The RoCKIn@Home Challenge

Sven Schneider[1], Bonn-Rhein-Sieg University of Applied Sciences, sven.schneider@h-brs.de, Germany
Frederik Hegger[1], Bonn-Rhein-Sieg University of Applied Sciences, frederik.hegger@h-brs.de, Germany
and Aamir Ahmad[2], Iman Awaad[1], Francesco Amigoni[5], Jakob Berghofer[4], Rainer Bischoff[4], Andrea Bonarini[5], Rhama Dwiputra[1], Giulio Fontana[5], Nico Hochgeschwender[1], Luca Iocchi[3], Gerhard Kraetzschmar[1], Pedro Lima[2], Matteo Matteucci[5], Daniele Nardi[3], and Viola Schiaffonati[5]

[1] Bonn-Rhein-Sieg University of Applied Sciences, Germany
[2] Lisbon Technical University, Portugal
[3] Sapienza Università di Roma, Italy
[4] KUKA Laboratories GmbH, Germany
[5] Politecnico di Milano, Italy

Abstract

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, to increase public awareness of the current state-of-the-art in robotics in Europe, and to demonstrate the innovation potential of robotics applications for solving societal challenges and improving the competitiveness of Europe in the global markets. RoCKIn develops two competitions, one for domestic service robots (RoCKIn@Home) and one for industrial robots in factories (RoCKIn@Work). The integration of benchmarking technology with the competition concept is one of the main objectives of RoCKIn.

1 Introduction

RoCKIn@Home [1] is a competition that aims at bringing together the benefits of scientific benchmarking with the attraction of scientific competitions in the realm of domestic service robotics. The objectives are to bolster research in service robotics for home applications and to raise public awareness of the current and future capabilities of such robot systems to meet societal challenges like healthy ageing and longer independent living.

2 The RoCKIn@Home User Story

The basic idea is that we have *an elderly person, named "Granny Annie", who lives in an apartment together with some pets. Granny Annie is suffering from typical problems of aging people: She has some mobility constraints. She tires fast. She needs to have some physical exercise, though. She needs to take her medicine regularly. She must drink enough. She must obey her diet. She needs to observe her blood pressure and blood sugar regularly. She needs to take care of her pets. She wants to have a vivid social life and welcome friends in her apartment occasionally, but regularly. Sometimes she has days not feeling so well and needs to stay in bed. She still enjoys intellectual challenges and reads books, solves puzzles, and socializes a lot with friends.*

For all these activities, RoCKIn@Home is looking into ways to support Granny Annie in mastering her life. The context for performing such activities by technical systems is set in the subsequent scenario description.

3 RoCKIn@Home Scenario

The RoCKIn@Home scenario description is structured into three sections, environment, tasks, and robots, which constitute the first part of the rules for the competition:

- The environment section specifies the environment in which the tasks have to be performed. This information is also relevant for building test beds and simulators.
- The tasks section provides some more detail on the tasks the participating teams are expected to solve through the use of one or more robots and possibly additional equipment. This information tells teams what to prepare for.
- The robot section specifies some constraints and requirements for participating robots, which mainly arise for practical reasons (size and weight limitations, for example) and/or due to the need to observe safety regulations.

3.1 RoCKIn@Home Environment

The goal of the RoCKIn@Home environment is to reflect an ordinary European apartment with all its environmental aspects, like walls, windows, doors or blinds as well as common household items, furniture, decoration and so on. The apartment depicted in Figure 1 serves as a guideline. More detailed specifications are given in the rule book.

Figure 1: Granny Annie's apartment.

The following embedded devices will be installed and can be used by teams:

- A networkable, camera-based intercom at the front door. It allows to see who is in front of the door, and to communicate with this person using speech.
- The ceiling lamps in the bedroom as well as the other lamps in this room are accessible and controllable via network.
- The shutters on the bedroom window are accessible and controllable via network.
- A networkable camera is installed under the exhaust hood above the stove in the kitchen, and another one is installed above the kitchen counter.

All networked devices are accessible within the apartment's WLAN network. In order to allow participation of teams having no or reduced manipulation capabilities, the consortium is also looking into devices that allow to remotely open doors and windows.

3.2 RoCKIn@Home Tasks

The following tasks have to be performed:

1. ***Catering for Granny Annie's comfort:*** This task is aimed at providing little kinds of help for Granny Annie throughout the day. After waking up in the morning, the robot is called by Granny Annie by touching an icon on her tablet computer. She wants the robot to lift the shutters, tilt the window, and switch off the lights. Then Granny Annie lets the robot know that she wants to read, but cannot find her reading glasses at the bedside table. She asks the robot to find them for her. The robot is expected to search for them at places where the glasses are likely to be, taking into account Granny Annie's habits. Information on these habits will be provided. Other comfort duties include lowering the shutters to block bright sunshine, bringing Annie a book, a cup of tea, or a glass of water.

2. ***Welcoming visitors:*** Granny Annie stays in bed because she is not feeling well. The robot will handle visitors, who arrive and ring the door bell, as follows:

- The Deli Man delivers the breakfast; the person is changing almost daily, but they all have a Deli Man uniform. The robot opens the door, guides the Deli Man to the kitchen, then guides him out again. The robot will always observe the stranger.
- An unknown person, trying to sell magazine subscription is ringing. The robot will tell him goodbye without letting the person in.
- Dr. Kimble is her doctor stopping by to see after her. He is a known acquaintance; the robot lets him in and guides him to the bedroom.
- The Postman rings twice and delivers mail and a parcel. The robot just opens the door, receives the deliveries, and farewells him.

If a visitor has been admitted, the robot guides him out after the visit and ensures the door is properly closed and locked. An additional task targets easier setup and operation of domestic service robots:

3. ***Getting to know my home:*** The robot is told to learn about a new environment. It is supposed to generate a semantic map of the apartment within a limited time frame. It is left to the teams how exactly they approach this task. For example, a team member may "demonstrate" the apartment by guiding the robot through the apartment, pointing to objects and speaking aloud their names. Alternatively, a robot may explore the environment completely autonomously. The robot may also interrogate a team member about the names of objects or places. At the end of the environment learning phase, the robot must provide a graphical presentation of the mapped environment and answer a set of questions, like "*Which furniture pieces are in the living room?*", "*How many chairs are around the dining table?*", or "*Is anything on the stove?*"

3.3 RoCKIn@Home Robots

Participating teams can use one or two robots to solve the tasks. The robots must fit through a door of 80cm width and weigh no more than 250kg. They must be fully autonomous, i.e. neither power supply via cable nor any kind of tele-operation is permitted. Each robot must be safe to operate in the environment. Robots polluting or damaging the environment or presenting a threat to humans in the environment are not allowed to particpate. A mechanism to stop the robots in case of emergencies must exist. Robots must be properly equipped to be able to solve the tasks at least in principle. For example, it is not permitted to substitute for lack of speech understanding by entering commands on the keyboard. Teams are not allowed to modify the environment, or to install their own embedded devices in the environment, e.g. additional sensors or actuators.

4 Task Benchmarks

For completeness and consistency, we provide complete task descriptions, even though the general idea has already been described above. The task benchmarking approach is described in more detail in [2].

4.1 *Catering for Granny Annie's comfort*

This benchmark assesses the performance of executing the comfort providing task.

The Task The robot aims to provide little kinds of help for Granny Annie throughout the day. After waking up in the morning, Granny Annie calls the attention of the service robot by touching a button on her tablet computer. After the robot has come to her bedside, Granny Annie gives subsequent task orders by spoken commands: She wants the robot to "*lift the shutters*", "*tilt the window*", and "*switch off the lights*". Granny Annie may give any subset of a set of possible commands in any order. Other comfort duties include lowering the shutters to block bright sunshine, bringing Annie a book, a cup of tea, or a glass of water.

A little later, Granny Annie lets the robot know that she wants to read, but cannot find her reading glasses. She asks the robot to find them for her. This task can occur in any room of the apartment. The robot is expected to search for the glasses at places where the glasses are likely to be, taking into account Granny Annie's habits. Information on these habits will be provided. Several pairs of glasses will be in the apartment, each of which will be clearly distinguishable from the others (different material and color of frames, different shape and color of glasses). Other objects Granny Annie may ask for are her keys, her watch, or her mobile phone.

Information Provided to the Team A set of possible user requests, like opening or closing doors, windows, or drawers, switching on or off or dimming lamps, operating shutters, etc. is provided to the team for each room in the apartment. The team is allowed to map (geometrically and semantically) the environment before this benchmark is executed. See also the Environment Learning Task Benchmark below.

A list of "likely" locations for the reading glasses (and possibly other objects) will be provided. The locations are described as text fragments such as "`on the kitchen counter`", "`on the bedside table`", and "`in the third row from the bottom of the cupboard`". The names used in these text fragments will be well defined in advance and designate objects in the environment such as furniture. The "likelihood" for a location is specified as a number between 0 and 100. It does not relate to probabilities, but may be understood such that if two places A and B have likelihoods 40 and 80, then your chance to find the item at B is roughly twice as good.

Expected Output or Behavior The robot is expected to attend the place where Granny Annie is located when she calls upon the robot's service. The robot should ask for orders in English language as speech output, and receive such orders as spoken commands in English. The robot should confirm orders in an appropriate way. The orders should be executed as expected.

When given the task to search for the glasses or some other object, the robot should acknowledge the order first. Then it may ask Granny Annie right away whether she remembers where she has used the glasses last. If Annie specifies a location, the robot will look there first, otherwise the robot will visit the likely locations in the order of their likelihood. The robot may deviate from this order if it can opportunistically shorten the overall search time of all locations.

Performance Criteria The performance criteria for ranking the quality of the solutions will take into account: the time needed to complete the task (including timings for completing certain subtasks or steps in the overall task), the quality of the speech communication (the robot immediately understands all speech by Granny Annie vs the robots keep asking for confirmation or it does not understand it at all), the quality of the search behavior (the robot explores all suggested/requested locations to find the glasses), the quality of perception (the robot can find the correct glasses), the quality of navigation (the robot does not bump into furniture or into a person present in the test bed), the quality of manipulation (the robot can grasp the glasses firmly without breaking them), and the degree of overall completion of the task (e.g. the robot stops after grasping the glasses or right before it delivers the glasses to Granny Annie).

Benchmarking Data While performing the task, the robot has to notify the benchmarking infrastructure about certain "events" in its internal operation and to log information to be postprocessed by the benchmarking system after the task execution. The information to be recorded has to be timestamped and will include: the event/command that the robot was called upon by Granny Annie, the raw audio signal of the request by Granny Annie out of which the command was extracted, a description of the place where the robot believes Annie to be (serving as the target for path planning, if any), the position the robot believes to be during the whole task execution, the path the robot planned to navigate to Annie's place, any unexpected event (such as the detection of an obstacle with the corresponding distance/position) during navigation to Annie's place, the event of arriving at Annie's place, any command spoken by the robot, any text output of speech recognition and the robot's association with it (e.g. classification as noise, user command, etc.), and other. A detailed list will be provided in the rule book.

4.2 *Welcoming visitors*

This benchmark assesses the robot's capability to interact effectively with humans.

The Task Granny Annie stays in bed because she is not feeling well. The robot will handle visitors, who arrive and ring the door bell, as follows:

- The Deli Man delivers the breakfast; the actual person is changing almost daily, but they all have a Deli Man uniform. The robot opens the door, guides the Deli Man to the kitchen, then guides him out again. The robot is supposed to always observe the stranger.
- An unknown person, trying to sell magazine subscription is ringing. The robot will tell him goodbye without letting the person in.
- Dr. Kimble is her doctor stopping by to see after her. He is a known acquaintance; the robot lets him in and guides him to the bedroom.
- The Postman rings twice and delivers mail and a parcel. The robot just opens the door, receives the deliveries, and farewells him.

The task involves handling several visitors arriving in any sequence, but separately from each other. The robot must be able to handle/interact with a video intercom and the door. If a visitor has been admitted, the robot guides him out after the visit and ensures the door is properly closed and locked.

Information Provided to the Team A list of known people will be provided to the team. For each of these, the name of the person, an associated image with the person's face, and a list of privileges will be specified, e.g. ("`Deli Man`" ("`open door`", "`allow to enter`", "`allow to kitchen`", "`allow deposit of boxes`", "`allow pickup of boxes`", "`allow to leave`" "`close door`")).

Expected Output or Behavior The robot can be explicitly activated for this task by a user order. Alternatively, the robot may act pro-actively after hearing the doorbell ring. The robot may request acknowledgement from Granny Annie that it should perform the task, but only if Granny Annie is awake.
The robot is expected to activate the door camera and to try identifying the visitor. If person identification is not possible, e.g. because the visitor is not looking towards the camera, the robot should use speech output, greet the visitor, ask to look at the camera, and to speak aloud his/her name. This should then result in a situation where the robot can either identify a known person or concludes that the person is not known. In the latter case, no further action is foreseen, and the robot tells the visitor good-bye.

In case the visitor is a known person, the robot is expected to take the appropriate actions, as informally indicated by the "privileges" associated with known visitors. For example, "allow to kitchen" involves guiding the visitor from the entrance door to the kitchen counter, while observing the visitor along the way. The required robot actions involved here include guiding and/or following a person, pointing to objects or places, and speech interaction. More detailed specifications on each privilege will be provided well before the competition.

Performance Criteria Performance criteria will include: the time needed to complete the task (including timings for completing certain subtasks or steps in the overall task), the quality of the speech communication (the robot immediately understands all speech by even unknown persons vs the robots keep asking for confirmation or it does not understand it at all), the quality of visitor handling behavior (the robot takes autonomously all the right decisions vs the robot has to involve Granny Annie each time), the quality of perception (the robot can correctly identify and recognize all visitors), the surveillance of unknown visitors (the robot successfully handles the visitor in the intented way while always maintaining a clear view on the visitor and his/her activities), the quality of manipulation (handover of objects like small parcels), and the degree of overall completion of the task.

Benchmarking Data While performing the task, the robot has to log information inlcuding: the event/command causing the activation of the robot, any attempts to detect and classify a person and its results, the audio signals of the conversations with the visitors, events like opening or closing the front door, guiding visitors into and around the apartment, or manipulating objects, etc. Details will be provided in the rule book.

4.3 *Getting to know my home*

This benchmark assesses a team's performance on executing a semantic mapping task and works as follows:

The Task Each team is asked to ensure that their robot has obtained a representation of the environment as it has been used before undergoing this benchmark. Teams may chose whatever means for this. The organizers will then change the environment in various ways and the robot is asked to learn about these modifications of the environment by adapting the semantic map of the apartment within a limited time frame. For practical reasons, the task may be constrained to apply only to parts of the environment, e.g. the living room, dining room, and kitchen.
It is left to teams how exactly they approach the environment learning/adaptation task. For example, a team member may "demonstrate" the apartment by guiding it through the apartment, pointing to objects and speaking aloud their names. Alternatively, a robot may explore the modified environment and discover the modifcations in a completely autonomous manner. The robot may also

interrogate a team member about the names of objects or places. Teams may not use any graphical or text-based tool to directly enter the relevant information to the robot's representation of the environment.

Information Provided to the Team In addition to any task constraints as described above, teams are given a list of names for spatial areas (rooms) affected by the modifications as well as of the objects removed, repositioned, or added to the environment. Only a subset of these objects may be involved in the changes to the environment.

Expected Output or Behavior At the end of the environment learning phase, the robot must provide suitable feedback about the modifications of the environment. Various means for providing this feedback are permissable, e.g. presenting a graphical representation of the mapped environment including markings of the modifications applied. Alternatively, some logic-based, text-based, or speech-based presentation of the learned modifications to the environment would be acceptable, for a example a list of the kind "*Two chairs have been removed from the dining table*", "*The newspaper has been moved from the coffee table to the bedside table*", or "*A coffee make is now on the kitchen counter*"

Performance Criteria Performance criteria reflected in the scoring system for this task will take into consideration: the number and percentage of modifications detected, the pose estimation errors for added or relocated objects, the time spent on mapping, etc.

Benchmarking Data During task performance, the robot has to log information as follows: the event/command that the robot should start learning about the modified environment, snapshots of the map taken at certain intervals during the mapping process, the detection, classification, and localization of objects, any kind of interaction with humans, like the reception of a semantically meaningful symbol after asking for it, and the presentation of the results at the end of the procedure.

5 Functionality Benchmarks

As part of RoCKIn@Home, we will also perform various benchmarks that assess the robot's performance with respect to particular functionalities following the approach described in [2]. These benchmarks aim to evaluate the functionalities embedded in a fully working system and on a real robot. In order to achieve this, the task to be performed will be narrowed to include, if possible, only a single functionality or the minimal subset of functionalities needed for performing the task. The focus is on repeating the task several times, often with varying the object playing a central role (the object to be perceived, manipulated, etc.) or the environment (varying lighting conditions, for example).

5.1 Object Perception Benchmark

Functionality A number of objects, selected from the list of RoCKIn@Home items, will be positioned, one at the time, on a table located directly in front of the robot. For each object presented, the robot has to perform the following activities:[1] i) Object detection: perception of the presence of an object on the table and association between the perceived object and one of the object classes (see "Information provided to the team"). ii) Object recognition: association between the perceived object and one of the object instances belonging to the selected class (see "Information provided to the team"). iii) Object localization: estimation of the 3D pose of the perceived object wrt to the surface of the table.

Information Provided to the Team The set of individual objects that will be presented to the robot during the functional benchmark is a subset of a larger set of available objects ("object instances"). All object instances are known to the team before the benchmark, but the team does not know what object instances will actually be presented to the robot during the functionality benchmark. More precisely, the team will be provided with the following information: descriptions/models of all the object instances, expressed according to a specified representation; categorical information on the object instances, i.e. object classes (for instance: boxes, mugs, cutlery); reference systems associated to the table surface and to each object instance (to be used to express object poses).

Expected Output or Behavior The robot has to estimate class, instance and pose of each object presented to it. The robot must communicate this information. Then the object is removed and a new object is set on the table. The functional benchmark ends as soon as one of the following situations occurs: (i) all the objects prepared for the functional benchmark have been presented, or (ii) the time available for the functional benchmark expires.

Performance Criteria The following criteria will be used for evaluation: number and percentage of correctly identified objects (correct class and instance), number and percentage of correctly classified objects (correct class, incorrect instance), pose error for all correctly identified objects, execution time (if less than the maximum allowed for the benchmark).

Benchmarking Data The following data will be collected to compute the performance criteria: number of objects presented to the robot; detection, recognition and localization data associated to the objects, provided by the robot; ground truth for object pose, object class, and object instance; sensor data used by the robot to perform classification.

[1] This functional benchmark corresponds to one of the functional benchmarks of RoCKIn@Work.

5.2 Object Manipulation Benchmark

Functionality This functional benchmark assesses the robot's capability to correctly operate switches and controls commonly found on domestic appliances. The objects to be manipulated include both digital (cf. ON/OFF) and analog controls, such as: digital latching or non-latching buttons (e.g. power button of a washing machine, power button of a PC), digital rocker switches (e.g. light switch), digital or analog sliders (e.g. lock switch of portable devices; dimmer of a lamp).
One or more *test panels* will be prepared, each of which is fitted with a set of controls of different category, size and type. A selected panel is affixed to a wall at a height chosen by the team. The team sets the robot on the floor in front of the panel; then, the robot receives an ordered list of controls to operate. For each digital control in the list, the robot is required to change its state. For each analog control in the list, the robot is given the direction of motion of the control: the robot must operate the control until notified that the required setting has been reached.

Information Provided to the Team Each panel will be provided with a unique marker (e.g. QR codes), compliant with publicly available localization software, which the robots are allowed to use to localize controls. Full specifications for the panels (including type, location, make and model of all switches, ID and location of the markers) will be distributed to the teams wth the specifications of the RoCKIn@Home test bed.

Expected Output or Behavior The robot has to correctly operate the controls specified in the list, in the correct order. Correct operation of a control includes the following phases: the robot uses its end effector(s) to bring the switch in the required final state, without damaging it; the robot notifies that such state has been reached; the robot releases the control. The functional benchmark ends as soon as one of the following situations occurs: i) all the controls in the list have been operated, or ii) the time available for the functional benchmark expires.

Performance Criteria During the functional benchmark the following will be evaluated: number and percentage of controls actually operated by the robot, among those in the ordered list provided to the robot; final state of these controls; number of controls accidentally operated by the robot (these include both controls not in the list and controls operated out of order); damages inflicted to the controls by the robot; time (if less than the maximum allowed for the benchmark).

Benchmarking Data During the functional benchmark the following will be collected to compute the Performance Criteria: notifications issued by the robot; initial and final state of all controls on the test panel; internal robot data referring to end effector position and target object position; external ground truth about panel position and end effector position.

5.3 Speech Understanding Benchmark

Functionality This functional benchmark aims at evaluating the ability of a robot to understand speech commands that a user gives in a domestic environment. A list of commands will be selected among the set of recognizable commands (i.e. commands that the robot should be able to perform in the competition), and they will be given to the robot as prerecorded utterances to the system or directly spoken by a user. The final representation for the recognized commands will have to respect a command/arguments structure where each argument is instantiated according to the arguments of command evoking verb. This representation is referred to as *Command Frame Representation* (CFR) (e.g. "go to the living room" will correspond to `GO( destination: "living room" )`).

Information Provided to the Team Each team will be provided with a knowledge base (Frame Knowledge Base, FKB) containing a set of *semantic frames*. Each frame corresponds to an action, or robot command. The FKB contains a description of each frame, in terms of allowed arguments (e.g. `destination` for a *motion* command), their names, the set of verbs evoking the frame and additional information on how to model the activated frame into the CFR. A specific lexicon for the home domain will be also released, including names of rooms and objects.

Expected Output or Behavior The robot should be able to understand a command starting from the speech input. The robot should correctly transcribe the user utterance and recognize the action to perform, resulting in the correct command frame (e.g. `GO`) and the arguments involved (e.g. the `destination` of a *motion* command). The output of the robot should provide the CFR for each command.

Performance Criteria During the functional benchmark, different aspects of the speech understanding process will be assessed: i) The Word Error Rate on the trascription of the user utterances, in order to evaluate the performance of the speech recognition process. ii) The number and percentage of correctly recognized command frames (both with and without considering arguments). The evaluation will be carried out in term of precision, recall and F-measure. iii) Time utilized (if less than the maximum allowed for the benchmark). An optional criterion is the number and percentage of morpho-syntactic and syntactic information, in terms of accuracy of POS-tagging and syntactic-parsing.

Benchmarking Data The following data will be collected to compute the performance criteria: i) Sensor data (in the form of audio files) used by the robot to perform

speech recognition[2]; ii) the set of all possible transcription for each user utterance iii) the final command produce during the natural language analysis process iv) intermediate information produced or used by the natural language understanding system during the analysis as, for example, syntactic information.

6 RoCKIn@Home Competition

The RoCKIn project runs from 2013 through the end of 2015. The RoCKIn competitions will be held twice, in Fall 2014 and Fall 2015. Both RoCKIn and RoboCup have the intention to eventually merge the competitions in a joint event for the years beyond 2015. The first RoCKIn event including the RoCKIn@Home competition will be held end of November 2014 at LAAS, Toulouse, France. Details on this competition will be supplied as they become known. Please refer to the web site of the RoCKIn project [3] and the web pages on RoCKIn@Home.

References

[1] S. Schneider and G. Kraetzschmar, "D2.1.1 – RoCKIn@Home – A Competition for Domestic Service Robots – Competition Design, Rule Book, and Scenario Construction," The RoCKIn Project (FP7-ICT-601012 grant agreement number 601012)," Public Deliverable, 2013.

[2] A. Ahmad, I. Awaad, F. Amigoni, J. Berghofer, R. Bischoff, A. Bonarini, R. Dwiputra, G. Fontana, F. Hegger, N. Hochgeschwender, L. Iocchi, G. Kraetzschmar, P. Lima, M. Matteucci, D. Nardi, V. Schaffionati, and S. Schneider, "D1.1 – Specification of General Features of Scenarios and Robots for Benchmarking Through Competitions," The RoCKIn Project (FP7-ICT-601012 grant agreement number 601012)," Public Deliverable, 2013.

[3] RoCKIn Project. RoCKIn – Robot Competitions Kick Innovation in Cognitive Systems and Robotics. The RoCKIn Project (FP7-ICT-601012 grant agreement number 601012). [Online]. Available: http://www.rockinrobotchallenge.eu/

[2]Speech files from all teams and all benchmarks (both Task benchmarks and Functional benchmarks) will be collected and used to build a public dataset. The audio files in the dataset will therefore include all the defects of real-world audio capture using robot hardware (e.g., electrical and mechanical noise, limited bandwidth, harmonic distortion). Such files will be usable to test speech recognition software, or (possibly) to act as input during the execution of speech recognition benchmarks.