Measuring the Performance of Urban Search and Rescue Robots

Elena R. Messina and Adam S. Jacoff
National Institute of Standards and Technology, Gaithersburg, MD
elena.messina@nist.gov, adam.jacoff@nist.gov

Abstract— *The Department of Homeland Security is sponsoring development of standards for urban search and rescue (USAR) robots. This program is being coordinated by the National Institute of Standards and Technology and will result in consensus standards developed through ASTM International. Robot deployment categories and performance requirements have been identified by emergency responders, These requirements are being translated into tests and metrics with which to measure the performance of robots. Several test methods have been entered into the standards process. Three major exercises have been held at US&R training facilities, in which responders work side-by-side with robot manufacturers to experiment with robot deployments in relevant scenarios and to refine and expand performance requirements. To date, over forty robots, including ground, aerial, and aquatic, have been involved. Supporting projects are developing an ontology of robot capabilities and situational constraints. In general, these efforts will enable responders to enhance their effectiveness while reducing risk to personnel during disasters through use of robotic assets.*

1. Introduction

The Department of Homeland Security and others have noted the need for comprehensive standards to support development, testing, and certification of effective robotic technologies for Urban Search and Rescue (USAR) applications. These standards will address all aspects of robot performance, such as robot mobility, sensing, communications, power, logistics, integration, and human-system interaction, as well as terminology. Such standards will allow DHS to provide guidance to local, state, and federal homeland security organizations regarding the purchase, deployment, and use of robotic systems for USAR applications.

USAR is defined as "the strategy, tactics, and operations for locating, providing medical treatment, and extrication of entrapped victims." [1] USAR teams exist at local and state levels. On the federal level, the Federal Emergency Management Agency (FEMA) has 28 Task Forces that structure local emergency responders into integrated disaster response forces. Well-known examples of events in which FEMA USAR Task Forces were deployed include the collapse of the World Trade Center and Hurricane Katrina.

In studies that focused on technology needs, robots have been specifically identified as holding promise for assisting responders in carrying out their search and rescue missions. USAR has been described as a domain "that is a very dangerous job for human rescuers, poses an almost infinitely difficult spectrum of challenges, and yet provides an opportunity for robots to play a pivotal support role in helping to save lives." [2] Currently, the state of robot technology overall is not yet very mature and there are significant shortcomings.[3] Purchase and maintenance costs are prohibitive for most response organizations. Therefore, a common set of requirements and measures for robot performance in a set of relevant categories is needed to help guide purchase decisions. Furthermore, the same set of information provides concrete development goals for robot manufacturers and researchers.

Since robots are not currently used by responders in USAR, new standard operating procedures must be developed. In order to refine the performance requirements and to experiment with deployment approaches for integrating robots into missions, NIST has been working with FEMA Task Forces to organize exercises at USAR Training Facilities. At these exercises, robot manufacturers and researchers work side by side with USAR Task Force Members to insert robots into missions within training scenarios. Both robot developers and responders benefit from the *in situ* collaboration. It is essential to have all stakeholders participate in the process of defining appropriate ways to utilize robots, measure their performance, and generate consensus standards.

More formal characterization of the disaster scenarios as well as the robot capabilities must be developed as well. Currently, there are no commonly accepted means of describing quantitatively (or even qualitatively) the type of building collapse. Similarly, there is no generic means of conveying robot capabilities in unambiguous terms that can be expanded and reused. To address these shortcomings, one component of the overall project involves defining classifications or taxonomies of both disasters/collapses and of robot capabilities. Eventually it is hoped that these two can be used in concert to help match appropriate robot capabilities when responding to different disasters.

This paper is organized as follows. A summary of the requirements definition process is found in Section 2. Section 3 details the standards development approach, which

1-4244-1053-3/07/$25.00 ©2007 IEEE.

Table 1: Robot Deployment Categories

Robot Category	Employment Role(s)
Ground: Peek Robots	Provide rapid audio-visual situational awareness; provide rapid HAZMAT detection; data logging for subsequent team work
Ground: Collapsed Structure--Stair/ Floor climbing, map, spray, breach Robots	Stairway & upper floor situational awareness; mitigation activities; stay behind monitoring
Ground: Non-collapsed Structure-- Wide area Survey Robot	Long range, human access stairway & upper floor situational awareness; contaminated area survey; site assessment; victim identification; mitigation activities; stay behind monitoring
Ground: Wall Climbing Deliver Robots	Deliver Payloads to upper floors; provide expanded situational awareness when aerial platforms are unavailable or untenable
Ground: Confined Space, Temporary Shore Robots	Adaptive, temporary shoring; provide stay behind monitoring; victim triage & support
Ground: Confined Space Shape Shifters	Search; provide stay behind monitoring
Ground: Confined Space Retrieval Robots	Retrieve objects from confined spaces; provide stay behind monitoring
Aerial:High Altitude Loiter Robots	Provide overhead perspective & sit. awareness; provide HAZMAT plume detection; provide communications repeater coverage
Aerial: Rooftop Payload Drop Robots	Payload delivery to rooftops; provide overhead perspective; provide communications repeater coverage
Aerial: Ledge Access Robot	Object retrieval from upper floors; crowd control with a loudspeaker object attached, provide situational awareness
Aquatic: Variable Depth Sub Robot	Structural inspection; leak localization/mitigation; object (body) recovery
Aquatic: Bottom Crawler Robot	Water traverse; rapid current station keeping; object recovery

is focused on creation of test methods that measure performance in the various requirements categories. An example of how a test method was developed is included in this section. Response robot exercises, which are held in order to promote communication among the stakeholders and to deepen understanding of how robots can be deployed, are described in Section 4. Section 5 briefly introduces the work being done to define disaster and robot taxonomies. Section 6 contains a summary and some initial conclusions based on the informal performance evaluations performed thus far.

2. Requirements Definition

The project began with a series of workshops in which FEMA Task Force members defined their expectations of robot performance. The definition process captured individual categories of requirements, along with how to measure robot performance against each requirement and objective (meaning desired) and threshold (minimum or maximum acceptable) values. The initial workshop process intentionally avoided inclusion of robot vendors and researchers so as not to bias the responders' thoughts. After three workshops, a preliminary set of requirements was produced.[4] Over one hundred individual requirements were grouped into higher-level categories such as Human-System Interaction, Logistics, Operating Environment, Safety, Chassis, Communications, Mobility, Payload, Sensing and Power. Example requirements are shown in Fig. 1.

Type:	**COMMUNICATIONS**
Sub-Type:	**N/A**
Requirement:	**RANGE – BEYOND LINE OF SIGHT**
Metric:	**METERS**
Description: This requirement captures the responders' expectation to project remote situational awareness into compromised or collapsed structures or to convey other types of information. They specifically noted that the robot should be able to ingress a specified number of meters into the worst case collapse, which was further defined as a reinforced steel structure. This requirement also covers operations around corners of buildings and other locations beyond line of sight. The responders made no distinction regarding tethered or wireless implementations to address this requirement.	

Type:	**HUMAN-SYSTEM INTERACTION**
Sub-Type:	**CONTEXT**
Requirement:	**PROTECTIVE CLOTHING**
Metric: SCALE 1-5 1 = No protection 3 = Minimum protection (threshold) 5 = Complete protection (objective)	
Description: This requirement captures the responders' expectation to be operating the system while wearing personal protective equipment such as gloves, helmet, eye protection, ear protection, etc. The operator should be able to maintain acceptable usability (discussed in greater detail in the Test Methods: Human-System Interaction section of this report) of the system while wearing the stated level of personal protective equipment	

Figure 1: Example of Performance Requirements. A category (type/sub-type) is identified, along with a metric to be used in measuring the robot's performance, which can be binary, length, time, or a scale defined by the responders.

An initial set of potential robot deployment categories within responses was also produced at the first workshops and included in [4]. A total of thirteen distinct categories were identified, along with concepts about deployment and tradeoffs in the design of each. The deployment categories do not necessarily map one-to-one to robot designs or categories; a particular robot may indeed serve in more than one category. The robot deployment categories are shown in Table 1. Due to space constraints, the deployment methods and tradeoffs are not shown. This categorization was intended to help characterize the domain in terms of robotic applications and to help the responders further define the performance requirements. USAR missions have wide variance in the types of structures, amount of collapse, surrounding conditions (e.g., is the area to be searched dry, wet, submerged, dusty, etc.), and regions within the zone to be searched (upper stories, below ground, adjoining vicinity, interior voids, etc.).

The next steps involve further analysis of each requirement in order to create a test method specifically designed to measure performance in that area. Given the large number of requirements, NIST is taking a staged approach to converting requirements to test methods. The ordering of requirements into "waves" is an ongoing process that is influenced by the widest applicability of requirements ("most bang for the buck" as it were) as well as the relative maturity of the technologies that would address a particular requirement. For instance, it would be premature to create a formal test method for the requirement for onboard generation of maps by the robot, since relatively few systems have such a capability that can be reliably deployed in the challenging environments that USAR robots must face.

To help identify which test methods to tackle in the initial wave of standards, NIST obtained further input from the responders about the applicability of requirements across different robot deployment categories. FEMA Task Force members indicated whether or not each requirement would apply to each of the thirteen categories. This produced a set of data that was used statistically analyzed to define which requirements should be included in the initial set of test method development. A set of 25 items was proposed for the initial wave of standardization.

Having such a breadth of response/deployment scenarios implies that the performance requirements (i.e., the objective and threshold values) will vary accordingly. For instance, the desired minimum speed over soft terrain for a peekbot is much lower than for a wide area ground survey robot. Similarly, the distance that the robot operator should be able to see through the onboard cameras will be much shorter for the peekbot (since it is expected to be in fairly confined spaces) versus the other two categories, which require long-distance situational awareness. Therefore, an additional component in refining the performance requirements is the definition of different performance ranges according to robot categories.

In later events, responders were exposed to a wide range of robots during exercises at FEMA USAR training facilities. Most of the thirteen categories were represented. Having had some experience with using these robots in different training deployment scenarios, responders were able to prioritize three robot categories above the others. These were Ground, Peekbots, Ground Wide-area Survey Robots, and Aerial Loiter/Survey Robots. Determining which robot categories were being targeted for the initial wave of standards submissions allowed the test method developers to design the tests to address the appropriate performance ranges.

3. Standards Development

The approach taken is based on assessing the end users' needs, ensuring that the process throughout is open to all stakeholders, and is built on a foundation of ongoing validation and refinement of requirements as well as the corresponding test methods. ASTM International was selected by the Department of Homeland Security to host the standards produced by this project. A Task Group was created within the Operational Equipment Subcommittee of the Homeland Security Committee (E54). ASTM International follows a consensus-based process for developing standards and strives for balanced representation among users (responders and their organizations), manufacturers (robot vendors and researchers), and others (government participants).

The work has been partitioned into several working groups, each focusing on one of the requirement categories. An additional working group was formed to develop terminology that is specific to US&R robots. Each working group has a leader who is well-versed in the technical and/or applied aspects of this category. Participation in the working groups is open to non-ASTM members, although only ASTM members are able to vote on ballots.

As described above, there are myriads of situations and conditions that fall under the umbrella of USAR, and no single robot platform is suitable to address all the various conditions and performance requirements. The goal is therefore to produce scales with which to measure robot performance, not to define a "standard USAR robot." Usage guides will be produced that will describe how to best match robots to given deployment scenarios. The usage guides will provide advice on interpreting the results from the test methods and suggest desirable performance ranges to look for in a given type of application.

We now provide a brief overview of the development process for the first test method that entered the balloting process (by which it becomes a standard test method). This test method is officially titled "Standard Test Method for Determining Visual Acuity and Field of View of On-Board Video Systems for Remote Operation of Robots for Urban Search and Rescue Applications." The test method addresses multiple performance requirements within the Sensing category and others as well as the expected ranges for the three targeted robot categories (small peekbots, ground survey robots, and aerial survey robots). The Sensing requirements addressed are: Real-time remote video system (near), Real-time remote video system (far), and Video Field-of-View. Also addressed is Chassis, Adjustable Illumination. All these requirements address the expectation that the robot will be mostly or totally teleoperatively driven by an operator who has to rely almost exclusively on the onboard video cameras. Environments being searched may have some ambient lighting, or may be totally dark, in which case the robot has to provide onboard lighting that is adjustable, so as to avoid washing out near-field items. The test measures what the operator can see through the entire system; starting with the onboard cameras, transmitting wirelessly or through fiber to the operator control station, and displayed on the screen that the operator views.

Several trial test method setups were produced and presented to the responders and vendors over the course of a year. Based on feedback, adjustments were made to the methodology and the data collection form (where the results of the measurements are noted). Early on, it was decided to use visual acuity as measured for human vision. Research into the types of charts used to measure vision led to the decision to use the "tumbling E" eye chart. Typical visual items that responders may need to recognize and read in conducting their searchers were correlated with character sizes on the eye charts. Fig. 2 shows hazardous materials labels next to the eye charts for comparison as seen on the operator screen, coming through the robot camera. Since the test methods focus on the initial three types of robots, the set up for this test included near-field (expected to be the range that the peekbot will need to see) and far-field (for both the ground and aerial survey robots).

One example of a modification to the test method since its initial design was the simplification of the robot positioning. Originally, the robot was to be driven to the location from which the eye charts were read. The robot was also placed in different orientations relative to the eye charts (it was positioned on non-flat terrain). These two aspects were removed from the test because they were found to involve other non-sensing components, such as the ability to navigate the robot and aim its cameras. Evaluating these capabilities is important, but they belong in their own test methods.

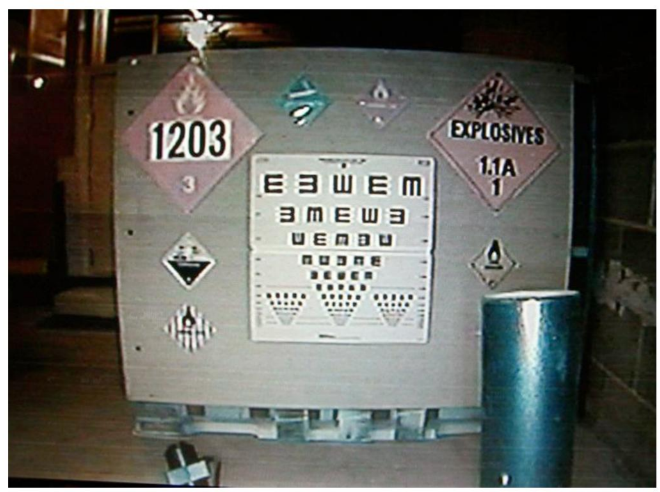

Figure 2: Visual Acuity Eye Charts and HazMat Labels Viewed Through Robot Camera at Operator Station

As a supporting effort for all the other standards working groups, a terminology group was formed, with the goal of having commonly-agreed upon definitions of terms that are applicable to USAR robotics. Their work is leveraging ongoing efforts by an ad hoc working group that is defining autonomy levels for unmanned systems and has published a set of definitions. [5, 6] An initial set of terms has been approved by the ASTM E54 committee. Additional terms are being introduced periodically.

4. RESPONSE ROBOT EXERCISES

Response robot evaluation exercises introduce emerging robotic capabilities to emergency responders while educating robot developers regarding the performance requirements necessary to be effective, along with the environmental conditions and operational constraints necessary to be useful. They also provide an opportunity to refine the emerging test methods and associated test artifacts being developed for submission as standards. Conducting these events in actual US&R training scenarios helps correlate the proposed standard test methods with envisioned deployment tasks and lays the foundation for the usage guides which will identify which robot categories appear best suited for particular response tasks.

To date, there have been three exercises held. Each one has had its own unique features and emphases. In each of these, NIST inserts simulated victims into scenarios and captures video of all missions. The simulated victims have a heat signature along with a combination of other features, such as human form (or partial form), sound (calling out, tapping), and movement.

The first one was held at the FEMA Nevada Task Force 1 Training Facility in August 2005. This exercise helped the organizers and participants understand the best way to approach the deployment of robots within training scenarios. This facility has a simulated freeway collapse prop, including trapped vehicles, that proved especially interesting for using serpentine as well as wall and ceiling climbing robots. Ground, aquatic, and amphibious robots participated.

The second exercise was held April 2006 at Disaster City®, which is the FEMA Texas Task Force 1 training facility, as well as a training resource used by national and international USAR teams. This 21 hectare (52 acre) facility features a wide array of scenarios, including multiple rubble piles, reconfigurable building collapses, a passenger and a cargo train derailments, and a pond. This site provided the opportunity to introduce aerial vehicles to the exercise, conducting initial overflights of the cargo train area. Versions of several Wave 1 test methods were piloted at Disaster City. A more detailed description of this exercise and the results is found in [7].

In August 2006, an exercise was held at the FEMA Maryland Task Force 1 training facility. This event provided an opportunity for initial integration of radiation sensors and other hazardous materials sensors within test methods and operational scenarios. Much more rigorous test method execution by all robots and extensive data capture was conducted. At the end of this exercise, several Wave 1 test methods were much better defined and ready to enter the balloting process.

Future exercises will be held on a regular basis in order to expand understanding of how robots can best fit into the responders' cache of tools and to gauge progress. New concepts for test methods (beyond the initial wave) will be piloted for review by responders and manufacturers. Eventually, it is expected that the exercises will become more focus on technology readiness level assessments, as the developers advance and mature the capabilities of robots.

5. CLASSIFICATIONS

Given the new and evolving nature of robots applied to challenging environments such as collapsed buildings, a means of organizing knowledge about the diverse population of robots and their components and capabilities may help users and developers better understand what is available, how to best apply different robot types, and where there are gaps. To this end, this program also encompasses efforts that will create classification structures for robots and their capabilities, as well as for buildings and collapse or disaster types. The robot and capabilities classification effort is defining the data structures to capture the pertinent characteristics of robots. A complementary effort will populate these structures with data about robots, such as the manufacturer's specifications of the robots as well as the results of standard performance tests. This "compendium" will be available in database format. It is envisioned that at some future time, it will be possible to map the particular mission (disaster type) to specific robot requirements. A decision support tool may be built that leverages the robot and disaster classifications.

The robot capabilities classification is being addressed through the use of ontologies. [8] In this context, an ontology can be thought of as a knowledge representation approach that represents key concepts, their properties, their relationships, and their rules and constraints. In general, ontologies make all pertinent knowledge about a domain explicit and are represented in a computer-interpretable fashion that allows software to reason over that knowledge to infer additional information.

An initial structure for the Robot Ontology has been developed. This initial structure can be broken down into the follow primary categories of knowledge:

- Structural Characteristics – describes the physical and structural aspects of a robot. Examples are size, weight, types of sensors, locomotion mechanism, and power source.
- Functional Capabilities – describes the behavioral features of the robot. Examples are locomotion capabilities (e.g. maximum speed, maximum slope, maximum stair climbing angle/speed) and sensory capabilities (e.g., minimum resolution, map building capability, self-localization).
- Operational Considerations – describes the interactions of the robot with the human and the interoperability with other robots. Examples of this are human-system interaction such as operator ratio, initial training required to gain proficiency and interactions with other robots.

To create a disaster taxonomy, researchers are developing a framework for integrating building classification, disaster type, and collapse type to provide general descriptions of probable operating environments with which the USAR robots will be confronted. Existing sources of building and collapse classifications are being studied. A major part of this effort involves capture of high-resolution three-dimensional data of environments through which the robots should have to traverse. Exteriors and interiors of rubble piles, for instance, are being digitally scanned using lasers. [9] Fig. 3 shows images of a rubble pile along with the three dimensional point cloud collected. Numerical analysis of the point data is being researched, so as to arrive at traversability classifications. Early results that point to possible approaches are described in [10]. This work will be extended to attempt matching robot dimensions and locomotion characteristics to the terrain types.

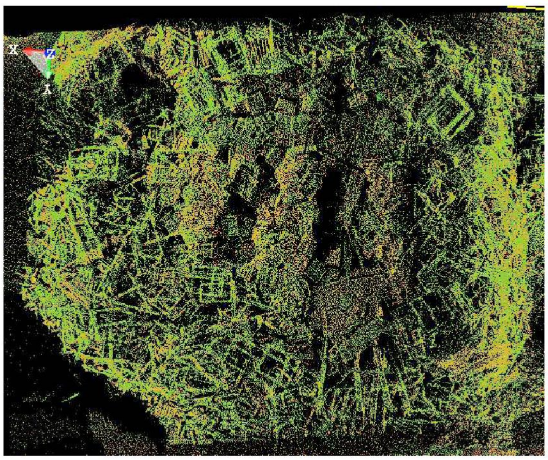

Figure 3: High Resolution Data Capture of a Wood Rubble Pile at Disaster City.

6. Conclusions and Future Directions

A solid foundation has been created for both evaluating and advancing the capabilities of robots applied to urban search and rescue. Working together, the stakeholders are formally defining the performance needs of robots for use in USAR missions and developing means of measuring robot capabilities in individual categories. Opportunities for robot technology improvements are being uncovered during response exercises [7] [11] and in experimental fieldings [12]. Researchers, developers, and manufacturers are rising to the challenge and starting to address the identified shortcomings. The community is also working together to develop concepts of operation in the deployment of robots during search and rescue. [13]

Test methods addressing all the individual requirements will continue to be developed and submitted into the standards process. However, the requirements also continue to evolve. New items, such as the need for robots to climb stairs, are being added. As the robot capabilities evolve and mature, we expect to see response organizations begin to purchase robots and start experimenting with fielding them. This will undoubtedly lead to another round of requirements definitions and refinements.

Acknowledgments

This work is supported by the Department of Homeland Security Science and Technology Directorate Standards Office. The authors would like to thank all the participants who are advancing this project and are too numerous to mention individually: the FEMA US&R Task Force members, the robot manufacturers and researchers, the ASTM Working Group Chairs and members, the technical experts who contribute to the science and engineering of measurement, and the rest of the NIST team.

References

1. "Urban Search and Rescue Response System In Federal Disaster Operations: Operations Manual." vol. 9356.1-PR, F. E. M. Agency, Ed., 2000.
2. J. Blitch, "Artificial Intelligence Technologies for robot assisted urban search and rescue," *Expert Systems with applications,* vol. 11, pp. 109-24, 1996.
3. R. Murphy, "Rescue robotics for homeland security," *Commun. ACM,* vol. 47, pp. 66-68, 2004.
4. E. Messina, A. Jacoff, J. Scholtz, C. Schlenoff, H. Hui-Min, A. Lytle, and J. Blitch, "Statement Of Requirements For Urban Search And Rescue Robot Performance Standards," http://www.isd.mel.nist.gov/US&R_Robot_Standards/Requirements%20Report%20(prelim).pdf, 2005.
5. H.-M. Huang, "Autonomy Levels for Unmanned Systems (ALFUS) Framework Volume 1: Terminology Version 1.0," National Institute of Standards and Technology Special Publication 1011, 2004.
6. H.-M. Huang, "Autonomy Levels for Unmanned Systems (ALFUS) Framework: Interim Results " in *Performance Metrics for Intelligent Systems*, NIST Special Publication 1062, Gaithersburg, MD, 2006.
7. A. Jacoff and E. Messina, "DHS/NIST Response Robot Evaluation Exercises," in *IEEE Safety, Security, and Rescue Robotics* Gaithersburg, MD, 2006.
8. C. Schlenoff and E. Messina, "A Robot Ontology for Urban Search and Rescue," in *Workshop on Research in Knowledge Representation for Autonomous Systems, part of the Association for Association for Computing Machinery Conference on Information and Knowledge Management*, Bremen, Germany, 2005.
9. A. Lytle, G. Cheok, K. Saidi, A. Jacoff, and E. Messina, "US&R Data Characterization Sets," in *2006 IEEE Safety, Security, and Rescue Robots* Gaithersburg, MD, 2006.
10. V. Molino, R. Madhavan, E. Messina, A. Downs, A. Jacoff, and S. Balakirsky, "Traversability Metrics for Urban Search and Rescue Robots on Rough Terrain," in *Performance Metrics for Intelligent Systems*, NIST Special Publication 1062, Gaithersburg, MD, 2006.
11. K. Remley, G. Koepke, E. Messina, A. Jacoff, and G. Hough, "Standards development for wireless communications for urban search and rescue robots," in *9th Annual International Symposium on Advanced Radio Technologies,* Boulder, CO, 2007.
12. R. Murphy and S. Stover, "Gaps Analysis for Rescue Robots," in *American Nuclear Society Joint Topical: Sharing Solutions for Emergencies and Hazardous Environments."* Salt Lake City, UT, 2006.
13. A. Ferworn, G. Hough, and R. Manca, "Expedients for Marsupial Operations of USAR Robots," in *IEEE Safety, Security, and Rescue Robotics,* Gaithersburg, MD, 2006.