

## Performance Evaluation and Benchmarking of Robotic and Automation Systems

Raj Madhavan ([raj.madhavan@ieee.org](mailto: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](mailto: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](mailto: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.

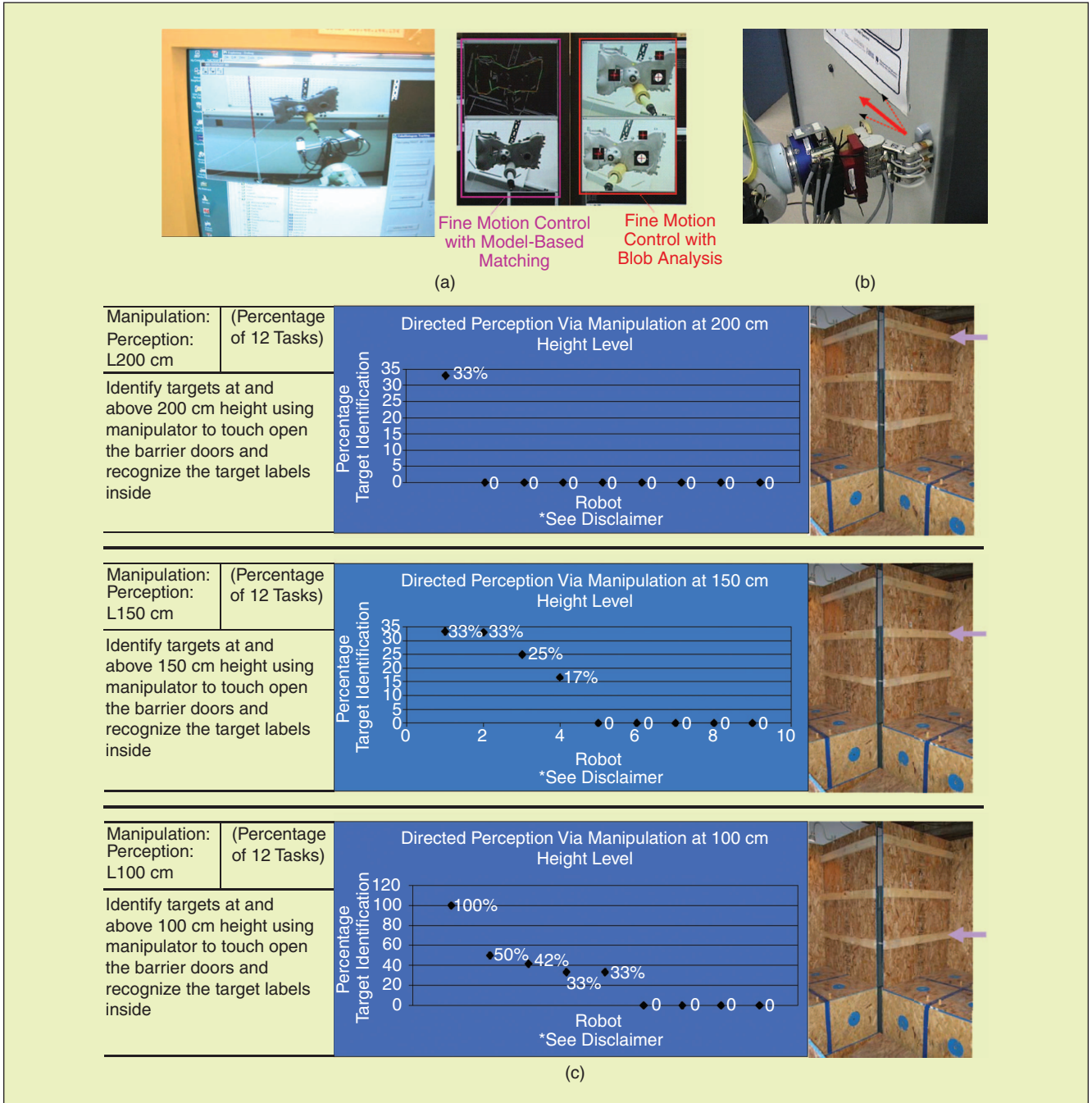
Some of the past and recent activities of this TC have focused on organizing workshops and publications such as journal special issues and books. A brief summary of these activities are as follows:

- ◆ 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.

- ◆ Angel del Pobil, Raj Madhavan, and Fabio Bonsignorio (Organizers), “Performance Evaluation and Benchmarking for Next Intelligent Robots and Systems,” in *IROS 2009 Full-Day Workshop*, St. Louis, MO, USA, Oct. 2009: Fourth in the series at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), the emphasis of the workshop was on principles, methods,

and applications expanding beyond the current limit of robotics applications in terms of cognitive capabilities and autonomy. More details on the speakers and the accompanying papers are available from <http://www.heronrobots.com/EuronGEMSig/GEMSIGIROS09Program.html>.

- ◆ *2009 Performance Metrics for Intelligent Systems (PerMIS'09) Workshop* September 2009, Gaithersburg, MD, USA: This



**Figure 1.** Performance evaluation of robots and robotic systems across different domains. (a) Measuring the performance of 6 degrees of freedom (DoF) motion tracking systems in manufacturing applications at NIST. (b) Mobile manipulation task-based evaluation at Universitat Jaume I. (c) Summary plot of performance by nine different robots on a directed perception test method used for response robots (data collected at NIST).

workshop is the only one of its kind dedicated to defining measures and methodologies of evaluating performance of intelligent systems. Started in 2000, the annual PerMIS series focuses on applications of performance measures to practical problems in commercial, industrial, homeland security, and military applications. It has proved to be an excellent forum for discussions and partnerships, dissemination of ideas, and future collaborations between researchers, graduate students, and practitioners from industry, academia, and government agencies. The main theme of the ninth iteration of the workshop, PerMIS'09 was "Does performance measurement accelerates the pace of advancement for intelligent systems?" Please refer to [http://www.isd.mel.nist.gov/PerMIS\\_2009/](http://www.isd.mel.nist.gov/PerMIS_2009/) for additional information on the workshop.

- ◆ Raj Madhavan, Edward Tunstel, and Elena Messina, Eds., *Performance Evaluation and Benchmarking of Intelligent Systems*, Springer-Verlag, ISBN: 978-1-4419-0491-1, Sept. 2009: The book presents a detailed and coherent picture of state-of-the-art, recent developments and further research areas in intelligent systems. The chapters in the book are a collection of expanded and revised papers presented at PerMIS'08. For a list of the chapters in the book, see <http://www.springer.com/computer/communications/book/978-1-4419-0491-1>.
- ◆ Fabio Bonsignorio, Angel del Pobil, and John Hallam (Organizers), "Good Experimental Methodology in Robotics," in *RSS'09 Full-day Workshop*, June 2009, Seattle, WA, USA: The focus of this workshop was on disseminating the work of good experimental robotics groups and using it to illustrate high-quality experimental work and reporting. Additional details are available from <http://www.heronrobots.com/EuronGEMSig/GEMSIGRSS09Program.html>.

Some of our envisaged future activities for the TC are as follows:

- ◆ We welcome discussions with other chairs and members on performance evaluation issues within the core subject areas of their TCs. Toward this, we are planning to organize a kick-off workshop to bring together the community and other TCs to identify areas of focus to define the future activities of TC-PEBRAS.
- ◆ Robotics competitions can be used as benchmarks and as a means to measure performance in targeted areas. Ongoing, yearly competitions can prove useful in quantifying progress and, through increasing complexity, directing research innovation in meaningful directions. Therefore, the TC activities include robotics competitions. Two of the cosponsored competitions of TC-PEBRAS are the Virtual Manufacturing Automation Competition (VMAC: <http://www.vma-competition.com>), which was initiated as a demonstration event at ICRA'08 and as a continuing robot challenge event for ICRA'09, ICRA'10, and the upcoming mobile microrobotics challenge

(<http://www.nist.gov/eel/semiconductor/mmc/>) to be held at ICRA'10 in Anchorage, Alaska.

The goal of the VMAC challenge is to provide an industrially relevant scenario and performance benchmarks to assess technologies in the areas of robot navigation in dynamic unstructured environments, including mixed palletizing operations and mobile manipulation. The competition's open-source policy is designed to encourage collaboration and the dissemination of ideas and algorithms. The objective of the microrobotics challenge is to inspire innovation in microrobot design and to evaluate the performance of the latest microrobotic technologies. Viewed under a microscope, the microbots are operated by remote control and move in response to changing magnetic fields or electrical signals transmitted across a playing field located on a microchip.

- ◆ To promote information sharing about robotics standards efforts world wide, we are establishing a wiki as a community resource (see <http://collaborate.nist.gov/twiki-robotstds>). This wiki is intended to cover evolving and existing standards for all types of robotics applications and includes performance standards. We invite the community to contribute to this wiki and use it as a resource to find relevant work and identify others who have similar interests.
- ◆ A steering committee is being assembled to facilitate interaction among interested parties. Please contact us if you would like to contribute to the efforts of this TC.

Even though the performance evaluation and benchmarking activities might be carried out under individual TCs pertaining to their specific area, it is our belief that TC-PEBRAS will serve all the existing TCs by providing a common forum to address this issue jointly and benefit from lessons learned. This will lessen the limitations currently faced by individual TCs and enable steps required to speed up the achievement of meaningful results by cross-germination and cooperation. TC-PEBRAS will concentrate on performance aspects of intelligent systems, including software and other influencing factors. We expect to learn from the experiences and insights of the members of the existing TCs, gained both through their theoretical development and practical implementation of intelligent systems, in a variety of diverse application domains.

More information on TC-PEBRAS and upcoming activities are periodically updated at the Web site <http://tab.ieee-ras.org/committeeinfo.php?tcid=35>. To join the TC or for sponsorship of related meetings, the cochairs can be reached via e-mail using [raj.madhavan@ieee.org](mailto:raj.madhavan@ieee.org), [pobil@uji.es](mailto:pobil@uji.es), and [elena.messina@nist.gov](mailto:elena.messina@nist.gov).

We look forward to working jointly with you on this time-critical endeavor!

## Acknowledgment

We thank Prof. Ken Goldberg and the TAB for their enthusiasm and continued support of this TC. A.P. del Pobil is supported by WCU program, National Research Foundation of Korea (R31-2008-000-10062-0).