Societal Acceptance of Intelligent Robotic Systems: The Role of Performance Evaluation, Benchmarking, and Standardization

Raj Madhavan, Ph.D.

Institute for Systems Research
Maryland Robotics Center
University of Maryland, College Park
&
Intelligent Systems Division
National Institute of Standards and Technology

madhavan@umd.edu/madhavan@nist.gov

ICAR'11 WS: ***Urban Service Robotics: Challenges and Opportunities***

Commercial equipment and materials are identified in this presentation in order to adequately specify certain procedures. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose. The views and opinions expressed are those of the presenter and does not necessarily reflect those of the organizations he is affiliated with.

MARYLAND ROBOTICS CENTER
THE INSTITUTE FOR SYSTEMS RESEARCH

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Outline

- Three Illustrative Examples
- Societal Acceptance and Role of Performance Evaluation and Standards
- ARRA-MSE Project
  - *World Modeling for Autonomous Navigation In Unstructured and Dynamic Environments: Performance Evaluation and Benchmarking*
  - Working with Industry, Developers, and End-users
- Quantitative Evaluation of the Quality of Robot-generated Maps
- Concluding Thoughts

Some Illustrative Examples

- **IEEE 802.11**
  - Wireless Local Area Network computer communication implemented by the IEEE LAN/MAN Standards Committee
- **HD DVD Vs Blu-ray Disc**
  - High definition optical disc format war
  - HD DVD (intially backed by Toshiba)
  - Adoption by Sony-PS3
- **Biometric Security**
  - RFID and Biometric Security methods (Chip implants)
  - Enablers: Convenience & Security
  - Inhibitors: Privacy, Personal Rights Protection, & Data Security
- **Societal Acceptance will heavily depend on:**
  - Consumer awareness & appreciation
  - Need (utility) & Cost (value-added)
  - Perception (in many cases IS Reality!)

A word about Standards and Standardization

- **New (Emerging) Vs Established (Mature) Fields**
  - Is it harder to generate standards in mature areas?
  - Corollarily, is it easier in emerging areas?
  - *De facto* standards (contrast with mandatory or *de jur*e standards)
- Do standards impede progress?!!
- In terms of societal acceptance, what are the implications?

- ***Performance Evaluation & Benchmarking → Innovation***

Measuring Performance of Intelligent Systems

- Performance Evaluation, Benchmarking, and Standardization are critical enablers for wider acceptance and proliferation of existing and emerging technologies
- Crucial for fostering technology transfer and driving industry innovation
- Currently, no consensus nor standards exist on
  - key metrics for determining the performance of a system
  - objective evaluation procedures to quantitatively deduce/measure the performance of robotic systems against user-defined requirements
- The lack of ways to quantify and characterize performance of technologies and systems has precluded researchers working towards a common goal from
  - exchanging and communicating results,
  - inter–comparing robot performance, and
  - leveraging previous work that could otherwise avoid duplication and expedite technology transfer.

Measuring Performance of Intelligent Systems

- The lack of ways to quantify and characterize technologies and systems also hinders adoption of new systems
  - Users don't trust claims by developers
  - There is lack of knowledge about how to match a solution with a problem

- Users may be reluctant to try a new technology for fear of expensive failure:
  - Think of the "graveyards" of unused equipment in some places

- …

Performance Evaluation of Intelligent Systems

Evaluation Philosophy

- To design and develop capable, dependable, and affordable robotic systems, their *performance* must be *measurable (quantitative)*
- Repeatable and reproducible test artifacts and measurement methodologies to capture performance data → focus research efforts, provide direction, and accelerate the advancement of mobile robot capabilities *(objective)*
- Only by involving users, developers and integrators in a coupled fashion, can meaningful solutions be produced that can stand the ever-varying requirements imposed by:
  - tasks that are either application or environment dependent,
  - hardware and software advancements/restrictions that affect the development cycle, and
  - budgetary constraints that interrupt and hamper sustained progress

ARRA-MSE Project Overview

- Commerce Department's American Recovery and Reinvestment Act (ARRA) Measurement Science and Engineering Research Grants
  - ***World Modeling for Autonomous Navigation In Unstructured and Dynamic Environments: Performance Evaluation and Benchmarking***
    - One of 27 projects funded (out of 1300 proposals) at higher-education, commercial, and nonprofit organizations in 18 states
    - Intended to "bolster U.S. scientific and technological infrastructure, increasing our nation's ability to innovate, compete, and solve scientific and technological problems"
    - Temple University and University of Maryland, College Park

Press Release

AGVs are an integral component of today's manufacturing processes; they are widely used on factory floors for intra-factory transport of goods. However, they require highly structured environments and reference markers installed throughout plants, which can carry prohibitively high maintenance and installation costs. AGVs could be much more widely used in manufacturing if they could cope with unstructured, dynamic environments and adapt to human-centered collaboration while still keeping humans out of harm's way. Having robots sense unstructured environments and automatically generate a sufficiently accurate world model is still an unsolved problem, and the solution requires a framework for generating accurate representations of the operational domain. This, in turn, requires scientifically sound and statistically significant metrics, measurement, and evaluation methodologies for quantifying intelligent systems' performance. To address these challenges, the researchers will create and experimentally validate a world modeling framework for unstructured manufacturing environments containing dynamic objects. The researchers also plan on creating reference data sets for end users, developers, and vendors, and to actively participate in standards efforts in this field.

Automated Guided Vehicles (AGVs)

- AGVs are workhorses of the manufacturing, warehousing, distribution, and other industries
- Widely used on factory floors for intra-factory transport of goods between conveyors/assembly sections, parts/frame movements, and truck-trailer loading/unloading
- *State-of-the-art*: Heavy dependency on Infrastructure
  - Laser (reflectors), Inertial/Grid (magnets), Buried wire guidance
- *Critical Enablers*:
  - Ability to cope with unstructured, dynamic environments (*smart*)
  - Keeping humans out of harm's way (*safe*)
  - Adaptability to human-centered collaboration (*flexible*)
- *And it is 2011!*

AGV-Clip

Goals & Anticipated Outcomes (1)

R & D

- ***Development of a world modeling framework for unstructured manufacturing environments with ability to cope with dynamic objects (moving vehicles and humans)***

Strengths of the proposed framework lie in the fact that

- – it can provide continual and simultaneous estimates of mobile robot (AGV) positions and features in the operating environment,
- – it is sensor-agnostic and can work with multiple sensor modalities compensating for individual deficiencies of single sensors, and
- – it is sufficiently generic that it can be extended for world modeling in other domains

Validation & Verification

- ***Experimental verification and validation of proposed framework in relevant environments*** contributing to the science of performance evaluation and benchmarking, via characterization of components at the system and sub-system levels of autonomous navigation of mobile robots in manufacturing domains

- ***Development of repeatable and reproducible reference test methods and measurement methodologies*** to design scientific experiments that will provide statistically significant results through field exercises

Goals & Anticipated Outcomes (2)

Standardization

- ***Bringing together end-users, developers, and vendors*** to gather requirements, discuss and devise an action plan to overcome existing barriers by providing reference data sets to work collaboratively in developing shared-solutions across different application areas

- Channeling efforts from lessons learnt towards ***standards-defining activities leading to the establishment of de facto standards***

Dissemination

- ***Scholarly dissemination of results*** by organization of workshops, publications in conferences, archival journals, trade magazines, newsletters, and

- ***Raising awareness*** among and encourage participation of the general public and students via competitions, seminars, and a web-portal.

What we are not doing …

- We are not promising an end-end mapping/navigation solution
- We are not competing but rather interested in collaborating
- Focus is on benefiting end-users with input from researchers/developers & industry but not on product development
- We will not undertake actual testing/evaluation but rather facilitate it (development of performance metrics/test methods)

Societal Acceptance of Intelligent Robots: Hurdles and Challenges

- Interesting discrepancy between academia and industry

  Scientists claim to have 'robust and fast' solutions even for seemingly more challenging tasks in robotics
  Vs
  Performance is assessed by academia only (peer evaluation of journal papers with stronger emphasis on theory)

  World Modeling in (static) indoor environments is often seen as a solved problem by academia

  Vs

  Industry strength algorithms are seldom implemented

- Do we have a
  - a research problem? (Current algorithms lack in their core abilities)
  - an engineering problem? (Current implementations lack robustness)
  - an acceptance problem? (Algorithms are ready but why change current practices?)

  - ALL OF THE ABOVE?

Quantitative Evaluation of the Quality of Robot-generated Maps

Mobile Robot Mapping 101

- Robot mapping is the process of creating an internal representation of the robot’s physical environment
- Critical component in (autonomous) mobile robot navigation (path-planning, self-localization)
- Typically, the following cycle is employed:
  - Observe (measurements from on-board sensors)
  - Predict (robot position, correlate observations)
  - Update (robot position, map estimates)
- A myriad of schemes in open literature with varying levels of success
  - Metric (e.g. occupancy grids), Topological (e.g. behavior-based) representations
  - 2D, 2.5D, 3D representations
  - Sensors (vision, rangefinders)
  - Probabilistic (Bayesian, SLAM), heuristics-based, perceptual (cognitive), …

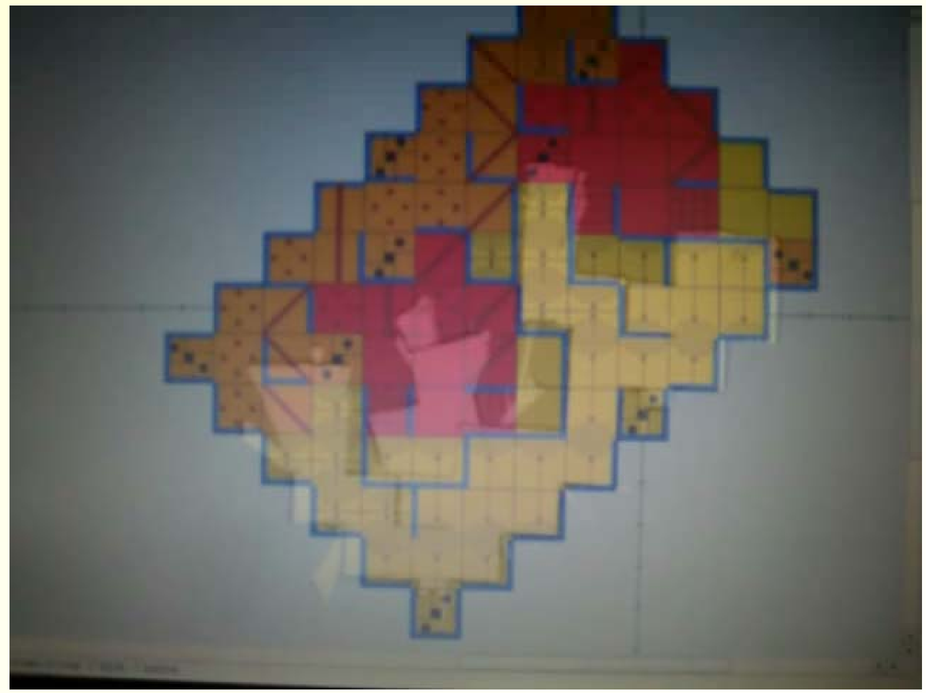
Motivation & Background

- ***(Quantitative) Map Quality*** is a performance measure of how well a robot or team of robots can explore, understand and interpret the operational domain; subsequently, indicative of the *utility* of the robot-generated map

State-of-the-art

- Qualitative comparison of resulting maps is used to assess performance, e.g. visual inspection
- Common practice in the literature to compare newly developed mapping algorithms with former methods by presenting images of generated maps
  - suboptimal, particularly when applied to large–scale maps
  - clearly not a good choice of evaluation
  - hard to inter-compare results
- Prevalent problem spanning multiple domains: rescue, manufacturing, military, service robotics, …

Case in Point: Two Examples



Maps produced by various teams at the RoboCupRescue Virtual League Competition (geotiff format)

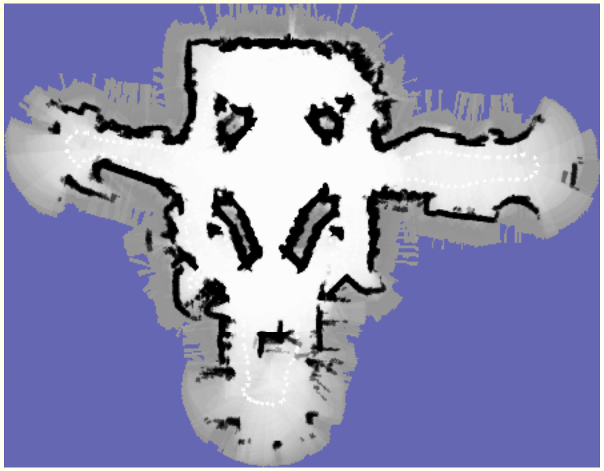
Generated map comparison with ground truth map (RoboCupRescue Physical League Competition)

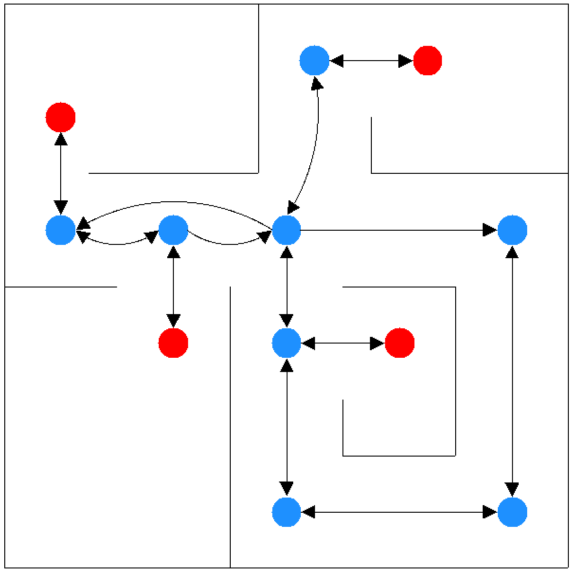
MARYLAND ROBOTICS CENTER
THE INSTITUTE FOR SYSTEMS RESEARCH

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Quantifying Robotic Mapping

- Two classes:
  - *Metric (Topographic) Maps:* Represent an environment in terms of geometric relations between objects in the environment and a reference frame
  - *Topological Maps:* Represent an environment as a *graph where* “nodes” represent places, “edges” represent adjacency and “arcs” correspond to actions for moving from one place to another
- Often referred to as ‘global correctness’ Vs ‘local accuracy’ and can be related to *Grid-based* and *Pose-based* approaches to map evaluation
- In robotic mapping, due to inherent task dependency, there is no ‘optimal general mapping’





Domain-specific Considerations

Manufacturing Environments

- **Automated Guided Vehicle (AGV) Navigation**
  - Critical Enablers
    - Adaptability to human-centered collaboration (flexible),
    - Ability to cope with unstructured, dynamic environments (smart), and
    - Keeping humans out of harm's way (safe)
  - Minimal Infrastructure World Modeling Framework
    - Minimizing dependency on external reference markers
    - Engineering & Maintenance of the operational domain
    - Coping with dynamic (unmodeled) events and obstacles
- **Human-Robot Interaction**
  - Cooperation Vs Collaboration (Master-Slave Vs Active Participant)
  - Next generation of robots working side-by-side with humans
- **Mobile Manipulation**
  - On-the-move manipulation requires greater flexibility and robustness

Search and Rescue Robotics Environments

- **Geometric accuracy** not as important as topological correctness
- **Situational awareness**
- **Tele-operated Vs Autonomous missions**
  - Bounded autonomy
- **Multi-robot mapping**
  - Map-merging issues
  - Co-operation & Co-ordination
- **Human-in-the-loop evaluation issues**
  - Operator skill-level
  - Influences on the operator (e.g. fatigue)

Human- and robot-centric representational challenges in evaluation of mapping systems

Robot Competitions

Competitions and field exercises are two different yet effective ways of systematically evaluating the performance of robotic systems

- **Virtual Manufacturing Automation Competition (VMAC)** (Organizers: Stephen Balakirsky and Raj Madhavan)
  - ICRA Robot Challenge (May'10 & May'09): International
  - NIST (April'09 & April'08): National
  - Striving to allow increased automation for small- and medium-manufacturers
  - Competition allows for the design of performance metrics so that current and potential end-users can competitively compare technology and promote innovation

IEEE ICRA 2010 Robot Challenge

Virtual Manufacturing Automation Competition (VMAC)

Team Description Papers due January 5th, 2010

Call for Participation

Entries are sought for the **2010 Virtual Manufacturing Automation Competition (VMAC)** to be held in conjunction with the 2010 IEEE Intl. Conference on Robotics and Automation. Participation in the competition will consist of a software demonstration of a team of robots solving one or more of the pre-defined challenge tasks in the USARSim simulation environment. This demonstration may be performed remotely or on-site in the simulated environment set up for the competition.

The goal of the challenge is to provide an industrially relevant scenario and performance benchmarks to assess technologies in the areas of robot navigation in dynamic unstructured environments, mixed palletizing operations, and mobile manipulation. The competition's open source policy is designed to encourage collaboration and the dissemination of ideas and algorithms.

Competition Format

The competition will include several "elemental" tests as well as a full scenario. Teams are encouraged to participate in as many events as possible, with winners in each of the individual events as well as an overall winner being announced. Since this is a simulated event, perfect ground truth is available on such items as vehicle locations, package locations, package types, etc. It is desired that teams utilize as little of this information as possible, however, the organizers realize that most teams will require some help from ground truth. Teams must disclose their ground truth needs in their Team Description Paper (described below).

Events

A **Mixed Palletizing** task will test a team's ability to autonomously fill orders that a distribution center receives. The teams will have up to 3 robotic arms available, and each arm will work simultaneously on up to 3 pallets of materials. This test will consist of a circular conveyor of items (speed controlled by the team) from which team's must assemble a stable configuration of contents on the pallets as rapidly as possible.

A **Mobility** task will have team's autonomously deliver completed pallets throughout an unstructured factory environment. Teams will have up to 3 robotic platforms available for delivering the packages. When a package becomes available, the team must dock a robot with a loading conveyor to pick-up the package, and then deliver the package to one of several locations.

A **Scenario** will combine the two elemental tests into a single event. Teams will prepare pallets for transport and then move them throughout the factory environment to their destination.

What's Next?

Potential participants should submit a team description paper (TDP) of their intended entry to roboSim@nist.gov. The TDP should contain which challenge events your team will participate in, a high-level description of the algorithms that you will employ, requirements that you will place on the simulation system (i.e. what robots, sensors, and infrastructure you expect to need), and references to the team's relevant work in the area.

More information can be found on the VMA Competition web page at http://www.vma-competition.com/

Contact Details:

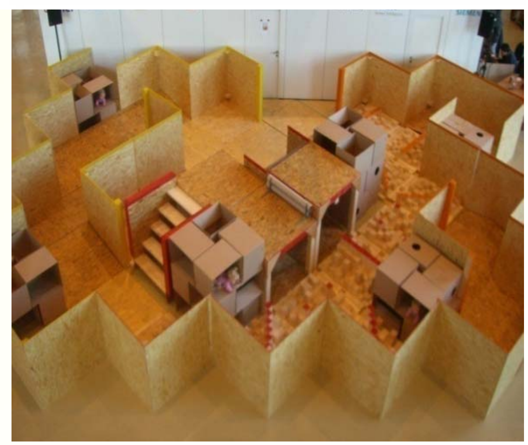
Dr. Stephen Balakirsky
Dr. Raj Madhavan
Intelligent Systems Division
National Institute of Standards and Technology (NIST)
Gaithersburg, MD 20899-8230

Call: 301-975-4791
Email: robosim@nist.gov

IEEE Robotics and Automation Society

IEEE

- **RoboCup Rescue League (Virtual & Real)**
  - Tying the real-world to research
  - Standard test methods are embedded



MARYLAND ROBOTICS CENTER
THE INSTITUTE FOR SYSTEMS RESEARCH

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

RoboCup Rescue League Competitions
Virtual League* (1)

- Motivation: Emergency responders must enter unknown space and quickly reach designated areas
- Features: Flat floor and sloped mazes, large featureless spaces, various lighting conditions
- Teaming: Up to 4 robots given 20 minutes to map environment
- *A priori* data: None
- Models: Realistic sensor noise models. Perfect radio coverage
- Resultant Map: Geotiff format

Map Evaluation

- Skeletal quality
- Metric quality
- Attribution/Annotation

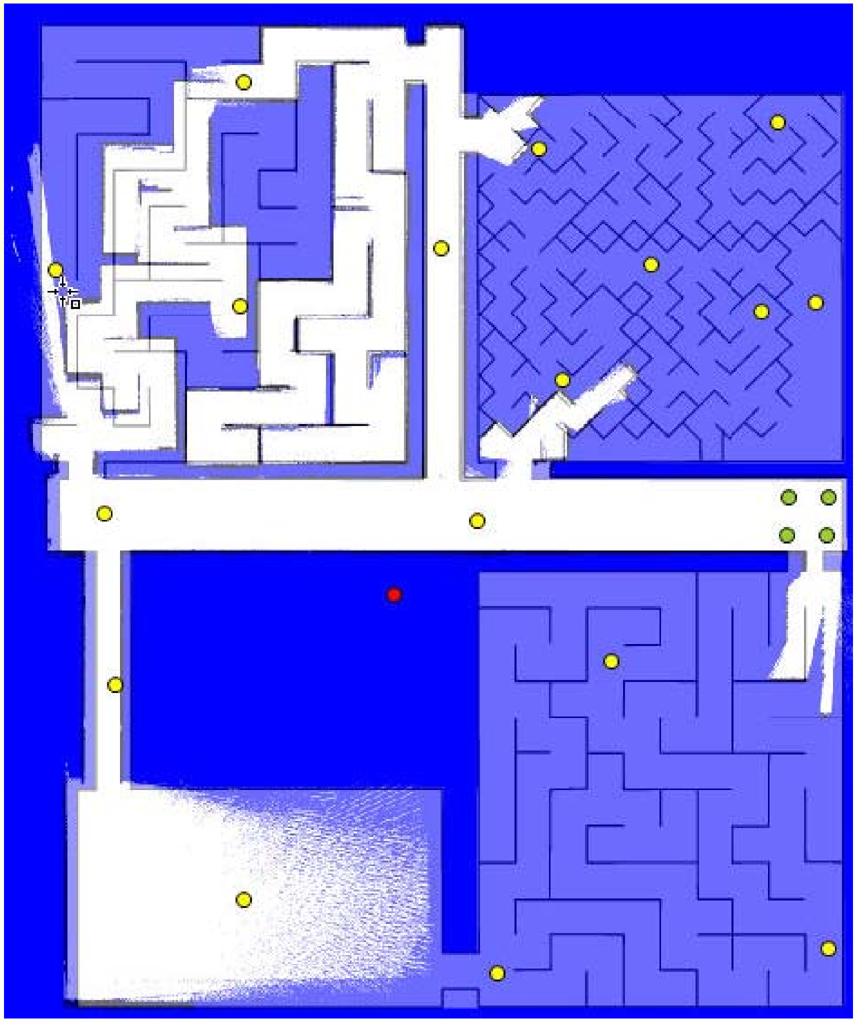
**Work by Dr. Stephen Balakirsky et al.*

MARYLAND ROBOTICS CENTER
THE INSTITUTE FOR SYSTEMS RESEARCH

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

RoboCup Rescue League Competitions
Virtual League (2)

- Task-based map assessment
- Technique examines if a path to given Points of Interest (POI) can be extracted correctly
  - POIs pseudo-randomly selected
  - Points awarded for each valid path
- Maps delivered in grid-based format
  - POIs mapped to team's map
  - For each POI in the team's explored area a standard path generating algorithm was run to produce a path
  - Topological features of path are then extracted and points awarded for correct paths



RoboCup Rescue League Competitions
Physical League*

Yearly competitions provide direct comparison of robotic approaches, objective performance evaluation, and a public proving ground for field-able robotic systems

- Motivation: Robot Mobility and Mapping Capabilities & Best in Class Events
- Features: Yellow, Orange and Red NIST Reference Test Arenas
- Teaming: Single Robot Timed Trials
- *A priori* data: None
- Models: Maze Geometry
- Resultant Map: 2D Grid Maps

Map Evaluation

– 2D Grid-based

– Metric quality evaluation based on ground truth

**Work by Adam Jacoff et al.*

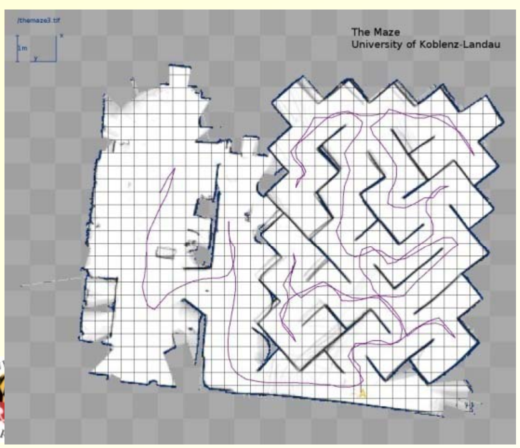
MARYLAND ROBOTICS CENTER
THE INSTITUTE FOR SYSTEMS RESEARCH

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Field Testing & Evaluation

- **Field Exercises**
  - Co-located with the Disaster City Response Robot Evaluation in College Station, Texas
  - Demonstrate sensing and mapping technologies for assisting responders in disaster scenarios
  - Test methods focused on evaluating limitations of mapping schemes
  - Generated Maps: primarily 2D with laser rangefinder data

- **Dissemination of Sensor Datasets**
  - Consists of real sensor datasets and simulated environments in addition to physical versions propagated internationally e.g. NIST maze dataset w/ ground truth



IEEE-RAS TC-PEBRAS

- ***TC on Performance Evaluation and Benchmarking of Robotic and Automation Systems***

  Co-Chairs: Raj Madhavan, Angel del Pobil, and Elena Messina
  - Approved during ICRA'09 TAB/AdCom meeting
  - TC-PEBRAS concentrates on performance aspects of intelligent systems including software and other influencing factors
  - *http://tab.ieee-ras.org/committeeinfo.php?tcid=35*

MARYLAND ROBOTICS CENTER
THE INSTITUTE FOR SYSTEMS RESEARCH

TCPEBI

Membe

Recen
- IEE
- ICR
  - –
  - –
- EC/
- 10tl 201
- IRC Rot

Future
- Inte Spr
- Rot
- ICR

T C S P O T L I G H T

Performance Evaluation and Benchmarking of Robotic and Automation Systems

Raj Madhavan (raj.madhavan@ieee.org), Computational Sciences and Engineering Division, Oak Ridge National Laboratory (ORNL), Intelligent Systems Division, National Institute of Standards and Technology (NIST); Angel P. del Pobil (pobil@uji.es), Engineering and Computer Science Department, Universitat Jaume I, Spain, and Department of Interaction Science, Sungkyunkwan University, Seoul, South Korea; Elena Messina (elena.messina@nist.gov), Intelligent Systems Division, NIST

The Technical Committee (TC) on Performance Evaluation and Benchmarking of Robotic and Automation Systems (TC-PEBRAS) was approved at the Technical Activities Board (TAB) meeting held at the 2009 International Conference on Robotics and Automation (ICRA) in Kobe, Japan. It is intended to serve as a forum to address performance evaluation and benchmarking issues pertaining to robotic and automation systems, in general.

Fueled by investments from the defense and industrial sectors, the availability of increased computing power, and advances in sensor systems, the development of robotic systems has progressed with a renewed vigor in recent years. In the coming decade, significant progress can be expected in manufacturing robotics and automation, automotive, service, and health care robotics, demonstrating the utility of robotic systems and, as a result, helping their societal acceptance. It is our belief that end-users' requirements should drive developers and integrators such that a resulting intelligent system is useful and affordable. Only by involving all of the three parties: users, developers, and integrators in a coupled fashion, can meaningful solutions be produced that can stand the ever-varying requirements imposed by: 1) tasks that are either application or environment dependent, 2) hardware and software advancements/restrictions that affect the development cycle, and 3) budgetary constraints that interrupt and hamper sustained progress.

To guarantee such requirements and ensure reliability and robustness of robotic and automation systems, it is crucial to quantify their performance via scientifically sound and statistically significant metrics, measurement, and evaluation methodologies. Currently, there is no accepted standard for quantitatively measuring the performance of such systems against user-defined requirements; there is no consensus on what objective evaluation procedures need to be followed to deduce the performance of these systems. The lack of reproducible and repeatable test methods has precluded researchers working toward a common goal from exchanging and communicating results, intercomparing robot performance, and leveraging previous work that could otherwise avoid duplication and expedite technology transfer. Furthermore, for robotics to be accepted as a scientific endeavor, repeatable and reproducible test methods are paramount to experimentally verify and validate technical methodologies. It is not an exaggeration to claim that interest in benchmarking and standardization of robotic systems has reached critical mass as witnessed by the increasing number of workshops, journal special issues, and publications dedicated to this topic. Researchers, developers, and end-users alike are aware of the problems, the importance of such efforts, and how it can be beneficial to them. Leaving emerging robotic technologies to proliferate in an unguided direction comes with a high price: synergistic opportunities remain unrealized and lack of cohesion in the community hinders the progress in many domains (Figure 1).

Quantifying performance via benchmarking and standardization will improve the utility of robotic and automation systems in already established application areas. Having agreed-upon measures of performance and merit is a critical prerequisite to wider acceptance and proliferation of emerging technologies. Reproducible experiments and benchmarks are a foundational tenet of the scientific method. Robotics as a discipline needs to adopt more of these practices to mature. It is our hope that the efforts of this TC will bring together fragmented attempts to provide a baseline for comparison and mechanisms for targeting specific aspects of a system, thus allowing researchers and practitioners to assess the performance of various systems in different scenarios and environmental conditions. The end-user communities can use this forum to monitor progress in emerging technologies and to provide input regarding their needs and requirements. Benchmark and challenging problems in specific technology areas can be submitted to this TC to stimulate progress in concrete and directed ways.

Some of the past and recent activities of this TC have focused on organizing workshops and publications such as journal special issues and books. A brief summary of these activities are as follows:

- R. Madhavan, C. Scrapper, and A. Kleiner, Eds., "Characterizing Mobile Robot Localization and Mapping," *Autonomous Robots* (Journal Special Issue), vol. 27, no. 4, Nov. 2009: The primary focus of the special issue is to quantify performance characteristics of various approaches to mobile robot localization and mapping in a variety of domains. The nine articles in the issue detail the capabilities and limitations of several approaches by the intercomparison of experimental results and development of schemes for ground truth generation as well as the underlying mechanisms used to formulate these solutions. See http://www.springerlink.com/content/j87h11r73031/?p=eb111fcc4c6f440c89ddc69fc41b5726&pi=0 for a list of accepted articles.

Digital Object Identifier 10.1109/MRA.2010.935811

IEEE-RAS Standing Com

Two Study Groups (SGs):

- **Map Data Representation:** The ob data representation(s) and discuss ho data for robot navigation via standard and proper use in robot software fram
- **Glossary/Ontology for Robotics a** to identify, develop, and document sa serve as a common reference for the

- Study Group Docs. was made avail formation of a Working Group is being by *June 2011*.

- A full day Standards Meeting/WS to (Shanghai) and 2011 IROS (San Fran

Raj Madhavan
Chair, Standards Cmte. & Associate VP:

"RAS-SCSA : History and Recent Activities", I Hui, Journal of the Robotics Society of Japan, Sp May 2011.

MARYLAND ROBOTICS CENTER
THE INSTITUTE FOR SYSTEMS RESEARCH

RAS Standing Committee for Standards Activities—An Update on Recent Activities

By Raj Madhavan

Standards are crucial for driving industry innovation and technology transfer. Benchmarking and standardization are vital to the development of robotic and automation systems in already-established application areas and are critical to wider (including societal) acceptance of emerging technologies. It is widely accepted within the robotics and automation community that leaving emerging robotic technologies to proliferate in an unguided direction comes with a high price: synergistic opportunities remain unrealized and a lack of cohesion in the community hinders the progress in many domains such as manufacturing, service, health care, and security, to name a few [1], [2]. For an in-depth discussion of the existing standards and worldwide efforts, an interested reader is referred to [3].

The Standing Committee for Standards Activities (SCSA) under the Industrial Activities Board (IAB) of the IEEE Robotics and Automation Society (RAS) is working together with the research and industrial communities and other standards developing organizations to help develop standards for robotics and automation. The scope of the activities of the RAS-SCSA is to formally adopt and confirm best practices in robotics and automation as standards. Within this scope, the RAS-SCSA is pursuing the following objectives [4]:

- promote common measures and definitions in robotics and automation
- promote measurability and comparability of robotics and automation technology
- promote integrability, portability, and reusability of robotics and automation technology.

Some of the previous work carried out by the Standards Committee can be found in [5] and [6]. At the 2010 International Conference on Robotics and Automation (ICRA 2010) held in Anchorage, Alaska, the SCSA hosted two meetings. One of these meetings was organized by the IEEE Standards Association (IEEE-SA) to better understand the procedures involved in the standards process. The second meeting was part of the SCSA's regular meeting series, which delved into identifying the suitable areas for standardization that are both crucial and achievable in the short term. These meetings served as an excellent forum to discuss and exchange ideas from professionals working on various aspects of robotics and automation and many with previous experiences in standards development. Two follow-up meetings were held at the 2010 International Conference on Intelligent Robots and Systems (IROS 2010) in Taipei, Taiwan, to further discuss the scope of the study groups (SGs) and to develop a timeline for the standards-defining activities.

On the basis of the discussions stemming from these meetings and with input and consensus from participating

Digital Object Identifier 10.1109/MRA.2011.940148
Date of publication: 14 April 2011

Performance Metrics for Intelligent Systems (PerMIS)

- 10th Anniversary PerMIS'10 WS
  - *Main theme: Key role of performance assessment of intelligent systems that can co-exist with humans*
  - NIST, DARPA, NSF, IEEE, ACM (Sponsors)
  - *5 Plenaries*: Ken Goldberg (UCBerkeley), Helen Grenier (Cyphy Works), Greg Dudek (McGill), Jon Bornstein (ARL), Herman Bryunincks (KUL)
  - *6 Special Sessions*
  - *http://www.nist.gov/mel/isd/permis2010.cfm/*

PerMIS'10

Performance Metrics for Intelligent Systems Workshop

September 28-30, 2010

http://www.nist.gov/mel/isd/permis2010.cfm/

NIST National Institute of Standards and Technology

DARPA NSF IEEE acm

General Chair
Elena Messina, NIST

Program Chair
Raj Madhavan, UMD-CP/NIST

Publicity Chair
B. Grabowski *MITRE USA*

Publications Chair
B. Weiss *NIST USA*

Program Committee
S. Balakirsky *NIST USA*
G. Berg-Cross *EM & I USA*
G. Blankenship *UMD-CP USA*
F. Bonsignorio *UC3M Spain*
R. Bostelman *NIST USA*
M. Childers *ARL USA*
M. Fields *ARL USA*
A. Godil *NIST USA*
J. Gunderson *GammaTwo USA*
L. Gunderson *GammaTwo USA*
S.K. Gupta *UMD-CP USA*
A. Kleiner *UFreiburg Germany*
R. Lakaemper *Temple USA*
M. Lewis *UPitt USA*
A. del Pobil *UJaume-I Spain*
E. Prassler *UAppSci-BRS Germany*
F. Proctor *NIST USA*
D. Prokhorov *Toyota USA*
C. Schlenoff *NIST USA*
C. Scrapper *MITRE USA*
J. Shi *GM USA*
M. Shneier *NIST USA*
E. Tunstel *JHU-APL USA*

Location
Hyatt Regency Baltimore

Sponsors
NIST, DARPA, NSF, ACM-SIGART, IEEE-RAS TC-PEBRAS, IEEE W/NV Sensors Council Chapter

Call for Papers

The 2010 Performance Metrics for Intelligent Systems workshop will be the tenth in a series dedicated to defining measures and methodologies of evaluating performance of intelligent systems. Started in 2000, the PerMIS series focuses on applications of performance measures to applied problems in commercial, industrial, homeland security, and military applications.

PerMIS'10 will emphasize the ***key role of performance assessment in developing intelligent systems that can co-exist with humans*** towards improving the quality of our lives intertwined with automation. Adaptability to human-centered collaboration, the ability to cope with unstructured, dynamic environments, and keeping humans out of harm's way have been widely accepted as critical prerequisites. Designing such flexible, smart, and safe systems requires that their performance be quantifiable thereby facilitating emerging technologies and societal acceptance.

Selected papers from PerMIS'10 will be considered for publication in a *special issue with the International Journal of Intelligent Control and Systems*. A National Science Foundation *Poster Session* is also being planned with travel support to undergraduate and graduate students. The Proceedings of PerMIS are indexed by INSPEC, Compendex, ACM Digital Library, and are released as a NIST Special Publication.

In relation to the main theme, topic areas include, but are not limited to:

- Defining & Measuring Aspects and Capabilities of a Co-X (e.g. Co-Workers, Co-Inhabitants, ...) System:
  - Human-Robot Interaction, Collaboration and Coordination
  - Mobility
  - Responsiveness, Reliability, Trustworthiness, Interchangeability, Durability
  - Variable Levels of Autonomy
  - Taxonomies
  - Biologically Inspired Models
- Evaluating Components within Intelligent Systems:
  - Sensing and Perception
  - Knowledge Representation, World Models, Ontologies
  - Planning and Control
  - Learning, Adapting and Reasoning
- Underlying Infrastructural Support for Performance Assessment:
  - Testing and Evaluation (including testbeds and competitions for inter-comparisons)
  - Instrumentation and Other Measurement Tools
  - Simulation and Modeling
- Technology Readiness Measures for Intelligent Systems
- Benchmarks and Applied Performance Measures in Various Domains:
  - Manufacturing, Logistics, and Industrial Systems
  - Service: Domestic, Mining, Agriculture, ...
  - Intelligent Transportation Systems
  - Defense and Security
  - Emergency Response Robots (e.g. search and rescue, bomb disposal)
  - Intelligent systems for Hazardous Environments (e.g. nuclear remediation)
  - Smart Grid
  - Space Robotics
  - Medical & Healthcare

Submission Information

Prospective authors are requested to submit a full paper (max. 8 pages) or an extended abstract (1-2 pages) for review. *Special session* proposals can also be submitted as papers but should contain 1) a session title and a brief statement of purpose, 2) name and affiliation of the organizer(s), and 3) a preliminary list of speakers. All submissions must be written in English, starting with a succinct statement of the problem, the results achieved, their significance, and performance evaluation of the results. Papers are to be submitted at the workshop website using the specified templates.

Important Dates

Submission of full papers	July 02, 2010
Proposal for special sessions	July 02, 2010
Notification of acceptance	July 30, 2010
Final papers due	August 27, 2010

Concluding Thoughts

- Societal Acceptance of Robots and Robotic Systems
  - Curb unrealistic expectations
  - Adoption and Acceptance will largely depend on convenience, cost and need

- ***Robotics as a scientific endeavor***
  - Scientific framework for performance evaluation of IS
  - How can we develop scientifically sound & statistically significant methodologies and design experiments/test methods to evaluate intelligent mobile robots (navigation, behaviors, …)?
    - Better developed by taking into account requirements imposed by end-users and domain specific constraints that are grounded in practicality

Thank you!

Questions?

Quantifying Autonomy of Robotic Systems: Performance Requirements & Metrics

- Perform certain missions
- In certain environments
- With certain accuracies
- Possess certain autonomy
- Under certain costs
- Technology Readiness
- Reliable
- Safe
- …

- Accuracy: { final state tolerance }
- Efficiency: { costs }
- Effectiveness: { %completeness }
- Repeatability: { %confidence }
- Reliability: { %reliability }
- Readiness: TRL, …

- **Quantitative Metrics**
- **Repeatable Process**
- **Consistent/Compatible with Practices**
- **Extensible**
- **Scalable**