia to the broader user and manufacturer community. In the process they have been critical in forming an interconnected community of researchers, manufacturers, users, and test developers. Further, the competitions have turned out to be ideal proving grounds for the development of standard test methods which has enabled the rapid passage of relevant standards through ASTM E54.09. This symbiotic relationship is a potential model other standard development efforts could use to educate, speed research and development, and disseminate standards through their communities.

ACKNOWLEDGMENTS

We would like to acknowledge the support of all of the mentors and students who participated in, supported, and contributed to this competition and initiative, without whom it would not be possible. We would also like to thank the volunteer referees and administrators who helped us to run the events and all of our collaborators under the ASTM International Subcommittee E54.09 on Response Robots (and the precursor Subcommittee E54.08.01) for their efforts in developing the underlying standard test methods. Finally we would like to thank our colleagues, particularly

William Billotte, Kevin Jurrens, Kenneth Kimble, and Elena Messina for their assistance in reviewing this paper. This work was performed in part under the financial assistance award 70NANB17H265 from the U.S. Department of Commerce, National Institute of Standards and Technology.

References

[1] RoboCup Federation, "The Objective of RoboCup," 27 June 2018. [Online]. Available: http://www.robocup.org/objective.

[2] R. Sheh, S. Schwertfeger and A. Visser, "16 Years of RoboCup Rescue," *KI - Künstliche Intelligenz,* vol. 30, no. 3-4, pp. 267-277, 2016.

[3] A. Jacoff, E. Messina and J. Evans, "Jacoff, Adam, Elena Messina, and John Evans. "Performance evaluation of autonomous mobile robots," *Industrial Robot: An International Journal,* vol. 29, no. 3, pp. 259-267, 2002.

[4] A. Jacoff, E. Messina and J. Evans, "A standard test course for urban search and rescue robots," NIST Special Publication SP, Gaithersburg, 2001.

[5] A. Jacoff, A. Downs, A.-M. Virts and E. Messina, "Stepfield pallets: repeatable terrain for evaluating robot mobility," in *PerMIS '08 Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, Gaithersburg, Maryland, USA, 2008.

[6] *ASTM Standard E2828-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Terrains: Symmetric Stepfields,* West Conshohoken, PA: ASTM International, www.astm.org, 2011.

[7] R. Sheh, B. Collidge, M. Lazarescu, H. Komsuoglu and A. Jacoff, "The response robotics summer school 2013: bringing responders and researchers together to advance response robotics," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, 2014.

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

[9] R. Sheh, "RAS Summer School Held in Australia [Society News]," *IEEE Robotics & Automation Magazine,* pp. 134-136, 3 2015.

[10] M. Buehler, K. Iagnemma and S. Singh, The DARPA Urban Challenge: Autonomous Vehicles in City Traffic, Springer, 2009.

[11] A. Eguchi, "Bringing Robotics in Classrooms," in *Robotics in STEM Education*, Springer, Cham, 2017, pp. 3-31.

[12] RoboCup Singapore, "CoSpace Robotics," 2015. [Online]. Available: http://cospacerobot.org/. [Accessed 27 June 2018].

[13] FIRST, "FIRST Inspires," 2018. [Online]. Available: https://www.firstinspires.org/. [Accessed 27 June 2018].

[14] New Energy Industrial Technology Development Organisation (NEDO), "World Robot Summit," 2018. [Online]. Available: http://worldrobotsummit.org/en/about/. [Accessed 27 June 2018].

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

[16] RoboCup Federation, "RoboCup Junior," 2018. [Online]. Available: http://junior.robocup.org/. [Accessed 27 June 2018].

[17] Arduino, "Arduino," 2018. [Online]. Available: https://www.arduino.cc/. [Accessed 27 June 2018].

[18] Raspberry Pi Foundation, "Raspberry Pi," 2018. [Online]. Available: https://www.raspberrypi.org/. [Accessed 27 June 2018].

[19] R. Sheh, "The Open Academic Robot Kit," 2018. [Online]. Available: http://www.oarkit.org/. [Accessed 27 June 2018].

[20] R. Sheh, A. Eguchi, H. Komsuoglu and A. Jacoff, "The Open Academic Robot Kit," in *Robotics in STEM Education*, Springer, Cham, 2017, pp. 85-100.

[21] Robotis, "TurtleBot 3," 2018. [Online]. Available: http://www.robotis.us/turtlebot-3/. [Accessed 27 June 2018].

[22] INRIA, "Poppy Project," 2016. [Online]. Available: https://www.poppy-project.org/en/. [Accessed 27 June 2018].

[23] Vorpal Robotics LLC, "Vorpal Robots," 2018. [Online]. Available: https://www.vorpalrobotics.com/. [Accessed 27 June 2018].

[24] R. Ma, L. Odhner and A. Dollar, "A Modular, Open-Source 3D Printed Underactuated Hand," in *IEEE International Conference on Robotics and Automation (ICRA)*, Karlsruhe, Germany, 2013.

[25] *ASTM Standard E2804-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Obstacles: Stairs/Landings,* West Conshohocken, PA: ASTM International, www.astm.org, 2011.

[26] *ASTM Standard E2826-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Terrains: Continuous Pitch/Roll Ramps,* West Conshohocken, PA: ASTM International, www.astm.org, 2011.

[27] *ASTM Standard E2827-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Terrains: Crossing Pitch/Roll Ramps,* West Conshohocken, PA: ASTM International, www.astm.org, 2011.

[28] "Subcommittee E54.09 on Response Robots," ASTM International, 2019. [Online]. Available: https://www.astm.org/COMMIT/SUBCOMMIT/E5409.htm. [Accessed April 2019].

[29] A. Jacoff, E. Messina, H.-M. Huang, A. Virts, A. Downs, R. Norcross and R. Sheh, "Guide for Evaluating, Purchasing, and Training with Response Robots Using DHS-NIST-ASTM International Standard Test Methods," 2013. [Online]. Available:

https://www.nist.gov/sites/default/files/documents/el/isd/ks/DHS_NIST_ASTM_Robot_Test_Methods-2.pdf. [Accessed 27 June 2018].

[30] *ASTM Standard E2802-11 Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Obstacles: Hurdles,* West Conshohocken, PA: ASTM International, www.astm.org, 2011.

[31] *ASTM Standard E2992/E2992M-17 Standard Test Method for Evaluating Response Robot Mobility: Traverse Sand Terrain,* West Conshohocken, PA: ASTM International, www.astm.org, 2-17.

[32] *ASTM Standard E2991/E2991M-17 Standard Test Method for Evaluating Response Robot Mobility: Traverse Gravel Terrain,* West Conshohocken, PA: ASTM International, www.astm.org, 2017.

[33] *ASTM Standard E2566-17a Standard Test Method for Evaluating Response Robot Sensing: Visual Acuity,* West Conshohoken, PA: ASTM International, www.astm.org, 2017.