2010 IEEE/RSJ IROS
Taipei, Taiwan
October 22, 2010

Performance Evaluation and Benchmarking for Intelligent Robots and Systems with Cognitive and Autonomy Capabilities

Fabio P. Bonsignorio
Angel P. Del Pobil
Raj Madhavan

2010 IEEE/RSJ IROS
Taipei, Taiwan
October 22, 2010

Program

08:20-08:30 Introduction: Towards Benchmarking and Replicable Robotics: Progress and Future Directions

08:30-8:55 Experiments on grasp acquisition
Thomas Wimböck*, Berthold Bäuml, Gerd Hirzinger
DLR (German Aerospace Research Center)
Germany
Jeff Trinkle
Rensselaer Polytechnic Institute
USA

8:55-9:20 Hierarchical structuring of manipulation benchmarks in service robotics
Rainer Jäkel, Sven R. Schmidt-Rohr*, Martin Lösch, Rüdiger Dillmann
Karlsruhe Institute of Technology
Germany

9:20-9:45 Benchmarking for grasping in human environments
Aaron Dollar
Yale University
USA

9:45-10:10 Evaluating Social Robots: Insights from the Diagnosis of Autism Spectrum Disorders
Bill Smart
Willow Garage
USA

10:10-10:20 Coffee Break

10:20-10:45 Standardization of Cleaning Robot Performance Measurement in IEC
Sungsoo Rhim
Kyung Hee University
Korea

Performance Evaluation and Benchmarking for Intelligent Robots and Systems with Cognitive and Autonomy Capabilities

2010 IEEE/RSJ IROS
Taipei, Taiwan
October 22, 2010

Program

10:45-11:10 Decision-Theoretic Probabilistic Methods for Benchmarking
Matthijs T. J. Spaan, Pedro U. Lima
Institute for Systems and Robotics, Instituto Superior Técnico
Portugal

11:10-11:35 Metrics for Assistive Robotics Brain-Computer Interface Evaluation
M Stolen A Jardon C Balaguer
Universidad Carlos III de Madrid
Spain
F Bonsignorio Heron Robots and Universidad Carlos III de Madrid
Italy

11:35-12:00 RoboCup@Home and RoboCup@Work: From Benchmarking Algorithms to Benchmarking Service Robots in Residential and Industrial Scenarios
Gerhard K. Kraetzschmar
Bonn-Rhine-Sieg University
Germany

12:00 Final Discussion

Performance Evaluation and Benchmarking for Intelligent Robots and Systems with Cognitive and Autonomy Capabilities

2010 IEEE/RSJ IROS
Taipei, Taiwan
October 22, 2010

Events

IEEE/RSJ IROS 2010 Workshop on Performance Evaluation and Benchmarking for Intelligent Robots and Systems with Cognitive and Autonomy Capabilities (Taipei, Taiwan, October 22, 2010)

PerMIS'10 Performance Metrics for Intelligent Systems Workshop (Baltimore, USA, 28-30 September, 2010

Workshop on Good Experimental Methodology in Robotics and Replicable Robotics Research at RSS'10 (Zaragoza,Spain, June 28, 2010).

Workshop on The Role of Experiments in Robotics Research at IEEE ICRA 2010 (Anchorage, Alaska, USA, May 3-8, 2010)
Workshop on Good Experimental Methodology and Benchmarking in Robotics Research and Applications - GEMBENCH10 at the Euron Annual General Meeting (Donostia-S.Sebastian, Spain, 10-12 March, 2010)

2010 IEEE/RSJ IROS
Taipei, Taiwan
October 22, 2010

Events

IEEE/RSJ IROS 2009 Workshop on Performance Evaluation and Benchmarking for NEXT Intelligent Robots and Systems (St. Louis, MO, USA, October 15, 2009)

PerMIS'09 Performance Metrics for Intelligent Systems Workshop (NIST Gaithersburg, USA, 21-23 September, 2009)

Workshop on Good Experimental Methodology in Robotics at RSS'09 (Seattle,WA, USA, June 28, 2009).

Events

Workshop on Good Experimental Methodology and Benchmarking in Robotics Research and Applications - GEMBENCH09 at the Euron Annual General Meeting (Leuven, Belgium, 6-8 April, 2009)

IEEE/RSJ IROS 2008 Workshop on Performance Evaluation and Benchmarking for Intelligent Robots and Systems (Nice, France, 26 September, 2008

Clawar 2008 Session on Benchmarking and standardization for clawar and mobile robots (Coimbra, Portugal, 8-10 September, 2008)

PerMIS'08 Performance Metrics for Intelligent Systems Workshop (NIST Gaithersburg, USA, 19-21 August, 2008)

2010 IEEE/RSJ IROS
Taipei, Taiwan
October 22, 2010

Events

Workshop on Experimental Methodology and Benchmarking in Robotics Research at RSS'08 (Zurich, Switzerland, June 28, 2008).

CogGemBench 2008 (Karlsruhe, Germany, April 1,2008)

GEMBenchForum 2008 (Prague, Czech Republic, March 25-26, 2008)

IEEE/RSJ IROS 2007 Workshop on Performance Evaluation and Benchmarking for Intelligent Robots and Systems (San Diego, USA, November 2, 2007).

IEEE/RSJ IROS 2006 Workshop on Benchmarks in Robotics Research (Beijing, China, October 10, 2006).

www.heronrobots.com/EuronGEMSig
www.robot.uji.es/benchmarks
www.nist.gov/el/isd/permis2010.cfm/

Performance Evaluation and Benchmarking for
Intelligent Robots and Systems with Cognitive and Autonomy Capabilities

- 'Look Ma, No Hands' syndrome?
- Replication of experiments
- Performance measure benchmarks to allow results comparison
- Needed to foster research advancement and enable practical application of research achievements

Both replication and benchmarking are needed to foster a cumulative advancement of our knowledge of intelligent physical agents and even to correctly appreciate disruptive innovation in the science (?) and technology of robots.

Should we take inspiration from biology and medicine ?

If robotics aims to be serious science, serious attention must be paid to experimental method.

What is an 'experiment' in robotics?

Replication&Falsification

There are different modulation of this concept, but whether we think we are in a cumulative phase in the development of a scientific field or in presence of a 'disruptive' creative paradigm shift, as somebody is claiming in nowadays robotics, a kind of widely accepted experimental methodlogy is needed in order to be able to ground the advancement of research on a shared quantitative language.

Replication&Falsification

It seems clear that in robotics the experimental methodology standards are currently in many cases weaker, and the syndrome 'it worked once, in my lab' could be more widespread than we may think.

Replication&Falsification

A limit to replication is given by the huge variability of robot machines.
Perhaps, following the biomedical analogy, we have to compare behaviors and performances of different 'animals'.

What about similar issues in Biology?

The definition of what should be considered a 'law of nature' in biology raises a number of issues. For reasons not very different from those raised from robotics research. The laws are usually not universal but apply to specific species: the Mendel laws apply to species with sexual reproduction, but not to all living species.
Almost every theoretical enunciate refer to a species or a set of species and has stochastic characteristics.

What about similar issues in Biology?

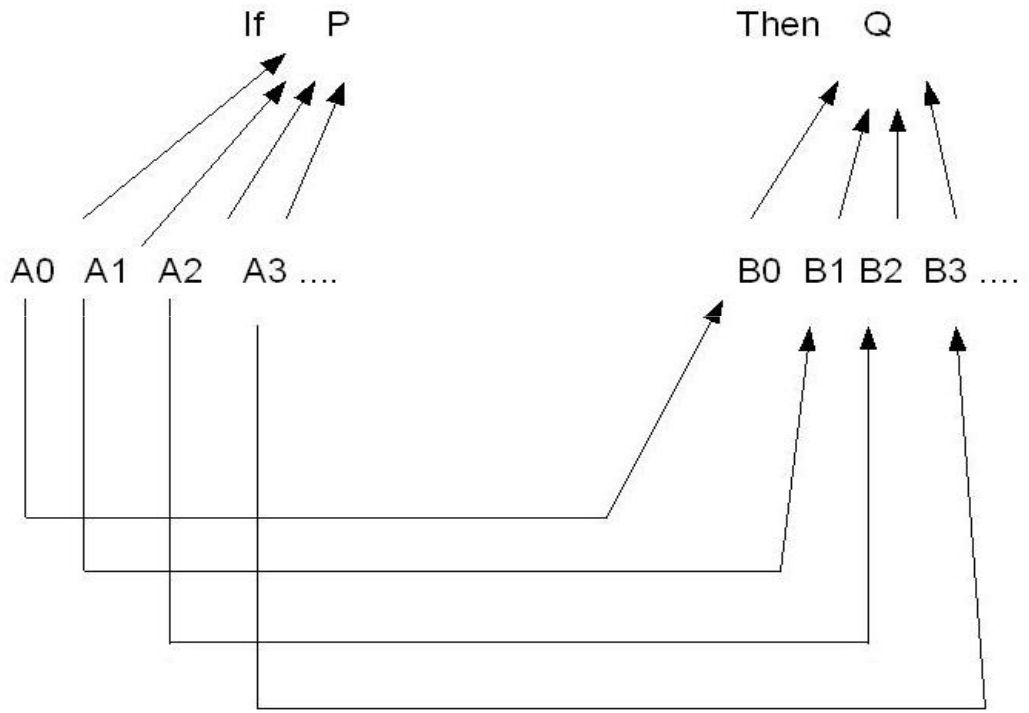
Systems are usually very complex, involve a huge numbers of variable and work in open ended stochastic environments. The same function, for example flight, can be performed in many different ways. The wing morphology and dynamics of a fly are quite different from those of a bird. On an other end, the wing of a penguin are used to stabilize swimming.

An interesting point is that the laws regarding a specific function in a species become true at a specific time, as a new function evolve, as depicted afterwards., and only if some initial conditions occur.

I	→	If P	then Q
t_0		t_1	t_2

Time dependence of biological 'laws'

Performance Evaluation and Benchmarking for
Intelligent Robots and Systems with Cognitive and Autonomy Capabilities



'Causality at different levels'.

Performance Evaluation and Benchmarking for
Intelligent Robots and Systems with Cognitive and Autonomy Capabilities

L1, L2, L3, ..., Ln covering laws

explanans

C1, C2, ..., Cn initial conditions

E

explanandum

Hempel-Oppenheim Schema

Visual servoing example

Visual servoing control the movement of robot (video assisted mobile robots or manipulators) on the basis of feedback coming from a video device, like a video camera.

This example is relevant because formal proof are very difficult if not impossible in many if not most cases, as a consequence experimental work is necessary to assess the potential of different approaches to control.

Visual servoing example
Assumptions

For a visual servoing systems there typical which must be detailed. A non exhaustive list is given here:

- the visual features
- scene 3D model
- the kinematics model of the robot.
- dynamics model of the robot.

Visual servoing example
Assumptions

Plus the list related to image processing:

- background characteristics (homogeneous or if not color and luminance distributions)
- lighting conditions
- robustness to outliers in feature detection
- others inherent to real life experimentation.

Visual servoing example
Performance criteria

Generally speaking these criteria measure the convergence of the system to a predefined goal.

Non exhaustive list:

- the time of convergence
- the trajectories of the visual features in the image plane
- the 3D trajectory of the robot computation time
- positioning error after convergence.

A special attention must be paid to stability and robustness against image noise, the errors in the models (object, camera, robot), and the control parameters.

Visual servoing example
Measured characteristics

An unequivocal procedure to derive the quantitative aspects of the system must be given. For example visual features can be directly obtained from the video camera.

For manipulators what is directly measurable are the generalized joint angles while the end effector 3D trajectory must be estimated by the (direct) kinematic model.

Calibration procedures for the robot relevant characteristics and camera must be described.

In experiments the visual features (at least) must be variated and the variation policy documented.

Visual servoing example Implementation Information

The information given above don't allow by themselves the replication of results.

There more data needs than in other kind of papers:

- Visual servoing system configuration environment (either real or simulation) should be described in detail: in-hand vs. external camera, etc.
- model and control parameters
- ground truth for robot positioning and the environment

Visual servoing example Implementation Information

- Technical specification of the hardware platform
- Technical specifications of the camera (model, frame rate, resolution, etc.).
- Computer specifications (at least, processor and amount of memory, o.s., relevant configuration details)
- SW libraries (they should be available at least as linkable components) list and configuration

Probably the adoption of widely known sw libraries like ViSP, VXL, OpenCV may ease replication.

Visual servoing example
Parameter and variable distribution

Statistical distributions of all relevant parameter must be given (as in an open ended stochastic environment results will have a probabilistic formulation). This is by the way quite common in clinical research (as noticed before)

Visual servoing example
Parameter and variable distribution

The list of findings in the discussion/conclusion section should be against a detailed list of criteria within a detailed list of conditions as recalled above.

For example better convergence speed, robustness /weakness against certain parameters, behavior with respect to current technology visual servoing systems:

- visual features moving of the field of view
- workspace and singularity issues

The findings listed in a paper might be negative: the given algorithm in our test conditions fail under the listed set of conditions with respect to the listed series of criteria.

Discussion

Why we need both replication AND benchmarking?

FACT: Benchmarking is more studied than Replication

- SLAM
- Mobile Robots' Motion Control
- Robot Obstacle Avoidance
- Grasping
  - Visual Servoing
  - Autonomy/Cognitive tasks

Discussion

Replication of robotics experiments

- Research Reporting in Biology and Medicine
- Evidence Based Medicine
- Early activities in robotics research replication

Discussion

The bare replication of experiments and the quantitative comparison of research results in robotics raise many challenging issues.
This is due to the variety of applications, tasks, mechanical structures, sensor sets, actuators, control system, software architectures, required levels of flexibility and autonomy, and so on.

Discussion

It is thought that in both these situations the epistemological model based on 'context' discussed above for biology and extended to robotics may provide a working framework.

A new kind of papers?

We may think of theoretical/concept papers, proof of concept papers, and experimental papers , as we have started to define here, as steps in a research idea 'life-cycle'. We believe that more paper of the 'experimental' kind would greatly help the research activities in robotics and the industrial exploitation of the results.

A new kind of papers?

- ‘description’ : a journal paper text+figures+ multimedia ….according to GEM Guidelines (or similar)

- Data sets (similar to IJRR ‘Data paper’

- Complete ‘code’ identifiers and or downloadable code (executables may be enough)

- ‘HW’ description or HW identifier (if it is identifiable)

…

2010 IEEE/RSJ IROS
Taipei, Taiwan
October 22, 2010

Thank you!

Performance Evaluation and Benchmarking for
Intelligent Robots and Systems with Cognitive and Autonomy Capabilities

2010 IEEE/RSJ IROS
Taipei, Taiwan
October 22, 2010

Performance Evaluation and Benchmarking for Intelligent Robots and Systems with Cognitive and Autonomy Capabilities

2010 IEEE/RSJ IROS
Taipei, Taiwan
October 22, 2010

Performance Evaluation and Benchmarking for Intelligent Robots and Systems with Cognitive and Autonomy Capabilities