Robotics and Autonomous Systems 78 (2016) 1–16

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Living with robots: Interactive environmental knowledge acquisition

Guglielmo Gemignani *, Roberto Capobianco, Emanuele Bastianelli, Domenico Daniele Bloisi, Luca Iocchi, Daniele Nardi

Department of Computer, Control, and Management Engineering, Sapienza University of Rome, Italy

HIGHLIGHTS

- A method for incremental and on-line semantic mapping based on HRI.
- A four-layered representation for semantic maps used to support robot task execution.
- Throughout description and evaluation of a fully semantic mapping system.

ARTICLE INFO

Article history:
Received 24 June 2015
Received in revised form
26 October 2015
Accepted 2 November 2015
Available online 14 January 2016

Keywords:
Semantic mapping
Knowledge representation
Human–robot interaction

ABSTRACT

Robots, in order to properly interact with people and effectively perform the requested tasks, should have a deep and specific knowledge of the environment they live in. Current capabilities of robotic platforms in understanding the surrounding environment and the assigned tasks are limited, despite the recent progress in robotic perception. Moreover, novel improvements in human–robot interaction support the view that robots should be regarded as intelligent agents that can request the help of the user to improve their knowledge and performance.

In this paper, we present a novel approach to semantic mapping. Instead of requiring our robots to autonomously learn every possible aspect of the environment, we propose a shift in perspective, allowing non-expert users to shape robot knowledge through human–robot interaction. Thus, we present a fully operational prototype system that is able to incrementally and on-line build a rich and specific representation of the environment. Such a novel representation combines the metric information needed for navigation tasks with the symbolic information that conveys meaning to the elements of the environment and the objects therein. Thanks to such a representation, we are able to exploit multiple AI techniques to solve spatial referring expressions and support task execution. The proposed approach has been experimentally validated on different kinds of environments, by several users, and on multiple robotic platforms.

1. Introduction

Robots are expected to be the next technological frontier, impacting our society in many sectors, including security, industry, agriculture, transportation, and services. In particular, robots are expected to become consumer products, that massively enter our homes to be part of our everyday life. This naturally brings back the view of the robot as an intelligent agent, capable of smoothly interacting with humans and operating in real life environments, whose features are undoubtedly challenging. In addition to the recent developments in robotics, other technological advancements make it more feasible to address the design of robots as intelligent agents. Specifically, new sensors allow for more effective approaches to perception, new devices support multi modal interaction with the user and novel speech recognition technologies enable the realization of actual conversational agents.

This notwithstanding, there is still a gap in terms of user expectations and robot functionality. A key limiting factor is the lack of knowledge and awareness of the robot about the operational environment and the task to be accomplished. In other words, the world models that robots embody do not support even simple forms of common sense knowledge and reasoning. Additionally, in order to support the implementation of simple commands such as to "*go near the fridge*" or "*check whether the TV set is on*", state of the art systems typically need a significant engineering effort in coding the knowledge about the specific operational environment. Consequently, it is impractical to scale-up, enabling the robot to

* Corresponding author.
E-mail addresses: gemignani@dis.uniroma1.it (G. Gemignani), capobianco@dis.uniroma1.it (R. Capobianco), bastianelli@dis.uniroma1.it (E. Bastianelli), bloisi@dis.uniroma1.it (D.D. Bloisi), iocchi@dis.uniroma1.it (L. Iocchi), nardi@dis.uniroma1.it (D. Nardi).

http://dx.doi.org/10.1016/j.robot.2015.11.001
0921-8890/

deal with different environments and different requests from the user.

To this end, in this paper we present an approach that allows a robot to incrementally learn the knowledge about the environment, by relying on a rich multi-modal interaction with the user. We specifically address the problem of acquiring the knowledge about the environment (i.e., the semantic map) and maintaining it. Compared with previous work, our approach can be seen as an *incremental on-line semantic mapping*, in which a rich and detailed representation of the operative scenario is built with the help of the user. The resulting integrated representation enables the robot to perform topological navigation, understanding target locations and the position of objects in the environment.

The approach described in this paper builds on previous work for off-line construction of semantic maps [1], based on two main components: (i) a component for Simultaneous Localization And Mapping (SLAM), that provides a metric map of the environment; (ii) a multi-modal interface that allows the user to point at the elements of the environment and to assign their semantic role. The novel contributions of this work are the following:

- The representation of the environment, which is automatically extracted from the metric map and is labeled through user interaction. On this representation different forms of symbolic AI techniques can be applied.
- The process of building and updating the representation, which is done incrementally during the deployment of a robot through a continuous interactive process;

By exploiting the representation of the environment built through this process, we are able to support several forms of reasoning, task execution and complex interactions.

The above listed features have been embedded into a prototype system that has been extensively used over months to validate the proposed approach, in substantially different environments, and with multiple users. This robotic system is a fully functioning prototype able to enter an unknown environment and incrementally create a semantic map that supports the execution of complex tasks in a partially dynamic environment. A semantic mapping approach similar to the one presented here could be deployed on robots that are currently entering the market, such as telepresence robots [2]. In this specific scenario, users could be allowed to give a tour of their homes or offices to the robot, teaching it about relevant objects or rooms in the environment, possibly in different moments. The knowledge acquired through this process could then be used to simplify the teleoperation of a robot by allowing both the remote and the nearby user to command the robot using natural language. For example, users could just instruct the robot to reach a previously learnt room or object, without the need to tele-operate it.

In this paper we focus on the representation, acquisition and use of the robot knowledge. We refer the reader to previous publications for a more detailed description of the robotic system and for an overall discussion of the natural language interaction [3–5]. The remainder of the paper is organized as follows. Section 2 describes the background and provides a suitable context for our work with respect to the state of the art. Section 3 outlines in detail the aim and the contributions of our work. The representation of the robot's knowledge is illustrated in Section 4. Section 5 describes the knowledge acquisition and maintenance, while the uses of the acquired knowledge in the behaviors of the robot and in the dialogs with the user are discussed in Section 6. We illustrate the experiments that we have carried out to validate the proposed approach in Section 7. Conclusions are drawn in Section 8.

2. Related work

Our work mostly relates to the literature on semantic map acquisition and especially on those approaches that rely on human–robot interaction in natural language. After the early-day works in semantic mapping [6], all the recent approaches can be grouped into two main categories, by distinguishing fully automatic methods from approaches involving a human user to help the robot in the semantic mapping process. The first category of *fully automatic* methods, in which human interaction is not considered at all, can be further divided into three different sets. The first set of techniques aims at acquiring features of the environment from laser based metric maps to support labeling and extract high-level information. In [7], for example, environmental knowledge is represented by augmenting a topological map (extracted by means of fuzzy morphological operators) with semantic knowledge using anchoring. The second set of techniques uses classification and clustering for automatic segmentation and labeling of metric maps. For example, the generation of 2D topological maps from metric maps is described in [8] (using AdaBoost), in [9] (using spectral clustering), and in [10] (using Voronoi random fields). Finally, the third set of techniques for object recognition and place categorization uses visual features, such as in [11], or a combination of visual and range information, provided by an RGBD camera, such as in [12]. Although significant progress has been made in fully automated semantic mapping [13], even the most recent approaches still do not scale up in terms of robustness and generality.

In the second category of approaches for *human augmented mapping*, the user role is exploited in grounding symbols into objects that are still autonomously recognized by the robotic platform. In this case, the human–robot interaction is generally uni-modal, and typically achieved through speech. In [14], a system to create conceptual representations of indoor environments is described. A robotic platform owns an *a priori* knowledge about spatial concepts and, through them, builds up an internal representation of the environment acquired through low-level sensors. The user role throughout the acquisition process is to support the robot in place labeling. In [15] the authors present a multi-layered semantic mapping algorithm that combines information about the existence of objects and semantic properties about the space, such as room size, share, and appearance. These properties decouple low-level information from high-level room classification. The user input, whenever provided, is integrated in the system as additional properties about existing objects. In [16] the authors present a mixed initiative strategy for robotic learning by interacting with a user in a joint map acquisition process. This work however considers only rooms, not enabling the system to embed objects in the semantic map. The approach proposed in [17] alternatively adopts human augmented mapping based on a multivariate probabilistic model to associate a spatial region to a semantic label. A user guide supports a robot in this process, by instructing the robot in selecting the labels. Finally, the method presented in [18] enables robots to efficiently learn human-centric models of their environment from natural language descriptions.

Few approaches aim at a more advanced form of human–robot collaboration, where the user actively cooperates with the robot to build a semantic map, not only for place categorization and labeling, but also for object recognition and positioning. Such an interaction is more complex and requires natural paradigms to avoid a tedious effort for non-expert users. For this reason, multi-modal interaction is preferred to naturally deal with different types of information. For example, the authors in [19] introduce a system to improve the mapping process by clarification dialogs between human and robot, using natural language. A similar construction of the representation is also addressed in [20], where the robot learns the features of the environment through the use of narrated