INTERNATIONAL JOURNAL OF INTELLIGENT CONTROL AND SYSTEMS
VOL.16,NO.2,JUNE 2011,37-39

EDITORIAL
International Journal of Intelligent Control and Systems Special Issue: Quantifying the Performance of Intelligent Systems

"It is both in a spirit of scientific enquiry and for pragmatic motivations that we embark on the quest for metrics for intelligence of constructed systems."

> From the White Paper explaining the goals of the Workshop: "Measuring Performance and Intelligence of Systems with Autonomy: Metrics for Intelligence of Constructed Systems," (Editors: E. Messina and A. Meystel), *Measuring the Performance and Intelligence of Systems: Proceedings of the 2000 PerMIS Workshop*, Gaithersburg, MD, U.S.A, August 14-16, 2000, NIST Special Publication 970.

As the new millennium was upon us in 2000, a group of researchers gathered for the first time seeking to address several issues pertaining to intelligent systems:

- How can we measure the current state of the science and assess progress in the field?
- How can users select among different candidate systems and decide which system will be most suited to their application?
- How can we break the cycle of re-invention and constant initiation of projects with blank slates and find ways to reuse existing components?

The first Performance Metrics for Intelligent Systems workshop brought together researchers, developers, and users from disparate academic disciplines and domains of application to share ideas about how to tackle the multi-faceted challenges of defining and measuring intelligence in artificial systems. The intelligent systems could take numerous forms: robots, factory or enterprise control systems, smart homes, decision support systems, etc. A community was formed, which evolved over the years. The workshop series carried on and became an annual event (with the exception of 2005). The 10^{th} workshop in the series focused on systems designed to work closely with humans. The theme of PerMIS'10 held at Hyatt Regency Baltimore from September 28-30, 2010 was 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.

Intelligent systems are becoming more of a reality with each passing year and the questions raised in the first workshop are still relevant. Additional questions have been raised, such as "how does one specify the requirements for the performance of an intelligent system?" and "how can concrete performance goals and good measures of performance help spur and focus innovation?" Over the years, the center of gravity of the program shifted more towards applied measures, rather than theoretical discussions about the general nature of intelligence. Many communities have availed themselves of the special sessions to focus on their particular interests and create mini-workshops. The concept of performance evaluation being an integral part of any research and acquisition program has become accepted. Many of the papers published in the PerMIS proceedings have been highly referenced and provide the communities with good starting points for establishing measurements for new projects and programs. We are extremely grateful to the numerous colleagues who have supported PerMIS throughout the years. Without their dedication and hard work, this series would not have survived for a decade.

In this vein, this journal special issue addresses performance evaluation and associated metrics for intelligent systems from a wide variety of domains. The articles in this issue are representative of the state-of-the-art and bring forward authors' perspectives on what aspects are missing and how this gap should be bridged. The special issue consists of selected revised and expanded articles from PerMIS'10 that underwent two rounds of rigorous peer-reviews before being accepted. Out of fifty nine papers presented at the 2010 workshop, twelve were selected based on their presentation at the workshop and feedback from the reviewers during the workshop paper acceptance process. After the journal review process, nine articles were accepted for publication in the special issue.

The first four articles in this issue deal with human-robot interaction or human factors studies. These issues that the authors raise in their articles have to be contended with when robots and humans interact and are increasingly relevant as we witness more and more robots being developed to work alongside humans. The next two articles focus on perception metrics for outdoor unmanned ground vehicle (UGV) navigation. The last three articles round out the special issue by reporting on the authors' experience in developing fielded systems and the lessons learned from deploying such technologies. A summary of the articles follow:

The submission by Katherine Tsui, Munjal Desai, Holly Yanco, Henriette Cramer, and Nicander Kemper discusses "Measuring Attitudes Toward Telepresence Robots". This topic hews to the theme of the 2010 PerMIS, which explored the performance measurement aspects of robots that co-exist with humans. The authors investigated whether an existing scale used to measure how humans feel about autonomous robots can also be used for robots that are operated with a human in the loop. The Negative Attitude towards Robots Scale (NARS) has been extensively used and validated as a means for predicting user interaction based on how humans score. This psychological tool asks participants to rate their feelings in three main areas: "situations of interaction with robots," "social influence of robots," and "emotions in interaction with robots". Specifically, the team studied the applicability of the NARS scale for telepresence robots, which are remotely controlled by a human to perform tasks in an office-like environment.

Also related to the issue of human interaction with robots is the contribution by Birsen Donmez and Mary Cummins in the article "Metric Selection for Evaluating Human Supervisory Control of Unmanned Vehicles". The perspective taken in this article is a higher-level one: the selection of appropriate metrics for accomplishing the evaluation of the human-automation system. The authors present a series of metrics evaluation criteria along with cost and benefit parameters that may be used by experiment designers. In their article, they present the results from an experiment in which subjects used two different multi-criteria frameworks (Analytic Hierarchy Process and Ranking Input Matrix) for establishing their metrics. The results can prove useful in human-automation metric selection in future experiments.

Intelligent semi-autonomous wheel chairs assist people with mobility concerns ranging from day-to-day use to specialized short-term needs. To benchmark the quality of such intelligent wheel chairs, Joelle Pineau, Amin Atrash, Robert West, and Julien Villemure, in the article "On the Feasibility of Using a Standardized Test for Evaluating a Speech-Controlled Smart Wheelchair", extend the existing Wheelchair Skills Test (WST) with their new evaluation paradigm especially with respect to user interface requirements. The authors share their experiences and results while identifying missing gaps and their continuing efforts to address them.

A thorough analysis of unmanned aerial vehicle (UAV) operator workload during remote piloting in cluttered low altitude environments is presented in the contribution "An Indoor Study to Evaluate a Mixed-Reality Interface For Unmanned Aerial Vehicle Operations in Near Earth Environments" from James Hing, Justin Menda, Kurtulus Izzetoglu and Paul Oh. They assert and test hypotheses related to remote-pilot performance enhancement using a virtual view from behind a UAV relative to performance using a more typical onboard camera view. Results from a set of UAV piloting experiments in a reduced-scale laboratory environment are presented wherein the authors make novel use of functional near-infrared brain sensor data as a measure of brain activity to assess cognitive workload of eleven test subjects during UAV piloting tasks. Statistical metrics are applied to quantify observed positive effects of the virtual view on UAV performance as well as reductions in cognitive workload.

In autonomous vehicle navigation applications in outdoor environments, the ability to reliably interpret sensor data is of paramount importance. Christopher Brunner, Thierry Peynot, and Teresa Vidal-Calleja propose image quality metrics to evaluate the quality of outdoor robotic perception systems in their article "Visual Metrics for the Evaluation of Sensor Data Quality in Outdoor Perception". The authors show the utility of the developed metrics in a monocular SLAM (Simultaneous Localization And Mapping) UGV navigation application. They demonstrate how the reliability of the pose estimate can be maintained by switching between sensing modalities depending on challenging environmental conditions (e.g. smoke).

Perception techniques for UGV autonomous navigation continue to improve and extend to address ever more complex outdoor operating environments. Computer vision-based perception is central to many effective solutions spanning a wide range of bases for motion control decisions. Suitable measures for evaluating their performance serve to enhance the progressive development and assessment of the technology. Arturo Rankin, Tonislav Ivanov and Shane Brennan address the uncommon perception task of detecting water bodies on the ground that could represent hazards to UGV mobility in "Methods for Evaluating the Performance of Unmanned Ground Vehicle Water Detection". They present methods to evaluate performance for image-based algorithms facilitated by a software tool to automatically generate ground truth data, thereby improving upon a manual process that required human operator involvement and introducing measures of localization accuracy. Their approach enables water detection performance evaluation in image space and evaluation of water body localization performance in map space.

Marshal Childers, Barry Bodt, and Richard Camden report on their experiences while assessing various perception and planning systems for autonomous ground vehicles from a variety of field trials carried over a number of years. Their article entitled "Assessing Unmanned Ground Vehicle Tactical Behaviors Performance" describes their insights and challenges faced in evaluating tactical behaviors of UGVs within the U.S. Army Research Laboratory Robotics Collaborative Technology Alliance program. Three specific assessment cases are included to illustrate how experiments were conducted and what subsequent lessons were learned in developing test methods, metrics, and assessment methodologies in close collaboration with developers.

Intelligent systems for military use are persistently advancing through development, acquisition, and deployment. At each stage there is a need for performance measures and metrics. Experiential accounts of technology evaluation are instructive to all developers, purchasers, and users. The article by Craig Schlenoff, Brian Weiss, and Michelle Steves entitled "A Detailed Discussion of Lessons Learned in Evaluating Emerging and Advanced Military Technologies" presents such accounts of two multi-year DARPA programs and the application of a systematic evaluation methodology that evolved to consider lessons learned. The technology programs involve sensor and information processing systems worn by combat soldiers (ASSIST) and systems for speech-to-speech language translation enabling communication in tactical situations in lieu of human interpreters (TRANSTAC). The authors describe lessons learned from evaluating advanced and emerging technology development prototypes of these complex systems. Among a number of instructive lessons discussed, the authors emphasize maximal early effort applied to evaluation design and logistical planning as a means to yield benefits as the evaluation progresses.

In order for embodied intelligent systems, such as robots, to function in the world, they need to reliably determine the position and orientation of objects in their surroundings. For numerous applications, the determination of the 6-dimensional pose of an object is critical. In other cases, the particular object must be tracked as it or the robot itself moves. The article "Real-Time Dynamic Pose Estimation Systems in Space: Lessons Learned for System Design and Performance Evaluation" by Chad English, Galia Okouneva, Pierre Saint-Cyr, Aradhana Choudhuri, and Tim Luu addresses the performance evaluation challenges specifically for demanding space applications. The authors provide an overview of their experiences in designing, developing, and deploying 6D pose estimation systems and present their perspective on performance metrics and methods.

We would like to extend our thanks to reviewers who read the submitted articles and provided their valuable comments which improved the overall quality of the articles in this issue. We hope that you find these articles useful and interesting. It is our sincere wish that the intelligent systems community invests more time and effort in evaluating and benchmarking their systems against user-defined requirements.

Guest Editors:
Raj Madhavan
Institute for Systems Research, University of Maryland, College Park &
Intelligent Systems Division, National Institute of Standards and Technology

Elena Messina
Intelligent Systems Division, National Institute of Standards and Technology

Edward Tunstel
Space Department, Johns Hopkins University Applied Physics Laboratory

May 2011
Maryland, U.S.A.