Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, NV, October 27-31, 2003.

Test Arenas and Performance Metrics for Urban Search and Rescue Robots

Adam Jacoff, Elena Messina, Brian A. Weiss[1], Satoshi Tadokoro[2], and Yuki Nakagawa[3]

1. Intelligent Systems Division
National Institute of Standards and Technology
Gaithersburg, MD
adam.jacoff@nist.gov

2. Department of Computer Science and Systems Engineering
Kobe University
Kobe, Japan
tadokoro@cs.kobe-u.ac.jp

3. Exhibition Development Group
National Museum of Emerging Science and Innovation
Tokyo, Japan
yuki@nkgw.com

ABSTRACT

In this paper, we discuss the development and proliferation of robot test arenas that provide tangible, realistic, and challenging environments for mobile robot researchers interested in urban search and rescue applications and other unstructured environments. These arenas allow direct comparison of robotic approaches, objective performance evaluation, and can ultimately provide a proving ground for field-able robotic systems such as those used at the World Trade Center collapse. International robot competitions using these arenas require robots to negotiate complex and collapsed structures, find simulated victims, and generate human readable maps of the environment. A performance metric is presented which quantifies several pertinent robot capabilities and produces an overall score used to evaluate and compare robotic implementations. Future directions for the arenas and the competitions are also discussed.

KEYWORDS:
autonomous mobile robots, urban search and rescue, sensory perception, knowledge representation, planning, mapping, collaboration, performance metrics.

1.0 INTRODUCTION

Reproducible and widely known challenges can help evolving fields by providing reference problems with measures of performance which allow researchers to compare implementations, communicate results, and leverage each other's work. The National Institute of Standards and Technology (NIST) developed the *Reference Test Arenas for Autonomous Mobile Robots* to focus research efforts, provide direction, and accelerate the advancement of mobile robot capabilities. These arenas, modeled from buildings in various stages of collapse, allow objective performance evaluation of robots as they perform a variety of urban search and rescue (USAR) tasks [Jacoff et al., 2000]. Robots explore the maze-like test course, negotiate obstacles, find simulated victims, and generate human readable maps of the environment. The NIST arenas have hosted the Rescue Robot Competitions at the American Association for Artificial Intelligence (AAAI) and the International Joint Conference on Artificial Intelligence (IJCAI) meetings since 2000. They are available year round to researchers in an effort to raise awareness of the challenges involved in search and rescue applications and provide a venue for collaboration.

Last year, Japan's National Museum of Emerging Science and Innovation (MeSci) replicated the arenas to host the RoboCup Rescue Robot League competition at RoboCup2002 in Fukuoka, Japan. The entire event drew 117,000 spectators to watch 188 teams (over 1000 participants) from 30 countries compete in the various leagues. The Rescue Robot League competition included 10 teams from 5 countries.

RoboCup's autonomous soccer playing robot leagues (small size, middle size, legged, and simulation) have evolved considerably since the initial event in 1997. Many teams now demonstrate sophisticated sensory perception, reactionary behaviors, and multi-robot collaboration. The relatively structured environment of the soccer field has provided a much needed reference environment allowing teams to focus on issues of localization and strategy while dealing with the dynamic difficulties created by an opposing team. The Rescue Robot League was initiated to move the focus away from highly structured environments toward unstructured environments by highlighting the challenges involved in search and rescue applications. In some cases, existing robotic capabilities in perception, localization, and mapping can be adapted to negotiate obstacles and locate victims in the rescue arenas. The least difficult section, representing modestly collapsed living spaces, is designed for immediate technology transfer of such implementations. The more difficult sections encourage collaboration between systems with sophisticated autonomous capabilities and more rugged, teleoperative implementations. Due to the complexity of search and rescue applications, fully autonomous robots are not yet practical. However, bounded and adjustable autonomous behaviors, along with improved operator interfaces, are essential to effectively field teams of collaborative robots in realistic situations.

The vision for the RoboCup Rescue Robot League is ambitious:

When disaster happens, minimize risk to search and rescue personnel, while increasing victim survival rates, by fielding teams of collaborative robots that can:

- *Autonomously negotiate compromised and collapsed structures*
- *Find victims and ascertain their conditions*
- *Produce practical maps of their locations*
- *Deliver sustenance and communications*
- *Identify hazards*
- *Provide structural shoring*

…allowing human rescuers to quickly locate and extract victims.

2.0 ARENA DESIGNS

The primary goal of the test arenas is to provide challenging and reproducible environments to evaluate mobile robot capabilities and behaviors. Collapsed structures found in the USAR domain provide a huge range of obstacles and features from which to model environments. In an effort to encourage autonomy, the arenas attempt to isolate and test typical sensors used by mobile robots while providing somewhat realistic challenges to robot agility.

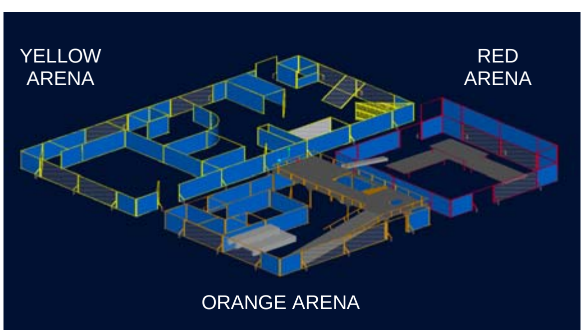

Figure 1: Model of the NIST Reference Test Arenas for Autonomous Mobile Robots

There are three separate indoor arenas, each labeled by a color, forming a continuum of difficulty for robots (Figure 1). The NIST arenas and the MeSci arenas are similar in design but display characteristics of their local cultures and building materials. These differences are important because structural collapses differ widely based on regional and local building materials and it is important to evaluate robotic systems in realistic situations.

The Yellow arena is the easiest to traverse (Figure 2a and 2b). Researchers using non-agile robots to test their sensory perception, mapping, or planning algorithms can explore the entirety of the Yellow arena. It consists of a planar maze with isolated sensor tests (tactile, audible, sonar, infrared, visual, ladar) in the form of obstacles or simulated victims. The maze is easily reconfigurable to form a variety of passages. It has doors, blinds, and simple collapses to block passages during missions, specifically challenging mapping and planning algorithms.

Figure 2a: A simulated bedroom and victim in the NIST Yellow arena

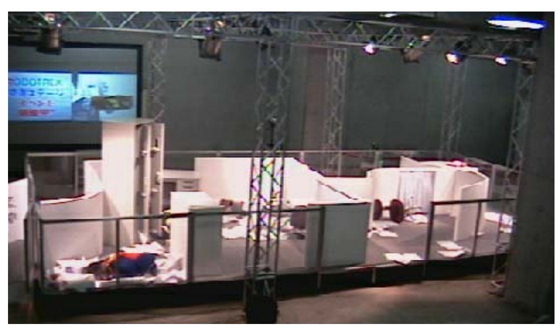

Figure 2b: An office environment in the MeSci Yellow Arena

The Orange arena provides more difficult challenges for both sensing and agility (Figure 3a and 3b). Assorted types of flooring materials are introduced. There is an elevated floor section, reachable via ramp, stairs, or ladder, requiring considerable agility to negotiate. Holes in the elevated flooring provide negative obstacles to avoid. Leaning collapses provide perceptual and physical obstacles to negotiate without causing a secondary collapse. Robots must consider the entirety of this three-dimensional maze to successfully map the environment and plan their way through the arena. The Orange arena is also reconfigurable in real-time using doors, blinds, and simulated collapses. Some of the cultural differences found in this arena include tile flooring and carpeting in the NIST arena versus typical tatami mats and wall materials in the MeSci arena.

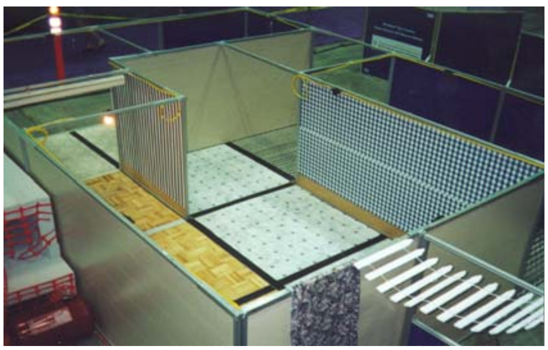

Figure 3a: Assorted flooring materials within the maze in the NIST Orange arena

Figure 3b: Maze and elevated section of the MeSci Orange arena

The Red arena provides the least structure and the most challenge to robot agility (Figure 4a and 4b). It is essentially a rubble pile with assorted debris throughout the arena and is a very difficult environment in which to sense and maneuver. The debris, which includes steel wire, gravel, plastic bags, pipes, etc., is very problematic for robot locomotion, and even harder for sensory perception algorithms. There are leaning and pancaked collapses (floors collapsed onto lower floors). Tactile obstacles such as unstable flooring may collapse under the weight of a heavy robot. Some cultural differences found in this arena include simulated concrete rubble in the NIST arena versus wooden building materials in the MeSci arena.

Figure 4a: Simulated rubble pile and unstable flooring in the NIST Red arena

Figure 4b: Collapsed housing materials in the MeSci Red arena

3.0 THE SIMULATED VICTIMS

Simulated victims provide the motivation for robots to fully explore and map the arenas (Figure 5). There are as many as thirty simulated victims placed throughout the arenas. Each victim displays up to five signs of life: human form (mannequins dressed as civilians and rescuers); body heat (heating pads and blankets under clothing); motion (waving arms and moving fingers); sound (audio tape recorded shouting, moaning, tapping, and locator beacons); and CO_2 emissions (leaking tanks concentrated in voids). These signs of life are used in various combinations to simulate victim states: aware, semi-conscious, or unconscious. The intent is to encourage use of multiple sensors to increase confidence in victim identifications, while challenging robots and their operators to accurately determine vital information regarding a victim's condition.

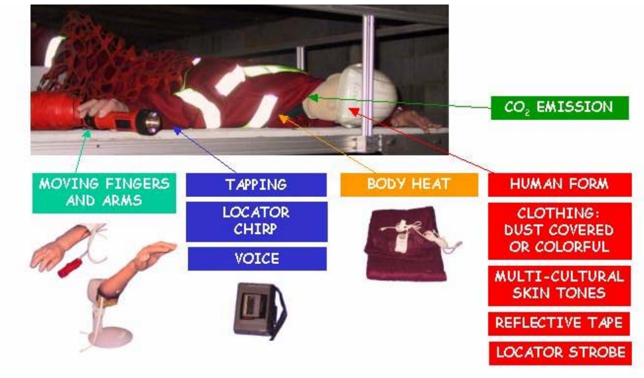

Figure 5: Simulated victim with form, motion, thermal, sound, and chemical signatures

Each simulated victim is placed in one of four typical rescue situations in urban search and rescue environments: surface, lightly trapped, void, entombed. Surface victims are relatively visible. Lightly trapped victims are partially occluded and trapped under debris. Void victims are occluded under leaning or pancaked collapses requiring robots to maneuver into particular viewing positions to identify victims. Entombed victims are fully occluded under collapsed structures or rubble requiring robots to probe the debris with various sensors to determine a victim's location and state.

These simulated victim placements encourage different combinations of robot mobility, sensing, localization, and operator interfaces, while focusing research efforts toward practical implementations that may ultimately prove effective in real disaster sites.

4.0 THE COMPETITIONS

The NIST arenas made their debut in Austin, Texas at the 2000 AAAI Mobile Robot Competition [Murphy, 2000][Schultz, 2001]. Their second deployment was in Seattle, Washington at the 2001 IJCAI meeting, where the RoboCup and AAAI Rescue Robot competitions were jointly held [Jacoff et al, 2002]. Since then, both organizations have held annual competitions in Fukuoka, Japan and Edmonton, Canada respectively. These competitions are classified as ranked competitions using objective scoring based on specified criteria [Yanco, 2001] and the rules for these competitions are based on those developed by the AAAI/RoboCup steering committee. Several refinements were implemented in advance of the 2002 competitions to increase the realism, simplify the judging, and add performance incentives. These changes are described below.

4.1 ZONES OF OPERATION

Vaguely defined in previous competitions, three zones of operation are now strictly identified [Murphy et al, 2002]. The hot zone includes the three arenas, representing the restricted area of an actual collapsed structure. The warm zone, representing the closest staging area to the collapsed structure, includes the operator station and the robot's starting point. The cold zone, representing the maintenance areas that are far away from the collapse, includes the team preparation area.

Each team is given minimal prior knowledge of the hot zone beyond its location relative to the warm zone. No operators or team members are allowed in the hot zone before or during the competition. The robot is placed in the warm zone at the start of each competitive mission.

4.2 COMPETITIVE MISSIONS

In the preliminary round of competition, each team is allowed three missions of twenty minutes each. The best two of three mission scores for each team are counted, and the highest scoring teams advance to the next round of competition. This initially allows each team one failed mission without consequence. Time permitting, each round begins with no score and is administered similarly, although in later rounds all scores count. Shorter competitions may carry over scores from previous rounds to reward stamina. Among the finalists, the highest three scores earn place awards, provided they score over a minimum scoring threshold based on finding a reasonable percentage of victims. This may vary based on the arena difficulty so is determined by the league chair at each competition.

4.3 OPERATORS AND ROBOTS

Experience deploying robots at the World Trade Center collapses highlighted the need to increase the ratio of robots to operators at the scene of a disaster [Murphy et al, 2002]. So mixed initiative approaches are encouraged, using both teleoperation and autonomy, to allow high-level management of multiple robots. For scoring purposes, operators are defined as all support personnel present in the warm zone during the competition, regardless of their interaction with a robot. Robots are counted in the scoring once they have found a unique victim (a victim that has not yet been identified). Multiple robot teams are encouraged, but collaboration between the robots is essential. Two robots finding the same victim does not double the score.

4.4 ROBOT RESETS

The robot may return to the warm zone for service at any time during a mission. Once there, the operator may repair the robot and send it back into competition without penalty. However, mission time does not pause for such events. If the robot is incapacitated for any reason within the hot zone, the operator may call 'reset' to have the robot returned to the warm zone. Once there, the operator may repair the robot and send it back into competition. The penalty for a 'reset' is equivalent to an additional operator in the scoring formula. The intent of this rule is to allow a robot to continue if it gets stuck or has minor problems.

4.5 VICTIM IDENTIFICATION and MAPPING

While searching the arenas, negotiating both physical and sensory obstacles, robots must find simulated victims to score points. When a robot finds a victim, the operator notes the perceived signs of life (form, motion, heat, sound, chemical), and maps the location of the victim. A victim may be counted only once (1.0 point per victim). At the end of the mission, the team must immediately produce a map (preferably sensor generated and automatically annotated) which indicates the location of all victims and other arena features. This map is used by the Judge to find and verify victim identifications. The accuracy of the mapped victim locations and the overall map quality are key elements in the performance metric.

4.6 PENALTIES

Search and rescue applications require robots to interact with the environment and victims in a controlled and non-destructive manner. So during competition, penalties may be assessed based on the robot's actions. Arena penalties are assessed for uncontrolled bumping or other undesirable contact with the arena that does not result in damage (-0.25 point deduction). A heavy damage penalty is assessed for undesirable shifting of arena features or damage to arena components (-0.75 point deduction). Victim bumping penalties are assessed for any robot contact with a victim (-0.25 point deduction). Any robot contact that repositions or 'harms' the victim is assessed a harsher penalty (-0.75 point deduction). The intent is for robots to clearly demonstrate controlled motion within the arenas, and to show they fully recognize victims by avoiding any contact. Penalties may compound if a robot causes heavy damage to an arena that results in victim harming. The performance metric provides that each penalty is weighted according to arena difficulty.

4.7 SCORING

The scoring procedure for each round requires one referee to monitor each robot and one judge playing the role of incident commander. For consistency in scoring across teams, these personnel maintain their respective responsibilities for an entire round of competition. Referees may be non-competing team members or other volunteers whose responsibilities include tracking the robot throughout the mission, noting victim identifications, and assessing penalties. The Judge for each round is typically an organizing official or a non-competing team leader whose responsibilities include maintaining the official time, interacting with the operator when potential victims are found, and final scoring. Following each mission, the Judge uses the map generated by the team to find each victim, determine the positional accuracy, assess the map quality, tally robot penalties, and calculate the mission score. Judges have final authority over any disputes.

5.0 PERFORMANCE METRIC

The performance metric used to score the competition is predominantly similar to that used at AAAI/RoboCupRescue 2001 [Jacoff et al, 2001], but contains some changes to include penalties, encourage more useful maps, and simplify the scoring procedure (Figure 6). Remaining intact is the core ratio of robots to operators, which encourages fewer operators and rewards autonomous implementations. The victim point weighting for each arena also remains intact. The greater the difficulty of the arena, the higher the weighting for victim identification. This year's changes in the performance metric are discussed below.

5.1 EMPHASIS

Several capabilities should be demonstrated prior to fielding robots at real disaster sites [Murphy et al, 2002]. The area around a structural collapse is tightly controlled to minimize the number of people in harm's way, so fewer operators should be able to deploy teams of robots if necessary. These robots must demonstrate controlled interaction with the environment or risk triggering secondary collapses. Finally, accurately locating victims and producing human readable maps is essential for the incident commander to deploy human resources effectively

RobotRescueScore =
(VictimsPoints (NumberOfRobots / (1+ NumberOfOperators)^3) AverageAccuracy

VictimsPoints **= (YellowVictimsFound - YellowPenalties) * (*YellowVictimWeighting*)**
+ (OrangeVictimsFound - OrangePenalties) * (*OrangeVictimWeight)*
+ (RedVictimsFound - RedPenalties) * (*RedVictimWeighting*)

[*YellowVictimWeighting = 0.50*]
[*OrangeVictimWeighting = 0.75*]
[*RedVictimWeighting = 1.00*]

NumberOfRobots **=** *Number of robots that find a unique victim*
NumberOfOperators **=** *Number of operators having touched the robot or are in the hot zone*
AverageAccuracy **=** *(positional accuracy + map quality) / Total victims found*

Figure 6: Performance Metric

toward rescue situations. So the current performance metric encourages the following robot team capabilities:

5.2 ADJUSTMENTS

The average accuracy term now includes two critical values: positional accuracy and map quality. The positional accuracy of a mapped victim refers to the mapped location of a victim relative to the entrance, walls and features within the arenas. If any part of a simulated victim is located within 1m of the location shown on a map, the positional accuracy for this found victim is 1.0 point. If the victim is mapped to an adjacent 1 meter cube, 0.5 point. If the victim is mapped to any other part of the arena or the position is not known, the positional accuracy for that victim is 0.25 point.

The other key element to average accuracy is map quality. Map quality values can range from 0 to 1.0 point based on the clarity and practicality of the maps that are generated for each arena. A map that conveys no indication as to where the robot searched is given 0 points for map quality. A map that displays a direction to a victim from the start position is map quality value of 0.2 points. A map that provides basic topographic information to get to the victim (for example, 1st right, 2nd left) is scored 0.4 points. Hand drawn maps that indicate victim locations while showing obstacles and features are scored 0.8 points. Sensor generated maps that produce an accurate victim location and note obstacles and features score 1.0 point. These categories will vary, as teams become more successful in localization and mapping. The intent is to identify the spectrum of possibilities and encourage the most effective implementations.

5.3 EFFECTIVENESS

The performance metric produced an even scoring distribution. Two teams at this year's RoboCup competition and three teams at the AAAI competition exceeded the minimum scoring threshold to receive place awards. The performance metric proved to be effective in deterring some obviously undesirable behaviors. But as expected, some teams found ways to exploit the performance metric without pushing the state of the art in mobile robot capabilities.

One example is parallel teleoperation of robots. Teams with multiple robots used multiple operators to go to different arenas concurrently. This operational model does not improve the 1:1 ratio of operators to robots. It simply rewards well-funded teams, and should be discouraged.

Another example is sequential teleoperation of robots. A single operator sequentially controls multiple robots, each attempting to find only one unique victim in an attempt to inflate their scoring. This clearly undermines the intent of the performance metric and should be discouraged. However, sequential teleoperation can be beneficial if robots are used to assist, or augment, one another's capabilities. This requires a careful balance of incentives and deterrents in the performance metric.

5.4 PROPOSED CHANGES

The performance metric emphasizes the importance of several key robot capabilities: localization, mapping, human interfaces, reliability. However, certain changes to the performance metric could enhance the overall fairness of the competitions, and further encourage desirable robotic capabilities while discouraging undesirable team strategies.

For example, the performance metric should reward robots that use multiple sensors to identify victims. In the field, multiple sensors would increase the confidence level that a victim has been found. Similarly, false positive identification of victims should be discouraged through penalties because in the field they could jeopardize rescuers unnecessarily.

The performance metric should also discourage reliance on radio communications. In a real disaster environment communication frequencies are already strained by the needs of emergency responders. Also, structural debris makes radio communication unreliable from within buildings or rubble piles. Limiting radio communications would encourage autonomous behaviors that would be very beneficial in actual deployment scenarios.

6.0 FUTURE DIRECTIONS

NIST's Reference Test Arenas for Autonomous Mobile Robots are currently available year-round in Gaithersburg, MD for robot researchers and developers interested in evaluating their robotic implementations. So are the arenas at the National Museum of Emerging Science and Innovation in Tokyo, Japan. As these arena sites proliferate, competition participation increases. This year, a replica of the NIST Orange arena will host the first RoboCup Rescue - American Open competition at Carnegie Mellon University in Pittsburgh, PA. Likewise, the MeSci arenas will host the first Japanese Open competition. Newly built replicas of the arenas will host this year's international RoboCup Rescue Robot League competition in Padua, Italy. After the competition, these arenas will be available year-round at a rescue training facility in Rome. This will be repeated each year in each host county that RoboCup visits, while the NIST arenas continue to host the AAAI/IJCAI Rescue Robot competitions.

Arena proliferation will help increase awareness of the challenges involved in search and rescue applications, provide testing in representative environments, and promote collaboration between researchers. As the robots begin demonstrating repeated successes against the obstacles posed in these arenas, the level of difficulty will be increased accordingly so that the arenas provide a stepping stone from the laboratory to the real world. Meanwhile, the yearly competitions provide direct comparison of robotic approaches, objective performance evaluation, and a public proving ground for field-able robotic systems.

ACKNOWLEDGEMENTS

The Reference Test Arenas for Autonomous Mobile Robots were developed through support from the Defense Advanced Research Projects Agency (DARPA).

REFERENCES

[1] Blitch, JG, "Artificial intelligence technologies for robot assisted urban search and rescue," EXPERT SYSTEMS WITH APPLICATIONS 11: (2) 109-124 1996.

[2] Jacoff, A., Messina, E., Evans, J., "A Standard Test Course for Urban Search and Rescue Robots," Performance Metrics for Intelligent Systems Workshop, August 2000, Gaithersburg, MD.

[3] Kitano, H., Asada, M., Kuniyoshi, Y., Noda, I., Osawa, E., Matsubara, H., "RoboCup: A Challenge Problem for AI," AI Magazine 18(1): Spring 1997

[4] Murphy, R., Casper, J., Micire, M., Hyams, J., "Assessment of the NIST Standard Test Bed for Urban Search and Rescue Competitions," Performance Metrics for Intelligent Systems Workshop, August 2000, Gaithersburg, MD.

[5] Schultz, A. "The 2000 AAAI Mobile Robot Competition and Exhibition," AI Magazine, 22(1), spring 2001, AAAI.

[6] Jacoff, A., Messina, E., Evans, J. "Performance Evaluation of Autonomous Mobile Robots," Industrial Robot: An Industrial Journal, Vol. 29 No. 3 Emerald Library, May 2002.

[7] Murphy, R., Blitch, J., and Casper, J., "AAAI/RoboCup-2001 Urban Search and Rescue Events," AI Magazine, 23(1): Spring 2002, 37-42

[8] Yanco, H. "Designing Metrics for Comparing the Performance of Robotic Systems in Robot Competitions," Performance Metrics for Intelligent Systems Workshop, September 2001, Mexico City, Mexico.