Benchmarking Through Competitions

Matteo Matteucci
Politecnico di Milano

Overview

- Competitions - Benchmarking - Experiments

- On the Benchmarking of Modules and Systems

- The RoCKIn Project
  - Benchmarking through competitions
  - Set up scientific robot competitions

Competitions & Benchmarking (1)

- Robotic competitions have positive effects …

  - They are appealing (people like to compete)

  - They take place with regularity and precise timing

  - They are showcases of current state of the art in research / industry

  - They switch the focus from specific subsystems towards complete systems and highlight the influence of integration

  - They promote critical analysis of experiments taking them out of labs

  - They share among participants the cost and effort of setting up complex experimental installations among many participants

  - …

Competitions & Benchmarking (2)

- … but competitions often lack of scientific grounding
  - They do not apply the so called "scientific method"
    - Comparison, Reproducibility/ repeatability, and Justification/explanation
  - They produce a ranking, but few insights on the motivations for this ranking
  - Their results cannot be used as benchmarking tools

The Benchmarking through Competition challenge:

"Designing competitions so to make them more scientifically grounded and suitable as benchmarks"

RoCKIn

Competitions as Experiments (1)

- Can competitions be treated as scientific experiments (despite the obvious differences)?

  "Challenge and competition events in robotics provide an excellent vehicle for advancing the state of the art and evaluating new algorithms and techniques in the context of a common problem domain. [...] treat competitions and challenges as repeatable experiments." (Anderson et al. - IEEE Robotics & Automation Magazine, 2011)

- Competitions should aim at providing benchmarks by adopting a scientific approach (both in goals and methods)

  "Scientific" means able to increase science and technology related knowledge by using rigorously experimental method

Competitions as Experiments (2)

- Reproducibility and repeatability should be guaranteed

    - ***Reproducibility*** is the possibility to verify, in an independent way, the results of a given experiment. It refers to the fact that other experimenters, different from the one claiming for the validity of some results, are able to achieve the same results, by starting from the same initial conditions, using the same type of instruments, and adopting the same experimental techniques.

    - ***Repeatability*** concerns the fact that a single result is not sufficient to ensure the success of an experiment. A successful experiment must be the outcome of a number of trials, performed at different times and in different places.

Benchmarking Modules & Systems (1)

- Competitions can challenge robots at two different levels (*ability* vs *capability* in SRA jargon?)
  - **Task Level**: evaluation of whole systems on a specific task (e.g., the "bring me a beer" tasks)
  - **Functionality Level**: evaluation of modules implementing, in a general manner, capabilities (e.g., grasping and manipulation)

- Benchmarking competitions allow independent evaluation at both levels
  - To encourage participation of people interested in specific aspects of robotics (e.g., object recognition)
  - To evaluate at what extent the Interplay among modules is relevant (e.g., the precision in positioning before grasping)
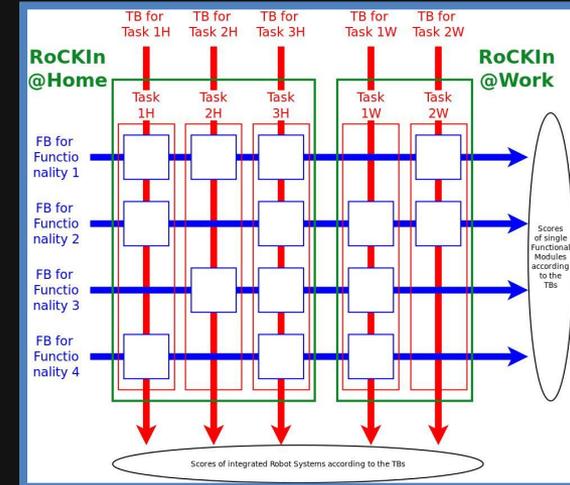
Proposal for RoCKIn Competitions (1)
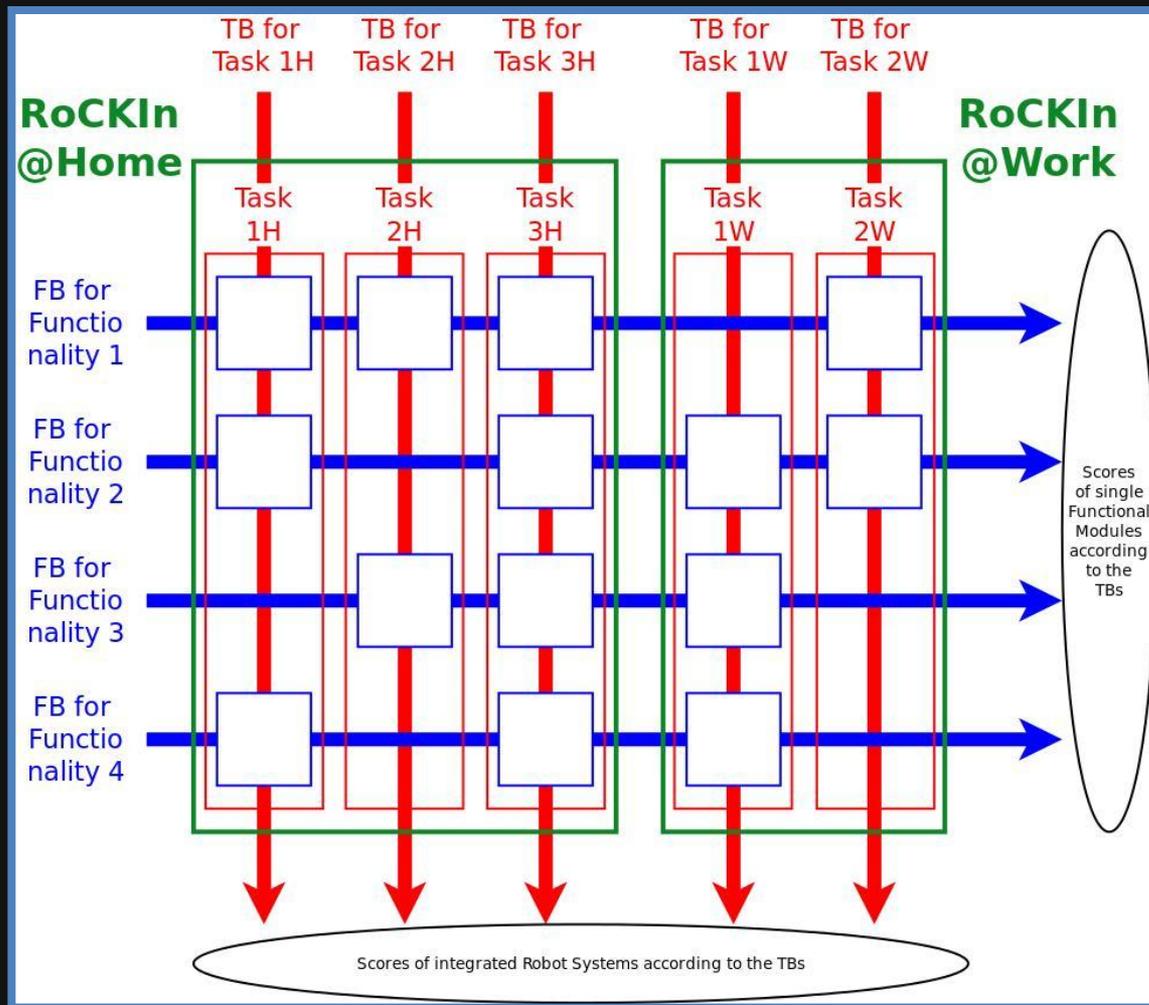
A RoCKIn "Fake" Example (1)

Functionalities 1 to 4 (out of those from WP1):

1. Autonomous navigation;
2. Object recognition;
3. Grasping and manipulation;
4. Processing of voice commands.

Task Benchmark 2H: "The Robot System is provided with a map of the environment. It must enter the Testbed, navigate through it to reach an object located in a predefined position, and pick it up".

Task Benchmark 1W: "The Robot System is located in a specified pose in front of a table. Over the table are located randomly (but according to suitable specifications) 5 identical mugs which differ only in their color. The Robot System must receive a voice command from a human, specifying the color of the mug to pick up, then pick up the required mug".

A RoCKIn "Fake" Example (2)

Functionalities 1 to 4 (out of those from WP1):

1. Autonomous navigation;
2. Object recognition;
3. Grasping and manipulation;
4. Processing of voice commands.
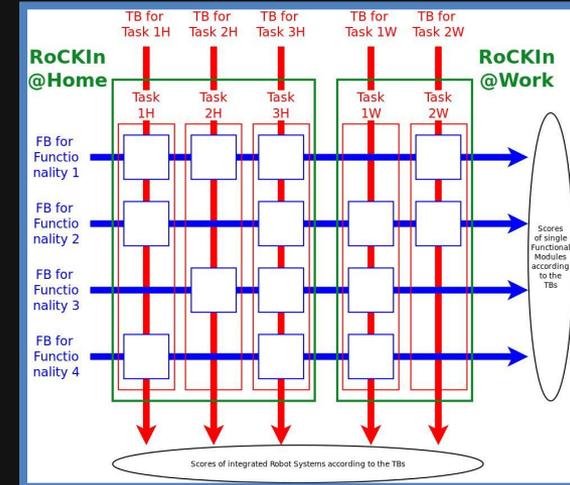
Functional Benchmark 2: "Recognize 10 objects, randomly selected out of all possible objects from RoCKIn@Home and RoCKIn@Work databases. Category, size , position, and color have to be returned."

Functional Benchmark 3: "Grasp and lift firmly10 different objects, randomly selected out of all predefined objects from RoCKIn@Home and RoCKIn@Work, in a given working space. The pose of each object is sent to the robot at the beginning of the test."

Proposal for RoCKIn Competitions (2)

- Design the competitions to
  - Allow people interested in specific functional benchmark to participate only to those
  - Stimulate people to tackle to tackle both functional/module and task/system benchmarks
  - Promote participation in both RoCKIn@Home and RoCKIn@Work benchmarks

- Both functionality and task scores should be considered for the challenge ranking(s)
  - As the final outcome of each RoCKIn competition we propose:
    - RoCKIn@Home Winner (or Top 3 teams)
    - RoCKIn@Work Winner (or Top 3 teams)
  - Additional awards (open also for specific modules/systems)
    - Functionality award for each functionality

Openings for Discussion

- Benchmarking through Competitions in robotics is a challenge

- A proposal has been made to tackle both levels of benchmarking (Module, Functionality)

  – Questions? Feedback?

- RoCKIn has just started: take the opportunity to participate in shaping the way RoCKIn benchmarking competitions will be

  – Comments? Suggestions?

Benchmarking Through Competitions

Matteo Matteucci
Politecnico di Milano