RoCKIn - Robot Competitions Kick Innovation in Cognitive Systems and Robotics
FP7-ICT-601012

Grant Agreement Number: 601012
Funding Period: 01.01.2013 - 31.12.2015
Instrument: Coordination and Support Action

Specification of General Features of Scenarios and Robots for Benchmarking Through Competitions

Aamir Ahmad, Iman Awaad, Francesco Amigoni, Jakob Berghofer, Rainer Bischoff, Andrea Bonarini, Rhama Dwiputra, Giulio Fontana, Frederik Hegger, Nico Hochgeschwender, Luca Iocchi, Gerhard Kraetzschmar, Pedro Lima, Matteo Matteucci, Daniele Nardi, Viola Schaffionati, Sven Schneider

---

Lead Contractor for this Deliverable:	Bonn-Rhein-Sieg University oAS
Due date of deliverable:	June 30, 2013
Actual submission date:	July 25, 2013
Dissemination level:	Public
Revision:	1.0

Executive Summary

RoCKIn is a EU-funded project aiming to foster scientific progress and innovation in cognitive systems and robotics through the design and implementation of competitions. An additional objective of RoCKIn is to increase public awareness of the current state-of-the-art in robotics and the innovation potential of robotics applications. From these objectives several requirements for the work performed in RoCKIn can be derived:

- The RoCKIn competitions must start from convincing, easy-to-communicate user stories, that catch the attention of relevant stakeholders, the media, and the crowd. The user stories play the role of a mid- to long-term vision for a competition. Preferably, the user stories address economic, societal, or environmental problems.
- The RoCKIn competitions must pose open scientific challenges of interest to sufficiently many researchers to attract existing and new teams of robotics researchers for participation in the competition. The competitions need to promise some suitable reward, such as recognition in the scientific community, publicity for a team's work, awards, or prize money, to justify the effort a team puts into the development of a competition entry. The competitions should be designed in such a way that they reward general, scientifically sound solutions to the challenge problems; such general solutions should score better than approaches that work only in narrowly defined contexts and are considred over-engineered.
- The challenges motivating the RoCKIn competitions must be broken down into suitable intermediate goals that can be reached with a limited team effort until the next competition and the project duration.
- The RoCKIn competitions must be well-defined and well-designed, with comprehensive rule books and instructions for the participants in order to guarantee a fair competition.
- The RoCKIn competitions must integrate competitions with benchmarking in order to provide comprehensive feedback for the teams about the suitability of particular functional modules, their overall architecture, and system integration.

This document takes the first steps towards the RoCKIn goals. After outlining our approach, we present several user stories for further discussion within the community. The main objectives of this document are to identify and document relevant scenario features and the tasks and functionalities subject for benchmarking in the competitions.

Contents

List of Figures

Chapter 1

Introduction

1.1 The RoCKIn Project in a Nutshell

Robot competitions have proved to be an effective instrument to foster scientific research and push the state of the art in a field. Teams participating in a competition must identify best practice solutions covering a wide range of functionalities and integrate them into practical systems by designing and implementing an appropriate architecture. These systems have to work in the real world, outside of the usual laboratory conditions. The competition experience helps to transfer the applied methods and tools to successful and high-impact real-world applications. Other effects of robot competitions are that young students are attracted to science and engineering disciplines, and that the relevance of robotics research is demonstrated to citizens. However, some limitations can emerge as competitions mature: the effort required to enter the competition grows and may present a barrier for the participation of new teams; a gap between benchmarking complete systems in competitions and benchmarking subsystems in research may develop and limit the usefulness of the competition results to industry.

The goal of RoCKIn is 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). Both challenges have been inspired by activities in the RoboCup community, but RoCKIn improves and extends them by introducing new and prevailing research topics, like natural interaction with humans or networked mobile robots with sensors in ambient environments, in addition to specifying concrete benchmark criteria for assessing progress. Another important role of RoCKIn is to increase public awareness on the state of the art in robotics and of the research activities of the European robotics community amongst the major industrial, societal, and political stakeholders in Europe.

The RoCKIn project

- formulates challenge problems for robotics motivated by two major future application domains, in particular domestic service robots and networked mobile manipulators for industrial applications,
- designs open domain testbeds for competitions targeting the two challenges and usable by researchers worldwide,
- develops methods for benchmarking through competitions that allow to assess both particular subsystems as well as the integrated system,

- organizes two robot competition events, each of them based on the two challenges and testbeds,
- organizes camps open to student participants, so as to help new teams getting involved in the competitions, and
- executes dissemination activities to target stakeholders in industry and academia, as well as the general public.

1.2 Structure of the Document

This document is structured as follows:

1. Chapter 1 gives a first overview over the RoCKIn approach (in Section 1.3)
2. Chapter 2 is supposed to initiate discussion on the competition design and presents several user stories. The competition scenarios, including information on the tasks to be solved, the environments, and the robots targeted in each scenario are described in a yet informal manner. For each of the @Work and the @Home domains, we specify several scenarios, which at this point are meant to stimulate discussions both within the consortium as well as with other stakeholders, like industry, research groups, and technical committees of competitions like RoboCup@Work and RoboCup@Home.
3. A review of past and current robot competitions is given in Chapter 3.
4. Chapter 4 defines numerous concepts and terms relevant for our work.
5. Chapter 5 then describes in detail the task features, testbed features, and robot features derived from the user stories of the competitions.
6. Additional information on benchmarking and benchmarking competitions elsewhere in science, as well as a survey of espcially scoring systems in a variety of sports is provided in the Appendix **??**.

1.3 Putting RoCKIn to Work

The RoCKIn project uses a systemic approach to achieve its goals. This chapter explains this approach in an informal manner, relying for now on an intuitive understanding of concepts like "scenario" and "benchmark". A later chapter of this document will provide more formal definitions of these concepts.

How can we achieve the objectives of RoCKIn? We take a stepwise approach, as illustrated in Figure 1.1.

(1) Formulation of challenge problems: As a first step, challenge problems need to be formulated. This is a difficult yet critical step, because the challenge problems need to satisfy various constraints:

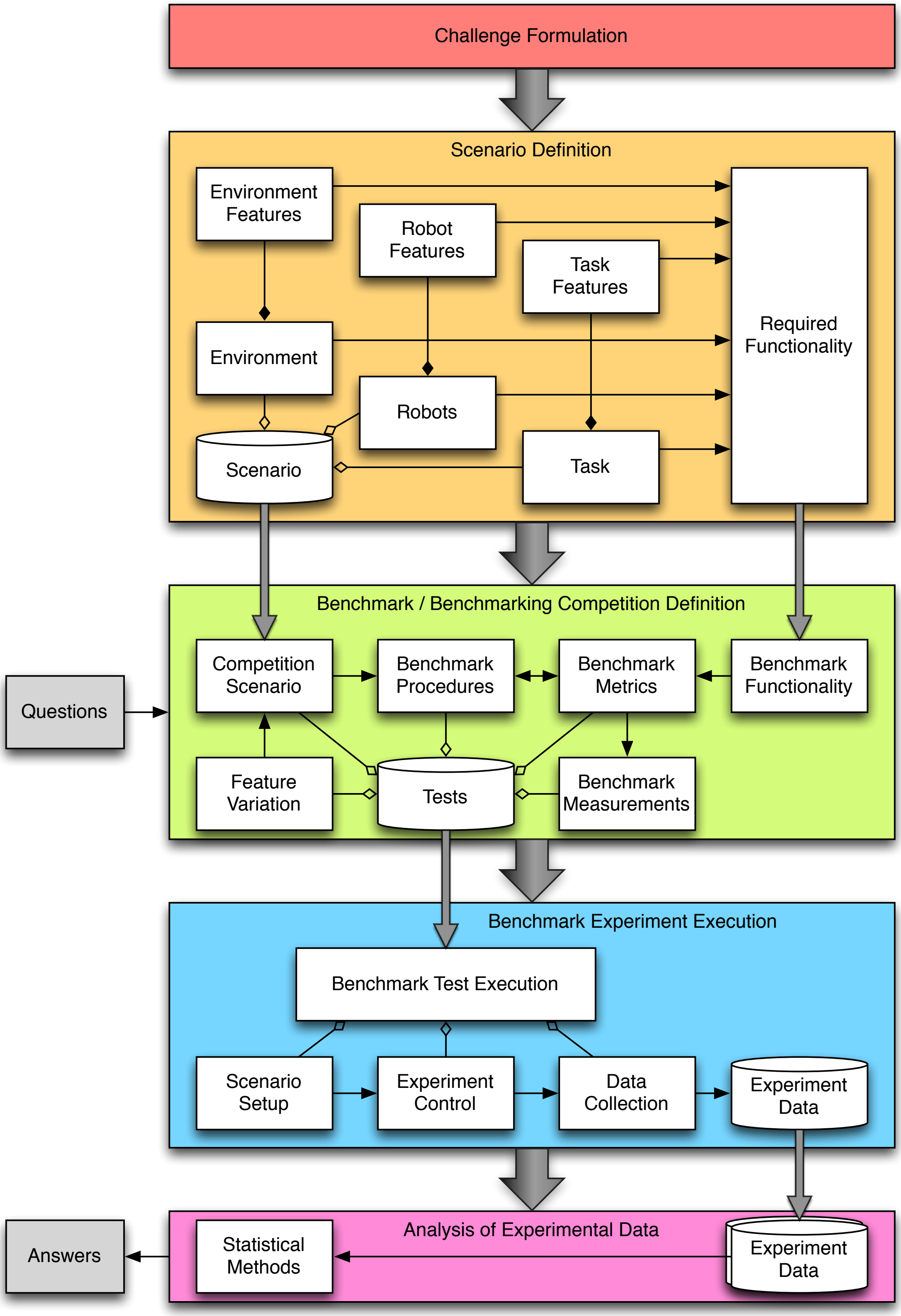


Figure 1.1: Overview of the RoCKIn approach.

- They should be new and interesting enough to raise the interest of the general public, the media, and relevant societal and political stakeholders. If a challenge problem addresses a major economic, ecological, or societal problem, this should usually be the case.

- The challenge problems should have an interesting application potential in domains with significant market potential to raise the interest of industry.

- The challenge problems should pose interesting and new research questions to raise the interest of the research community, but solutions need to be within reach in a foreseeable period of two to five years based on the current state of the art in order to expect solutions ready to be tested in competitions.

- Last but not least, the challenge problems need to be such that a competition, including all aspects of the scenario and scientific benchmarking, can be designed and implemented given the resources and constraints of the project.

(2) Definition of a scenario: Once a challenge problem has been identified and described, we need to define a scenario for it. A scenario description includes all relevant aspects of the *task* to be performed, of the *environment* in which the task is to be performed, and of the *robots* to be used.

A task usually involves achieving a particular state of affairs ("*Clean the table!*", "*Set up work station 22 for an assembly of product number 5511!*") or performing a particular kind of activity ("*Go to the post office and deliver this parcel!*", '*Paint the planar surface of work piece #388*"). Transforming such informal task descriptions into more concrete task specifications suitable for benchmarking and scientific competitions implies specifying the initial situations, the activities to be performed (at least on an abstract level), and the state of the affairs in the final situation.

A task description may include its decomposition into subtasks. If tasks or subtasks involve interaction between humans, task descriptions may include exemplifying dialogues and all information necessary to determine what acceptable interactions of robots with humans are and what not. If a task involves interaction with other robots, the information to be exchanged as well as the protocols for the communication need to be defined. Performance or quality criteria may be specified for tasks and subtasks, as well as any kind of constraints that must hold during task execution or for the final outcome of the task.

The environment description includes all aspects that enable or constrain the behaviour of the robots or somehow else influence the performance of the task. Examples are the rooms (and their properties) in which the task is to be performed, the objects and subjects in the rooms, the objects, tools, and substances to be grasped, transported, or otherwise manipulated.

The description of robots is meant to define a set of minimal criteria that robots performing the specified tasks must meet. This could include e.g. size, weight, and speed constraints or requirements concerning the availability of particular sensors or actuators.

For each of these three major aspect categories, a set of relevant features[1] are identified. For each feature, a feature domain, i.e. a range of possible instances of the feature, is defined. Features can be mandatory or optional in a scenario. Thus, while the set of

[1]Most of the features can be viewed as *parameters* in a scientific experiment.

features describes the *structure of a scenario*, the feature domains span a space of possible parameter values and thereby define the *variability of the scenario*.

An interesting aspect of scenario definitions is their relationship with the functionalities that the robots are required to have. One of the major objectives of the RoCKIn project is to better understand how the performance of particular functional subsystems, e.g. for mapping, navigation, grasping, taking orders from humans, etc. correlates with the overall performance of the robot executing a particular non-trivial task, e.g. "*Deliver this parcel to the post office!*". To that extent the scenario definition should include a list of *abstract* functionalities which seem to be implied by the specification of the scenario and its major subcategories task, environment, and robot. Examples of such abstract functionalities are:

- **Task planning** functionality is needed if the task is specified in a way, where the robot must itself decompose the task into simpler activities, if it needs to determine which activities it must perform in order to achieve the goal of the task, or needs to determine the order in which these activities need to be performed.

- **Path planning** functionality is needed when a task requires a mobile robot to move between different places in the environment and a path between these places is not known a priori, or may need to be modified due to the occurrence of obstacles.

- **Grasp planning** functionality may be required if the task requires the robot to grasp and manipulate objects, and grasping cannot be achieved with e.g. purely haptic feedback.

- **Visual object recognition** is required e.g. if the task requires the robot to fetch a particular object, which must be recognised in a scene where several objects appear at the same time in an image produced by one of the robot's cameras.

- **Visual person recognition and identification** is required if the task requires the robot to interact with people, especially if commands are to be taken from a constrained, well-defined set of people the robot needs to identify before accepting commands from them.

- **SLAM** — simultaneous localization and mapping — functionality is needed if the task requires the robot to move in the environment, e.g. to follow a previously planned path or trajectory, and if the trajectory following module requires that the robot maintains a position estimate of itself within a map constructed and maintained by the robot itself. Sometimes localization only may be needed, when working with a priori maps, but in practice these cases can be considered rare.

- **Speech understanding** functionality is required if the task involves taking commands or similar information (cf. names of objects or persons) from a human user.

- **Speech generation** or *speech production* functionality is needed if the robot is required to communicate with a human user by voice or otherwise produce auditory feedback.

This list will be made more specific and comprehensive in later chapters.

Please note, that the functionalities referred to in this document are always *abstract* functionalities, which are normally implied by at least one of the three major aspect categories of scenarios, i.e. task, environment, or robots. This document does not make any assumptions whether and how these abstract functionalities are implemented by a particular instance of a robot, e.g. via a 1-to-1 mapping to specific software modules, by a particular subsystem itself consisting of several modules working together, or by any other means of implementing the functionality.

(3) Definition of a benchmarking competition: While scenario definitions are an excellent means to characterise a particular application domain, they usually leave an immense variability in various aspects of task, environment, and robot. When performing a benchmark or scientific competition we like to make it as much a scientific experiment as we can.

As for any good scientific experiment, there should be a question or hypothesis first. What do we expect to learn from a benchmark or scientific experiment? Which question is it supposed to answer? Which hypothesis do we want to verify/falsify?

This question should allow us to identify which functionalities are in the focus of the benchmark. Suitable measures with associated metrics need to be identified to assess both the performance of the robot with respect to these functionalities as well as with respect to the overall task. We now know which measurements we need to collect in order to answer the question, and can design the experiment in order to get "good" data. A suitable scenario can be selected and properly constrained. The constraints will eventually be determined by the scientific question to be answered and which kind of dependencies between the performance measures and confounding factors (usually a subset of the scenario features) we would like to investigate. The constraints also determine the remaining variability of the scenario features. Some examples:

- Assume we want to assess the ability of a single robot to perform a certain variety of tasks in a well-specified environment. An appropriate scenario for such a benchmark would fix all features of the environment and the robot, but leave variability in the task specification. A very simple example would be to assess the navigation ability of a robot, by giving it a set of goal destinations the robot should visit in the shortest time possible.

- In another situation we might want to assess the ability of a single robot to perform a task under certain variations of environmental conditions, e.g. recognising objects from a set of 100 objects under varying lighting conditions. This would involve setting up an experiment where the amount of lighting will be controlled during the execution of an experiment, and the robot would have to recognise all 100 objects in each of the lighting situations.

- A standard situation for competitions is to assess and compare the performance of different robots (or robot teams), all performing the same task in the same environment. While it is not so difficult to find comparative criteria for overall robot performance, it is usually an open problem to assess after a competition why a robot has performed better or worse than another, and how this different performance relates to particular implementations of the functionalities.

- An even more difficult situation is when we want to assess and compare the performance of different robots on a variety of tasks and under varying environment conditions, i.e. we are interested to assess criteria like flexibility, reliability, scalability, and robustness.

The latter two cases will be those of particular interest to the planned RoCKIn competitions. The first example, simply considered as "testing the system" by many, usually applies to a work group who has implemented some robot system and wants to explore or assess its capabilities. The second example would specifically target the robustness of a system design.

Aside of identifying the functionalities, defining measures and metrics, and suitably constraining the scenario, we need to specify precise procedures for executing the experiments. In scientific competitions this includes timing, setup of the scenario, initiation of a test run, any kind of interaction with the robots during a run, controlling any environment or task features (e.g. supplying an object at a particular instant of time), scoring, measuring and recording of data, etc.

If several functionalities are involved, as is frequently the case in our targeted application domains, the benchmark may be broken down into a set of tests. Each test would impose additional constraints on the scenario features or even look only at much smaller subtasks in order to reduce the number of confounding factors, and look at more direct dependencies between scenario features and measured performance of functionalities and overall robot performance. The RoboCup@Work competitions as held in 2012 and 2013 exemplify this kind of approach:

- The Basic Navigation Test (BNT) focuses exclusively on the navigation ability of the robots without involving any kind of manipulation.
- The Basic Manipulation Test (BMT) focuses almost exclusively on manipulation ability and requires only minimal navigation. (This minimal navigation was included to avoid that teams focus on approaches requiring extensive a priori calibration procedures.)
- The Basic Transportation Test (BTT) combines the prior two tests and assesses the robot's ability to interleave navigation and manipulation in a reliable manner.
- In all of the previous three tests only a single team with a single robot performed in the testbed. The Competitive Transportation Test (CTT), which was defined but never executed, would then have added significant unpredictability by having two robots compete simultaneously, giving rise to situations where robots present mutual dynamic obstacles to each other.

The set of tests defined by a benchmark or scientific competition is the basis for actually running experiments, described below.

(4) Execution of the benchmarking competition: Once a benchmark or scientific competition has been defined and a well-specified set of tests is provided, the tests need to be executed. This requires the actual setup of the scenarios specified in the tests, carefully executing the procedures for experiment control, and diligently recording all required data.

(5) Analysis of the results: After execution of the experiments, the recorded data can be analysed by applying the appropriate statistical methods in order to provide the answers that motivated the benchmark or competition.

Chapter 2

User Stories

The purpose of this chapter is to **give an initial idea** of the kind of robots we target, the environments in which we want them to be embedded, and the tasks we want them to perform. The project has chosen two major categories of application domains: *domestic service robot applications* (→ RoCKIn@Home) and *new industrial applications* (→ RoCKIn@Work), in particular applications that exploit recent developments in mobile manipulators and bi-manual or multi-manual manipulation. In the following two sections, we describe several user stories each for RoCKIn@Work and RoCKIn@Home. They are meant as inspiration for discussions, and no implications should be derived pertaining to the competitions eventually adopted by RoCKIn.

2.1 User Stories for RoCKIn@Work

The European robotics industry is convinced that recent developments in robotics, especially more lightweight manipulators, more dextrous hands, mobile manipulators, compliant manipulators, and bi-manual or multi-manual manipulators allow for many new and innovative industrial applications of their products in new markets. In order for these markets to develop, sufficiently realistic demonstrators need to be developed and publicised which show the capabilities of these new robot systems. In the following subsections we describe several scenarios for such innovative industrial robot applications.

2.1.1 Manufacturing Logistics and Assembly Support

Motivation: It is commonly believed that practically all manufacturing processes, which can be effectively/technically and efficiently/economically automated, are already automated. Vice versa, if such production processes are not yet automated, there exist either unsolved technical problems or it was not economical to do so. So far. With the advent of new robot technology, like small mobile manipulators (e.g. the KUKA youBot, Figure 2.1a), lighter-weight and compliant robot arms (like the KUKA LWR, see Figure 2.1b, the BioRob arm [1], or the manipulators by Universal Robots [2]) and bi-manual robot systems (see cf. the Tecnalia robot and the Baxter robot by Rethink Robotics [3], both in Figure 2.2) this situation may dramatically change. Using such robot systems new applications of robots in manufacturing can be designed. Areas of particular interest are manufacturing logistics and pre-assembly of parts for subsequent production steps.

The economic potential and importance of these applications are enormous (see e.g. [4], especially because the application of robots may become interesting in many small and

(a) KUKA youBot.

(b) KUKA LWR

Figure 2.1: Examples of innovatice robots by KUKA.

medium-sized enterprises, which are a economic stronghold of many European countries, but face constantly increasing competition due to globalization. Production in such enterprises is often characterized by much smaller lot sizes, higher variability of produced goods, and more frequent changes in the production schedule due to market demand. These companies could not economically use robots so far.

Scenario Task Idea: The environment consists of several workstations, in which either human workers or robots assemble a variety of different goods. A central scheduler gets orders for these goods and assigns them dynamically to the workstations. Once it is known where a product item will be assembled, the parts required for its assembly need to be available at the workstation sufficiently on time such that no delays in the production process are caused. Each type of product requires a different set of component items for the assembly phase. Some of these items, such as screws, nuts, and bolts, rings, etc. (*commodity items*) are used in all products, albeit in different numbers, while others are specific to the product type being assembled (*specific items*). All parts are kept in suitable storage containers and are delivered into adequately-sized boxes at the workstations The task to be solved by the robots fall into two categories:

(a) Tecnalia Robot[5]

(b) Baxter Robot[3]

Figure 2.2: bi-manual manipulators by Tecnalia and Rethink Robotics.

- The first category is *manufacturing logistics*. The robots have to observe the production schedule for each workstation and ensure that required specific parts will be there on time and placed in the correct, pre-specified places. Furthermore, the robots must ensure that a sufficient supply level for all commodity items is maintained.

- The second category is *pre-assembly*, for example arranging various component items into a tray, from which a subsequent, usually automated assembly process will take them.

This scenario can come in a lot of variations, e.g. in the number of parts required, the proportion between commodity and specific items, the number of workstations, the number of robots available for performing logistics tasks, the storage arrangements for the component items, and the scheduling strategy of the job shop scheduler, to name just a few. While it is possible to start this idea with significantly constrained scenarios involving only a single robot, it should be clear that more realistic scenarios are bound to be inherently multirobot. These robots will then not only be networked with each other, but also with other items in the environment, such as the scheduler, conveyor belts, electronic scales, item dispensers, etc.

It is possible to extend this scenario later to include the subsequent assembly process, such that semi-finished or finished products are produced as the outcome of this application scenario.

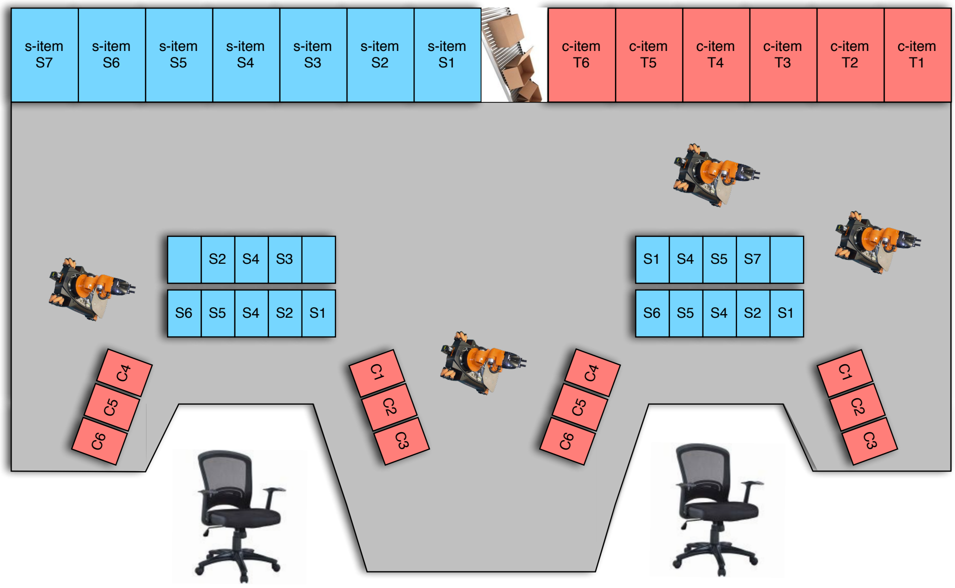


Figure 2.3: An environment for the manufacturing logistics scenario. Two workstations for human workers are shown on the bottom. The grey area above is all table area. To the left and right of each workplace there are boxes with commodity items (red). A revolving battery of boxes to be filled with specific items (blue) is right in front of the workplace. The robots maneuver on the free table area above the work places and fetch the requested items from storage areas with larger container boxes on the top. A conveyor belt in the center top carries completed items into another area for delivery or storage.

Targeted Robot Platforms: The scenario primarily targets all kind of mobile manipulators, such as e.g. a KUKA youBot or similar. Size and weight restrictions may apply. Additionally, mobile platforms without manipulators may be used for transportation tasks. For the pre-assembly task, non-mobile manipulators may be used in some workstations.

The question of what a suitable sensor arrangement for such a scenario might be remains to be discussed. While the use of several overhead cameras may be sufficient, the use of onboard sensors may be more flexible in the long run.

Environment: A small example environment is depicted in Figure 2.3.

2.1.2 Return Shipments Handling

Motivation: Internet sales of goods of all kinds has seen enormous market growth in the past decade and the major players in the business are now multinational economic enterprises. According to news reports, however, a very high ratio of parcels shipped is returned, for various reasons not of further interest here. All these parcels must be opened, the contents checked, and handled accordingly. For example, items that were damaged during shipment will be discarded, items returned in resellable state will have to go to storage, and other items will be handled specially.

(a) [6] (b) [7] (c) [8] (d) [9]

Figure 2.4: Some images of parcel handling in warehouses.

Scenario Task Idea: Opening the returned parcels usually involves using a knife, which bears the risk of injuries. The objective is to have a robot doing the opening part of the parcel. An extended capability would be that the robot is actually taking out the items in the parcel, and puts them on item-specific conveyor belts, e.g. one for clothes, one for books, one for electronic devices, etc., which carry the items to work stations where they will be inspected further and handled by humans.

Targeted Robot Platforms: This task usually involves bi-manual manipulation, or a combination of a single-arm manipulator and a fixating device (below referred to as a *in-place-holder*), which holds the parcel in place, such that the cutting operation can be successfully applied. Suitable robot platforms could be any dual-arm manipulator, e.g. a Baxter robot by Rethink Robotics, a combination of two manipulators cooperating appropriately, e.g. two KUKA LWR manipulators, or a combination of a manipulator and an in-place-holder.

Environment: A minimum environment requires a device supplying parcels (see Figure 2.4 for some examples), a work place where the robot operates, and a device to deliver the opened parcel. The latter can be a box or a conveyor belt. In the extended capability version, the delivery device must be replaced by several devices.

2.1.3 Pasta Food Factory

Motivation: The restaurant business is tough and competitive. Kitchen workers are difficult to find. For example, in the UK the "government is floating the idea of creating a curry college to train British people of all backgrounds to work in the [restaurant] industry" [10][11] and to meet the ever-growing demand for Chicken Tikka Massala, by now the most popular food in the UK. Maintaining quality level under stress situations is also difficult. Consider a large Italian restaurant specializing on freshly cooked pasta dishes (aside of possible other dishes like pizza, carne, pesce, etc.). A pasta dish usually involves cooking the pasta *à point* — *al dente* — nor over-cooked nor under-cooked and preparing a sauce. Some sauces, like a plain tomato sauce, can (or might even have to) be prepared in larger quantities in advance to fully develop its flavor, while others (any sauce containing cheese or cream) will have to be cooked just before serving the dish. If the restaurant has to meet a demand of, say, more than 200 pasta dishes per hour, and if the pasta dishes involve several different types of pasta, each of them having a different cooking time ranging from just one or two minutes to 12 or 14 minutes, then cooking the pasta can be a real challenge. An opportunity for robotics?

(a) Restaurant pasta cooking environment

(b) Pasta making machine

(c) Pasta cooking equipment

(d) Top view of pasta cooking equipment

Figure 2.5: Some images of professional pasta preparation and cooking equipment.

Scenario Task Idea: Build a robot system that supports an Italian restaurant in cooking pasta (only the pasta, not the sauce!). For cooking the pasta the restaurant has a device with boiling water into which sieves with the pasta can be hung. The robot can get the pasta from a suitable storage device. It needs to weigh or somehow otherwise determine the required amount of pasta. Interaction is necessary if the work load gets high enough such that not all constraints can be satisfied.

Targeted Robot Platforms: The basic scenario should be doable with any suitable mobile manipulator. If a dual-arm robot system is available, extended scenarios can be devised.

Environment: Some example environment elements are depicted in Figure 2.5.

2.1.4 Construction Industry Robot Assistant

Motivation: The idea is to look into various activities on construction sites. Of particular interest are activities relating to large, flat surfaces: cleaning, spraying, tiling, painting, grouting, silicone jointing, etc. (for some examples, see Figure 2.6 a and b), which involve compliant motion across surfaces and meeting other criteria like coverage, smoothness, or the precise following of a trajectory (e.g. along a joint).

(a) Floor tiling. [12] (b) Silicone jointing bridges. [13] (c) A robot painter. [14]

Figure 2.6: Exemplary scenes showing construction site activities.

Scenario Task Idea: A robot is to be designed which spray-paints a surface. The surface can initially be a comparatively small patch of floor or wall (e.g. two square meters) which the robot needs to smoothly cover with just the right amount of paint. The environment can be made more difficult by including wall edges or areas like windows and doors that need to be excluded from painting. Another variation can involve the robot painting particular patterns, e.g. using different colors for the upper and lower half of the wall and a small band in a third color separating them.

A yet more difficult task would be to design a robot for grouting tiles and for silicon jointing.

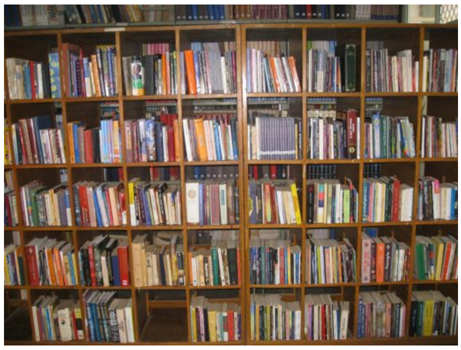
Targeted Robot Platforms: Somewhat unclear. Any currently available platform would need new end effector devices for handling the mortar, filler, silicone glue, or paint. Furthermore, these should be easily exchangeable.

Environment: Remains to be decided.

2.1.5 Library Assistant Robot

Motivation: Public libraries (see Figure 2.7 for some illustrations) get a large number of book returns every day, which must be manually sorted back into the bookshelves at the right place. The idea is to have a robot perform this task.

(a) A typical university library.



(b) Close-up of the bookshelves [15].

Figure 2.7: Illustrations of bookshelves in large public libraries.

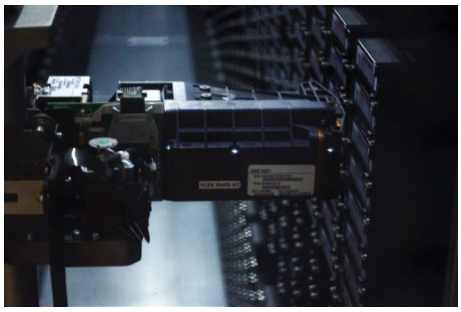
Scenario Task Idea: At the library, there is a special desk for book returns. Returned books are put onto a simple robot with a larger container for books on top. The returned books carrier robot cannot navigate on its own; it can either be manually operated and steered around or it can be guided/tele-operated by another robot. When a container is almost full, the desk staff will send an email to a reshelfing robot, which comes to pick up the carrier robot and guide it to the bookshelf area, where the reshelfing robot returns the books into the right place. The reshelfing robot can get a list of all the returned books from the terminal at the book return test and use this for various optimizations.

Whenever the reshelfing robot is idle, it is supposed to check the bookshelves for misplaced books, missing books, etc.

Targeted Robot Platforms: For example, a Care-O-bot 3 with KUKA LWR and Schunk hand. See Figure 2.8 for some examples of existing technology.

(a) Library robot. [16],

(b) Robot hand picking a book. [17],



(c) Media storage robot. [18]

Figure 2.8: Some examples of robot technology for library robots.

Environment: A public library.

2.2 User Stories for RoCKIn@Home

A slightly different approach than for RoCKIn@Work will be taken for RoCKIn@Home. All three user stories for RoCKIn@Home share a common environment and, possibly, but not necessarily, the same set of one or more robots. These robots will have to provide a range of services to support the inhabitant of the environment.

Common Motivation for RoCKIn@Home: Granny Annie is a nice but slightly seasoned lady. After a fulfilled work life she is already in retirement for well over ten years. Sadly, she has lost her husband a few years ago and now lives alone in her apartment, which features a kitchen, a combined dining and living room, a bedroom, a bathroom and a toilet, all connected by a hallway. Age is slowly taking its toll, and is recently showing in various situations. Getting out of bed or sitting down in the couch is getting harder, her eyesight is suffering a bit more than it used to, and she sometimes forgets to take her medicine timely. Many household chores are getting a lot harder for her to do.

Luckily, Granny Annie could get into a new program sponsored by her health and social security insurances. In this program, elderly people are supplied with household and elderly care robots. The robots are to assist the elderly people to manage and master their daily lives, to longer live an independent life in their known surroundings, and to better maintain social contacts, which is known to have very positive influence on the mental and bodily health of elderly people. An initial and extensible list of tasks these robots are eventually supposed to perform includes the following activities:

- lend an arm: help Granny out of bed and or the couch
- serve the breakfast (cereal, orange juice, tea/coffee, toast, jam, butter, boiled eggs)
- clean the breakfast table
- supply Granny Annie's medicine
- take medical tests (blood pressure, blood sugar, etc.)
- bring books or operate media (entertainment)
- make up the bedroom
- play a round of checkers with Granny Annie
- set the lunch table for the bridge round
- order the food
- welcome the guests, including following/guiding people
- serve food and drinks (sodas, coffee, tea) (cocktail party)
- adjust settings: shades, light, heat
- clean the table
- clean the kitchen
- serve drinks during the bridge round
- set the bathroom for taking a bath
- assist during bathing
- clean the bathroom

This list could be easily extended, but the reader should be able to get the general idea.

Common Targeted Robot Platforms for RoCKIn@Home: The basic scenario should be doable with a suitably-sized mobile robot featuring at least one manipulator (arm with gripper or dexterous hand). If a dual-arm robot system is available extended scenarios can be devised.

Common Environment for RoCKIn@Home: The RoboCup@Home environment will initially serve as an environment for RoCKIn@Home. The project will gradually extend the testbed to allow for applications involving multiple and networked robots, as well as sensors and actuators embedded in the environment (ambient intelligence).

2.2.1 Setting the Breakfast Table

Scenario Task Idea: It is 7am. Granny Annie is waking up and wants to have breakfast. The robot has to set the breakfast table (see Figure 2.9). Before doing so, the robot must ask Granny Annie what she would like to have for breakfast, e.g. cereal, yoghurt, fruit, orange juice, tea, coffee, jam, bread, rolls, butter, sausages, fried, scrambled, or cooked eggs, etc. Once the robot knows what Granny Annie wants to have for breakfast, a list of items required items can be determined, which the robot needs to fetch and set on the table in a well-defined arrangement. For example, a light breakfast for Granny Annie may require setting only a bowl for cereal, a spoon, and a mug for tea or coffee, while Granny Annie's more savory Sunday breakfast may include a plate, a tea cup with saucer, knife, fork, tea spoon, an egg cup and egg spoon, butter dish, two different glasses of jam, a drinking glass, a jug of orange juice, a basket of bread, rolls, and toast, a bowl with freshly sliced fruit, and a plate with hams and cheeses.

(a) A breakfast table setup. [19]

(b) Two household assistant robots. [20],

Figure 2.9: Images illustrating the breakfast table setup scenario.

As one of the subtask of this scenario, the robot should learn (by demonstration) a semantic map about the environment, such that it knows where to find all required items.

2.2.2 Bridge Round and Tea Party

Scenario Task Idea: Despite her age, Granny Annie is still diligently maintaining her social contacts. Twice a week, she hosts a couple of her friends for a round of bridge, which usually ends in a tea party (see Figure 2.10). The robot is supposed to set the table for the bridge round. When the guests arrive, the robot needs to welcome them at the door, ask whether it can receive and store any items (umbrella, hat, hand bag), and guide them to the bridge table. The robot serves drinks for the guests during the bridge round. When a glass or cup is emptied, the robot will ask if it should serve more. Finally, just before tea time, the robot will set the table for the tea party, with freshly brewed tea, and serve pastries and light sandwiches it has ordered before from a delivery service. The robot will prepare the items for the tea party on a cart, which it can push or pull from the kitchen towards the dining table.

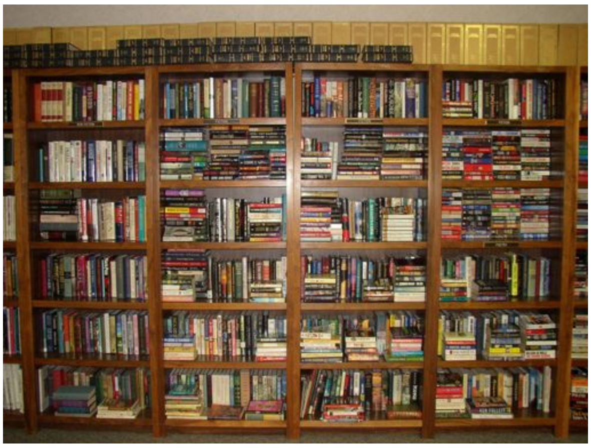
(a) Grannie Annie playing bridge. [21]

(b) Table setup for the tea party [22].

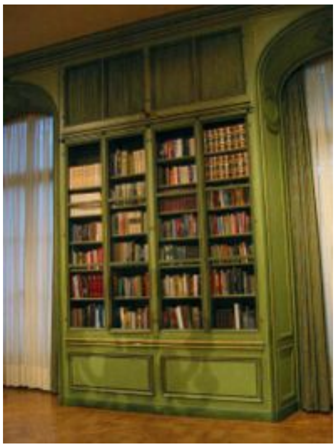
Figure 2.10: Images illustrating the tea party scenario.

2.2.3 Maintaining the Bookshelf

Scenario Task Idea: When Granny Annie is all alone, she is not at all up to being bored: she loves to read! 40 years of membership in the book club have bought her an awesome collection of books, which she is very proud of, and with new books still coming every month. And she has a major project milestone left in life: she wants to read again every single book she owns; after all, it has been decades since she had read most of them for the first time. Unfortunately, keeping track of what she has already reread is becoming more difficult. Sometimes she cannot find a book she is sure she owns. Is it just hiding somewhere in the bookshelf ? Or has Granny Annie lent it to one of her Bridge ladies?



(a) A private library.



(b) Private bookshelf.[23],

Figure 2.11: Some possible library bookshelves of Granny Annie.

2.2.4 Cleaning the Bathroom

Scenario Task Idea: Cleaning has become a very difficult task for Granny Annie, as she is suffering from gradually constrained mobility. The bathroom is difficult in particular, as Granny Annie is very hygiene-conscious and wants her bathroom cleaned diligently twice a week. Wouldn't it be nice if the robot can give a hand?

We start with having the robot to assist in cleaning various planar surfaces in the bathroom, e.g. a mirror or a tiled wall. Edges and corners will be added later. A sink and a bath tub will follow, first without, later with cleaning the faucets. For each item to be cleaned, the robot must apply the appropriate cleaning tools and cleaning agents. Eventually, the robot can clean a complete bathroom.

(a)

(b)

Figure 2.12: Some example bathrooms.

(a) A modern shower cabin made of glass.

(b) A typical bathroom mirror with lights and tray.

Figure 2.13: Bathrooms present challenging problems for robotics, especially for perception and manipulation. Two examples are illustrated in this figure.

2.3 From User Stories to Benchmarking Competitions

After having outlined a variety of user stories for both RoCKIn@Work and RoCKIn-@Home, we would like to emphasize again, that these user stories **do not yet present scenarios usable in scientific competitions**. This was not the role foreseen for the user stories. Rather, they are meant to *inspire a discussion in the community* on a range of aspects that will influence the decision, whether and how RoCKIn will take up any, some, or all of these ideas and turn them into scientific competitions. Some of these aspects include:

- Which of the tasks or subtasks can actually be performed by robots?
- Do we have all the hardware technology available in order to build robots capable of performing these tasks? If not, in which time frame can such technology be developed?
- Do suitable software components exist which provide the functionalities required to perform the targeted task? Can these software components be integrated into a coherent system performing the task in a sufficiently flexible and robust manner? If such software components are missing, in which time frame can they be developed?
- Can the robots reach sufficient performance levels such that practical applications become feasible?
- Can we build suitable environments in which to test and benchmark the performance of the robots?
- What are suitable assessment criteria, performance measures, and metrics to evaluate and compare robots performing the targeted tasks?

We hope for an active participation of the community and a lively discussion leading to interesting new competitions soon.

Chapter 3

Experiences from Other Robot Competitions

Before defining any new robot competition, it makes sense to look at other competitions, both inside and outside of robotics and science, to understand how they work, how they assess performance of participants, and what lessons can be learned from the experience made in these competitions.

Robotic competitions differ from many of the previously described scientific competitions in at least two aspects:

1. Most of the time whole robotic systems are considered and compared instead of single functionalities (e.g. the ability to classify objects or persons).

2. The robotic systems under consideration interact with the real world, e.g. physical objects in the environment or with humans. This real-world interaction introduces a number of issues, for example because of exogeneous events like stacked objects falling down or natural lamnguage interaction with humans, which can usually be ignored in most purely computational settings.

3.1 The AAAI Mobile Robot Competition and Exhibition

In 1992, the AAAI conference held in San Jose, CA, organized its first Mobile Robot Competition (MRC) [53]. It was a major event in which many top U.S. schools with AI groups working on some robotics-related topics participated. From today's point of view, the tasks to be performed, e.g. navigating to designated places in an environment with obstacles, were quite simple, but given the state of technology and the computational power available at that time, they were really challenging. Ranking of the teams was performed based on evaluating the subjective scores given by a panel of judges.

In the following years, the competition developed well and quickly developed interesting scenarios, including tasks ranging from "*Find Life on Mars*" to "*Scavenger Hunt*" and "*Hors d'Oeuvres Anyone?*"[54]. The latter task, a challenge posed in 1998 [55], implied robots performing in the same space as humans and interacting directly with human users. The aspect of robots interacting with humans in the same environment was very early taken up by AAAI MRC and pursued continually over many years. A nice survey of the kind of robots used by participating teams, with many pictures, can be found in [53] and [56].

We want to look a bit deeper into how the AAAI MRC worked and take the 2005 competition as an example. The competition included a contest called *The Robot Challenge*, which was posed as follows [57]:

> *The goal of the Robot Challenge is to work toward the development an interactive social robot. Toward that end, the Challenge requires a robot to participate in the AAAI conference. Aspects of conference participation goals include: locating the conference registration desk, registering for the conference, perform volunteer duties, and present talk (and answer questions) at a prescribed time and location. Additionally, the robot should socially interact with other conference participants. Navigational technical challenges include dynamic crowded environments, natural landmark detection, direction understanding and following, and map reading. Social interaction challenges may include natural conversation regarding the robot and the conference and personalization of conversation with recognized individuals (by name, badge, or face). All of these things should be done in as close to the normal environment as possible.*

Another citation, taken from a report on the 2006 event, also reveals interesting insights [58]:

> *The regular audience at AAAI is becoming increasingly habituated to robots wandering around, and tend to not pay them much attention any more. You will get individuals coming up to your robots and "kicking the tires" a bit (hopefully figuratively, but sometimes literally). You will want to make sure your robot can grab attention. Visitors to the conference will tend to crowd the robot in groups as they come through (particularly during breaks in the conference talks), so your robot ought to be able to handle a press of people and deal with the situation robustly. For mobile and wandering robots, try to keep some distance so that it doesn't look like you are shepherding or controlling the robot. Do make sure to have somebody on hand to talk to the audience and answer questions (and to step in if anything goes wrong!), but it is important that your entry be able to stand on its own without need for explanation.*

The problems of how robots can get the attention of people, if these people are already quite used to the presence of robots, in a manner such that the humans do not get annoyed, can still be considered unsolved. This is especially true in environments where many humans are present. Also, the navigation problem becomes significantly harder and may become even unsolvable without the use of appropriate human-robot interaction. This should lead robot navigation researchers to rethink their architectures. Last but not least, environments crowded by many humans pose hard challenges for safety and security. Because it is important for our consideration later on, we cite significant parts of the information on judging of the 2005 event[59] in the following pararaphs:

> *Judging will be a combination of ratings from audience members, other teams, and the judge panel. Certificates will be awarded for outstanding or creative examples of different types of AI and social interaction. Final judging policies will be discussed prior to the event.* [59]

We recognize that direct comparison of potentially very different entries is not easy. However, the judges will base their fundamental assessments on the extent and success of the spatial reasoning that each system demonstrates, given the naturalness and perceived difficulties of its operating conditions. Minimum requirements are a mobile robot. Each judge on the panel will give a subjective score between 1 and 10 for each entry. Their scores will be averaged to produce a final score. The following criteria will be used as a guide for the judges' considerations. [60]

- **Autonomy and shared autonomy:**

  We welcome a variety of teams to enter with one or more robots and/or human operators, though every entrant must demonstrate AI techniques during the competition. In particular, we encourage urban search and rescue teams with AI components to consider joining this event. Approaches resulting in systems with shared autonomy or full autonomy will be considered on equal footing. In shared autonomy systems, judges will consider the naturalness of both the interface and the delegation of tasks to the robotic system and its human assistants. In fully autonomous systems, the extent of that autonomy will be evaluated. [60]

- **Environmental modification:**

  Ideally, an entry would interact with the conference environment without modification, or by modifying the environment itself. By default, the staging area for the scavenger hunt will be a foyer of the conference venue, along with its accompanying hallways and rooms. The environment will not be engineered for the event, except that the density of people will be relatively low, so that crowding around a robot will not be allowed. Participants are also welcome to demonstrate capabilities under restricted conditions. In such cases the nature and extent of the restrictions should be well understood and conveyed to the judges. [60]

- **Unexpected, dynamic, and/or human interactions:**

  A key aspect of the scavenger hunt competition is having robots interact with people present in the environment. This category will assess systems' ability to handle unmodeled activity or changes in the environment. Robustness to such phonomena is a hallmark of intelligent spatial reasoning. As with the other judging criteria, participants may request onlookers and judges to keep to specific types of interactions. Robotic systems which make such requests for themselves will be judged even more favorably. [60]

- **Accuracy:**

  In order to convey its reasoning about the environment, each scavenger hunt entry should create and convey one or more representations of its surroundings. Many such "maps" are possible, e.g., traditional dense maps, sparse, loosely connected collections of landmark locations, networks of learned parameters, or other summaries of the systems' spatial input *data. Novel representations or approaches integrating diverse facets of AI are welcome. Judges will consider both the accuracy and utility of these representations in the demonstration and challenge phases of the competition.* [60]

- **Range and completeness:**

  Judges will assess the subset of the conference environment which each system can cope with, especially in light of the particular sensors available to each entry. For example, a system equipped with a laser range finder would be expected to reason about a larger swath of area than one with only a set of IR sensors. "Completeness" considerations include the variety of sensory modalities supported and their extent. For example, can a system locate objects not on the floor? Can a system distinguish objects using visual, auditory, direct range sensing, or other means? [60]

- **Speed:**

  ... is desirable, but it is not as important as a system's ability to interact with and reason about the (relatively) unmodified conference environment. [60]

On the overall scoring, the documents provide only the following information [60]:

> *The contestants will be evaluated on overall success as well as on any particular abilities they incorporate into their solutions, such as: technical Innovation, novel spatial reasoning approaches, mapping and navigation strategy, object recognition, object manipulation, and multiagent cooperation.*

We will discuss the implications of this scoring system later.

AAAI continued to organize Mobile Robot Competitions until 2007 [61], with repeatedly changing scenarios, rules, and robots. For some reports on the events and results, see [54], [62], [53], [63], [64], and [61]. Interestingly, it seemed that over the years participation came mainly from lesser-known university and colleges, while the big robotics labs did not get involved. From 2008 onwards, AAAI did not organize robot competitions any more and replaced the event with competitions focusing on agents utilizing various AI techniques, like trading agents, poker playing agents, or general game playing agents [65].

3.2 RoboCup

The currently largest scientifc robot competition is RoboCup. On its web site [66], an article on the history of RoboCup describes how it was invented and set up:

> *During the International Joint Conference on Artificial Intelligence (IJCAI-95) held at Montreal, Canada, August, 1995, the announcement was made to organize the First Robot World Cup Soccer Games and Conferences in conjunction with IJCAI-97 Nagoya. At the same time, the decision was made to organize Pre-RoboCup-96, in order to identify potential problems associated with organizing RoboCup on a large scale.*

...

Pre-RoboCup-96 was held during International Conference on Intelligence Robotics and Systems (IROS-96), Osaka, from November 4 – 8, 1996, with eight teams competing in a simulation league and demonstration of real robot for middle size league. While limited in scale, this competition was the first competition using soccer games for promotion of research and education. The first official RoboCup games and conference was held in 1997 with great success. Over 40 teams participated (real and simulation combined), and over 5,000 spectators attended. [66]

Since 1997, the football/soccer activities of RoboCup have steadily grown and improved. RoboCup currently features two soccer simulation leagues (2D and 3D) and four robot soccer leagues (Small-Size, Middle-Size, Standard Platform, and Humanoid). An impressive video covering 15 years of development in RoboCup soccer has been compiled by Manuela Veloso, Peter Stone, and others [67] and has been one of the finalists in the 2012 IROS Video Award Competition.

RoboCup did not stop at soccer, however. Ideas for extending the competitions with leagues targeting disaster management and search and rescue robot technology existed in the minds of the RoboCup inventors from the very beginning, as documented by the following quotation:

The trigger for the RoboCup Rescue project was the Great Hanshi-Awaji earthquake which hit Kobe City on the 17th of January 1995. [68]

RoboCup 2001 in Seattle included then RoboCup Rescue as a major new activity with two leagues, Rescue Simulation League and Rescue Robot League [69]. Information on recent rulebooks and environments can be found at [70] and [71]. RoboCup Rescue has been cooperating with Adam Jacoff and his group at NIST since the very beginning and has been constantly driving benchmarking issues.

In 2006, RoboCup added the RoboCup@Home League, a competition targeting classical service robot scenarios in domestic environments. The league uses a competition format based on three stages, each of which consists of several *tests*: The first round consists of tests requiring simpler functionalities, while the second round features tests where robots need to combine and integrate several simpler functionalities in more complex environments and tasks. The individual tests are scored either based on score sheets for each particular test or by a jury panel. The final round is an open competition, where the participating teams can decide on their own which tasks in which environment they want to demonstrate. The performance is judged by a jury panel. Details on current rules and competition arenas can be found at [72] and [73].

With the advent of advanced manipulators, such as the KUKA LWR, or mobile manipulators, such as the KUKA youBot, the community developed interest in a RoboCup league targeted towards new industrial robot applications. This led to the inception of the RoboCup@Work Demonstration League, organized for the first time in Mexico City in 2012. Information on the 2013 version of the competitions can be found in [74] and [75]. Almost simultaneously Festo Didactics decided to change the theme of their own sponsored competition, which previously was akin to a robot hockey game, into a production logistics scenario (see [76] and [77]).

3.3 The DARPA Challenges

Inspired by the growing success of robot competitions, DARPA decided to adopt and modify the idea in order to tackle so-called *grand challenges*. In 2002, the first DARPA Grand Challenge was announced [78], with the final competition scheduled for 2004. The scenario addressed the design and implementation of unmanned ground vehicles [79], which were supposed to travel autonomously a distance of 300 miles on roads as well as off-roads (mainly through desert ares) from Los Angeles to Las Vegas in less than 10 hours [80]. No human interaction (such as remote control or tele-operation) was allowed during the task [81]. Each entry could perform individually, without other participants interfering. External waypoints were provided that had to be visited while staying within defined boundaries of the track. At certain intermediate stations the vehicles could be repaired or refueled. In the 2004 competition, none of the participants could complete the task [80].

The Grand Challenge was repeated in the following year (2005) with some modifications [82]. The task was reduced to travel a distance of 132 miles over desert terrain. Again, unmanned ground vehicles were permitted as entries to the challenge [83], and the teams had to complete the task in less than 10 hours. In 2005, five teams could complete the task, and the shortest time needed to complete the task was used to determine the overall winner of the challenge, Stanley (see [84]).

After the first DARPA Grand Challenge was solved within just 2 years, DARPA announced a modified version, known as DARPA Urban Challenge, in 2006 [85][86]. The task in the Urban Challenge was to autonomously drive in traffic a distance of about 60 miles and included handling of situations like merging into traffic, parking or navigating intersections while always obeying the applicable traffic rules (e.g. speed limits) [85] [87]. The competition took place in 2007. The robot platforms were vehicles which had to meet certain requirements [88], mainly for safety reasons. Other manned and unmanned cars were present and acting in the same environment while a competition entry was performing. Several participating teams could successfully complete the challenge.

In 2012, DARPA announced its Robotics Challenge [89]. The challenge involves building (humanoid) robots capable of performing the following tasks:

1. Robots should operate in (destroyed) areas that have been designed for humans
2. Robots should use tools and devices that have been designed for humans from screwdrivers to cars
3. Robots should be operated/guided by humans

The Robotics Challenge has several stages starting in simulation to real-world robots [89]. Several sample scenarios are presented in [90]. Teams are evaluated based on the following criteria [91]:

1. Number of waypoints passed in 15 runs
2. Amount of transmitted data
3. Time left after the runs

The Robotics Challenge is still ongoing.

With its series of grand challenges, DARPA has demonstrated that it is possible to tackle and solve hard problems in science and engineering in a relatively short period of time. However, the DARPA way of implementing grand challenges involves significant financial budgets both on the side of the organizers as well as by the participating teams. Thus, this competition format is not easily and frequently reproducible in different settings. Policy makers and stakeholders should keep in mind, though, that sometimes a significant investment is necessary if bold steps are asked for.

Chapter 4

The RoCKIn Approach: Concepts and Terminology

4.1 Rationale of the RoCKIn Project

The rationale of the RoCKIn project is to provide benchmarking through competitions, i.e. to combine the merits of competitions with the scrutiny of scientific benchmarking. The following quotation indicates both the difficulties associated with this endeavor as well as the potential synergetic effects:

> *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. [...] Competitions and experiment-based challenges have complementary roles to play. Competitions are about inspiring the community to go beyond the boundaries of what appears to be possible or feasible. In contrast, experiments are about codifying our knowledge that currently exists for future refinement of our methods and generation of new scientific questions.* [94].

Competitions are appealing to many, for a variety of reasons beyond the scope of this document. Competitions usually take place with some regularity and at dates well known in advance. Having an externally defined deadline often helps to stimulate researchers to reach their goals. Competitions are good showcases for the current state-of-the-art, and they provide an opportunity to observe and evaluate it. Finally, competitions usually turn the focus from single subsystems towards integrated systems, thereby helping participants to take a broader view of the subject, and promote critical analysis of laboratory experiments in different contexts outside of the home lab (usually not in true real life conditions, but at least closer).

For robotics, another key advantage of competitions is that the (often large) costs and effort of setting up complex experimental installations for robot testing can be shared among many participants, or are in some cases sustained by sponsors. In the context of a competition, even very sophisticated and expensive experimental setups can become viable. Being a collectively managed and shared structure, the accessibility of the competition infrastructure to each participating team may be limited, though. For the same reason, limitations usually exist on the number of times a single experiment can be repeated and/or on the overall time that the testbed is accessible to each team. This may have an impact on the usefulness of the experimental data that can be obtained during the competition. However, this is usually perceived as an acceptable price to pay in order to use an otherwise inaccessible type of experimental infrastructure.

The robotics community is running numerous competitions every year[1], some of which have been surveyed in Chapter 3. The activities are sufficiently solid to warrant treatment in scientific journals. The AI Journal started recently a dedicated section, called "*The Competition Section*" [95] devoted to the description of competitions, their criteria and results, including comparisons to previous competition rounds, and why they are interesting to the AI community. Also, the IEEE Robotics and Automation Magazine has recently introduced a section dedicated to competitions.

So far, these competitions usually produce rankings among participants. They have the effect of stimulating participation, and thereby to face challenging problems, but they do not present the characteristics of a scientifically sound experimental approach. One way to promote a better scientific approach is to treat competitions as scientific experiments. However, combining the competition concept with scientific experimentation or benchmarking is neither easy nor devoid of methodological pitfalls.

A scientific experiment is usually designed to test a hypothesis and executed and documented such that the experiment is repeatable and its results can be reproduced by anyone else who is able to set up and execute the experiment. *Repeatability* requires a thorough description, including measurement data, of all features relevant to the experiment, such that an equivalent setup and execution of the experiment by an independent researcher is possible. Note that often experiments themselves consist of the execution of a number of trials with exactly the same conditions, usually in order to avoid or minimize the effects of chance or measurement errors. By *reproducibility* we usually refer to the ability to produce the same results as obtained in an initial experiment by the independent execution of a second experiment. The results to be reproduced can be the product of an analysis process on the data obtained through experiment execution. In some science disciplines, experiments must have been successfully reproduced by an independent research group before the results can be published. Ensuring reproducibility and repeatability in the context of robot competitions is an extremely complex task. For this reason, most existing robot competitions tend to neglect the issue. RoCKIn tries to openly face these problems, in an effort to bring a more solid experimental foundation to robot competitions.

It is evident that competitions and experiments differ under many respects. Competitions are designed to produce a ranking at a specific moment, while experiments are aimed at verifying some property. An experiment should be repeatable, while a competition is either held once or not meant to be repeated under exactly the same conditions. The results of an experiment should be reproducible, while the specifications of competitions are often (probably intentionally) vague. An experiment evaluates a specific hypothesis, while a competition usually evaluates general abilities to perform a task. An experiment requires a detailed description of the whole system, while in competitions the systems are not necessarily known, they only have to perform. An experiment requires the explanation about why a result has been obtained, a competition often provides only a ranking of competitors. In general competitions push to the development of solutions, experiments to result sharing, and to knowledge increase.

[1] According to http://robots.net/rcfaq.html hundreds per year, although information provided on this web site does not always withstand scrutiny.

Benchmarking can be defined as an objective performance evaluation of a system/subsystem under reproducible conditions. Benchmarking plays a fundamental role in robotics. It enables an objective comparison of different systems on a common, predefined, setting, and promotes reproducibility and repeatability. Benchmarking provides a set of scores (e.g., numerical values or pass/fail) together with a proper interpretation to perform an objective evaluation. It can be used to certify robot properties and functionalities.

We believe there are advantages in using competitions for benchmarking, as this endows competitions with an enhanced methodological foundation while allowing benchmarking to partake in the many favourable characteristics of competition. Benchmarking competitions need to be designed from the beginning to properly support the benchmarking activity. This is the reason why the benchmarking competitions in RoCKIn will be devised as a specific way of performing experimental evaluation. This means that RoCKIn competitions will provide quantitative evaluation metrics, be repeatable, and enable the comparison of different systems as well as different solutions to the same problem.

While driving scientific experimentation in competitions is one of the guiding forces of the RoCKIn project, it is not the only one. It is considered equally important to maintain the motivating elements of competitions, like "excitement" and "fun", which are key reasons for robot competitions being so successful and appealing both to researchers and to the general public. At any stage of the work of RoCKIn, a balance will be sought between these two aspects.

4.2 Concepts and Terminology

This section is dedicated to the definition of a common set of concepts and terms, to be used throughout subsequent work of RoCKIn to:

- frame properly the activity of the project,
- ensure clear communication within the project,
- communicate effectively with people already involved in robotic competitions, and
- provide a proper wording for public announcements.

Some concepts and terms have already been defined, sometimes implicitly, in the RoCKIn Proposal; additionally, definitions are available in the RoboCup@Home and RoboCup-@Work rule books, in the rule books of other competitions, and in scientific papers. This section draws from these sources, merging their contribution and expanding on them from the viewpoint of benchmarking through competitions in general, and of the RoCKIn project in particular.

4.2.1 Overview of Concepts

Given that a fairly complex hierarchy of concepts will be defined, for clarity we start by presenting the most important among them. In the following diagram, each concept is represented by a rectangular box. If the box associated to concept C2 is (fully or partially) contained by the box associated to concept C1, it means that concept C2 can be defined only in the context set by C1 (for instance, C2 may be a component of C1, or a special case of C1).

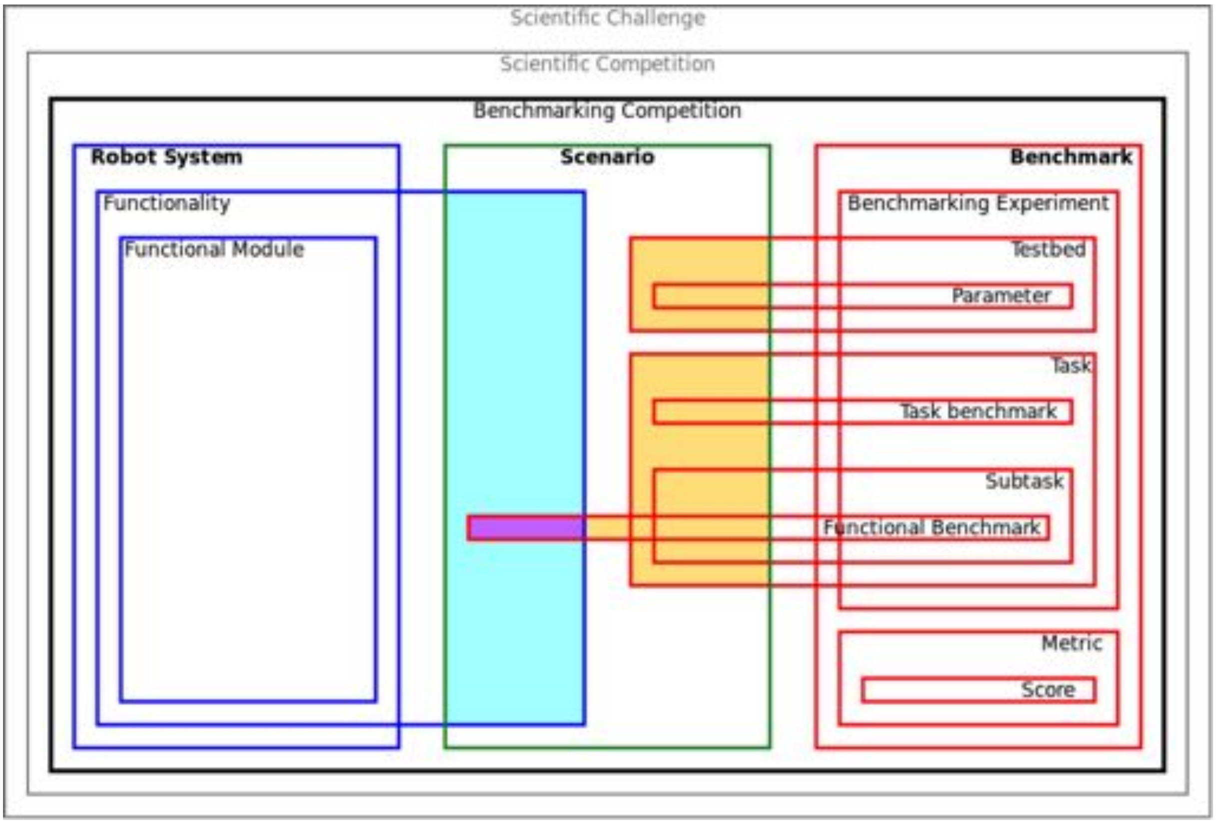


Figure 4.1: Overview of the main concepts relevant for RoCKIn.

4.2.2 Definition of Terms

This section provides definitions for the concepts which the methodological framework proposed by RoCKIn is based on. Since most of the terms are not unique to RoCKIn, the specific definition given in this section will be used for them in the project.

This section describes the conceptual level of the RoCKIn framework, while in the following of the document its contents will be implemented to the more practical level of competition design.

Definition 1 (Scientific Challenge:)**.** *A wide-ranging scientific and technological problem (often stated with intentionally vague terms) that can only be solved in the long term, which is explicitly defined for the purpose of pushing forward and directing the state of the art in scientific and technological research.*

For instance, the *scientific challenges* of RoCKIn are:

1. domestic service robots
2. innovative robot applications in industry

The rules of a *scientific competition* are designed to capture some of the key elements of the scientific achievements required to successfully tackle the class of problems considered by the *scientific competition*. The outcome of a *scientific competition* takes the form a ranking of the participants according to its rules.

Definition 2 (Benchmarking Competition, or Competition:)**.** *A* scientific competition *where the rules are designed in such a way that the rankings also take the role of measurements of the performance of participants, according to objective criteria.*

In the context of RoCKIn, measurements are part of a strategy that includes providing methods and tools for benchmarking that are used to assess each competitor's performance in a fair, as general as possible, and scientifically-driven way.

Definition 3 (Robot System:). *Any apparatus possessing the abilities and the characteristics required to participate to a given* benchmarking competition.

Such apparatus can be a single robot, a multirobot system, or a system composed of robots and external devices or services interacting with the robots (examples include, but are not limited to: sensors or actuators; providers of data, such as localization systems; networking gear; devices designed for use by humans, such as household appliances). Depending on the specific *competition*, each of the devices included in the *robot system* may be provided either by the participants and/or by the organizers.

Definition 4 (Functionality:). *One of the basic abilities that a* robot system *is required to possess in order to be subjected to a given* experiment.

These may include, for instance: self-localization, grasping, mapping or obstacle avoidance.

Definition 5 (Functional Module:). *The (hardware and/or software) components of a* robot system *that are involved in providing it with a specific* functionality.

Definition 6 (Scenario:). *All the aspects of the context where a* benchmarking competition *takes place.*

Such aspects include physical settings, environmental features (e.g. lighting, dynamic and static elements, ...), events or sequences of events that may occur, presence of robots/people/objects, and so on. The concept of *scenario* is too general to define how a *challenge*, that refers to such *scenario*, is put into practice. Such details constitute, instead, the elements of an *experiment* (see definition below).

Definition 7 (Testbed:). *A physical installation which sets a platform for scientific and technological experimentation in the context of a* benchmarking competition *by including the elements of the environment that the participating* robot systems *interact with.*

A *testbed* is not a completely passive element, in the sense that some of its components may react to the output of the robot or provide the robot with input. In this context, "output" and "input" refer to a wide range of possible ways of interaction, including, but not limited to: mechanical (opening a drawer or pushing a button); network-based (communicating with a server); acoustic (issuing or receiving voice commands); visual (showing or perceiving images). A *testbed* may include human beings and/or devices intended for human use. Additionally, a *testbed* is provided with data-collection systems that are used to generate benchmarking information, either autonomously by the *testbed* or through request of data that the *robot system* has to provide to the *testbed* according to specified methods and protocols.

In order to promote reproducibility, RoCKIn will publish detailed and unambiguous plans to make physical realization of copies of its *testbeds* feasible.

Definition 8 (Parameter:). *A* parameter *is an aspect of the specifications for a* testbed *that can take different possible configurations over a specified, discrete or continuous, range.*

Parameters are used to introduce elements of controlled variability into a *testbed*. For example, a *testbed* may specify that location of item B is always over the top surface of item A (this is a predefined condition, i.e., not a *parameter*), while the exact pose of B relative to A is a *parameter* of the *testbed*, with an associated range (e.g., three specific values). Whenever the *testbed* is used, the position of B relative to the top of A can take one of the values of the range associated to the *parameter*, according to the test conditions.

Definition 9 (Task:)**.** *An operation or set of operations that a* robot system *is required to perform in order to participate to a* benchmarking competition*.*

These operations, their expected results, the way they must be executed, and the features of the environment where the operations occur can be specified more or less precisely; for instance, type, position, and outlook of potentially useful objects can be known, partially known, or unknown before the execution of the *task*.

Definition 10 (Subtask:)**.** *A single operation or set of operations that a* robot system *has to perform in order to execute a* task*, but which by itself is not sufficient to achieve the final results of the* task*.*

Definition 11 (Metric:)**.** *A precisely defined, quantitative criterion to assess one or more aspects of the performance of a* robot system *or* functional module *in the context of the execution of a* task*.*

Operatively, a *metric* requires the application of a precisely defined algorithm to suitable experimental data describing the execution of a *task*, or comparison – according to a precisely defined algorithm – with the performance of some "reference" system or subsystem. In the latter case, the reference is actually part of the *metric*, so it should be made available along with all the information required for its setup and use.

While the algorithm defining the *metric* must always be quantitative and precisely defined, its input data can include subjective judgements by humans. This is necessary, for instance, for some human-robot interface issues, when subjective judgement of the *robot system*'s behaviour is the only available tool (e.g., when investigating if some movements are perceived as threatening).

Definition 12 (Benchmarking:)**.** *The process of evaluating the performance of a given* robot system *or one of its* functional modules*, according to a specified* metric*.*

In the context of the RoCKIn project, benchmarking is performed through *benchmarking competitions*, the rules of which are oriented towards benchmarking objectives.

Definition 13 (Benchmarking Experiment, or Experiment:)**.** *The composition of*

- *a* task *or Subtask that has to be performed by a* robot system
- *the* testbed *where the* robot system *performs the test.*

The term *experiment* is preferred with respect to "test" to underline the scientific aspects of the activity and the fact that its results will be measured and analyzed according to precise references that do not rely on subjective judgement (see *benchmarking*).

Definition 14 (Benchmark:)**.** *The union of*

- *one or more* benchmarking experiments
- *a set of* metrics *according to which the course and the outcome of the experiments – described by suitable data acquired during the experiments – will be evaluated.*

By combining the same *experiment* with different sets of *metrics*, different *benchmarks* can be defined. This allows the evaluation of the same *experiment* according to different purposes.

Definition 15 (Functional Benchmark:)**.** *A* benchmark *which aims at evaluating the quality and effectiveness of a specific* functional module *of a* robot system *in the context of one or more* scenarios*.*

A *functional benchmark* is a *benchmark* where the *experiment* component includes a *subtask* (i.e., not a full *task*). *Functional benchmark*s are aimed at evaluating the performance of single modules, independently from the other modules of the *robot system* that they are part of.

Definition 16 (Task Benchmark:)**.** *A* benchmark *which aims at evaluating the quality of the overall execution of a* task *by a* robot system *in the context of a single* scenario*.*

In a *task benchmark*, the *experiment* component includes a full *task* (i.e., not a *subtask*). *task benchmark*s are aimed at evaluating the overall performance of a *robot system*, without separating the contributions of its constituent parts.

Definition 17 (Score:)**.** *The result obtained when a* robot system *is subjected to a* benchmark *(*task benchmark *or Functional* benchmark*).*

In the context of RoCKIn, *scores* are numerical results produced according to objectively-defined *metrics*.

4.2.3 On Specifications and Realisations

Please note that all the terms defined in the preceding part of this section refer to physical realizations, i.e., their definitions make reference to actual, physically occurring objects or procedures, not to descriptions of objects or procedures. For instance, a testbed is a physically existing object, and an experiment is the actual execution of a procedure involving the operation of a physical robot. It will often be necessary to speak about the description of an instance of one of the previously defined entities. In this case, the additional terms description and specification will be used. For instance, a testbed specification is a complete set of directions to build an instance of a testbed, and all instances built according to these directions should be exact physical copies.

Chapter 5

General Scenario Features

The purpose of this chapter is to describe *(i) general features* of the scenarios and to define an *(ii) initial list of functional modules* as candidates for the functional benchmarks referred to in previous chapters.

The general features described in the next three sections below illustrate the spectrum of aspects relevant for defining the scenarios of RoCKIn competitions. Particular scenarios may warrant *a priori* constraints on the range of values for a feature; they may be derived directly from the intented application scenarios. For example, a robot supposed to perform a variety of household chores must be able to navigate domestic environments with doors and different kinds of floors. These features of the environment impose constraints on the features of robots usable in such environment; for example, on size, weight, and the suitable type of mobility. Any feature not explicitly mentioned in a scenario description is assumed to be unconstrained. Note, however, that leaving features unconstrained can be very problematic for the organizers of a competitions. Participating teams may arrive with robots obviously not fitting into the environment, damaging the environment during operation, or, worst of all, harm people attending the competition. Underspecified task features can be a reason for very hot rule debates during the competition; a situation organizers want to avoid.

The list of functional modules given in Section **??** is, despite being initial and remaining open for discussion, extension, modification, and reduction, important as the desire to benchmark certain functionalities is likely to have a strong influence on the scenario design, rule definition, and scenario implementation.

5.1 Task Features

The tasks used in the competitions should be described by the following features:

Task Preconditions Which state of affairs must be present in order for the task to be executable?

Task Achievement Criteria How do we know that the task is completed? Which state of affairs is holding after successful completion of a task?

Accuracy: What is the required accuracy of task execution?

Repeatability: If the task is repeated, how precisely should the same result be achieved?

Resource Constraints: Which constraints on the use of resources apply? Resources include time available for task execution, amount of energy needed to perform the task, etc.

Sensing Requirements: Which sensing capabilities must an agent have to perform the task?

Effector Requirements: Which actuating capabilities must an agent have to perform the task?

Navigation Complexity: How difficult is the navigation aspect of the task, if any?

Manipulation Complexity: How difficult is the manipulation aspect of the task, if any?

Interaction Complexity: How difficult are the necessary interactions with other agents, if any?

Heterogeneity: What is the range and distribution of subtasks to be solved in order to solve this task?

Context Sensitivity: How strong do environment features and other circumstances influence task performance?

Safety Constraints: Which safety regulations must be obeyed? What are their implications for task execution and its performance?

Nota bene: All these aspects refer purely to features of the task, not to the ability of a particular agent to perform the task in a certain way.

5.2 Environment Features

The environment for the RoCKIn competitions need to be thoroughly described. Below, we give two lists of features, one pertaining mainly to objects of various kinds, while the other addresses general features of the environment directly. For any object constituting the environment, including all floors, walls, ceilings, windows, doors, elevators, lights, all objects in the environment like furniture, any objects that can be (passively) moved by some other agent, and objects that can move themselves, including humans, animals, and robots, the following list of features should apply:

Type/Class: What is the type or class of the object?

Uniqueness: Is the object unique (in the environment) or do several perceptually ambiguous objects exist?

Size: What is the size of the object? It is not yet clear how to specify this in general. For convex objects, the size could possibly be specified by an enclosing cuboid or ellipsoid. An object's size imposes constraints on robots that are supposed to manipulate it.

Weight: What is the weight of the object? An object's weight impoises constraints on robots that are supposed to manipulate it.

Shape/Form: What is the shape or form of the object? Where can a robot grasp it, if necessary? Which grasps are suitable for which kind of activity? Aside of referring to simple common shapes, like cuboids, cylindric objects, spheres, etc. it is not clear how complex the language for describing shapes must be to meet the demands of robotics.

Material Features: What material is the object made of? What characteristics or features of the material, like color, texture, reflectivity, translucency, elasticity, fragility, surface friction, or aggregate state may be relevant?

Movability: Is the object movable? An object itself may seem to be movable (given its shape, size, weight, etc.), but may in fact not be movable if fixated (nailed, screwed, glued) to another (nonmovable) object.

Mobility: Can the object move by itself? If yes, by what means? How fast? How frequent does this happen? How can such movement (or the intention to perform such movements) be recognized? This feature applies for other robots and machines in the environment, as well as for animals and humans.

At some later point in the project, we might define defaults for these features, and leave out any of these features if it has the default value for the object on hand.

The environment description must then consist at least as a list of such objects. Note, that the above description does not yet contain any information on geometry, like position and orientation of objects in the environment – this requires suitable reference frames – or on spatial relationships between objects. Most likely, we will look at well-established technology, e.g. COLLADA [96], or a suitable extension, for describing environments, or at least its geomtric aspects.

Furthermore, the following physical properties of the environment are or may be of relevance:

Temperature: What is the temperature in the environment? The temperature can influence the ability to perform a particular task or the likelihood of an event to happen. Leaving elderly people or children in very hot or very cold environment can have very unhealthy effects.

Air Pressure: What is the air pressure in the environment? Air pressure does have non-trivial influence on the performance of sonars, for example.

Humidity: What is the humidity in the environment? This may influence the ability for grasping and holding objects.

Lighting Intensity: Measured in Lux.

Lighting Temperature: Measured in Kelvin. Could be important to estimate the effect of a particular light source on e.g. color perception.

Noise Level: The noise level may have significant influence on the ability to interact with humans.

Note, that if any of there features does not have constant values across the whole space of the environment, it should be specified using a suitable distribution.

5.3 Robot Features

The robot description of a scenario should consist of the following features:

Robot Type/Class What type of robot is foreseen? This feature could make use of a suitable classification scheme for robots. A widely acceptable scheme of this kind does not yet exist, but various categorizations are being used in practice, for example, in the Strategic Research Agenda by the euRobotics aisbl. For RoCKIn, it is probably sufficient to provide type or class characterization just for the targeted types of robots. In a benchmarking competition sponsored by a company the type of robot might be constrained to a particular model of robot provided or sold by this company. Robot competitions where this is the case include the Festo Logistics League or the FIRST Lego League.

Mobility Subsystems Is the robot mobile at all? Which type of mobility does it have? Some possibilities for mobility include wheeled, tracked, legs (1, 2, 4, 6, 8, n), brachiation, aircraft (quadcopter, airship), underwater. A competition for domestic robots might constrain mobility types to $\{none, wheeled, humanoid\}$.

Manipulation Subsystems Is the robot capable of manipulating objects or the environment? What kind of manipulation system does it have? Possible manipulation systems include a robot arm, a dual-arm subsystem, a forklift, a kicking device, etc.

Grasping Subsystems Does the robot have any devices for grasping objects? Which kind of grasping devices does the robot have? Some possible grasping devices include a parallel gripper, a three-finger hand, a five-finger anthropomorphic hand, or a pneumatic vacuum cup.

Sensor Subsystems Which sensing abilities does the robot have? Which kind of capabilities do they have? Where are these sensors mounted? Some possibilities include sonars, infrared sensors, 2D and 3D laser scanners, unidirectional, omnidirectional, stereo, 3D and infrared cameras, touch and force sensors, gyros, accelerometers, etc.

Communication Subsystems Can the robot be networked? Which kind of technical communication devices does it have? For most mobile robots, only wireless communication facilities (WiFi, Bluetooth, 3G network, etc.) will be of interest, while non-mobile robots may use a much wider range of communication interfaces (Ethernet, EtherCAT, USB, CAN).

Power Supply How is the robot supplied with power? Does it have a battery and is energy-autonomous? If yes, for how long? If no, what are the specifics of the required power supply?

Computational Subsystems Does the robot have some kind of computational device (microcontroller, embedded PC)? If yes: How many? What kind? How connected?

Robot Functionality : A list of functional modules the robot may employ. The next section provides an initial repository of functional modules which can be used here.

In addition to this list, all features applicable to an object are relevant for a robot as well: uniqueness, size, weight, shape/form, material features, movability. The object features type/class and mobility are redundant, as they are already in robot feature list.

5.4 Towards a Functional Reference Platform

5.4.1 Motivation for a Functional Reference Platform

Some capabilities of robots can be measured and benchmarked by observing the behavior of the robots. For example, the accuracy of object manipulation can be benchmarked by measuring the position of picked-and-placed objects. Other capabilities can be measured and benchmarked only by considering data generated by the robot. For example, accuracy (defined in terms of metric precision) of maps used for path or trajectory planning can be benchmarked by evaluating features of the maps generated and stored by the robot.

The functional reference platform (FRP) aims at defining functionalities needed by robots participating in the RoCKIn benchmarking competitions and, in particular, at defining functionalities that require generated data to be benchmarked (and the skeleton of the interfaces to get access to these data) and at defining functionalities that require observed data to be benchmarked. The functionalities are defined according to their corresponding functional modules, which are illustrated in the following.

Besides defining the information required to benchmark functionalities considered in the RoCKIn competitions, the RoCKIn functional reference platform aims at:

- defining ways for considering the performance of a specific functional module as "reference point" against which performance of other modules that perform the same functionality can be compared,
- promoting the exchange and reusing of functional modules between different teams,
- promoting the sharing of winners' functional modules to develop the community, and ease the access of new teams,
- possibly using information provided by the testbed (e.g., the pose of the robot) instead of information generated by robot modules, to reduce exogenous interference in functional benchmarks.

The definition of the RoCKIn FRP is an activity pertaining to Task 1.2, described by the DoW of the RoCKIn project as follows:

> *Functional specifications will not be targeted at the implementation level, where differences among different systems and architectures are too large to allow meaningful definitions; they will, instead, be focused on the definition of higher-level functional building blocks and their interactions. A functional specification could identify a module for each high-level task defined in Task 1.1 as part of scenarios. Different robots compliant with the specifications published by Task 1.2 will have a similar functional structure, thus opening the way to the compared evaluation of the contributions of single capabilities of such robots on the overall result of the robots in system-level competitions. Task 1.2 will produce these functional specifications in the form of a RoCKIn functional reference platform composed of the functional modules that are relevant for the RoCKIn scenarios.*

5.4.2 Description of Functional Modules

In the RoCKIn functional module repository, the description of functional modules is focused on defining specific capabilities of a robot that may participate to benchmarking experiments. While such capabilities often correspond to concepts widely used in robotics, it is important to note that the FRP is only a methodological tool for the design of benchmarking competitions. While it is possible to find contact points between the RoCKIn FRP and commonly used robot architectures, the FRP is not a proposal for a "standard architecture", or even a formalization thereof. In this way, the FRP focuses on the functionalities that the functional modules expose externally, and on ways to describe and (when possible) benchmark them; on the other hand, the FRP does not pose constraints on the internal elements of robots that are necessary to provide such functionalities.

A functional module does not necessarily correspond to a single element of a robot, to whole elements of it, or even to actually existing elements. The functional module corresponds instead to whatever elements of the robot contribute to make it capable to provide the functionality associated to the module. For this reason, a functional module is described using features of the associated functionality. Precisely:

1. **Functionality Description :** A description of the functional module in terms of the functionality that it provides the robot with.

2. **Available Information:** The information provided to the functional module, not necessarily generated internally by the robot. In fact such information can be provided (fully or partially) by the testbed: when this is the case, a description of the interfaces between the robot and the producers of such information (data types, formats, channels, rates, ...) is provided.

3. **Data for Benchmarking:** Specific pieces of information consequent to the operation of the robot to implement the functionality, not necessarily coming from the functional module itself (they can be generated, fully or partially, by the testbed). When the data for benchmarking comes from the functional module, a description of the interfaces between the robot system and the users of such information (data types, formats, channels, rates, ...) is provided.

Functionality descriptions usually correspond to established categories for robot capabilities.

The *available information* to a functional module may be provided by other functional modules of the robot, by the RoCKIn testbed, or both. For instance, the information available to a self-localization module may include data from onboard sensors, data from external sensors provided by the testbed, and/or data from other functional modules. In some cases, the same type of information may be made available to a module, alternatively or jointly, both by the testbed and by other modules of the robot. When the available information includes data provided by the testbed, the robot is required to be capable of receiving and processing such data in a proper way.

Depending on the specific Benchmarking Experiment, Data For Benchmarking is produced by the functional module or not. In the first case, the Robot System is required to provide such data through the defined interfaces; in the second case, the Testbed generates and provides the data (e.g., as output of an external measurement system).

It is important to note that the definition of a functional module of a Robot System includes elements that depend on how the robot has been designed and realized, but also elements that depend on how the Benchmarking Experiments that the Robot System will be subjected to have been set up: for instance, what external devices, e.g., cameras, are accessible to the Robot System. This is coherent with the idea of functional modules as conceptual tools for Benchmarking Competitions, and sometimes it makes them unsuitable for the description of a Robot System as isolated from the Testbed.

By separating the concept of functional modules from actual architectural components of robots, the fact that in some cases data used by some functional module (e.g., the mapping functional module) takes a form that is strictly linked to the algorithm used to produce it becomes unimportant. To benchmark such a module, in fact, RoCKIn will define the Data For Benchmarking so that it does not include the raw output of the architectural module. On the contrary, Data For Benchmarking will consist of derived data: easily computed data, but independent from internal data format, that can be used to gauge the quality of the performance of the module. For instance, the Data For Benchmarking required to evaluate the map generated by a mapping functional module might be composed of the distances between the elements of a set of predefined/closest features of the environment, and not include the entire map.

If computation of the derived data is easier – whatever the form taken by the module output – than the conversion of the full output into a standardized form, the practical complexity of setting up Benchmarking Experiments for such module is reduced while maintaining repeatability and reproducibility.

5.4.3 Towards the RoCKIn FRP

Being a general specification, the functional reference platform necessarily takes a rather abstract stance towards typical robotic problems. So, in order to give to the following description a more concrete grounding, we will start by showing how the modular structure that the FRP systematizes already emerges from the requirements provided by the current robot competitions.

We will consider two example tasks: one from RoboCup@Home (2013 draft rulebook), and one from RoboCup@Work (2012 rulebook). The choice of RoboCup is due to the fact that it is among the most successful and internationally recognized robot competitions: on one side, this means that a significant part of the robotics community already participated (or considered to participate) to its Challenges and/or to their definition and shaping; on the other side, it ensures that the requirements of the RoboCup Challlenges are close to the current state of the art. An additional reason for the choice of theese examples is that the scope of RoCKIn in terms of scenarios is similar to that of RoboCup@Home and RoboCup@Work: therefore the capabilities needed by participating Robot Systems are similar as well.

Example 1: RoboCup@Home We will consider here the test called "Cocktail Party" in RoboCup@Home rulebook. In this test, the robot has to learn and recognize previously unknown persons, and deliver drinks. 5 people are distributed in a room, either sitting or standing. The robot waits for being called by the persons (through voice or waving). When a person calls the robot, such person introduces her/himself by name and orders a specific type of drink. For each drink, the robot has to navigate to another room, grasp the correct drink, return to the party room, find the person (who may have changed her/his place) and finally deliver it.

In terms of capabilities (which correspond to the concept of functionalities that RoCKIn uses to model them), the RoboCup@Home Cocktail Party requires that the robot is capable of: detecting, identifying, and recognizing humans and objects; manipulating and transporting objects; navigating without collisions; understanding spoken natural language from distant speakers (i.e., not wearing microphones). The test arena (i.e., the Testbed) and the range of objects and features that it can include are loosely specified. Positions of objects and obstacles are not known. However, access to the arena is allowed to participating teams before the actual tests, so that they have the possibility to perform adjustments to their robots (e.g., define and load a map).

Example 2: RoboCup@Work One of the tests defined by RoboCup@Work competition is the Basic Navigation Test (BNT), which we already referenced to in this document. According to the 2012 rulebook, a robot subjected to the BNT must, in accordance to the commands specified by an unambiguous well formatted task string that it receives at the start of the test: enter the arena through a gate; move to the places specified in the task string, in the order as specified by the string, orient itself according to the orientation given, pause its movement for the time in seconds as specified by the pause length, and finally leave the arena through the gate. Specifications of the arena are known beforehand. Position of the obstacles is not known, but they can only occur in predefined areas. The arena is provided with a predefined configuration of visual markers to spport localization.

The robot capabilities required by the RoboCup@Work BNT are: recognizing the opening (gate) of the arena and navigate to a goal pose avoiding obstacles. To work correctly, this capabilities also require an additional one, namely self-localization. Not strictly required, e.g., purely reactive behaviors, but useful are the capabilities to use a map of the environment and to plan how to reach a specified pose.

As we will see shortly, for most of the capabilities required by the RoboCup tests there exists a corresponding functionality, and therefore an associated functional module of the RoCKIn functional Reference Platform. This correspondence is not surprising and supports the partition in modules of the RoCKIn FRP, which follows best practices in the design and analysis of robot systems.

5.4.4 Functional Modules of the RoCKIn FRP

We have now reached the stage where a general idea about the type of capabilities that can be useful to define has been provided. Therefore, we proceed with a description of functional modules – each of which representing one functionality – that compose the RoCKIn functional Reference Platform.

Please note that the RoCKIn FRP does not describe all the possible functional modules that a Robot System participating to the RoCKIn Competitions must possess. Conversely, the FRP it is only concerned with the modules that implement benchmarked functionalities: i.e., those that RoCKIn deems feasible and/or interesting to assess. For instance, the FRP does not include any Sensor modules, though sensory capabilities are obviously needed by Robot Systems participating to the RoCKIn Competitions. For the same reason, the set of functional modules described in the following should not be regarded as definitive: in fact it will evolve, when necessary, following the process of design of the actual RoCKIn Competitions that is the objective of Work Package 2.

RoCKIn explicitly tackles the issue of benchmarking networked and/or distributed Robot Systems. Thus, the FRP poses no limitations on the actual physical location of the elements of a Robot System that provide it with a functionality. These elements can (as is the case of most current autonomous robots) be physically contained in a single unit, but they can also be distributed over more than one mobile or fixed units. In particular, as already said, modules or sub-modules can be part of the Testbed where the Competition is taking place: e.g., in the case of sensors which are part of the environment.

Here follows a description of functional modules comprised by the RoCKIn functional modules. Such description, as already explained, is composed of three elements: Description, Available Information and Data For Benchmarking.

Functional Module: self-localization

- ***Description:*** determine the pose of one or more links of the robot w.r.t. the surrounding physical environment.
- ***Available Information:*** available sensor data ; map of the Testbed.
- ***Data For Benchmarking:*** estimated pose of the links in the reference system of the map.

Functional Module: mapping

- ***Description:*** build a map, i.e., and internal representation, of the physical environment surrounding the robot, usable to perform self-localization and/or trajectory planning.
- ***Available Information:*** available sensor data; positions of the Robot System and of the sensors external to it, expressed in a given reference system.
- ***Data For Benchmarking:*** map of the environment or derived data obtained from it .

Functional Module: trajectory planning

- ***Description:*** determine (if they exist) one or more feasible trajectories for one or more specified links of the robot, from a starting pose to a target pose.
- ***Available Information:*** map of the Testbed; initial and target poses of the links in the reference system of the map.
- ***Data For Benchmarking:*** an ordered sequence of feasible poses of the links, in the reference system of the map.

Functional Module: trajectory following

- ***Description:*** movement execution for one or more links of the robot to follow a specified trajectory, in the absence of obstacles.
- ***Available Information:*** trajectory to be executed and current pose of the robot, expressed in a given reference system.
- ***Data For Benchmarking:*** an ordered sequence of poses of one or more links of the robot (in the same reference system of the initial pose).

Functional Module: **obstacle avoidance**

- ***Description:*** movement execution for one or more links of the robot following a specified trajectory, in response to unpredicted obstacle, to avoid them.
- ***Available Information:*** available sensor data; map; trajectory to be executed and current pose of the robot, expressed in the reference system of the map.
- ***Data For Benchmarking:*** an ordered sequence of poses of one or more links of the robot (in the reference system of the map), collision data .

Functional Module: **object grasping**

- ***Description:*** grasp and hold an object, of known shape, consistence, and pose, without damaging it.
- ***Available Information:*** available sensor data; description/model of the object.
- ***Data For Benchmarking:*** success of the grasping action as measured by a specified procedure (e.g., holding the object for X seconds or holding the object firmly against a force of given intensity and direction).

Functional Module: **HRI: exchange of information**

- ***Description:*** interact with human beings through exchange of information of a specified type (e.g., voice signals).
- ***Available Information:*** available sensor data; semantics, in terms of symbolic transcription, of the phenomenon (i.e., voice) which generated sensor data.
- ***Data For Benchmarking:*** interpretation of sensor data according to the specified semantics.

Functional Module: **HRI: exchange of objects**

- ***Description:*** similar to object grasping Functional Module, the difference being that the object to be grasped is, respectively, initially held by a human or to be accepted by a human.
- ***Available Information:*** available sensor data; description/model of the object.
- ***Data For Benchmarking:*** success of the grasping action as measured by a specified procedure (e.g., holding the object for X seconds or holding the object firmly against a force of given intensity and direction).

Functional Module: **Semantic Map**

- ***Description:*** A semantic map is a 2D/3D/topological map augmented with labels about landmarks in the environment.
- ***Available Information:*** perceptions during a robot run in the environment (e.g., odom + laser + Kinect), a priori information about objects and locations
- ***Data For Benchmarking:*** 2D/3D/topological map + a vector containing a set of objects. Each object is characterized by the following attributes: position, orientation, category, unique ID.

Functional Module: **Simple Command Interpretation and Execution**

- ***Description:*** Evaluate the capability of the robot of interpreting and executing simple commands
- ***Available Information:*** Spoken command, semantic map of the environment, known objects, templates of simple behaviors (e.g., go to <X>, grasp <Y>, . . .)
- ***Data For Benchmarking:*** Instantiation of the corresponding robot behavior in a predefined format (to be defined, e.g. RoboFrameNet) – execution on simulator and on real robot

Functional Module: **Robot-Person-Object Interaction**

- ***Description:*** Execution of tasks that involve robot-person-object interaction
- ***Available Information:*** Command templates (e.g. bring <object> to <person>, look for <person> in <location>), DB of known people, DB of known objects and locations (map of the environment) Specific command (e.g., look for John in the kitchen)
- ***Data For Benchmarking:*** Execution of the corresponding behavior

Functional Module: **Decision making**

- ***Description:*** In the context of multiple cooperating robots, the robots need to decide their roles in the team. In order to achieve the desired task in the limited amount of resources (time/energy and computation), the decision making system must be adaptive, robust and fast. This module tests a multi-robot system's decision making optimality.
- ***Available Information:*** known cooperative task, map of the environment, self and teammate poses, obstacle and other object's positions in the environment.
- ***Data For Benchmarking:*** (open for discussion)

Functional Module: **Communication**

- ***Description:*** In the context of multiple cooperating robots, the inter-robot communication needs to be extremely robust. The benchmarking for this module aims to test the communication robustness in terms of i) communication delays and ii) communication losses, within the presence of various network-jamming equipments, constrained and cluttered physical spaces, distance between communicating robots and so on.
- ***Available Information:*** Available sensor information, a set of defined data to be communicated among the robots, pose information of all the robots, map of the environment and obstacle positions.
- ***Data For Benchmarking:*** (open for discussion)

Functional Module: **Object/Human detection and recognition**

- ***Description:*** detection and recognition of known objects as well as human faces.
- ***Available Information:*** available sensor data; known object's color/size/surface pattern, etc., (object templates), database of human faces
- ***Data For Benchmarking:*** the estimated pose/position of the object/human face.

Functional Module: **Object/Human tracking**

- ***Description:*** tracking of known objects as well as human faces.
- ***Available Information:*** available sensor data; known object's color/size/surface pattern, etc., (object templates), database of human faces
- ***Data For Benchmarking:*** the estimated poses/positions of the object/human face over the full trajectory of the object/human's movement.

Functional Module: **Random color/texture object detection/recognition**

- ***Description:*** tracking objects of random (not pre-specified) color and texture under various lighting conditions. Only a general classification of the object is pre-specified, e.g, mugs, cans, bottles, boxes, dishes and so on.
- ***Available Information:*** available sensor data; object's general classification.
- ***Data For Benchmarking:*** the estimated poses of the object.

Appendix A

Survey of Benchmarking Competitions in Science

One of the main objectives shared between most competitions is to find the *best* competition entry with respect to some task. This requires a method for *assessing and comparing the performance* of the competition participants in executing the task, usually within a well-defined environment. A prerequisite for performance assessment usually is a method for *checking the correctness* of the task execution or of the outcome of the task execution.

As a simple example, let us assume we want to compare CPUs with respect to their capability for floating point computations (i.e. the FPU):

- As a first step, we would be interested to check the correctness of the FPU, i.e. whether it is able to produce (arithmetically) correct results within the domain specified for the FPU operations. It can already be quite difficult to come up with a comprehensive statement on correctness, and solving this problem is part of software testing, esp. wrt. to coverage.

- Assuming we have several CPUs for which testing has proven that they produce correct results, we would interested in comparing their performance. At least two criteria are be of interest, *precision* and *runtime*.[1] The precision is usually specified for CPUs and can be easily assessed, while for the runtime we may have to run experiments to measure them.

An important design issue for a benchmark is how to combine different assessment criteria (or performance measures) in a suitable way to come up with an appropriate ranking of the benchmarked CPUs. This problem aggravates, if the benchmark encompasses (as is frequently the case) a multitude of test cases, each of which may assess the performance of executing a particular single operation or a particular combination of operations. Having a larger range of test cases covering many different situations can also be helpful to reduce *overspecialization* or *overengineering*, which describes situations where the systems under consideration are designed and engineered such to optimize its performance on particular benchmarks.[2] If a benchmark is supposed to help making the best choice among several

[1] We use runtime here just for simplicity; real CPU benchmarks often specify the CPU cycles required for particular FLOPS, accounting for the fact that most CPUs can be run at different clock cycles.

[2] Overengineering is not necessarily a bad idea, but it depends on the task domain under consideration. If a system performs excellently on the whole domain relevant for the task, "overengineering" simply yields excellent solutions. Overengineering becomes a problem, if a system is designed such that it performs well mainly on the situations covered by the benchmark, but performs very badly in other situations relevant

alternatives, then the tests used in the benchmark should reflect similar computation situations as needed in the targeted application system.

If a benchmark is "standardized"[3], then it is not necessary to perform the benchmark for all (or even several) objects under consideration at the same time and place. A new system appearing can be benchmarked at any time and inserted into a continuously updated ranking according to the results achieved. The computer hardware community is maintining such ranking e.g. for supercomputers (see [24]). Some examples of well-known benchmarks and benchmark tools include:

- Computer hardware benchmarks: Whetstone [25] [26], Dhrystone [27] [28], SPEC [29] [30] [31], Coremark [32] [33], LINPACK [34] [35], PARSEC [36] [37]
- Network performance benchmarks: NetPerf [38] [39], NetSpec [40], PassMark [41]

In many scientific disciplines, competitions have become popular in the recent past. Some examples include:

- **Computer Security:** For example, the Capture the Flag (CTF) competition [42] at the DEF CON conferences. [43]
- **Information Retrieval:** See e.g. the Text Retrieval Conference (TREC) [44].
- **Computer Games:** Various competitions at the Conference on Computational Intelligence in Games (CIG) [45].
- **Computer Vision:** See e.g. the Workshop on Performance Evaluation of Tracking and Surveillance (PETS) [46].
- **Machine Learning and Pattern Recognition:** See e.g. the machine learning challenges posted on the TunedIt platform [47].
- **Automated Planning:** See especially the International Planning Competition [48] held biannually from 1998 to 2008, since then tri-annually at ICAPS conferences, with three competition tracks: deterministic, learning, and uncertainty.
- **E-Commerce:** In particular, the Trading Agent Competitions [49], [50], [51], held since 2002 and recently extended by a variant specializing on energy markets [52].

Many of them are organized such that each competition entry must perform well-defined tasks, while certain performance measures are taken and used to rank the participants. These competitions can actually be viewed as benchmarks executed for a number of competition entries at a particular point in time.

to the targeted applications. Benchmarking competitions should try to avoid such over-engineering, but this is not easy to achieve.

[3]This does not necessarily imply a formal standardization by a norming institution such as ISO or others. A wide adoption of a benchmark in the community may establish it as *de facto* standard.

Appendix B

Survey of Scoring and Ranking in Sports

The survey of prior experiences would not be complete without having a deeper look into the world of sports. Most of the sports considered in the following sections are well known, and we do not need to describe in detail the scenario, as participants, environment, and tasks in most cases. However, it is interesting to survey the various scoring and ranking systems applied in different sports.

B.1 Track and Field

A classical and very simple sports competition is the 100m sprint: In a single run, up to eight athletes compete at the same time. The environment is the same for all athletes. The task is to run a distance of 100m as fast as possible. The performance assessment is simply done by measuring the time needed to complete the run. This competition has all ingredients for a very successful competition, including the direct, simultaneous performance of several athletes. Not surprisingly, the 100m sprint is scheduled as a highlight event at every track and field meeting including the Olmypics.

It is interesting to see how the character of the competition changes with increasing distance: Distances of 100m, 200m and 400m are considered as sprints; the time needed by top athletes is so short (less than 10 and 20 seconds, respectively) that athletes fully concentrate on themselves and try to give their best performance over the whole distance. Distances of 800m and 1500m are considered medium-distance, which athletes cannot fully sprint. Tactics come into play. In top track-and-field events, pacemakers are hired who drop out after a round or two of 400m, but are supposed to make the overall race faster. This is especially often the case if a top athlete announces to run a new world record. Running longer distances, like the 3000m steeplechase, the 5000m and 10000m have a quite different character, where runners try different tactics during a large part of the run and a run is often decided in the very last round.

While the track events always involve simultaneous competition between athletes, the field events, like long jump, high jump, pole vault, shotput, discus, javelin, and hammer throw involve an environment which can be used by a single athlete only at any time. The tasks for each of these are obvious; the measure is usually the length or height jumped or the distance the object was thrown by the athlete. In order to ensure almost equal environment conditions, the participating athlete compete in a rapid sequence determined by a draw. Each athlete usually has several trials, e.g. six.

All of the track-and-field events usually allow only a limited number of participants. If a larger number of athletes participate, as is the case in the Olmypics, then athletes must qualify for the finals in preliminary and qualification rounds. The running events use timing equipment to measure the times of athletes. All events use referees in order to measure distances and to judge the correct execution of the athletes' performances (early start in sprints, over-stepping in shotput, javelin, etc.).

B.2 Tennis and Golf

Tennis is a sport played in tournaments consisting of a sequence of pairwise matches between players (singles) or teams (double and mixed). Task and environment are shared. A match consists of sets (best-of-3 for women, best-of-5 for men), which themselves consist of so-called games. Interestingly, a tennis match can be won although loosing the majority of games, e.g. (0:6, 6:4, 6:4), giving 12:14 in games but 2:1 in sets, or (0:6, 7:6, 7:6), giving even 14:18 in games but 2:1 in sets. The player loosing a match is out, the winner proceeds to the next round. Only a global winner (and runner-up) are determined by a tournament, and there is no overall ranking of all participating players.

In order to avoid that two very strong players meet each other early in the tournament, some players are seeded according to their ranking in a globale player ranking list (WTA, ATP). These rankings typically take into account all performances of the player in the last 12 months. Despite this system, it happens quite often that even top-ranked players loose a match and drop out early in a tournament. Therefore, a tournament is viewed as reflecting more a momentary snapshot rather than a comprehensive assessment of the players' strengths.

Tournament golf is in many aspects similar to tennis, but usually uses a direct measure of performance: the number of shots needed to complete a round of 18 holes on the course.[1] The player performances of four consecutive days are added, and the overall winner is determined as the player using the fewest shots to complete all four rounds of the course. The environment is shared (all participants play on the same course), but weather conditions can of course vary quite a bit between early-morning and late-afternoon rounds. Direct match-ups between players have no direct influence on the game, except for mostly psychological effects arising from the particular player match-up and the attention these players draw amongst spectators.

B.3 Biathlon and Show Jumping

Biathlon and show jumping are interesting sports because they combine a time-based performance measure — the time needed to perform the course — with a penalty scheme for committing faults. In show jumping, participants are pairs of horses and riders who have to complete a course of obstacles without knocking them down within an alloted period of time. Faults like knockdown of a an obstacle or refusal to jump are penalized with fault points. Exceeding the alloted time permitted to complete the course is converted in extra fault points. The ranking is based primarily on the lowest fault points collected by the participants. Ties between participants having the same fault points are often resolved based on the time needed to complete the course. As it is practically impossible

[1]Some special tournaments, like the Ryder's Cup, are played in a different mode.

in show jumping to compensate for a fault by a faster completion of the course, avoiding faults is the prime concern of participants.

In Biathlon, the situation is somewhat different. Participants have to complete a course of varying length (from 6 km in women's relay to 20 km in men's individual) in cross-country skiing with intermersed rifle shooting rounds. Either 2 or 4 shooting rounds have to be completed, depending mainly on the overall distance of the race. Half of them are in prone position, the other half in standing position. Each shooting round consists of shooting at five targets in a distance of 50 m. Targets are sized 45 mm for shooting in prone position and 115 mm for standing position. The overall time needed to complete the course is used to determine the ranking of the participants. Thus, there is an incentive to perform the shootings quickly. However, missed shots are penalized, depending on the particular competition format, in one way or another: Most frequently, a missed shot results in having to do an extra penalty round of 150m, which for top athletes amounts to a time penalty of 20 to 30 seconds. In some events, every missed shot directly adds one minute to the clock for this athlete. Sometimes, the athlete can use extra cartridges to complete missed shots, and only targets not hit after shooting off all cartridges result in penalty rounds. Contrary to show jumping, it happens quite frequently that athletes can compensate for a few missed shots by a much better skiing performance.

B.4 Football and Basketball

Football[2] is the most popular sport in the world. Two teams of eleven players each play a match of 90 min tyring to score goals while defending their own goal; the team scoring more goals wins. Matches tied after 90 min occur frequently; the tie is not resolved except for matches in special tournament situations (FIFA World Cup, UEFA EURO Championships, UEFA Champions League, etc.). National championships are usually played in a double round-robin tournament lasting a whole season, where each team plays against each other team once at home and once in an away game. The scoring system of a match is simply the score of goals achieved for each team. A team is awarded 3 points for a win and one point for a tie; the team with the most points at the end of the season is the national champion.[3] The double round-robin is considered as turning out as champion the team which is the most consistent and well-playing over a whole season.

Football tournments like the FIFA World Cup or the UEFA Champions League are organized in a different fashion: an initial qualification period, a single or double round-robin series of matches in small groups of teams, and single or double playoff elimination matches. Seeding systems are applied in several stages to avoid that strong teams match and eliminate each other too early in the tournament.

Basketball is similar in many regards to football. It is a high-scoring game compared to football, and score difference is not considered as a reliable indicator to resolve ties in the standings. Most basketball leagues now determine the national champion in a series of playoff matches, where each playoff match-up between teams is implemented in a best-of-3, best-of-5, or best-of-7 fashion.

[2]American: Soccer

[3]Tie resolution is based on goal difference, then more goals scored.

B.5 Figure Skating and Diving

Figure skating and diving are sports working quite differently from the previously described ones. Either individuals or pairs of athletes compete one at a time. In both sports, athletes strive to achieve an excellent, error-free, artistic performance of various difficulties, like ice skating synchronized to some piece of music and performing various kinds of spins and jumps (toe-loops, axels, flips, Lutzes, etc.), or diving from a 10 m platform while performing various rotations or revolutions. Both sports rely on panels of expert judges responsible for awarding the scores for each performance, which eventually determine the final ranking of the athletes.

The use of human judgement is prone to bias, and especially in figure skating this has led to several controversies in the past, eventually leading to major revisions of the judging system. The system currently applied in major tournaments like the Olympics or World Championships is the ISU judging system, which is summarized in Wikipedia [92]:

> *Under the system, points are awarded individually for each skating element, and the sum of these points is the total element score (TES). Competitive programs are constrained to have a set number of elements. Each element is judged first by a technical specialist who identifies the specific element and determines its base value. The technical specialist uses instant replay video to verify things that distinguish different elements; e.g., the exact foot position at take-off and landing of a jump. The decision of the technical specialist determines the base value of the element. A panel of twelve judges then each award a mark for the quality and execution of the element. This mark is called the grade of execution (GOE) that is an integer from* -3 *to* $+3$*. The GOE mark is then translated into another value by using the table of values in ISU rule 322. The GOE value from the twelve judges is then processed with a computerized random selection of nine judges, then discarding the high and low value, and finally averaging the remaining seven. This average value is then added to (or subtracted from) the base value to get the total value for the element.* [93]

Despite such elaborate judging systems, it is not always easy for the general audience to understand why an athlete is judged better than another one. Compared to the previously described sports, the judges have a much more direct influence on an athletes ranking than the referees in other sports.

B.6 Summary of Sports and Scoring Systems in Sports

Sports competitions all share the idea that the (individual or team) participant performing some kind of task best is to be found. There is a lot of variation in how the competitions are organized, ranging from a sequence of individual performances (shotput, show jumping, figure skating), direct match-up of competitors (sprint, tennis, golf), tournament formats of various kinds all the way to formats consisting of long sequences of matches stretching over a whole season (football, basketball). The determination of match-ups is often based on draws. Seeding schemes, usually based on a ranking of the participants, are used to avoid early elimination of high-ranking participants or teams.

Sport competitions are valuating fairness to a high degree. Referees are used in practically all sports to ensure that the competition rules are obeyed and are supposed to identify and sanction fouls, which in some sports can be very difficult to judge for humans. Therefore, some sports make use of technical equipment (e.g. in tennis) or allow video replay (American football) to aid judges in their decision making. Fairness is also established by having separate competitions for female and male participants, and in some sports by having different competition classes according to athletes' weight (boxing, karate).

For the performance assessment, sports use a range of evaluation methods:

- **Timings:** The time needed to complete the task is used in sports like *track* and *alpine skiing*.
- **Distance Measurements:** Many *field* sports use a distance measurement, e.g. how far the athlete jumped or an object (javelin, hammer) has been thrown.
- **Timing with Penalties:** A variation of pure timings is sometimes used e.g. in *biathlon*, where missed shots are transformed into time penalties.
- **Scores and Timing:** A combination of (penalty) scores and timings is used in *show jumping*.
- **Objective Scores:** Many sports directly involve direct scoring as part of the game, e.g. goals in *football* and points scored in *basketball*. This type of scoring is easy to observe and usually undisputable, therefore tagged as objective.
- **Subjective Scores:** Sports like *figure skating*, *diving*, or *snowboarding* use scores awarded by jury panels; therefore this type of scores is considered subjective.
- **Match Scores:** Outcomes of direct match-ups are transformed into match scores, which can then be added in order to computer league standings. For example, in *football* a match winner is awarded three points, the loser none. In case of a tie, bioth teams get one point. Similar scoring systems are used in many team sports.
- **Match Series:** The playoff series in *basketball* or *hockey* are decided as soon as a team has achieved a majority of wins in a series of 3, 5, or 7 matches.
- **League Standings:** Most team sports, including *football* and *basketball* keep league standings as means for evaluating season performance. League standings accumulate match scores.
- **Rankings:** In sports like *tennis*, *golf*, *chess* rankings are kept. Some ranking systems aggregate just the outcomes of the matches played, while others take into account the ranking of the opponent teams as well.

Bibliography

[1] B. R. GmbH. Biorob arm. [Online]. Available: http://www.bionic-robotics.de/en/company/biorob-arm.html

[2] U. Robots. Collaborative robot solutions. [Online]. Available: http://www.universal-robots.com/GB/Products.aspx

[3] R. Robotics. Meet your new lab partner. [Online]. Available: http://www.rethinkrobotics.com/files/8713/6694/8748/BRR_004.20.13.pdf

[4] K. Systems, "ecommerce," http://www.kivasystems.com/industries/ecommerce, 2013. [Online]. Available: http://www.kivasystems.com

[5] Factoroffuture.jpg. [Online]. Available: http://www.ros.org/news/2013/01/29/FactorOfFuture.jpg

[6] D. Logistics. (2006) Parcels on conveyor belt. http://www.diamondlogistics.co.uk/wp-content/uploads/2012/05/parcels-on-conveyor-belt.jpg. Diamond Logistics. [Online, accessed 18-07-2013]. [Online]. Available: http://www.diamondlogistics.co.uk/wp-content/uploads/2012/05/parcels-on-conveyor-belt.jpg

[7] N. Nurmi. (2003) Logistics, parcels on conveyor belts. http://www.tekes.fi/imageserver/images/en_content/picture_bank/products_and_productionk/conveyorbelts_0.jpg. Tekes. [Online, accessed 18-07-2013]. [Online]. Available: http://www.tekes.fi/imageserver/images/en_content/picture_bank/products_and_productionk/conveyorbelts_0.jpg

[8] M. Oliveira. (2012) A package lands on the conveyer belt to be sorted after arriving at a fedex facility in toronto. http://thechronicleherald.ca/sites/default/files/imagecache/ch_article_main_image/articles/B97118368Z.120121209182006000GKK1TIKV.11.jpg. [Online, accessed 18-07-2013]. [Online]. Available: http://thechronicleherald.ca/sites/default/files/imagecache/ch_article_main_image/articles/B97118368Z.120121209182006000GKK1TIKV.11.jpg

[9] Subaru-cargo-photo_2.jpg. [Online]. Available: http://www.plasticpals.com/http://www.plasticpals.com/wp-content/uploads//2009/10/Subaru-cargo-photo_2.jpg

[10] P. P. R. International. British cry foul over shortage of trained curry chefs in the u.k. [Online]. Available: http://www.pri.org/stories/politics-society/british-cry-foul-over-shortage-of-trained-curry-chefs-in-the-u-k-7298.html

[11] anonymous. Chef shortage heats up uk curry crisis. [Online]. Available: http://www.reuters.com/video/2013/06/03/chef-shortage-heats-up-uk-curry-crisis?videoId=243112928

[12] floor-tiling.jpg. [Online]. Available: http://www.forcebeyondcapacity.com/wp-content/uploads/2012/06/floor-tiling.jpg

[13] 400px-771.11_bridge_joint_sealing_-_silicone3.jpg. [Online]. Available: http://epg.modot.org/files/thumb/9/92/771.11_Bridge_Joint_Sealing_-_Silicone3.jpg/400px-771.11_Bridge_Joint_Sealing_-_Silicone3.jpg

[14] senseless-bot-robot-kanno-so-takihiro-yamaguchi.jpg. [Online]. Available: http://www.tokyoartbeat.com/tablog/entries.en/wp-content/uploads/2011/10/senseless-bot-robot-kanno-so-takihiro-yamaguchi.jpg

[15] K. L. S. G. I. of Technology. (2013) Books. http://git.edu/library/images/stories/books.jpg. Karnatak Law Society's Gogte Institute of Technology. [Online, accessed 18-07-2013]. [Online]. Available: http://git.edu/library/images/stories/books.jpg

[16] M. Porter. (2006) Bleep. bloop. i am a library robot. http://www.flickr.com/photos/libraryman/292274709/. [Online, accessed 18-07-2013]. CC BY-NC-ND 2.0. [Online]. Available: http://www.flickr.com/photos/libraryman/292274709/

[17] A. Morales, M. Prats, P. Sanz, and A. P. Pobil, "An experiment in the use of manipulation primitives and tactile perception for reactive grasping," in *Robotics: Science and Systems (RSS). Workshop on Robot Manipulation: Sensing and Adapting to the Real World*, 2007.

[18] D. Coetzee. (2011) File:closeup of robotic arm in storagetek tape library at nersc (1).jpg. http://commons.wikimedia.org/wiki/File:Closeup_of_robotic_arm_in_StorageTek_tape_library_at_NERSC_%281%29.jpg. [Online, accessed 18-07-2013]. [Online]. Available: http://commons.wikimedia.org/wiki/File:Closeup_of_robotic_arm_in_StorageTek_tape_library_at_NERSC_%281%29.jpg

[19] breakfast-table.jpg. [Online]. Available: http://www.google.de/imgres?um=1&safe=off&hl=de&biw=1615&bih=891&tbm=isch&tbnid=ZDfLghIQTpHyNM:&imgrefurl=http://www.hoteligieapalermo.com/photo-gallery/p-g-5.htm&docid=8THAg4ZzPtyIHM&imgurl=http://www.hoteligieapalermo.com/images/breakfast-table.jpg&w=678&h=450&ei=9vzoUf-mG4TP4QS2qIDgCA&zoom=1&iact=rc&dur=352&page=1&tbnh=139&tbnw=217&start=0&ndsp=42&ved=1t:429,r:1,s:0,i:83&tx=132&ty=68

[20] robots_top.png. [Online]. Available: http://asset2.cbsistatic.com/cnwk.1d/i/tim/2011/06/13/robots_top.png

[21] Ammerland_bad_zwischenahn_1_2f1843d2-988e-40bb-ac87-64dcb8d5869a–498x337.jpg. [Online]. Available: http://www.nwzonline.de/rf/image_online/NWZ_CMS/NWZ/2011-2013/Produktion/2012/09/17/AMMERLAND/BAD_ZWISCHENAHN_1/Bilder/AMMERLAND_BAD_ZWISCHENAHN_1_2f1843d2-988e-40bb-ac87-64dcb8d5869a--498x337.jpg

[22] tea-party-whole-table2.jpg. [Online]. Available: http://onecharmingparty.com/wordpress/wp-content/uploads/2009/09/tea-party-whole-table2.jpg

[23] Green-bookshelf.jpg. [Online]. Available: http://nationalfurnituresupply.com/blog/wp-content/uploads/2012/12/Green-Bookshelf.jpg

[24] H. Meuer, E. Strohmaier, J. Dongarra, and H. Simon. Top 500 supercomputer sites, june 2013. [Online]. Available: http://www.top500.org/lists/2013/06/

[25] H. Cumow and B. Wichman, "A synthetic benchmark," *Computer Journal*, vol. 19, no. 1, pp. 43–49, 1976.

[26] Wikipedia. Whetstone (benchmark). [Online]. Available: http://en.wikipedia.org/wiki/Whetstone_(benchmark)

[27] R. Weicker, "Dhrystone: A synthetic systems programming benchmark," *Communications of the ACM*, vol. 27, no. 10, pp. 1013–1030, 1984.

[28] Wikipedia. Dhrystone. [Online]. Available: http://en.wikipedia.org/wiki/Dhrystone

[29] K. Kant, *Introduction to computer system performance evaluation*, ser. McGraw-Hill computer science series. McGraw-Hill, 1992. [Online]. Available: http://books.google.de/books?id=Ht5QAAAAMAAJ

[30] S. P. E. Corporation. Spec's benchmarks and published results. [Online]. Available: http://www.spec.org/benchmarks.html

[31] Wikipedia. Standard performance evaluation corporation. [Online]. Available: http://en.wikipedia.org/wiki/Standard_Performance_Evaluation_Corporation

[32] E. E. M. B. Consortium. Coremark. [Online]. Available: http://coremark.org

[33] Wikipedia. Coremark. [Online]. Available: http://en.wikipedia.org/wiki/Coremark

[34] J. Matlis. (2005, May 30) Sidebar: The linpack benchmark. [Online]. Available: http://www.computerworld.com/s/article/102050/Sidebar_The_Linpack_Benchmark

[35] Wikipedia. Linpack. [Online]. Available: http://en.wikipedia.org/wiki/Linpack

[36] C. Bienia, "Benchmarking modern multiprocessors," Ph.D. dissertation, Princeton University, January 2011.

[37] P. University. Parsec: A unit of measure. [Online]. Available: http://parsec.cs.princeton.edu

[38] R. Jones. Netperf homepage. [Online]. Available: http://www.netperf.org/netperf/

[39] Wikipedia. Netperf. [Online]. Available: http://en.wikipedia.org/wiki/Netperf

[40] R. Jonkman. Netspec: A tool for network experimentation and measurement. [Online]. Available: http://compnetworking.about.com/gi/o.htm?zi=1/XJ&zTi=1&sdn=compnetworking&cdn=compute&tm=469&f=00&tt=14&bt=8&bts=8&zu=http%3A//www.ittc.ku.edu/netspec/

[41] P. Software. Network benchmark - test your network speed. [Online]. Available: http://www.passmark.com/products/pt_advnet.htm

[42] Wikipedia. Capture the flag. [Online]. Available: https://en.wikipedia.org/wiki/Capture_the_flag

[43] ——. Def con. [Online]. Available: https://en.wikipedia.org/wiki/DEF_CON_(convention)

[44] NIST. Text retrieval conference (trec). [Online]. Available: http://trec.nist.gov

[45] anonymous. Ieee conference on computational intelligence in games (cig): Competitions. [Online]. Available: http://geneura.ugr.es/cig2012/competitions.html

[46] ——. Fourteenth ieee international workshop on performance evaluation of tracking and surveillance. [Online]. Available: http://pets2012.net

[47] ——. Machine learning and data mining algorithms: Automated tests, repeatable experiments, meaningful results. [Online]. Available: http://tunedit.org

[48] ——. Icaps competitions. [Online]. Available: http://www.icaps-conference.org/index.php/Main/Competitions

[49] W. Ketter, J. Collins, P. P. Reddy, and C. M. Flath, "The power trading agent competition," Soial Science Research Report, ERIm Report Series Ref No ERS-2011-011-US, May 17 2011.

[50] anonymous. Trading agent competition. [Online]. Available: http://tac.sics.se/page.php?id=2

[51] ——. Yearly archives 2013. [Online]. Available: http://tradingagents.eecs.umich.edu/2013/

[52] ——. Power agent trading competition. [Online]. Available: http://www.powertac.org/node/4

[53] T. Balch and H. A. Yanco, "Ten years of the aaai mobile robot competition and exhibition. looking back and to the future," *AI Magazine*, vol. 23, pp. 13–22, 2002, [Online; accessed 16-July-2013]. [Online]. Available: http://www.aaai.org/ojs/index.php/aimagazine/article/download/1603/1502

[54] "Seventh annual mobile robot competition and exhibition: Results and awards," http://www.aaai.org/Conferences/AAAI/1998/aaai98robotwinners.pdf, 1998, [Online; accessed 16-July-2013]. [Online]. Available: http://www.aaai.org/Conferences/AAAI/1998/aaai98robotwinners.pdf

[55] "Call for participation: Seventh annual mobile robot competition and exhibition," http://www.aaai.org/Conferences/AAAI/1998/aaai98robotwinners.pdf, 1998, [Online; accessed 16-July-2013]. [Online]. Available: http://www.aaai.org/Conferences/AAAI/1998/aaai98robotwinners.pdf

[56] H. A. Yanco, "The 1997 aaai mobile robot exhibition," *AI Magazine*, vol. 19, pp. 39–48, 1998, [Online; accessed 16-July-2013]. [Online]. Available: http://www.cs.uml.edu/~holly/publications/PDF/AIMag19-03-011.pdf

[57] (2005) Aaai 2005 robot challenge event. http://web.archive.org/web/20060716104948/http://palantir.swarthmore.edu/aaai05/robotChallenge.htm. The American Association for Artifical Intelligence. [Online; accessed 17-July-2013]. [Online]. Available: http://web.archive.org/web/20060716104948/http://palantir.swarthmore.edu/aaai05/robotChallenge.htm

[58] (2006) Aaai 2006 human-robot interaction event. http://web.archive.org/web/20060719124435/http://www.nd.edu/~mscheutz/humanrobotinteraction.html. The American Association for Artifical Intelligence. [Online; accessed 17-July-2013]. [Online]. Available: http://web.archive.org/web/20060719124435/http://www.nd.edu/~mscheutz/humanrobotinteraction.html

[59] (2005) Aaai 2005 open interaction event. http://web.archive.org/web/20060716105002/http://palantir.swarthmore.edu/aaai05/openInteraction.htm. The American Association for Artifical Intelligence. [Online; accessed 17-July-2013]. [Online]. Available: http://web.archive.org/web/20060716105002/http://palantir.swarthmore.edu/aaai05/openInteraction.htm

[60] (2006) Aaai 2006 scavenger hunt event. http://web.archive.org/web/20060506031138/http://www.cs.hmc.edu/~dodds/aaai06/. The American Association for Artifical Intelligence. [Online; accessed 17-July-2013]. [Online]. Available: http://web.archive.org/web/20060506031138/http://www.cs.hmc.edu/~dodds/aaai06/

[61] "Call for participation: Sixteenth annual aaai mobile robot competition and exhibition," http://www.aaai.org/Conferences/AAAI/2007/aaai07robotcall.pdf, 2007, [Online; accessed 16-July-2013]. [Online]. Available: http://www.aaai.org/Conferences/AAAI/2007/aaai07robotcall.pdf

[62] (1999) Eight annual mobile robot competition and exhibition: Results and awards. http://web.archive.org/web/20010419184104/http://www.aic.nrl.navy.mil/~schultz/aaai99/awards.html. The American Association for Artifical Intelligence. [Online; accessed 17-July-2013]. [Online]. Available: http://web.archive.org/web/20010419184104/http://www.aic.nrl.navy.mil/~schultz/aaai99/awards.html

[63] (2005) Fourteenth annual mobile robot competition and exhibition: Results and awards. http://web.archive.org/web/20060717204415/http://palantir.swarthmore.edu/aaai05/index.htm. The American Association for Artifical Intelligence. [Online; accessed 16-July-2013]. [Online]. Available: http://web.archive.org/web/20060717204415/http://palantir.swarthmore.edu/aaai05/index.htm

[64] "Call for participation: Fifteenth annual aaai mobile robot competition & exhibition," http://www.aaai.org/Conferences/AAAI/2006/aaai06robotcall.pdf, 2006, [Online; accessed 16-July-2013]. [Online]. Available: http://www.aaai.org/Conferences/AAAI/2006/aaai06robotcall.pdf

[65] (2008) Aaai-08 chicago. http://www.aaai.org/Conferences/AAAI/aaai08.php. Association for the Advancement of Artificial Intelligence. [Online; accessed 16-July-2013]. [Online]. Available: http://www.aaai.org/Conferences/AAAI/aaai08.php

[66] (2013) A brief history of robocup. http://www.robocup.org/about-robocup/a-brief-history-of-robocup/. The Robocup Federation. [Online; accessed 16-July-2013]. [Online]. Available: http://www.robocup.org/about-robocup/a-brief-history-of-robocup/

[67] M. Veloso and P. Stone, "Video: Robocup robot soccer history 1997 – 2011," October 2012, available from http://www.robocup.org/2012/10/robocup-video-finalist-for-best-video-award-at-iros-2012/.

[68] (2013) Robocup rescue website. http://www.robocuprescue.org/. RoboCup Rescue League. [Online; accessed 17-July-2013]. [Online]. Available: http://www.robocuprescue.org/

[69] A. Jacoff, "Robocup rescue robot league - progress update and future directions," http://www.isd.mel.nist.gov/projects/USAR/2009/RoboCupRescue_Robot_League_Overview_(2007-2008).pdf, 2009, [Online; accessed 17-July-2013]. [Online]. Available: http://www.isd.mel.nist.gov/projects/USAR/2009/RoboCupRescue_Robot_League_Overview_(2007-2008).pdf

[70] J. Pellenz. (2013) Robocup rescue rulebook. http://wiki.ssrrsummerschool.org/doku.php?id=rrl-rules-2013. RoboCup Rescue League. [Online; accessed 17-July-2013]. [Online]. Available: http://wiki.ssrrsummerschool.org/doku.php?id=rrl-rules-2013

[71] (2013) Robocup rescue field descriptions. http://dl.dropboxusercontent.com/u/21976613/Detailed%20FIELD%20DESCRIPTION.pdf. RoboCup Rescue League. [Online; accessed 17-July-2013]. [Online]. Available: http://dl.dropboxusercontent.com/u/21976613/Detailed%20FIELD%20DESCRIPTION.pdf

[72] J.-D. Dessimoz, P. F. Dominey, M. R. Elara, D. Gossow, D. Holz, L. Iocchi, G. Kraetzschmar, F. Mahmoudi, D. Nardi, S. Olufs, C. Rascon, J. R. del Solar, P. E. Rybski, J. Savage, S. Schiffer, J. Stückler, K. Sugiura, T. van der Zant, S. Wachsmuth, T. Wisspeintner, J. Xie, and A. Yazdani. (2013) Robocup@home rulebook. http://purl.org/holz/2013_rulebook.pdf. RoboCup@Home league. [Online; accessed 17-July-2013]. [Online]. Available: http://purl.org/holz/2013_rulebook.pdf

[73] (2013) Robocup@home website. http://www.robocupathome.org/. RoboCup@Home League. [Online; accessed 16-July-2013]. [Online]. Available: http://www.robocupathome.org/

[74] (2013) Robocup@work website. http://www.robocupatwork.org/. RoboCup@Work League. [Online; accessed 17-July-2013]. [Online]. Available: http://www.robocupatwork.org

[75] G. Kraetzschmar, W. Nowak, N. Hochgeschwender, R. Bischoff, D. Kaczor, and F. Hegger. (2013) Robocup@work rulebook. http://www.robocupatwork.org/download/rulebook-2013-06-08.pdf. RoboCup@Work League. [Online; accessed 17-July-2013]. [Online]. Available: http://www.robocupatwork.org/download/rulebook-2013-06-08.pdf

[76] (2013) Robocup logistics league website. http://www.robocup-logistics.org/. RoboCup Logistics League League. [Online; accessed 17-July-2013]. [Online]. Available: http://www.robocup-logistics.org/

[77] (2013) Robocup logistics rulebook. http://www.robocup-logistics.org/rules/rulebook2013.pdf?attredirects=0&d=1. RoboCup Logistics League. [Online; accessed 17-July-2013]. [Online]. Available: http://www.robocup-logistics.org/rules/rulebook2013.pdf?attredirects=0&d=1

[78] D. A. R. P. Agency. (2004) Grand challenge schedule/events. http://archive.darpa.mil/grandchallenge04/schedule.html. Defense Advanced Research Projects Agency.

[Online; accessed 16-July-2013]. [Online]. Available: http://archive.darpa.mil/grandchallenge04/schedule.html

[79] ——, "Grand challenge rules," http://web.archive.org/web/20040516171932/http://www.darpa.mil/GrandChallenge/rules.pdf, 2004. [Online]. Available: http://web.archive.org/web/20040516171932/http://www.darpa.mil/GrandChallenge/rules.pdf

[80] ——. (2004) Darpa grand challenge - defense advanced research projects agency. Defense Advanced Research Projects Agency. [Online; accessed 16-July-2013]. [Online]. Available: http://archive.darpa.mil/grandchallenge04/overview_pres.pdf

[81] ——, "The darpa grand challenge. a historic demonstration of autonomous robotic vehicles," http://archive.darpa.mil/grandchallenge04/sponsor_toolkit/brochure.pdf, 2004, [Online; accessed 16-July-2013]. [Online]. Available: http://archive.darpa.mil/grandchallenge04/sponsor_toolkit/brochure.pdf

[82] ——. (2007) Darpa grand challenge '05. overview. http://archive.darpa.mil/grandchallenge05/overview.html. Defense Advanced Research Projects Agency. [Online; accessed 16-July-2013]. [Online]. Available: http://archive.darpa.mil/grandchallenge05/overview.html

[83] ——, "Darpa grand challenge 2005 - rules," http://archive.darpa.mil/grandchallenge05/Rules_8oct04.pdf, 2004. [Online]. Available: http://archive.darpa.mil/grandchallenge05/Rules_8oct04.pdf

[84] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, K. Lau, C. Oakley, M. Palatucci, V. Pratt, P. Stang, S. Strohband, C. Dupont, L.-E. Jendrossek, C. Koelen, C. Markey, C. Rummel, J. van Niekerk, E. Jensen, P. Alessandrini, G. Bradski, B. Davies, S. Ettinger, A. Kaehler, A. Nefian, and P. Mahoney, "Winning the darpa grand challenge," *Journal of Field Robotics*, 2006, accepted for publication.

[85] D. A. R. P. Agency. (2007) Darpa urban challenge. welcome. [Online; accessed 16-July-2013]. Defense Advanced Research Projects Agency. [Online]. Available: http://archive.darpa.mil/grandchallenge/

[86] ——. (2007) Schedule. http://archive.darpa.mil/grandchallenge/schedule.asp. Defense Advanced Research Projects Agency. [Online; accessed 16-July-2013]. [Online]. Available: http://archive.darpa.mil/grandchallenge/schedule.asp

[87] ——, "Urban challenge - technical evaluation criteria," http://archive.darpa.mil/grandchallenge/docs/Technical_Evaluation_Criteria_031607.pdf, 2006. [Online]. Available: http://archive.darpa.mil/grandchallenge/docs/Technical_Evaluation_Criteria_031607.pdf

[88] ——, "Urban challenge - rules," http://archive.darpa.mil/grandchallenge/docs/Urban_Challenge_Rules_102707.pdf, 2007. [Online]. Available: http://archive.darpa.mil/grandchallenge/docs/Urban_Challenge_Rules_102707.pdf

[89] ——. (2013) Darpa robotics challenge. http://www.theroboticschallenge.org/aboutprogram.aspx. [Online; accessed 16-July-2013]. [Online]. Available: http://www.theroboticschallenge.org/aboutprogram.aspx

[90] A. Jacoff, R. Sheh, A. Virts, D. Schmitt, A. Moore, J. Quinby, and J. Pippine, "Darpa robotics challenge - drc trials 2013 - initial task descriptions," http://www.theroboticschallenge.org/local/documents/DRC%20Trials%20Initial%20Task%20Descriptions%20DISTAR%20Case%2021473.pdf, 2013. [Online]. Available: http://www.theroboticschallenge.org/local/documents/DRC%20Trials%20Initial%20Task%20Descriptions%20DISTAR%20Case%2021473.pdf

[91] D. A. R. P. Agency, "Virtual robotics challenge rules. release 2 on march 30, 2013. distar case 21064," http://www.theroboticschallenge.org/local/documents/VRC%20Rules%20Release%202%20DISTAR%20Case%2021064.pdf, 2013. [Online]. Available: http://www.theroboticschallenge.org/local/documents/VRC%20Rules%20Release%202%20DISTAR%20Case%2021064.pdf

[92] anonymous. Figure skating. [Online]. Available: http://en.wikipedia.org/wiki/Figure_Skating

[93] I. S. Union. Summary of isu judging system. [Online]. Available: http://www.isu.org/vsite/vcontent/page/custom/0,8510,4844-152094-169310-31825-132302-custom-item,00.html

[94] M. Anderson, O. C. Jenkins, and S. Osentoski, "Recasting robotics challenges as experiments," *Robotics Automation Magazine, IEEE*, vol. 18, no. 2, pp. 10–11, June 2011.

[95] A. G. Cohn, R. Dechter, and G. Lakemeyer, "The competition section: A new paper category," *Artificial Intelligence*, vol. 175, no. 9–10, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0004370211000609

[96] M. Barnes and E. L. Finch, "COLLADA – digital asset schema release 1.5.0," Sony Computer Entertainment Inc., Tokyo, Japan, Specification, April 2008. [Online]. Available: http://www.khronos.org/files/collada_spec_1_5.pdf