

RoCKIn and the European Robotics League: Building on RoboCup Best Practices to Promote Robot Competitions in Europe

Pedro U. Lima¹(✉), Daniele Nardi², Gerhard K. Kraetzschmar³,
Rainer Bischoff⁴, and Matteo Matteucci⁵

¹ Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
`pedro.lima@tecnico.ulisboa.pt`

² Università di Roma La Sapienza, Rome, Italy
`nardi@dis.uniroma1.it`

³ Bonn-Rhein-Sieg University, Sankt Augustin, Germany
`gerhard.kraetzschmar@h-brs.de`

⁴ KUKA Roboter GmbH, Augsburg, Germany
`Rainer.Bischoff@kuka.com`

⁵ Politecnico di Milano, Milano, Italy
`matteo.matteucci@polimi.it`

Abstract. This paper describes activities that promote robot competitions in Europe, using and expanding RoboCup concepts and best practices, through two projects funded by the European Commission under its FP7 and Horizon2020 programmes. The RoCKIn project ended in December 2015 and its goal was to speed up the progress towards smarter robots through scientific competitions. Two challenges have been selected for the competitions due to their high relevance and impact on European societal and industrial needs: domestic service robots (RoCKIn@Home) and innovative robot applications in industry (RoCKIn@Work). RoCKIn extended the corresponding RoboCup leagues by introducing new and prevailing research topics, such as networking mobile robots with sensors and actuators spread over the environment, in addition to specifying objective scoring and benchmark criteria and methods to assess progress. The European Robotics League (ERL) started recently and includes indoor competitions related to domestic and industrial robots, extending RoCKIn's rulebooks. Teams participating in the ERL must compete in at least two tournaments per year, which can take place either in a certified test bed (i.e., based on the rulebooks) located in a European laboratory, or as part of a major robot competition event. The scores accumulated by the teams in their best two participations are used to rank them over an year.

Keywords: Robot competitions · Benchmarking · Domestic robots · Industrial robots

1 Introduction

In 2012, under its 7th Framework Programme (FP7), the European Commission (EC) launched a first Call for Coordination Actions to foster research in robotics and benchmarking through robot competitions. A consortium composed of six partners (five of which represented by the authors of this abstract) applied and got funding for a three-year project that ended successfully in December 2015.

The goal of RoCKIn¹ (Robot Competitions Kick Innovation in Cognitive Systems and Robotics) was to speed up the progress towards smarter robots through scientific competitions. Two challenges have been selected for the competitions due to their high relevance and impact on Europe's societal and industrial needs: domestic service robots (RoCKIn@Home) and innovative robot applications in industry (RoCKIn@Work). As it is clear from the designations, both challenges were inspired by activities in the RoboCup community, but RoCKIn extended them by introducing new and prevailing research topics, such as networking mobile robots with sensors and actuators spread over the environment, in addition to specifying objective scoring and benchmark criteria and methods to assess progress. RoCKIn goal was to bring back to RoboCup some of these newly introduced aspects, and this is currently happening through negotiations with the RoboCup@Home and RoboCup@Work technical committees. Moreover, at least three new RoCKIn teams that had not participated in RoboCup@Work and RoboCup@Home before applied and qualified to RoboCup 2016.

The RoCKIn project has taken the lead in boosting scientific robot competitions in Europe by (i) specifying and designing open domain test beds for competitions targeting the two challenges and usable by researchers worldwide; (ii) developing methods for scoring and benchmarking through competitions that allow to assess both particular subsystems as well as the integrated system; and (iii) organizing camps whose main objective was to build up a community of new teams interested to participate in robot competitions.

Within the project lifetime, two competition events took place, each of them based on the two challenges and their respective test beds: RoCKIn Competition 2014 (in Toulouse, France) and RoCKIn Competition 2015 (in Lisbon, Portugal), with the participation of 12 teams and more than 100 participants. Three camps were also organized, in 2013 (Eindhoven, together with RoboCup 2013), 2014 (Rome) and 2015 (Peccioli, Italy, at the ECHORD++ Robotics Innovation Facility).

A significant number of dissemination activities on the relevance of robot competitions were carried out to promote research and education in the field, targeting the research community, in industry and academia, as well as the general public. The potential future impact of the benchmarking methods developed on robotics research was recognized by many researchers worldwide, and an article on the topic was published in the Special Issue on Replicable and Measurable Robotics Research of the IEEE Robotics and Automation Magazine in 2015 [1].

¹ <http://rockinrobotchallenge.eu>.

RoCKIn’s success was a consequence of the experience accumulated over the years in RoboCup by the consortium members as organizers and participants of the respective RoboCup competitions. The lessons learned during RoCKIn paved the way for a step forward in the organization and research impact of robot competitions. The authors expect many of these to be brought back to RoboCup, improving the quality and impact of the @Home and @Work leagues.

A continuation of RoCKIn is already under way, based on the European Robotics League (ERL) concept, again funded by the EC under the new Horizon2020 (H2020) programme. Teams participating in the ERL must compete in at least two tournaments per year, which can take place either in a certified test bed (i.e. based on ERL/RoCKIn’s rulebooks) located in a European research laboratory, or as part of a major robot competition event. The scores accumulated by the teams in the best two participations are used to rank them, leading to the awarding of prizes during the annual European Robotics Forum event. This brings extra visibility to robot competitions and to RoboCup in Europe, since we plan to hold some of the major tournaments as part of RoboCup world or regional events.

This paper highlights the main contributions of RoCKIn to benchmarking robotics research through robot competitions: task and functionality benchmarks, as well as the corresponding benchmarking methods, are introduced in Sect. 2. Scoring methods and metrics are described in Sect. 3, rulebooks, test beds and datasets in Sect. 4. Camps and competition events major outcomes are referred in Sect. 5. Section 6 concludes the paper with an outlook to the new concept for robot competitions currently underway in Europe: the European Robotics League.

2 Task and Functionality Benchmarks

RoCKIn’s approach to benchmarking experiments is based on the definition of two separate, but interconnected, types of benchmarks [2]:

- Functionality Benchmarks, which evaluate the performance of hardware and software modules dedicated to single, specific functionalities in the context of experiments focused on such functionalities.
- Task Benchmarks, which assess the performance of integrated robot systems facing complex tasks that usually require the interaction of different functionalities.

Functionality Benchmarks are certainly the closest to a scientific experiment from among the two. This is due to their much more controlled setting and execution. On the other side, these specific aspects of Functionality Benchmarks limit their capability to capture all the important aspects of the overall robot performance in a systemic way. More specifically, emerging system-level properties, such as the quality of integration between modules, cannot be assessed with Functionality Benchmarks alone. For this reason, the RoCKIn Competitions integrate them with Task Benchmarks.

In particular, evaluating only the performance of integrated system is interesting for the application, but it does not allow to evaluate the single modules that are contributing to the global performance, nor to put in evidence the aspects needed to push their development forward. On the other side, the good performance of a module does not necessarily mean that it will perform well in the integrated system. For this reason, the RoCKIn benchmarking competitions target both aspects, and enable a deeper analysis of a robot system by combining system-level and module-level benchmarking.

System-level and module-level tests do not investigate the same properties of a robot. Module-level testing has the benefit of focusing only on the specific functionality that a module is devoted to, removing interferences due to the performance of other modules which are intrinsically connected at the system level. For instance, if the grasping performance of a mobile manipulator is tested by having it autonomously navigate to the grasping position, visually identify the item to be picked up, and finally grasp it, the effectiveness of the grasping functionality is affected by the actual position where the navigation module stopped the robot, and by the precision of the vision module in retrieving the pose and shape of the item. On the other side, if the grasping benchmark is executed by placing the robot in a predefined known position and by feeding it with precise information about the item to be picked up, the final result will be almost exclusively due to the performance of the grasping module itself. The first benchmark can be considered as a system-level benchmark, because it involves more than one functionality of the robot, and thus has limited worth as a benchmark of the grasping functionality. On the contrary, the latter test can assess the performance of the grasping module with minimal interference from other modules and a high repeatability: it can be classified as module-level benchmark.

Let us consider an imaginary, simplified RoCKIn Competition including five tasks (T_1, T_2, \dots, T_5). Figure 1 describes such imaginary competition as a matrix, showing the tasks as columns while the rows correspond to the functionalities required to successfully execute the tasks. For the execution of the whole set of tasks of this imaginary RoCKIn Competition, four different functionalities (F_1, \dots, F_4) are required; however, a single task usually requires only a subset of these functionalities. In Fig. 1, task T_X requires functionality F_Y if a black dot is present at the crossing between column x and row y . For instance, task T_2 does not require functionalities F_2 and F_4 , while task T_4 does not require functionality F_1 .

The availability of both task and functionality rankings opens the way for the quantitative analysis of the importance of single functionalities in performing complex tasks. This is an innovative aspect triggered by the RoCKIn approach to competitions. To state the importance of a functionality in performing a given task, RoCKIn borrows the concept of Shapley value from Game theory [5]. Let us assume that a coalition of players (functionalities in the RoCKIn context) cooperates, and obtains a certain overall gain from that cooperation (the Task Benchmark scoring in the RoCKIn context). Since some players may contribute

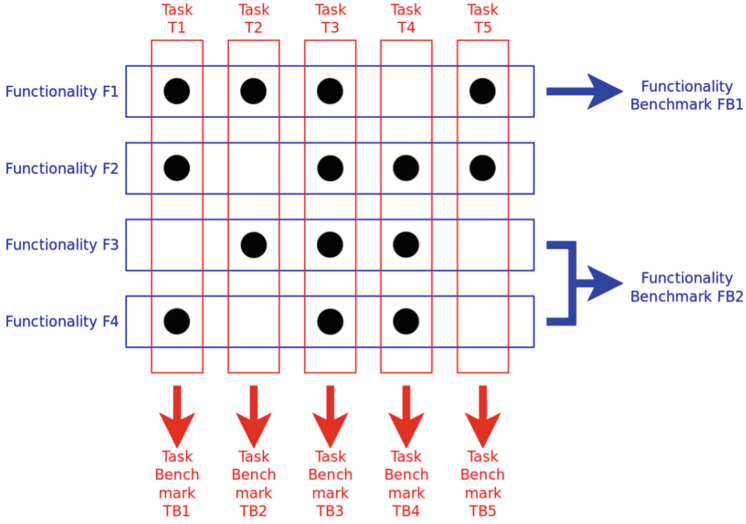


Fig. 1. Functionalities vs Tasks Matrix.

more to the coalition than others or may possess different bargaining power (for example threatening to destroy the whole surplus), what final distribution of generated surplus among the players should arise in any particular game? Or phrased differently: how important is each player to the overall cooperation, and what payoff can (s)he reasonably expect? Or in the RoCKIn jargon: how important is each functionality to the reach a given performance in a Task Benchmark?

Assuming that all scores are expressed according to the same scale, the Shapley values of the single functionalities can be calculated as:

$$\Phi_i = \frac{1}{n!} \sum_{\pi} [\nu(C_{\pi}(i) \cup i) - \nu(C_{\pi}(i))] \tag{1}$$

where i is a functionality, n is the total number of functionalities, π is a permutation of the n Functionality Benchmark scores, $C_{\pi}(i)$ is the set of functionalities that precede i in the permutation π , and $\nu()$ is the score of the set of functionalities specified as argument. Examples of the application of Shapley values to task benchmarking can be found in RoCKIn deliverable D1.2 [2].

3 Scoring Methods and Metrics

The scoring framework for the evaluation of the task performance in RoCKIn competitions is the same for all tasks of RoCKIn@Home and RoCKIn@Work, and it is based on the concept of performance classes used for the ranking of robot performance in a specific task.

The performance class that a robot is assigned to is determined by the number of achievements (or goals) that the robot reaches during its execution of the

task. Within each class (i.e., a performance equivalence class), ranking is defined according to the number of penalties assigned to the robot. These are assigned to robots that, in the process of executing the assigned task, make one or more of the errors defined by a task-specific list associated to the Task Benchmark. More formally:

- The ranking of any robot belonging to performance class N is considered better than the performance of any robot belonging to performance class M when $M < N$. Class 0 is the lowest performance class.
- Among robots belonging to the same performance class, a penalization criterion is used to define ranking: the robot which received less penalties is ranked higher.
- Among robots belonging to the same class and with the same number of penalties, the ranking of the one which accomplished the task in a shorter time is considered the highest (unless specific constraints on execution time are given as achievements or penalties).

Performance classes and penalties for a Task Benchmark are indeed task-specific, but they are grouped according to the following three sets (of which here we define the semantics; the actual content is specific to each Benchmark):

- set DB = disqualifying behaviors, i.e. things that the robot must not do;
- set A = achievements (also called goals), i.e., things that the robot should do;
- set PB = penalizing behaviors, i.e., things that the robot should not do.

Once the content of each of the previous sets is provided as part of the specifications of the relevant Task Benchmark, the following 3-step sorting algorithm is used to apply the RoCKIn scoring framework:

1. if one or more of the disqualifying behaviors of set DB occur during task execution, the robot gets disqualified (i.e., assigned to class 0, the lowest possible performance class), and no further scoring procedures are performed for it;
2. the robot is assigned to performance class X , where X corresponds to the number of achievements of set A which have been accomplished by the robot;
3. a penalization is assigned to the robot for each behavior of the robot belonging to set PB that occurs during the execution of the task.

One key property of this scoring system is that a robot that executes the required task completely will always be placed into a higher performance class than a robot that executes the task partially. In fact, penalties do not change the performance class assigned to a robot and only influence intra-class ranking.

It is not possible to define a single scoring framework for all Functionality Benchmarks as it has been done for Task Benchmarks in the previous chapter. These, in fact, are specialized benchmarks, tightly focused on a single functionality, assessing how it operates and not (or not only) the final result of

its operation. As a consequence, scoring mechanisms for Functionality Benchmarks cannot ignore how the functionality operates, and metrics are strictly connected to the features of the functionality. For this reason, differently from what has been done for Task Benchmarks scoring methodologies and metrics are defined separately for each Functionality Benchmark of a Competition. In RoCKIn, Functionality Benchmarks are defined by four elements:

- Description: a high level, general, description of the functionality.
- Input/Output: the information available to the module implementing the functionality when executed, and the expected outcome.
- Benchmarking data: the data needed to perform the evaluation of the performance of the functional module.
- Metrics: algorithms to process benchmarking data in an objective way.

RoCKIn Deliverable D1.2 [2] provides more details and examples on scoring and ranking team performance in task and functionalities, as well as methods to combine task rankings to determine the competition winner.

4 Rulebooks, Test Beds and Datasets

The RoCKIn@Home test bed (see Fig. 2) consists of the environment in which the competitions took place, including all the objects and artefacts in the environment, and the equipment brought into the environment for benchmarking purposes. An aspect that is comparatively new in robot competitions is that RoCKIn@Home is, to the best of our knowledge, the first open competition targeting an environment with ambient intelligence, i.e. the environment is equipped with networked electronic devices (lamps, motorised blinds, IP cams) the robot can communicate and interact with, and which allow the robot to exert control on certain environment artefacts.

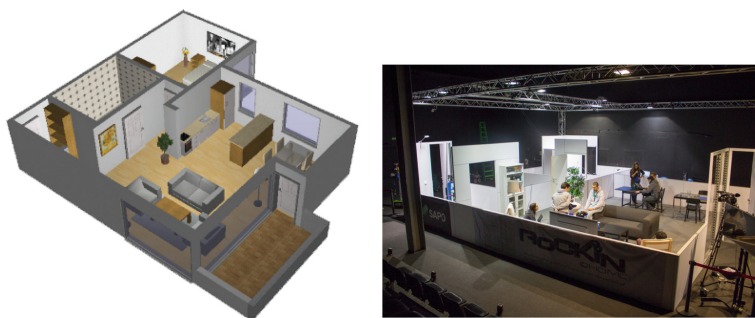


Fig. 2. RoCKIn@Home test bed: left - 3D layout; right - real setup

The RoCKIn@Home rulebook [3] specifies in detail:

- The environment structure and properties (e.g., spatial arrangement, dimensions, walls).

- Task-relevant objects in the environment, split in three classes:
 - Navigation-relevant objects: objects which have extent in physical space and do (or may) intersect (in 3D) with the robots navigation space, and which must be avoided by the robots.
 - Manipulation-relevant objects: objects that the robot may have manipulative interactions (e.g., touching, grasping, lifting, holding, pushing, pulling) with.
 - Perception-relevant objects: objects that the robot must only be able to perceive (in the sense of detecting the object by classifying it into a class, e.g., a can; recognizing the object as a particular instance of that class, e.g., a 7UP can; and localizing the object pose in a pre-determined environment reference frame.)

During the benchmark runs executed in the test bed, a human referee enforces the rules. This referee must have a way to transmit his decisions to the robot, and receive some progress information. To achieve this in a practical way, an assistant referee is seated at a computer and communicates verbally with the main referee. The assistant referee uses the main Referee Scoring and Benchmarking Box (RSBB). Besides basic starting and stopping functionality, the RSBB is also designed to receive scoring input and provide fine grained benchmark control for functionality benchmarks that require so.

The RoCKIn@Work test bed (Fig. 3) consists of the environment in which the competitions took place (the RoCKIn’N’RoLLIn medium-sized factory, specialized in production of small- to medium-sized lots of mechanical parts and assembled mechatronic products, integrating incoming shipments of damaged or unwanted products and raw material in its production line), including all the objects and artefacts in the environment, and the equipment brought into the environment for benchmarking purposes. An aspect that is comparatively new in robot competitions is that RoCKIn@Work is, to the best of our knowledge, the first industry-oriented robot competition targeting an environment with ambient intelligence, i.e. the environment is equipped with networked electronic devices (e.g., a drilling machine, a conveyor belt, a force-fitting machine, a quality control camera) the robot can communicate and interact with, and which allow the robot to exert control on certain environment artefacts like conveyor belts or machines.

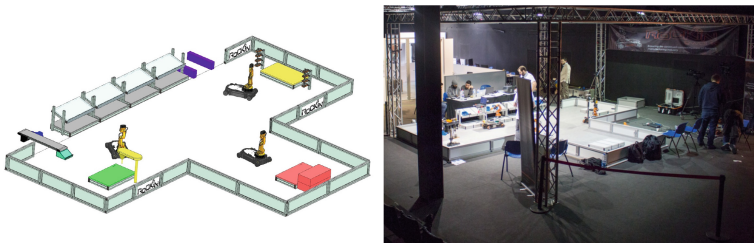


Fig. 3. RoCKIn@Work test bed: left - 3D layout; right - real setup.

The RoCKIn@Work rulebook [4] specifies in detail:

- The environment structure and properties (e.g., spatial arrangement, dimensions, walls).
- Typical factory objects in the environment to manipulate and to recognize.

The main idea of the RoCKIn@Work test bed software infrastructure is to have a central server-like hub (the RoCKIn@Work Central Factory Hub or CFH) that serves all the services that are needed for executing and scoring tasks and successfully realize the competition. This hub is derived from software systems well known in industrial business (e.g., SAP). It provides the robots with information regarding the specific tasks and tracks the production process as well as stock and logistics information of the RoCKIn’N’RoLLIn factory. It is a plug-in driven software system. Each plug-in is responsible for a specific task, functionality or other benchmarking module.

Both RoCKIn test beds include benchmarking equipment. RoCKIn benchmarking is based on the processing of data collected in two ways:

- internal benchmarking data, collected by the robot system under test;
- external benchmarking data, collected by the equipment embedded into the test bed.

External benchmarking data is generated by the RoCKIn test bed with a multitude of methods, depending on their nature. One of the types of external benchmarking data used by RoCKIn are pose data about robots and/or their constituent parts. To acquire these, RoCKIn uses a camera-based commercial motion capture system (MCS), composed of dedicated hardware and software. Benchmarking data has the form of a time series of poses of rigid elements of the robot (such as the base or the wrist). Once generated by the MCS system, pose data are acquired and logged by a customized external software system based on ROS (Robot Operating System): more precisely, logged data is saved as bagfiles created with the `rosviz` utility provided by ROS. Pose data is especially significant because it is used for multiple benchmarks. There are other types of external benchmarking data that RoCKIn acquires; however, these are usually collected using devices that are specific to the benchmark. Finally, equipment to collect external benchmarking data includes any server which is part of the test bed and that the robot subjected to a benchmark has to access as part of the benchmark. Communication between servers and robot is performed via the test bed’s own wireless network.

During RoCKIn competitions and events, several datasets have been collected to be redistributed to the Robotics community for further analysis and understanding about the Task level and Functional level performance of robotics systems. In particular, data from the Object Perception (@Home and @Work) and Speech Understanding Functional Benchmarks was collected during RoCKIn Competition 2014 and RoCKIn Field Exercise 2015. The datasets are available and will continue to be updated in the RoCKIn wiki².

² <http://thewiki.rockinrobotchallenge.eu/index.php?title=Datasets>.

RoCKIn Deliverables D2.1.3 [3] and D2.1.6 [4] provide the full rulebooks for the two Challenges, including details of the RSBB and CFH referee boxes and pointers to the deliverables where details of the MCS and benchmarking system are available.

5 RoCKIn Camps and Competitions

Within the project lifetime, two competition events took place, each of them based on the two challenges and their respective test beds:

- RoCKIn 2014, in La Cité de L’Espace, Toulouse, 24–30 November 2014: 10 teams (7 @Home, 3 @Work) and 79 participants from 6 countries.
- RoCKIn 2015, in the Portugal Pavilion, Lisbon, Portugal, 17–23 November 2015: 12 teams (9 @Home, 3 @Work) and 93 participants from 10 countries.

Organizing each of the competition events followed and improved established RoboCup best practices for the organization of scientific competitions:

1. issuing the Call for Participation, requiring teams to submit an application consisting of a 4-pages paper describing the team research approach to the challenge, as well as the hardware and software architectures of its robot system, and any evidence of performance (e.g., videos);
2. selecting the qualified teams from among the applicants;
3. preparing/updating and delivering the final version of the rulebooks, scoring criteria, modules and metrics for benchmarking about 4–5 months before the actual competition dates, after an open discussion period with past participants and the robotics community in general;
4. building and setting up the competition infrastructure;
5. setting up the MCS for ground-truth data collection during benchmarking experiments, listing all data to be logged by the teams during the competitions for later benchmarking processing, and preparing USB pens to store that data during the actual runs of the teams robot system;
6. preparing several devices and software modules required by the competition rules (e.g., referee boxes, home automation devices—remotely-controlled lamps, IP camera, motorised blinds—and device network, factory-mockup devices—drilling machine, conveyor belt—objects for perception and manipulation, visitors uniforms and mail packages, audio files and lexicon);
7. establishing a schedule for the competitions and their different components;
8. establishing the adequate number of teams awarded per competition category and preparing trophies for the competition awards;
9. realizing the event, including the organization of visits from schools, and the availability of communicators who explain to the audience what is happening, using a simplified version of technically correct descriptions.

Three camps were also organized:

- RoCKIn Kick-off Camp, in Eindhoven, the Netherlands, 28 June till 1 July 2013, during RoboCup2013: 12 participants. The camp consisted of several

lectures by the partners, on RoCKIn challenges and activities, covering subjects such as: principles for benchmarking robotics; raising awareness and disseminating robotics research; as well as discussion on developing robotics through scientific competitions like RoboCup. In addition to the lectures, attendees got first-hand experience of demo challenges, tests, and hardware and software solutions during the RoboCup@Home and RoboCup@Work practical sessions.

- RoCKIn Camp 2014, in Rome, Italy, 26–30 January 2014: 19 teams (11 @Home, 8 @Work), corresponding to a total of 63 students and researchers from 13 countries. This Camp was designed to support the preparation of (preferably new) teams to participate in RoCKIn@Home and RoCKIn@Work competitions, and featured guest lectures on vision-based pattern recognition, object and people detection, object grasping and manipulation, and Human-Robot Interaction in natural language.
- RoCKIn Field Exercise 2015, in Peccioli, Italy, at the ECHORD++ Robotics Innovation Facility, 18–22 March 2015: 42 participants divided in 9 teams (4 @Home, 5 @Work). The Field Exercise has been designed as a follow up of the previous RoCKIn Camp 2014, where most of the RoCKIn Competition 2014 best teams displayed their progresses and all participants improved their interaction with the RoCKIn scoring and benchmarking infrastructure.

6 Future Outlook: The European Robotics League

The novel European Robotics League (ERL³) competitions format has been introduced in the H2020 RockEU2 project. It aims to become a sustainable distributed format (i.e., not a single big event) which is similar to the format of the European Football Champions League, where the role of national leagues is played by existing test beds (e.g., the RoCKIn test beds, but also the ECHORD++ project Robotics Innovation Facilities/RIFs), used as meeting points for “matches” where one or more teams visit the home team for a Local tournament. This format will exploit also arenas temporarily available during major competition events in Europe (e.g., RoboCup) allowing the realization of Major tournaments with more teams.

According to this new format, teams are scored in a given challenge for each tournament they participate to, and they get ranked based on scores accumulated over the year in their two best participations. The top ranked team(s) per Task and Functionality Benchmark are awarded prizes delivered during the European Robotics Forum in the year after. Travel support will be provided to selected teams based on criteria that will take into account research quality, financial needs and team technology readiness. Teams will be encouraged to arrive 1–2 weeks before the actual competition/event so to participate in integration weeks where the hosting institution provides technical support on using the local infrastructure (referee boxes, data acquisition and logging facility, etc.), ensuring a higher team technical readiness level (TTRL). TTRL concerns the ability of

³ https://eu-robotics.net/robotics_league/index.html.

a team to have its robot(s) running without major problems, using modular software that ensures quick adaptation and composition of functionalities into tasks, and to use flawlessly the competition infrastructure, whose details may change from event to event.

Local tournaments will take place in currently available test beds at Instituto Superior Técnico premises in Lisbon, Portugal, at the ECHORD++ RIF in Scuola Superiore Sant’Anna, Peccioli, Italy, for ERL Service Robots (ERL-SR, former RoCKIn@Home); at Bonn-Rhein-Sieg University labs in Sankt Augustin, Germany, for both ERL-SR and ERL Industrial Robots (ERL-IR, former RoCKIn@Work). Major tournaments will be part of RoboCup2016 (Leipzig, Germany), and possibly RoboCup GermanOpen 2017 and the RoboCup PortugueseOpen in 2017. RockEU2 will provide a certification process to assess any new candidate test beds as RIFs for both challenges, based on the RoCKIn rulebook specifications and the implementation of the proper benchmarking and scoring procedures. This will enable the creation of a network of European robotics test beds having the specific purpose of benchmarking domestic robots, innovative industrial robotics applications and Factory of the Future scenarios.

ERL and the RoboCup Federation established an agreement that includes the sharing of tasks between the corresponding challenges in the two competitions, starting with different scoring systems, but that may possibly converge in the future. RoboCup is also starting to use benchmarking methods that were introduced during RoCKIn lifetime and that will be used in the ERL.

Acknowledgments. The RoCKIn project was funded under the EC Coordination Action contract no. FP7-ICT-601012. The RockEU2 project is funded under the EC Coordination Action contract no. H2020-ICT-688441.

References

1. Amigoni, F., Bastianelli, E., Berghofer, J., Bonarini, A., Fontana, G., Hochgeschwender, N., Iocchi, L., Kraetzschmar, G., Lima, P.U., Matteucci, M., Miraldo, P., Nardi, D., Schiaffonati, V.: Competitions for benchmarking: task and functionality scoring complete performance assessment. *IEEE Robot. Autom. Mag.* **22**(3), 53–61 (2015)
2. RoCKIn Deliverable D1.2: General evaluation criteria, modules and metrics for benchmarking through competition. http://rockinrobotchallenge.eu/rockin_d1.2.pdf
3. RoCKIn Deliverable D2.1.3: RoCKIn@Home Rule Book. http://rockinrobotchallenge.eu/rockin_d2.1.3.pdf
4. RoCKIn Deliverable D2.1.3: RoCKIn@Work Rule Book. http://rockinrobotchallenge.eu/rockin_d2.1.6.pdf
5. Shapley, L.S.: A value for n-person games. In: Kuhn, H.W., Tucker, A.W. (eds.) *Contributions to the Theory of Games. Annals of Mathematical Studies*, vols. II and 28, pp. 307–317. Princeton University Press, Princeton (1953)