

# Comprehensive Standard Test Suites for the Performance Evaluation of Mobile Robots

Adam Jacoff, Hui-Min Huang, Elena Messina, Ann Virts, Anthony Downs  
National Institute of Standards & Technology  
100 Bureau Drive MS 8230  
Gaithersburg, MD 20899  
+1.301.975. 3427  
hui-min.huang@nist.gov

## ABSTRACT

Robots must possess certain sets of capabilities to suit critical operations such as emergency responses. In the mobility function, ground robots must be able to handle many types of obstacles and terrain complexities, including traversing and negotiating positive and negative obstacles, various types of floor surfaces or terrains, and confined passageways. Additional mobility requirements include the ability to sustain specified speeds and to tow payloads with different weights. Standard test methods are required to evaluate how well candidate robots meet these requirements. A set of test methods focused on evaluating the mobility function has been collected into a test suite. Likewise, in other functions such as sensing, communication, manipulation, energy/power, Human-System Interaction (HSI), logistics, and safety, corresponding test suites are required. Also needed are test suites for aerial and aquatic robots. Under the sponsorship of DHS, NIST researchers are collaborating with others to establish such a collection of test suites under the standards development organization ASTM International. Apparatuses must be set up to challenge specific robot capabilities in repeatable ways to facilitate direct comparison of different robot models as well as particular configurations of similar robot models.

## Categories and Subject Descriptors

J.2 [physical sciences and engineering] unmanned systems performance

## General Terms

Measurement, Performance, Design, Human Factors, Standardization, Verification

## Keywords

energy, environment, goal, human-system interaction, HSI, measure, metrics, mobility, power, radio communications, robot, performance, sensor, standard, task, terminology, test, test method, test suite

## 1. INTRODUCTION

U.S. emergency responders face extremely dangerous or hazardous environments when responding to natural or man-made disasters. Urban search and rescue (US&R), bomb disposal, and law enforcement are a few of these critical operational areas. Major efforts have been underway to improve the effectiveness of the emergency responses. The Department of Homeland Security (DHS) Federal Emergency Management Agency (FEMA) and the National Institute of Justice (NIJ) co-sponsored an effort to identify and define functional requirements for new and/or improved technologies that meet the needs of both urban search and rescue teams as well as law enforcement agencies. The report [8] stated needs included “Reliable non-human, non-canine search and rescue systems - robust systems that combine enhanced canine/human search and rescue capabilities without existing weaknesses (i.e., robots).”

The National Response Framework<sup>1</sup> states that “Governments at all levels have a responsibility to develop detailed, robust, all-hazards response plans.”

It would be extremely helpful for the successfully tested robots to be made available to the emergency responders.

Under the sponsorship of the DHS, NIST has embarked on an effort for the research and development of the performance evaluation methodology of the response robots since 2005. Earlier papers [5,6] described some initial results. This paper provides an update as well as an overall structure of the standard test methods.

## 2. APPROACH

DHS and NIST adopted an iterative, user-oriented approach for developing the robotic performance evaluation standards. See Figure 1. The process starts with collecting operational requirements, which must be provided by the emergency responders. The project objective is, then, to be able to employ robots to accomplish the required tasks.

For example, one requirement is for a robot to sustain its speed when navigating in an obstacle-rich environment. A

This paper is authored by employees of the United States Government and is in the public domain. PerMIS'10, September 28-30, 2010, Baltimore, MD, USA. ACM 978-1-4503-0290-6-9/28/10.

<sup>1</sup> National Response Framework, U.S. Department of Homeland Security, January 2008, [www.fema.gov/NRF](http://www.fema.gov/NRF).

test method to characterize how well a candidate is able to do so is developed. This test method would entail sets of apparatus, metrics, and procedure, as indicated in Figure 1.

The draft test methods evolve through several validation steps, where they are applied to candidate robots to see how effective they are. The validation can be conducted through organized robotic exercises or competitions. The matured test methods are submitted for standardization. They are also proliferated for wide application, for the purpose of verifying that the subject testing is reproducible, and for responder proficiency training of the robotic tools. Beyond validating test methods, robot exercises further serve to educate developers about the domain requirements, as well as to allow responders to experiment with deploying robots in realistic scenarios.

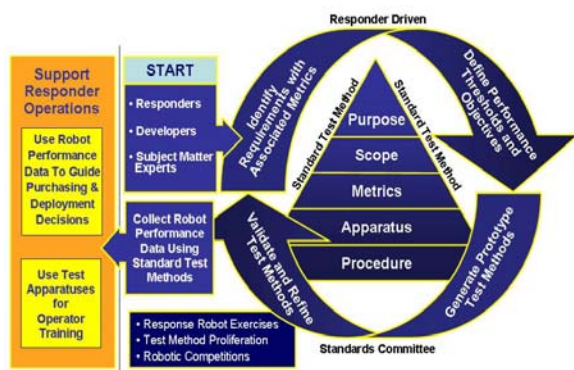


Figure 1: Test Method Development Cycle<sup>2</sup>

## 2.1 REQUIREMENTS CAPTURE

The original intent of this standardization effort was to enable the Department of Homeland Security to make well-informed decisions about the application of robots to urban search and rescue missions. A major requirement capture process was conducted in 2005 through a series of workshops attended by FEMA US&R task force members [7]. These requirements have been the foundation for the standards development efforts. Over the years, the focus of the work has expanded to include other civilian response applications, such as bomb disposal. The requirements definition process for the latter has been less formal and has also benefitted from prior studies, in particular funded by the National Institute of Justice. [8]

Over the years, the requirements have been refined and updated as the responders have become increasingly familiar with robotic capabilities. They gain familiarity through the continued participation in the standards development and evaluation process, especially during test exercises. Responders evaluate the test methods and provide constructive feedback throughout the development

process. Their input serves to ensure that the test methods measure real world functionality in meaningful ways. The objective is for the robots to meet the requirements before they can be adopted by and integrated into the agencies' operating processes.

## 2.2 Test method standardization

The NIST team has joined the ASTM International for developing the standard procedures, test methods, and metrics to fully to address the requirements. This effort is under Committee on Homeland Security (E54), Subcommittee on Operational Equipment (.08), Robotics Task Group (.01). In other words, the designation for this response robot performance evaluation effort is (ASTM E54.08.01).

According to the ASTM classification, the following are the relevant types of standards:

- terminology
- practice: a definitive set of instructions for performing one or more specific operations that does not produce a test result
- test method—a definitive procedure that produces a test result

From these, we establish the following major milestones for the standardization effort. The various test methods evolve at different paces. Currently, three standards have been approved by ASTM [1, 2, 3]. About 20 test methods are in the balloting, validating or prototype status.

1. Prototype: when a test procedure and apparatus is conceived, built, and under evolution.
2. Validating: when a prototype has progressed enough after going through several sets of tests by robots. A Work Item description might have been submitted to ASTM to indicate that the standardization balloting process may be ready in months.
3. Standards:
  - a. Standard Practice: the balloted and approved test procedure; meanwhile, the team continues collecting test data to fine tune the metrics, the evaluation form, and the accompanying performance repeatability issue.
  - b. Standard Test Method: the goal of the prototype; once the team has gathered sufficient data to prove the performance repeatability, the metrics/evaluation form will be finalized and added into the previously approved standard practice to be re-balloted as a standard test method.

## 3. TEST METHOD

Each of the test methods corresponds to the requirements as specified by U.S. emergency responders and additional constituents. A robot's performance in this test is indicative of its capabilities needed in such operations as emergency responses. ASTM has a standard style guide for

<sup>2</sup> Figure based on original by Dr. Bert Coursey, Department of Homeland Security Standards Executive.

the test methods. Parts of the essential information generated by the test method development team are:

- **Metrics:** We identify the characteristics for measuring the corresponding aspect of the robotic performance that addresses a particular requirement or subset of requirements. For example, the emergency responders may require a robot to be able to drive around, on, or through particular obstacles or challenging terrains. We must identify the characteristics (sizes, severity) of the obstacles to be measured.

Associated with the metrics is the issue of performance requirements, in other words, what measured values are acceptable. These values can be derived from the requirements.

- **Apparatus:** To measure the performance, we must design and develop the testing setup such that the metrics can be applied for the performance evaluation. The apparatuses associated with the test methods challenge specific robot capabilities in repeatable ways to facilitate direct comparison of different robot models as well as particular configurations of similar robot models.

The apparatus can contain either notional or operational objects or setups. With notional objects or setups, we can easily standardize the design characteristics, such as size, weight, surface type, color, etc., of the apparatus. This would facilitate reproducing the apparatus. The opposite approach, using operational objects or objects with operational flavors, might have the benefit of close to reality but might suffer the drawback of difficult to standardize.

- **Procedure:** A procedure is generated based on the apparatus and how we want to exercise the metrics to measure the performance.
- **Test Form:** Corresponding to a test method and contains fields for recording the testing results and the associated information, including:
  - Metrics and corresponding measurement scales and ranges;
  - Any additional testing features such as those reflecting performance proficiency;
  - Important notes to be recorded during the test, including particular fault conditions that occurred, the reason the robot developer abstained from participating in this test (if this was the case), any observations by the test administrator that could augment the recorded results in either positive or negative ways, or any comments that the robot's operator requests to be put on the form;
  - Testing administrative information; including: names for the involved personnel,

organizations, and robot; the testing date(s) and time; version number of the form; and the testing conditions on the environment, apparatus, and robotic configuration (tether versus radio communication, for example). If audio/video recording is done during the testing, the file names should be recorded on the form.

- **Repeatability analysis:** Tests must be conducted with a statistically significant number of repetitions to establish the reliability of the testing method and the associated confidence levels.
- **Reproducibility analysis:** A test method must be reproduced at multiple locations to verify that similar levels of reliability and confidence can be obtained.

#### 4. TEST METHOD ORGANIZATION

The entire set of requirements and the corresponding testing standards are organized into the following categories:

- Terminology
- Robotic Subsystems
  - Mobility
    - Ground Locomotion
    - Aerial Maneuvering
    - Aquatic Maneuvering
  - Energy and Power
  - Sensing
  - Communications
  - Manipulation and Other Payloads
  - Chassis
- Human-System Interaction (HSI)
- Logistics
- Safety/Operating Environment

A collection of test methods is to be developed for each of the subsystems. These test methods are called a test suite for the subsystem. The test methods in a test suite are intended to collectively characterize a robot's performance in a particular functional area.

For example, a set of test methods has been developed for the Mobility subsystem. A robot with a larger size may be better suited for the gap crossing test but may be more constrained in the confined space traversal test. Therefore, the collective test evaluation, as opposed to individual test methods, should give a comprehensive perspective for the robot's mobility capability in the response environments.

Note that a lot of the requirements may involve multiple robotic subsystems. Particularly, the Mobility subsystem is involved in most of the other subsystem requirements. For example, a radio communication requirement is irrelevant unless the robot can navigate to a location of concern. The endurance requirement for the battery is relies on the robot's mobility capability, since the test method entails

having the robot drive a prescribed pattern over a designated apparatus terrain repeatedly. The Mobility subsystem is therefore considered the enabling subsystem and a heavy focus has been placed upfront on test methods for this subsystem.

## 5. TEST SUITES AND TERMINOLOGY

### 5.1 Terminology

Terms must be formally defined to facilitate proper communication among all the test method development efforts. Given that this effort covers multiple communities, terms might be used with different meanings. Consistent terminology is crucial.

The applicability of the terms varies, ranging from: test method specific, common to a test suite, and common to all response robots standard test methods. As such, the terms are identified and defined correspondingly throughout the whole test method organization.

### 5.2 Cache Packaging Test Suite

The following standards are being developed to evaluate the cache packaging performance:

Standard Practice for Establishing the Test Configuration and Associated Cache Packaged Weight and Volume of Emergency Response Robots for

- FEMA Urban Search and Rescue Teams (ASTM E2592-07)
- Federal/State/Local Bomb Squads (P)<sup>3</sup>

### 5.3 Mobility Test Suite

Suitable ground robots must be able to handle many types of obstacles and terrains. Standard test methods are required to evaluate whether candidate robots meet these requirements. The following test methods are being prototyped and validated:

Standard Test Method for Evaluating the Mobility Capabilities of Emergency Response Robots Using

- Terrains: Flat/Paved Surfaces (V)<sup>4</sup>
- Terrains: Continuous Pitch/Roll Ramps (V)
- Terrains: Crossing Pitch/Roll Ramps (V)
- Terrains: Symmetric Stepfields (V)
- Terrains: Sand (P)
- Terrains: Gravel (P)
- Terrains: Mud (P)
- Obstacles: Inclined Planes (V)
- Obstacles: Gap Crossings: Static, Horizontal, Parallel (V)
- Obstacles: Gap Crossings: Dynamic, Horizontal, Parallel (P)
- Obstacles: Pipe Steps (V)

<sup>3</sup> P - Indicates the development status as being Prototyped

<sup>4</sup> V - Indicates the development status as having completed the Prototyping stage and being Validated.

- Obstacles: Stair/Landings (V)
- Towing Tasks: Grasped Sleds (V)

See Figure 2, Figure 3, Figure 4 for the apparatuses of three of the mobility test methods.



**Figure 2: Paved Surface Terrain-for Sustained Speed and Towing Test Methods**



**Figure 3: Crossing Ramps Terrain Test Method**



**Figure 4: Stairs Obstacle Test Method**

### 5.4 Energy/Power Test Suite

The following test methods are being developed to evaluate the energy/power subsystem performance:

Standard Test Method for Evaluating the Energy/Power Capabilities of Emergency Response Robots Using

- Endurance Tasks: Terrains: Continuous Pitch/Roll Ramps (V)
- Peak Power Tasks: Obstacles (P)

See Figure 5 for the apparatus of the endurance test method.



Figure 5: Endurance Test Method

## 5.5 Radio Communications Test Suite

The following test methods are being developed to evaluate the radio communications subsystem performance:

Standard Test Method for Evaluating the Radio Communication Capabilities of Emergency Response Robots Using

- Control and Perception Tasks: Line-of-Sight Environment (V). See Figure 6.
- Control and Perception Tasks: Non-Line-of-Sight Environment (V)
- Control and Perception Tasks: Structure Penetration Environment (P)
- Control and Perception Tasks: Interference Signal Environment (P)
- Control and Perception Tasks: Urban Canyon Environments (P)

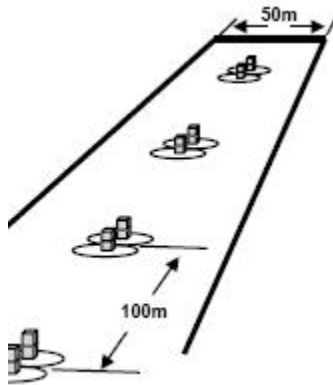


Figure 6: Light-of-Sight Radio Comms Test Method

## 5.6 Sensor Test Suite

The following test methods are being developed to evaluate the sensor subsystem performance:

Standard Test Method for Evaluating the Sensor Capabilities of Emergency Response Robots Using

- Video Acuity Charts and Field of View Measures (ASTM E2566-08) (See Figure 7.)
- Video Directed Search Tasks: Complete (V) (See Figure 8.)
- Video Directed Search Tasks: Rapid (V)
- Audio Rhyming Words and Loudness Measures (at the Operator and Robot) (V)
- Audio Spectrum Tones (at the Operator and Robot) (P)
- Laser Ranging Targets and Spatial Resolution Measures (P)
- Localization and Mapping Tasks: Hallway Labyrinths with Complex Terrain (P)
- Localization and Mapping Tasks: Wall Mazes with Complex Terrain (P)
- Localization and Mapping Tasks: Sparse Feature Environments (P)
- Localization and Mapping Tasks: Tunnel Mazes (P)



Figure 7: Video Acuity Charts and Field of View Measures Test Method



Figure 8: Video Directed Search Tasks: Complete Test Method

## 5.7 Manipulation Test Suite

The following test methods are being developed to evaluate the manipulator subsystem performance:

Standard Test Method for Evaluating the Manipulation Capabilities of Emergency Response Robots Using

- Directed Perception Tasks in Elevated Shelves: Open Access (V)
- Directed Perception Tasks in Elevated Shelves: Reach-Over Access (P)
- Directed Perception Tasks in Elevated Shelves: Reach-Under Access (P)
- Gasping Dexterity Tasks in Elevated Shelves: Open Access (V) (See Figure 9.)
- Gasping Dexterity Tasks in Elevated Shelves: Reach-Over Access (P)
- Gasping Dexterity Tasks in Elevated Shelves: Reach-Under Access (P)
- Door Opening and Traversal Tasks (V)



**Figure 9: Gasping Dexterity Tasks in Elevated Shelves: Open Access Apparatus**

## 5.8 Human-System Interaction (HSI) Test Suite

The following test methods are being developed to evaluate the HSI subsystem performance:

Standard Test Method for Evaluating the Capabilities of Emergency Response Robots Using

- Navigation Tasks: In Unknown Environments with Complex Terrain (V)
- Search Tasks: In Unknown Environments with Complex Terrain (V) (See Figure 10.)
- Search Tasks: Under-Body Voids with Complex Terrain (V)



**Figure 10: HSI Search Test Method Using a Random Maze Apparatus**

## 6. TESTING POLICY

A testing policy has evolved to ensure consistent testing efforts. The main points include:

- All tests are conducted with the robot operator stationed remotely from the robot. The robot must be out of sight of the operator and ideally out of sound of testing apparatus.
- The operator can choose to abstain (withdraw from) the test, which causes the result to be not reported. By doing so, the robot developer acknowledges the omission of the performance data while the test method was available at the test time. The operator typically abstains when the robot configuration is not designed nor equipped to perform the tasks as specified in the test method. The abstention should be granted only before the test. The testing authority should make a consistent policy about the timing.
- Testing is conducted by a test administrator. She/he is to ensure the readiness of the apparatus, the test form, and any required measuring devices.
- The test administrator ensures that the specified or required environmental conditions are met.
- She/he will also explain the test to the robot operator. This includes fault conditions before the test starts.
- She/he will inform the operator when the safety belay (if needed to protect the robot from damage) is available and ensure that the operator has either decided not to use it or assigned a person to handle it properly.
- The administrator is to call the operator to start and end the test and record the performance data and any notable observations during the test.
- Verbal communication between the operator and the administrator regarding the performance of a test repetition is not allowed other than instructions on when to start and notification of faults and any safety related conditions. The operator has the full responsibility to determine whether the robot has reached a test goal.
- Operator is allowed to have as much practice time as practical before entering a test.

## 7. TESTING

### 7.1 Testing Results

Extensive testing has been conducted for the purposes of prototyping and validating the test methods as well as supporting the project sponsor's objectives. Some representative testing results are shown in Figure 11, which shows that two of the nine testing robots successfully traversed 45° stairs and Figure 12 shows the endurance test distances of the participating robots. The endurance traversing distances for the nine participating robots range from 6915 m to 345 m.

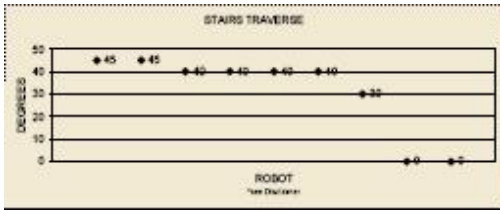


Figure 11: Stair Traversing Tests for Nine Robots

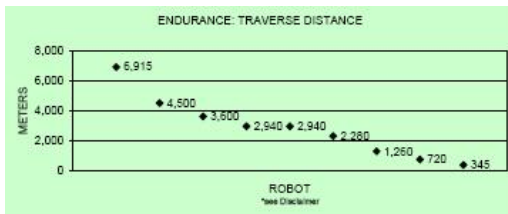


Figure 12: Endurance Test Results for Nine Robots

## 7.2 Operational Testing

The NIST team also developed a collection of test operational scenarios that require combinations of the aforementioned test suite capabilities. The scenarios include suspected package on a bus (for bomb squads), which requires the mobility, manipulation, and sensing capabilities. Other scenarios include aerial post-disaster assessment and search in a rubble pile (Figure 14), which are conducted in a Texas Engineering Extension Service (TEEX) training facility called Disaster City [9].



Figure 13: Test Scenario of Suspected Packages on Bus



Figure 14: Rubble Pile Scenario

### 7.1 Test Site Proliferation

The project aims at utilizing or implementing various testing resources, not only in the U.S. but also in the world.

Testing has been extensively conducted at the NIST site, the TEEX Disaster City site, Montgomery County training facility in Maryland, the Southwest Research Institute (SwRI) in Texas [10], etc. Plans are being made to implement the testing apparatuses in Asia and Europe. Ultimately, selected test sites will be certified within the United States and internationally and will be responsible for conducting robot testing.

## 8. SUMMARY

The key principles of the program are:

- user-focused requirements capturing process
- easily reproducible standards evaluation processes
- individually developed but collectively presented standard performance evaluation test suites
- teleoperation based testing [4]
- human level sizes and weights for robots and objects

As of September 2010, 3 standards have been approved and published, 4 are being balloted within ASTM, and over a dozen test methods have been validated and are in the process of final review and formatting so that they can be submitted for balloting. About 10 additional test methods are being prototyped as the next wave of the validation. Once approved, the resulting test suites will form a foundation for characterizing the performance of robots functionality. This will facilitate informed purchase decisions for response organizations and foster growth in the market and further innovations in the robots' capabilities.

## 9. ACKNOWLEDGEMENTS

This work was supported by the Department of Homeland Security, Science and Technology Directorate, Office of Standards, the National Institute of Justice, and the NIST Office of Law Enforcement Standards.

## 10. REFERENCES

- [1] ASTM Standard E 2521 – 07a, *Standard Terminology for Urban Search and Rescue Robotic Operations*, ASTM International, West Conshohocken, PA.
- [2] ASTM Standard E 2592 – 07, *Standard Practice for Evaluating Cache Packaged Weight and Volume of Robots for Urban Search and Rescue*, ASTM International, West Conshohocken, PA.
- [3] ASTM Standard E2566, *Standard Test Method for Determining Visual Acuity and Field of View of On-Board Video Systems for Teleoperation of Robots for Urban Search and Rescue Applications*, ASTM International, West Conshohocken, PA.
- [4] *Autonomy Levels for Unmanned Systems (ALFUS) Framework, Volume II: Framework Models Version 1.0*, NIST Special Publication 1011-II-1.0, Huang, H. et al., Ed., National Institute of Standards and Technology, Gaithersburg, MD, December 2007.

- [5] Jacoff, A., and Messina, E, "Urban Search and Rescue Robot Performance Standards: Progress Update," SPIE Defense and Security Conference 2007.
- [6] Messina, E. and Jacoff, A. S. "Measuring the Performance of Urban Search and Rescue Robots," IEEE Conference on Homeland Security Technologies, 2007.
- [7] Messina, E. R., et al., Statement of Requirements for Urban Search and Rescue Robot Performance Standards, NIST Draft Report, May 2005
- [8] Urban Search and Rescue Technology Needs Identification of Needs, DHS/FEMA and National Institute of Justice Report, June 2004.
- [9] <http://www.teex.com/index.cfm>
- [10] <http://www.swri.org/>