Interactive Semantic Mapping: Experimental Evaluation

Guglielmo Gemignani, Daniele Nardi,
Domenico Daniele Bloisi, Roberto Capobianco and Luca Iocchi

Dept. of Computer, Control, and Management Engineering
Sapienza University of Rome, Italy
<lastname>@dis.uniroma1.it

Abstract. Robots that are launched in the consumer market need to provide more effective human robot interaction, and, in particular, spoken language interfaces. However, in order to support the execution of high level commands as they are specified in natural language, a semantic map is required. Such a map is a representation that enables the robot to ground the commands into the actual places and objects located in the environment. In this paper, we present the experimental evaluation of a system specifically designed to build semantically rich maps, through the interaction with the user. The results of the experiments not only provide the basis for a discussion of the features of the proposed approach, but also highlight the manifold issues that arise in the evaluation of semantic mapping.

Keywords: Cognitive Robotics, Human Robot Interaction, Knowledge Representation and Reasoning, Semantic mapping

1 Introduction

As robots are targeting the consumer market, the need for developing suitable interaction paradigms and interfaces for consumers is increasing. In this scenario, spoken language interaction plays a key role, as also demonstrated by other consumer products, such as cell phones and cars. However, in order to provide a system with the ability of interacting with the user using natural language, the robot must be able to interpret high level commands, such as "go to the printer near the secretary office". For executing such a command, the system must understand not only the meaning of the terms used by the user, but also to *ground* them into its world model (i.e. the representation of the operational environment).

To address this problem, several researchers have been developing *semantic maps*, that, according to the definition given in [1], should be able to integrate symbolic knowledge into the representation of the environment used by the robot. Although significant progress has been made in the last years, the semantic maps that robots can acquire and deploy are still limited. On the one hand, the acquisition of semantic knowledge by state-of-the-art approaches to perception is challenging, on the other hand, a systematic approach that exploits the interaction with the user, to build semantically rich representations of the environment has been only partially addressed.

The goal of our work is to rely on the interaction with the user, according to the paradigm of symbiotic autonomy [2] in order to build a representation of the environment that can allow a mobile robot to interpret and execute user commands that refer to places and objects in the environment. Specifically, we have developed a system that builds a layered semantic map through a multi modal interaction with the user that relies on the use of a simple pointer device [3]. The system has been deployed on four different robotic wheeled platforms and has been used to successfully build the semantic map of office and home environments. The system at an earlier stage has been presented in [4] and the present paper is specifically addressing the experimental evaluation of the proposed approach. To this end, we have reviewed the literature on semantic mapping to identify a proper methodology for a quantitative evaluation of the proposed approach. The outcome of our survey shows that there are no established methodologies for a quantitative evaluation of semantic mapping. In fact, several methods are adopted, each one covering a specific aspect of the proposed approach. Consequently, we have defined an experimental setting for each system component and evaluated their performance in isolation. Moreover, we have run several experiments aiming at the evaluation of the overall system. The results of this evaluation, that are discussed in detail in the paper, show that the proposed system has an overall very interesting performance. Moreover, since the representation of semantic knowledge requires several forms of approximations, our system shows a good trade-off between accuracy and ability to deal with high level semantic notions. This notwithstanding, several key issues remain to be addressed by the research on semantic mapping, to make possible the deployment of robots that are able to incrementally acquire and keep up-to-date the knowledge about the operational environment in the face of changes.

The paper is structured as follows. In the next section we review the state of the art on semantic mapping; then, we present a quick overview of our system (Section 3). The rest of the paper is devoted to discussing the experimental evaluation of the system (Section 4), by first analyzing the approaches found in the literature and then presenting a detailed evaluation of our system. A summary of the contributions of the proposed approach and hints for future work conclude the paper.

2 Related Work

The acquisition of the semantic knowledge needed to suitably interpret the commands given by a user to a robot is typically achieved through a process called *semantic mapping* [5]. The literature about such research topic can be divided into two main categories, by distinguishing automatic methods from the so called “human-in-the-loop” approaches, where a user is asked to help the robot in the acquisition process, as proposed also by [2].

As an example of automatic approaches, in Galindo *et al.* [6] environmental knowledge is represented by augmenting a topological map (extracted by means of fuzzy morphological operators) with semantic knowledge using anchoring. In Goerke *et al.* [7], in Brunskill *et al.* [8], and in Friedman *et al.* [9], instead, a set of techniques are used to automatically classify and cluster metric maps. Finally, in Mozos *et al.* [10] visual features are used for object recognition and place categorization. Although sig-

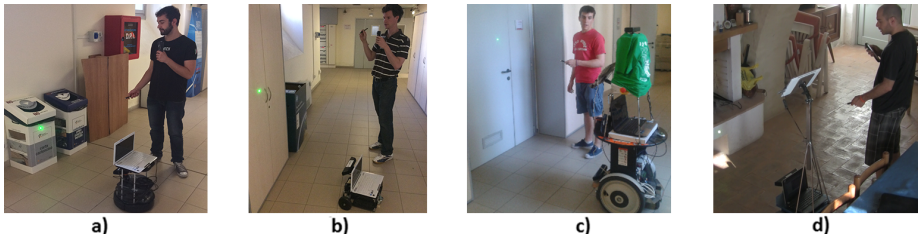

Fig. 1: Robots on which our system has been deployed. a) Turtlebot. b) MARRtino, a mobile base built by our students. c) Mobile base derived from Segway. d) Videre Design platform.

nificant progress has been made in fully automated semantic mapping [11], even the most recent approaches still lack of robustness and generality.

Therefore researchers in the AI and Robotics community have started to enclose the human in the semantic acquisition process, trying to overcome the limitations that the current robotic systems have. As an example of "human-in-the-loop" approach, in [12] the authors describe a system for the creation of conceptual representations of indoor environments. In this work, a priori knowledge about spatial concepts is provided to the robotic platform, which produces an internal representation of the environment acquired through low-level sensors with the help of the user for place labeling. In [13], instead, an approach that uses heterogeneous modalities for a comprehensive multi-layered semantic mapping algorithm, aiming at place categorization and topological map construction, is presented. This system builds a probabilistic representation that includes information about the existence of objects and properties of space. Such a representation is used in order to estimate room labels. The user input, whenever provided, is integrated in the system as additional properties about existing objects. While in the latter described approach the support of the user does not play a central role, in [3] the authors propose a rich multi-modal interaction, including speech, gesture, and vision. Such an approach enables the system to perform a semantic labeling of the environment, without many pre-requisites on the features of the environment itself. In this system however, the authors do not attach any additional semantic information to the landmarks other than their position.

Compared with the related work, our approach, initially proposed in [4], improves the construction of semantic maps through the interaction with the user, aiming not only at representing objects as points in the metric map, but at creating a semantic map that holds manifold information of the objects (e.g., dimensions, colors, 3D models), which is needed by the robot for task execution and reasoning.

3 System Overview

The proposed system is built for wheeled robots that are capable of mapping the environment through an off-line slam technique and, afterwords, can also navigate in it through the conventional ROS *movebase* module[1]. In our experiments we have used a

[1] `http://wiki.ros.org/move_base`

Turtlebot (Figure 1a), a MARRtino[2], a mobile base built by our students (Figure 1b), a mobile base derived from Segway (Figure 1c), and a Videre Design platform (Figure 1d). In addition to the navigation component, the robot is equipped with a kinect that can perform several functions, including the ability to detect a laser dot produced by a laser pointer, that the user exploits to point at the objects that the system should store in the map. The system can also acquire the image and the point cloud associated with the objects pointed by the user, later used to recognize previously seen objects.

The user can interact with the system using natural language through the use of a suitable human-robot interface. This component is implemented as a separate subsystem that can be deployed on different robotic platforms; it includes a speech processing component and a natural language processing chain that provide an interpretation of the user command in terms of frames, representing the commands executable by the robot. The knowledge acquired by the robot through the interaction with the user is stored in a multi-layered knowledge base, which contains the semantic knowledge about the environment, structured according to an abstract representation that is automatically built from the conventional 2D map.

The process of building the representation of the robot's knowledge is composed by the *Metric Map and Instance Signatures Construction Phase*, where a 2D metric map is generated through a SLAM module and the initial knowledge is extracted, and by the *Semantic Grid Map and Topological Graph Generation Phase*. In this latter phase, starting from the 2D metric map, a grid-based topological representation (*Semantic Grid Map*) is obtained, later used to produce the topological graph needed by the robot to perform high level behaviours.

More in detail, in the first phase, the robot is used to navigate the environment in order to acquire the 2D map (using a Graph-based SLAM approach [14]) and to register the positions of the different objects of interest. During the robot exploration the user can in fact tag a specific object by using a commercial laser pointer. While the object is pointed through the laser, the user has to name it, so that a label can be assigned to it and its image and point cloud can be memorized. The registered object poses with the corresponding labels are processed to create the Semantic Grid Map and the Topological Graph. The Semantic Grid Map contains a high-level description about the regions, structural elements, and objects contained in the environment. The algorithm used to generate such a map, rasterizes the metric representation of the map into a grid-based topological representation, automatically labeling the areas of the environment (using contour closure and region filling techniques) and including representations of the objects described by the user. In the final step of the knowledge building process, a topological graph is created in order to represent the information needed by the robot for navigating and acting in the environment. The constituting nodes of this graph are locations associated to cells in the Semantic Grid Map, while the edges are connections between these locations (for a more detailed description about the representation and its building process we refer to [4]).

[2] http://www.dis.uniroma1.it/∼spqr/MARRtino

4 Experimental Approaches Analysis

Analyzing the literature on semantic mapping, no standard references for performing a correct evaluation of a system can be found. Due to this fact, in order to better explain why the evaluation part has been carried on as described in Section 5, this section will be dedicated to the problem of evaluating a semantic map.

The standard evaluation methods that can be found in the literature on Simultaneous Localization and Mapping (SLAM) consist in testing a particular system by processing a set of raw sensors data and then comparing the obtained output with a ground truth (see for example the Victoria Park Dataset [15]). Such a comparison is feasible thanks to the standard output generated by every SLAM algorithm. This is not the case for semantic mapping systems. In this particular research area, in fact, the output of each system is bound to subsets of the world model whose semantics is defined in an ad-hoc way; such an output is therefore hardly comparable with other systems. Due to this fact, research in semantic mapping has often focussed the evaluation on particular aspects of the proposed system.

An initial and probably the most simple evaluation approach adopted in the literature of semantic mapping consists in giving a qualitative evaluation of the output (usually a labelled metric map), by comparing it with a hand made ground truth (see for example [3]). While this method gives an idea of how well a system can perform and it is used to focus the evaluation process on the metric output, it can not be used to compare two different semantic mapping approaches. Moreover, while it is possible to compare at least qualitatively the metric output of the system, an evaluation of the semantic information stored in the map is usually not available.

Another testing method adopted to evaluate how well a system can acquire semantic information about an environment has been inspired by the literature on classifiers and consists in testing the system in a task of environment classification, by measuring the percentage of correctly classified places during a variable number of runs (see [7] and [10]). This approach raises two issues: it implicitly assumes that it is possible to classify a place by the objects enclosed in it (e.g., it is not clear how to evaluate the classification of a room with a stove and a bed in it); it reduces the semantic mapping problem to a specific classification problem, not evaluating the system ability in acquiring other types of knowledge outside the ones needed for classification (e.g., spatial properties of the environment, objects' affordances, positions, and dimensions, etc.).

An alternative evaluation for the capability of acquiring semantic information consists in measuring the benefits gained from the addition of the acquired knowledge during a typical task executed by a robot. For example, semantic information is sometimes used to improve the performance of SLAM tasks [16]. Evaluating the improvements achieved by acquiring semantic information during a test run can indeed be used to get an idea of how well a system performs the semantic mapping task. This approach is typically the most complete evaluation approach; however, it is still not clear to the research community what are the types of tasks that should be considered to perform a full evaluation of an arbitrary semantic mapping technique.

An additional test that can be performed on systems that enclose the human in the mapping process consist in user evaluation studies. In this kind of tests, a system is

tested to see how well it can interact with a user and how effectively it handles multiple, complex and dynamic interactions with a user.

As shown by the above analysis, no clear methodology is available to evaluate the performance of a semantic mapping approach. Consequently, we have chosen to analyze each system component separately in a quantitative way and, in addition, to test our system as a whole, both quantitatively and qualitatively during task execution. Such testing evaluation is described in the next section.

5 Experiments

In this section we discuss the experiments performed to validate the system and the results gathered from them over the last months, mainly focusing on the developments obtained after [4]. In order to validate the approach discussed in this abstract, we have tested both the single constituting components and the whole system. Specifically, we evaluated the spoken interaction, the Semantic Grid Map generation, the object segmentation and the spatial reasoning performed by the system. Since the spoken interaction and the Semantic Grid Map generation have already been tested in detail in [17] (the grammar based approach) and [18], respectively, in this section we will briefly report the results obtained for these two components, referring to the original articles for a more accurate evaluation and focussing our analysis on the object segmentation and the spatial reasoning components.

5.1 System Component Evaluation

Spoken Interaction The speech component has been designed mainly as a support for the *Augmented Mapping* task experiment described in this paper. For this part, we aimed at having a robust system, covering a controlled language with a low error rate in terms of transcription ability, instead of trying to deal with a wide range of linguistic phenomena. We therefore evaluated the performance of the Speech component with respect to the quality of the transcription of the user utterances and the command interpretation process. The former has been evaluated in terms of the Word Error Rate (WER) [19], obtaining a value of 0.258 on the transcription of commands uttered during the experiments. The second measurement has been carried out in terms of Precision (P), Recall (R) and F1-Measure (F1), as defined in [20]. The results obtained are reported in Table 1. Overall, the system satisfies the usability requirement, showing an acceptable performance during the interactions with the user, although covering a limited range of linguistic phenomena. Ongoing work is thus being carried out in order to improve this specific system component.

Table 1: Performance of the Speech Recognition component.

Metric	**P**	**R**	**F1**
Action Recognition	89.47	80.63	84.82
Full Command Recognition	75.43	67.98	71.51

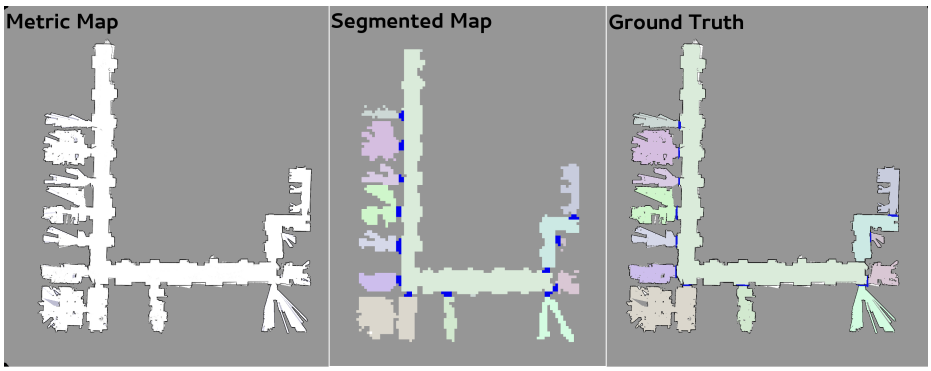

Fig. 2: Representation obtained for a metric map and respective ground truth.

Semantic Grid Map Generation In order to evaluate the Semantic Grid Map representation, a detailed set of experiments has been conducted. During these experiments, a set of 10 different metric maps has been processed by our system to get a qualitative evaluation of the capabilities of the system. An example of processed map and its output representation is shown in fig 2. When the objects are placed in the Semantic Grid Map, errors in their positions and dimensions are introduced because of the discretization of this map. To this end, we performed an additional evaluation by considering 11 instances of 3 different categories of objects in our department in order to measure their position and size errors with respect to a manually built ground truth. The results obtained are reported in Table 2 and 3.

Table 2: Comparison between the pixels of each processed metric map and the cells of the corresponding Semantic Grid Map.

Map	**Pixels**	**Cells**
BelgioiosoCastle	768 792	11 600
dis-B1	1 080 700	10 290
dis-B1-part	501 840	7372
dis-Basement	992 785	13 455
FortAPHill	534 520	7878
Freiburg	335 248	4794
HospitalPart	30 000	285
Intel	336 399	4473
scheggia	92 984	1116
UBremen	831 264	10 962

In general, even if the error for the object area can reach values around 3, loosing precision is still acceptable from the point of view of the task execution, since after reaching the desired location on the semantic map, an accurate localization of the objects is performed through perception. Overall, the data acquired show that the proposed

Table 3: Average error evaluation for the width (W), depth (D) and area (A) of the objects in the Semantic Grid Map (SGM), normalized with the ground truth values.

Object	**Avg. SGM cells**	**Avg. Err.** e_W	**Avg. Err.** e_D	**Avg. Err.** e_A
Cabinets	4.2	0.31	0.22	0.44
FireExtinguishers	1	1.13	0.67	2.6
RecycleBins	4	0.64	0.82	2.02

Fig. 3: Three classes of objects have been selected for the quantitative evaluation of the object segmentation module: fire extinguisher, cabinet, and recycle bin. In the first column the images of the objects are reported, while the manually obtained ground truth images for the silhouettes of the objects are shown in the second column. The third column contains the results of the segmentation process.

representation substantially decreases the computational load, providing an acceptable approximation of the objects' position and size that suitably supports task execution.

Object Segmentation A quantitative evaluation for the object segmentation process has been carried out by considering the same objects used for the Semantic Grid Map evaluation. In particular, we have evaluated the accuracy of our approach in segmenting multiple instances of three different classes of objects in our knowledge base (i.e., fire extinguishers, cabinets, and recycle bins) as shown in Figure 3.

Table 4 reports the results of the image segmentation process in terms of Detection Rate (DR) and False Alarm Rate (FAR), computed as follows:

$$DR = \frac{TP}{TP + FN} \qquad FAR = \frac{FP}{TP + FP}$$

where TP are the true positives, i.e., correctly segmented pixels, FN are the false negatives, i.e., the number of object points detected as background, and FP are the false

Table 4: Error for the Object Segmentation module in terms of Detection Rate (DR) and False Alarm Rate (FAR).

Object	DR	FAR
Cabinet1	0.865	0.055
Cabinet2	0.946	0.010
Cabinet3	0.622	0.000
Cabinet4	0.841	0.037
Cabinet5	0.911	0.022
FireExtinguishis1	0.621	0.151
FireExtinguishis2	0.677	0.151
FireExtinguishis3	0.795	0.280
RecycleBin1	0.892	0.195
RecycleBin2	0.839	0.119
RecycleBin3	0.900	0.502
RecycleBin4	0.628	0.022

Table 5: Error in extracting the width (W size) of the tagged object.

Object	Ground Truth W	Detected W	Err. e_W
Cabinet1	100 cm	96.56 cm	0.034
Cabinet2		76.03 cm	0.239
Cabinet3		79.16 cm	0.208
Cabinet4		138.20 cm	0.382
Cabinet5		80.50 cm	0.195
FireExtinguishis1	15 cm	11.29 cm	0.247
FireExtinguishis2		11.72 cm	0.218
FireExtinguishis3		15.71 cm	0.047
RecycleBin1	38 cm	44.30 cm	0.165
RecycleBin2		29.25 cm	0.230
RecycleBin3		79.30 cm	1.086
RecycleBin4		34.85 cm	0.082

positives, i.e., the number of background points detected as object points. Low values for DR are mainly caused by holes in the depth data, especially along the borders of the objects. High values for FAR are mainly caused by a slight misalignment between the RGB image and the depth map provided by the sensor. The highest FAR value is obtained in the case of *RecycleBin3* since part of a cabinet alongside the tagged recycle bin is incorrectly segmented as part of it.

Since the final goal of our framework is to acquire knowledge for generating an accurate semantic map, we evaluate also the precision of our segmentation method in extracting the width (W size) of the tagged objects. The results are reported in Table 4. The error e_W is calculated as follows:

$$e_W = \frac{|detected_W - GT_W|}{GT_W}$$

where $detected_W$ is the width detected by our segmentation algorithm and GT_W is the ground truth width. The analysis of the results suggests that the proposed approach can recover the W size of the tagged objects with an acceptable error e_W. The highest e_W value is caused by the erroneously segmented *RecycleBin3*. It is worth noticing that in such a case the system memorizes the tagged object. However, since the W value for *RecycleBin3* is not coherent with the object properties stored in the conceptual KB, a clarification dialog has been implemented to flag this error.

Spatial Reasoning Several tests have been conducted in order to demonstrate the improvements that qualitative spatial reasoning can determine in grounding the commands given by the users to a robot, as well as the efficacy of implementing such an approach on a real robot. Our validation work has been therefore focused on two different kinds of experiments.

The purpose of the first experiment was to evaluate the impact of a qualitative spatial reasoner on an agent whose amount of knowledge continuously grows, as well as

the influence of the already available knowledge on such a reasoning. Such an evaluation has been carried out by considering the number of unambiguous and ambiguous commands (i.e., commands referring to more than one object with a specific spatial property) grounded by the agent. Indeed, when full knowledge about the environment is available, grounding ambiguous commands would mostly lead to the execution of the wrong action with respect to the user expectation, while all the unambiguous commands are supposed to be correctly grounded. We therefore analyzed first the impact of the presence or absence of the qualitative spatial reasoner (QSR) and then the impact of the amount of knowledge available to the agent. In detail, we first asked to 26 students to provide a set of 3 commands containing spatial relations between objects, by looking at pictures of the test environment. Then, from the 78 acquired commands, we extracted two types of tasks: 28 ambiguous and 50 unambiguous. By gradually adding knowledge about the objects inside the knowledge base of the agent, we therefore measured how many commands were grounded. We repeated the experiment for both categories of commands, with or without the qualitative spatial reasoner. Since the curves depend on the order of the objects inserted in the knowledge base, the experiment has been performed five times in order to obtain its average trend (Fig. 4). In case the QSR was not present (red curve), only the objects in the environment, whose category has a unique member, were correctly identified. For example, since we had two cabinets in the test environment, there was no way of distinguish them without exploiting spatial relations. By comparing the two curves in the image, it can be noticed that the presence of the QSR does not greatly affect their trend when a little amount of knowledge is available, due to the absence of exploitable spatial relations between objects. On the contrary this is not true when substantial environmental information is accessible. Note that, when a complete knowledge about the relevant elements of the environment is known by the robot, the number of grounded commands, as expected, is equal to the number of unambiguous phrases (50 commands) present in the adopted set of commands.

The second experiment performed aimed at understanding the limitations of the proposed approach. To this end, we measured the agreement between the user expectations and the grounding performed by the robot. In particular, we first produced a Semantic Grid Map by driving the robot on a tour of the environment and tagging 23 objects within an office environment, as well as the doors and the functional areas in it. Then, we asked 10 different non-expert users to assign 10 distinct tasks to the robot, additionally asking them to evaluate whether the robot correctly grounded their commands, meeting their expectations. The commands have been directly acquired through a Graphical User Interface, in order to avoid possible errors due to misunderstandings from the speech recognition system. In detail, the users had the possibility to choose the action to be executed by specifying the located object, the reference object and one of the 10 spatial relations implemented in our reasoner. Table 6 shows that approximately 80% of the given commands have been correctly grounded. The remaining 20% of wrongly grounded commands where due to two different phenomena: *(i)* the command given was ambiguous, requiring other properties, in addition to direction and distance, to identify the object; *(ii)* the users did not behave coherently during the interaction with the robot, by varying their concept of vicinity or by adopting different reference frames.

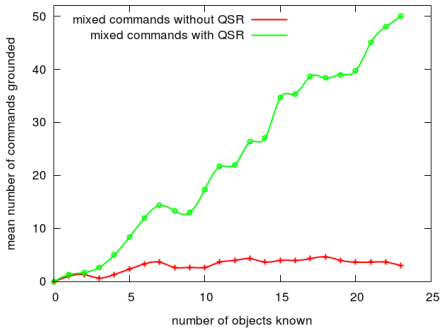

Fig. 4: Mean number of grounded commands with respect to the number of objects known in the environment, added in a random order.

Table 6: Number of correctly and wrongly grounded commands with respect to the expectations of the users.

User	Correctly Grounded Commands	Wrongly Grounded Commands
1st	7	3
2nd	8	2
3rd	10	0
4th	6	4
5th	8	2
6th	8	2
7th	10	0
8th	7	3
9th	9	1
10th	8	2
Total	81	19

5.2 Whole-system Evaluation

For evaluating the system as a whole, three kinds of experiments have been performed, two qualitative and one quantitative. A first set of tests has been carried out to verify the mapping procedure and the automatic construction of the representation of different kinds of environments. The main focus of this first set of experiments has been on demonstrating how a robot, being deployed in an unknown environment, can be endowed with the ability of acquiring specific knowledge of the environment and later using it to accomplish motion tasks. For this type of qualitative validation, two different kinds of environments have been taken into consideration: homes and offices. More specifically, as described in Section 3, we have deployed our system on four different mobile bases in the office spaces of our department and in two different houses. During these tests, several non-expert users have been asked to guide the robot in discovering the environment and the objects in it. After having acquired the specific information about the environment, the users have also been asked to assign simple tasks to the robot through natural language, such as “move in front of the couch next to the tv-set”, in order to test the consistency of the produced environmental representation. In particular our system has been tested in:

- the basement and the first floor of our department. In this environment we mapped four different laboratories and ten offices, as well as the corridors that connects them and asked several non-expert users to tag multiple objects during an open day of our lab.
- the ground floor of a house of one of the authors. With a couple of hours of work we were able to enter an unknown environment, extract a metric map of it and create a semantic map usable to fulfill the commands uttered by a user. In particular, a small environment composed by a kitchen and a living room was mapped and 41

Fig. 5: Domestic environment mapped by the students during the First Örebro Winter School on "Artificial Intelligence and Robotics". The Topological Graph is depicted on top of the Semantic Grid Map and the objects in it. The metric map is also depicted in the background.

different objects were successfully and easily added in the robot's knowledge with the aid of multiple users.

– a domestic environment used at Örebro University for domotic applications. During the First Örebro Winter School on "Artificial Intelligence and Robotics"[3], we created a representation of the apartment composed of a kitchen, a living room, a bed room and a dining-room. As part of their practical activity during the course, the students that participated in the school were invited to help the robot acquiring the knowledge about the objects in the environment. 15 different objects were tagged during this process. An image of the semantic map gathered during the school can be seen in Figure 5.

The second set of tests was performed in order to validate the system in a long-run. We are in fact interested in understanding whether the developed approach is suitable for long-life learning and how well the produced representation can be consistently updated over time. To this end, we developed an on-line mapping experiment, where the segway and the Videre design robot were deployed for three weeks in our department. During this period, the robots interacted with multiple users in order to keep track of the objects that could change position over time. Twenty different object types that changed position over time were thus tagged and stored in the semantic map of the environment. Videos of some of the experiments and several data acquired during them can be found at `http://www.dis.uniroma1.it/~gemignani/Articles/iser14.html`.

The goal of the final quantitative experiment was to evaluate the whole system in a real environment during a typical task executed by the robot. For this reason we deployed our robot in an office environment and we asked both expert and non-expert

[3] `http://aass.oru.se/Agora/Lucia2013/`

Table 7: Result obtained from the test performed on the whole system. The position of the tagged objects is compared with the one obtained from a manually generated ground truth by calculating the distance between the two points.

Distance Thresholds	Average Percentage	Experts Percentage	Non-Experts Percentage
≤ 0.1m	18%	20%	16%
≤ 0.2m	42%	37%	47%
≤ 0.3m	48%	46%	50%
≤ 0.4m	76%	72%	80%
≤ 0.5m	88%	94%	82%

users to drive the robot around using the vocal interface and to tag the various objects present in the environment. To test the robustness of our system in a noisy environment, we carried out a data collection during a public opening of our department asking 10 visitors, in addition to all of the authors of this paper (for a total of 16 users), to take part in the following experiment. The robot started with no knowledge about the objects enclosed in the environment and each user, after being explained for a minute the commands understood by the robot, had to drive, using the vocal interface, the mobile platform in front of a desired object and teach the robot its position and name. Having memorized different objects, the user had to ask the robot to move in front of them in order to demonstrate that the learning process had been carried out successfully. In this experiment all the users have been able to successfully memorize an object, thanks to the behaviors implemented on the robot that allowed to overcome the system components' limitations. After collecting the data needed, we calculated the distance between the position of the centroid of the learned objects with the one belonging to a ground truth manually created. The result of such a comparison is shown in Table 7. From the table it can be noticed that almost 90% of the objects were placed with an error less than 50 cm. The remaining objects were placed at a distance between 50 cm and 1.5 m due to errors deriving from the object segmentation component, the Semantic Grid Map Generator and the robot pose localizer. It can also be noticed that the precision seems not to vary between expert and non-expert users, thus suggesting that this system does not require a specific training to be used. Overall, the evaluation of the performance shows that the system can effectively acquire knowledge about the environment, allowing for the representation in the semantic map of a wide variety of elements. The evaluation also shows that several aspects of the system could be improved. In our view, the most critical improvement would arise from a tighter integration between state of the art techniques for object detection and categorization. Finally, the results of the final experiment with the users show that the approximations that have been introduced in the representation do not affect the execution of the task, thus providing some evidence of a good balance between abstraction and accuracy reached in our representation.

6 Conclusion

The experiments performed with our system show that our semantic mapping approach can be effectively deployed to build, represent and process environmental knowledge, acquired through the aid of the user. Indeed, this approach clearly supports the thesis

that symbiotic autonomy [2] can help to make a step forward in the current robotic capabilities. Moreover, as it has been demonstrated by the deployment of different robotic platforms, the proposed approach is both independent from the chosen robotic platform and also independent from the user interacting with it. Such features allow for an easy deployment of various mobile bases over different experimental scenarios.

Summarizing, a simple, yet effective interaction with the user allows to build a semantic representation of the environment that is much richer and more accurate than existing automatic and user-guided approaches to semantic mapping. Indeed, the proposed approach can be substantially empowered by exploiting some of the state of the art approaches to automatically classify spaces, or to detect and classify objects. Specifically, the robot can take a more proactive role in handling knowledge that can be autonomously acquired through perception either by adding it in the semantic map or by querying the user about it, further developing the approach towards symbiotic autonomy. As a matter of fact, the proposed approach shows a different perspective on the implemented robot capabilities: the system performs intelligent behaviors (or it has an improved performance) not by fully relying on general knowledge, rather by acquiring specific knowledge about the operational environment. This shift of viewpoint, that is enabled by the interaction with the user, is applicable not only to the knowledge about the environment, but also in the knowledge about the tasks to be performed and also about the users of the system.

A second outcome of the proposed experimental setting is the notion of online semantic mapping. This should not be regarded just as a natural extension of the off-line procedure, that enables the robot to accumulate knowledge during operation; more generally, an online semantic mapping capability is needed to enable the robot to continuously adapt to the environment that changes over time. In this respect, our experiments on long-term performance of the robot brought up several interesting research challenges:

- update of the knowledge about objects in the face of new knowledge acquired either through perception or from the user (or different users);
- learn the spatio-temporal relations among the objects in the environment;

Our future research will focus on experiments that encompass the deployment of the robot for long periods of time, thus allowing to investigate the above issues.

Acknowledgements We would like to thank Joachim Hertzberg for insightful discussions on the experimental evaluation of semantic mapping. Moreover, we acknowledge the contribution of Emanuele Bastianelli and Taigo M. Bonnani to the implementation of the system. This work is part of the activities in the RoCKIn Coordination Action[4], which is focussing on benchmarking of home robots through competitions.

References

1. Nüchter, A., Hertzberg, J.: Towards semantic maps for mobile robots. Robot. Auton. Syst. **56**(11) (2008) 915–926

[4] `http://www.rockinrobotchallenge.eu`

2. Rosenthal, S., Biswas, J., Veloso, M.: An effective personal mobile robot agent through symbiotic human-robot interaction. In: Proceedings of 9th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS). (2010) 915–922
3. Bastianelli, E., Bloisi, D.D., Capobianco, R., Cossu, F., Gemignani, G., Iocchi, L., Nardi, D.: On-line semantic mapping. In: Advanced Robotics (ICAR), 2013 16th International Conference on. (Nov 2013) 1–6
4. Hertzberg, J., Saffiotti, A.: Using semantic knowledge in robotics. Robotics and Autonomous Systems **56**(11) (2008) 875–877 Semantic Knowledge in Robotics.
5. Galindo, C., Saffiotti, A., Coradeschi, S., Buschka, P., Fernández-Madrigal, J., González, J.: Multi-hierarchical semantic maps for mobile robotics. In: Proceedings of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS). (2005) 3492–3497
6. Goerke, N., Braun, S.: Building semantic annotated maps by mobile robots. In: Proceedings of the Conference Towards Autonomous Robotic Systems. (2009) 149–156
7. Brunskill, E., Kollar, T., Roy, N.: Topological mapping using spectral clustering and classification. In: Proceedings of IEEE/RSJ Conference on Robots and Systems (IROS). (2007) 3491–3496
8. Friedman, S., Pasula, H., Fox, D.: Voronoi random fields: Extracting the topological structure of indoor environments via place labeling. In: Proceedings of 19th International Joint Conference on Artificial Intelligence (IJCAI). (2007) 2109–2114
9. Mozos, O.M., Mizutani, H., Kurazume, R., Hasegawa, T.: Categorization of indoor places using the kinect sensor. Sensors **12**(5) (2012) 6695–6711
10. Gunther, M., Wiemann, T., Albrecht, S., Hertzberg, J.: Building semantic object maps from sparse and noisy 3d data. In: Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on, IEEE (2013) 2228–2233
11. Zender, H., Martínez Mozos, O., Jensfelt, P., Kruijff, G., Burgard, W.: Conceptual spatial representations for indoor mobile robots. Robotics and Autonomous Systems **56**(6) (2008) 493–502
12. Pronobis, A., Jensfelt, P.: Large-scale semantic mapping and reasoning with heterogeneous modalities. In: Proceedings of the 2012 IEEE International Conference on Robotics and Automation (ICRA'12). (2012) 3515–3522
13. Guivant, J., Nebot, E.: Simultaneous localization and map building: Test case for outdoor applications. In: IEEE Int. Conference on Robotics and Automation. (2002)
14. Nüchter, A., Wulf, O., Lingemann, K., Hertzberg, J., Wagner, B., Surmann, H.: 3D Mapping with Semantic Knowledge. In: RoboCup 2005: Robot Soccer World Cup IX. (2005) 335–346
15. Bastianelli, E., Castellucci, G., Croce, D., Basili, R., Nardi, D.: Effective and robust natural language understanding for human robot interaction. In: Proceedings of the 21st European Conference on Artificial Intelligence, in press. (2014)
16. Capobianco, R., Gemignani, G., Bloisi, D., Nardi, D., Iocchi, L.: Automatic extraction of structural representations of environments. In: Proceedings of the 13th International Conference on Intelligent Autonomous Systems, in press. (2014)
17. Popović, M., Ney, H.: Word error rates: Decomposition over pos classes and applications for error analysis. In: Proceedings of the Second Workshop on Statistical Machine Translation. StatMT '07, Stroudsburg, PA, USA, Association for Computational Linguistics (2007) 48–55