



FP6-507728

EURON

European Robotics Network

Network of Excellence
Information Society Technologies Priority

DR 2.5
Second EURON Workshop on Benchmarks

Due date of deliverable: 30 April 2007

Actual submission date: 4 June 2007

Start date of the project: 1 May 2004

Duration: 48 months

Organisation name of lead contractor for this deliverable: UJI

Revision: Revised version, June 2007

Dissemination level: PU

Report on

2007 EURON Workshop on

Experimental Practices and Benchmarking

Rationale

On March 28th, 2007 this workshop was held in the context of the 2007 EURON Annual Meeting in Chania (Greece). The workshop was co-organized by Angel P. del Pobil from Universitat Jaume I in Spain, John Hallam from the University of Southern Denmark, and Fabio Bonsignorio from Heron Robots (Italy).

After some preliminary discussions it was decided to focus the topic of the event on *Experimental Practices and Benchmarking*. The rationale was that a good definition for a benchmark is that it is an experiment with:

- an agreed interesting problem
- an agreed standard procedure, including agreement on:
 - choice of parameters
 - what to measure
 - how to measure it
 - how to analyse the results
- an agreed index of performance

Consequently, a good way to promote a culture for benchmarking in the research community is by promoting good experimental methodologies, since it was felt by many that a common body of knowledge concerning experimental methodology was missing in the robotics research community, i.e., a corpus of specifications on how to design and conduct good experiments in robotics. This fact is reflected in the frequent publication of scientific articles that report poor experimental work, few or no replications or tests, dubious comparison between algorithms, lack of suitable quantification of performance and its variability, and conclusions which, while may be correct, are unjustified by the reported experimental work. The goal of the workshop was to address this situation.

Program

The workshop was preceded by a plenary session on the same topic that served as an introduction. Since it was felt that the topic was not still mature enough, no formal presentations were held at the workshop, but instead it was arranged in a brainstorming style in which the main questions proposed in the plenary session were discussed in an informal atmosphere. The structure of the discussions was as follows:

- What makes a problem hard?
 - What are the key properties?
 - What are similar problems?
 - How is it hard?
 - Hard compared to what?
 - How does it scale?
 - How can 'hard' be quantified?

- Requirements for a proposed solution
 - Give a Complete Method Description
 - Identify Key Assumptions
 - Identify key Parameters

- Requirements for a demonstration
 - Can I reproduce this?
 - Evaluate not Demonstrate
 - Measures of performance
 - Comparison with baseline methods
 - Maybe you were lucky?
 - Maybe you were unlucky?
 - How does it fail?

- Generality of proposed solutions
 - Replicability: 'It worked in your lab. at least once under demo conditions'
 - How well does it work (wrt other techniques)?
 - What are its limits?
 - What are its failure modes?
 - How does it scale?
 - 'Will it work for me on my problem?'
 - What precisely may we conclude from the result?

- Actions to be taken
 - Education:
 - Courses in Methodology
 - Summer Schools / Workshops
 - Web materials on Methodology
 - Reviewing guidelines
 - Educate our own students

Actions

The main planned action was to propose a *Special Interest Group* with the focus of increasing the quality of experimental methodology practiced in robotics by sharing good practice via educational workshops, summer schools, email discussion and web presentation; by providing assistance to journal and conference reviewers and editors concerning what constitutes experimental robotics and good practice in that sub-discipline; by encouraging the principled replication and comparison of results; and by encouraging the development and use of appropriate systems benchmarks and standard evaluation procedures.

Concretely, the following actions were considered:

- Make a document enumerating recommended quality criteria for Robotics Journal/Conference reviews, including a one page quickguide summary, that the community expects a proper highquality experimental robotics paper have to satisfy. This could be formulated as a reviewer checklist and circulated to Journal Editors and Conference Programme Committees.
- Develop a Web site as part of the EURON web facility that includes information on planned meetings and activities, educational material on good experimental practices, recommendations and outcomes from discussions, and a compendium of other efforts to promote good experimentation and quantitative methods in robotics, both in the EU and internationally.
- Organize/support workshops on good experimental methodology and Benchmarking at major robotics conferences
- Develop initiatives (meetings and information interchange) to collaborate with other experimental methodology, benchmarking and standards efforts (e.g. RoSta, PerMIS, RawSeeds).
- Invited sessions: to present and disseminate results at major robotics, control, communications, and computer science conferences. Including recognized domain experts.
- A repository for tools and data sharing and a workflow environment for collaborative work over the Internet (through restricted access/public blogs)
- Demo cases illustrating the importance of the issues targeted; community-endorsed examples of best practice.
- Support/start the organisation of summer/winter school(s) on experimental methods for robotics (high quality teaching on experimental design, methods, reporting etc.

with examples of best practice and poor technique; preferably hands-on training as well as lecture material, and instruction in basic statistical techniques.)

- Encourage and facilitate the publication of replications of robotics results, either in existing robotics journals or by establishing a high quality open access web-based journal that specifically encourages the publication of replications of published experimental results, rational reconstructions of systems, and similar.