Knowledge-Based Reasoning on Semantic Maps

R. Capobianco, G. Gemignani, D. Nardi, D. Bloisi, L. Iocchi

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

Abstract

Robotic systems should have a deep and specific knowledge about the environment they live in to properly interact with people and effectively perform the requested tasks. To this end, a suitable representation of the environment is needed, including both metric spatial information and semantic representations of locations and objects of interest. In this paper a representation of spatial and environmental knowledge, as well as a method for reasoning on it are presented. More specifically, the representation method is designed to properly integrate metric information about the environment and semantic information provided by the user, allowing for an effective knowledge-based reasoning. The result is a qualitative high-level representation of the environment that embodies all the knowledge required by a robot to actually reason on it and execute complex tasks.

Introduction

Service robotics concerns the use of semi- or fully autonomous systems to perform activities, aiming at achieving the well-being of humans (International Federation of Robots 2013). Since the majority of tasks that robots have to face are inherently complex, human-robot interaction (HRI) can be exploited to overcome such difficulties. Out of all the existing HRI interfaces (e.g., haptic or gesture interfaces), natural language offers a rich and intuitive mechanism by which human users can easily interact with robotic platforms, expressing rules and commands in a very concise way.

However, natural language based interaction presents two main challenges: 1) Correctly interpreting the commands given by the users; 2) Grounding such commands. The first challenge involves the problem of fully understanding the commands that the robot is asked to perform, while the latter one derives from the difficulty of univocally assigning a relation between a physical element and its formal representation.

The aim of this paper is to describe how an effective environmental knowledge representation for command grounding can be automatically built by means of a semantic map. As underlined by (Galindo et al. 2008), existing methods for reasoning on semantic maps suffer of an insufficient expressivity for the environmental representation. On one end, a limited amount of knowledge can be learnt by the system, while on the other, existing approaches do not fully exploit the potential that arises from semantic information, focussing only on the problems of recognizing and locating objects in a map.

This work is a first step towards building a rich knowledge base that can lead to a better command interpretation and to an easier and more effective reasoning process. In particular, we focus on how to obtain a precise and reliable grounding of the commands given by the user. When processing a command, it is crucial to understand both the actions that the robot is required to accomplish and the arguments related to them. For example, if the command is "go to the closet", then the key point is to establish where the robot has to move to. Indeed, it is possible to understand the command and to carry out the appropriate actions only if a suitable representation of the environment is available. The proposed method for automatically building a semantic map has been qualitatively evaluated by processing several publicly available metric maps, as well as maps generated by our robots. A quantitative evaluation of the improving factor that the knowledge representation allows to obtain is being carried out by performing several sessions of experiments, in which different users are asked to interact with our robotic platform through the use of natural language.

The remainder of this paper is organized as follows. Section II discusses the background and the related work. The representation of the robot's knowledge is presented in Section III, while the method for its construction is described in Section IV. The experimental evaluation of the approach is shown in Section V and the conclusions are drawn in the last section.

Related Work

The literature about environmental knowledge representation and acquisition, and in particular about acquiring and representing semantic maps, can be divided into two main categories, by distinguishing automatic processing from the so called "human-in-the-loop" approaches, where a user is asked to help the robot in the acquisition process.

A first set of works attempted a fully automated approach in order to construct environmental maps, by classifying functional areas. In the early stages of research in this field, many methods focused on the extraction of room attributes. For example, (Fabrizi and Saffiotti 2002) and (Buschka and Saffiotti 2002) presented a local technique that uses range information to detect room-like data during navigation, through a virtual sensor. Starting from the idea of dividing occupancy grid into regions separated by local narrowings, as proposed by (Thrun 1998), they worked on an algorithm for partitioning the space in an incremental way. In particular, computing a certain number of parameters for each region, they produced a local-topological map. Similarly, (Anguelov et al. 2004) focused on doorways detection and object-based modeling in order to define a probabilistic model of corridors containing doors and walls. Specifically, they parametrized each object by its shape, color, and motion model in order to obtain the distribution over possible observations of the robot. Moreover, in (Galindo et al. 2005), environmental knowledge is represented by augmenting with semantic knowledge a topological map, which is extracted with fuzzy morphological operators.

A second set of techniques make use of classification and clustering for the automatic segmentation and labeling of metric maps. For example, in (Nüchter et al. 2005), environmental knowledge is extracted by labeling 3D points through the gradient difference between neighboring points which are then classified as floor-points, object-points or ceiling-points. (Mozos, Stachniss, and Burgard 2005) and (Martinez Mozos et al. 2007), instead, extracted simple geometric real-valued features from scans, and classified them through an Adaboost multi-classifier obtained by arranging several weak binary classifiers in a decision list. A similar approach has been proposed by (Goerke and Braun 2009), in which a learning classifier is used to build semantic annotated maps from laser range measurements. Various alternative approaches have been proposed, based, for example, on spectral clustering (Brunskill, Kollar, and Roy 2007) or Voronoi random fields (Friedman, Pasula, and Fox 2007). Only few works, however, consider the acquisition and classification of objects.

Recently, techniques for object recognition and place categorization, based on visual features (Wu, Christenseny, and Rehg 2009), or a combination of visual and range information provided by an RGB-D camera (Mozos et al. 2012) have been proposed. For example, (Pangercic et al. 2012) investigate the representation and acquisition of Semantic Objects Maps (SOMs) in kitchen environments, with low-cost RGB-D sensors by using vision and active manipulation actions such as opening drawers and doors.

As shown by these works, a significant progress has been made in fully automated semantic mapping, however, relying on fully automated interpretation of sensor data has multiple limitations in the knowledge acquisition process. Indeed, sensor data interpretation often introduces errors or uncertainties and it is non-trivial to deal with them in an automatic way.

In order to overcome the above mentioned limitations, several researchers suggested to use *human augmented mapping*, in which the user actively supports the robot to acquire the required knowledge about the environment. For example, in (Diosi, Taylor, and Kleeman 2005) an interactive SLAM procedure and a watershed segmentation are employed in order to create a contextual topological map. While (Zender et al. 2008) describes a system for the creation of conceptual representations of human-made indoor environments. In their 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. The user role, during the acquisition process, is to support the robot in the activity of place labeling, while the obtained representation is also used for human-robot dialogue.

A more general approach to human-robot collaboration for semantic mapping is taken by (Kruijff et al. 2006), where clarification dialogues are used to improve the quality of the adopted representation. Through the use of natural language, therefore, their system increases the robots robustness when dealing with uncertainties or incomplete information. Moreover, (Nieto-Granda et al. 2010) adopt human augmented mapping based on a multivariate probabilistic model to associate a spatial region to a semantic label. In particular, the user supports this activity directly teaching the labels to the robot, which is taken on a tour of the environment. Then, taking a laser scan measurement and fitting a Gaussian to the resulting points, the mean and the covariance are stored in the map along with the label provided by the human. During navigation, the robot computes the Mahalanobis distance from its position to the mean of each Gaussian and, if not sufficiently confident its position in a region with a known label, it prompts the user to input the name of the current region. (Pronobis and Jensfelt 2012), instead, use heterogeneous modalities for a comprehensive multi-layered semantic mapping algorithm, aiming at place categorization and topological map construction. Their system builds a probabilistic representation that includes information about the existence of objects and properties of space, such as room size, shape and appearance. 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 process the support of the user does not play a central role, (Randelli et al. 2013) 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. However, a suitable representation of the acquired knowledge into an expressive semantic map is missing.

In our work, the knowledge acquisition process and the construction of the semantic map integrates the two approaches described above: from one side, a user guides the robot through the environment describing objects and locations of interest with a multi-modal interaction scheme (this part is better described in (Randelli et al. 2013)), from the other side the construction of the knowledge base from data acquired through the sensors of the robot and from information provided by the user is obtained through an automatic process, that is described in this paper. This integration allows both for acquiring rich knowledge about the environment and for automatically building a symbolic representation that is used by the reasoning module of the our system in order to solve spatial referring expressions.

Symbolic Representation

In this section we describe the representation of the robot's knowledge, while the description of how this knowledge is actually built is provided in the next section.

The representation of the robot's knowledge is divided into two layers: the *World Knowledge*, that represents the specific knowledge about the environment that the system acquires and in which the robot is operating, and the *Domain Knowledge*, which is a general knowledge about a certain type of domains. It is important to point out that, while the two components may recall the extensional and intensional components of a classical knowledge base (KB), in this work they are used in a different manner. The World Knowledge, in fact, may be inconsistent with the Domain Knowledge, which is used to support the action of the robot only when specific World Knowledge is not available. For example, when the user asks the robot to reach a specific object, if there is no specific knowledge about the position of the target, the system will refer to the general Domain Knowledge to find out a possible location, asking the user to confirm its location. In this way, the exceptions that are typical of each environment can be explicitly stored in the World Knowledge (e.g., a printer could be in a restroom), without considering them in the Domain Knowledge.

In previous work, the Domain Knowledge has typically been characterized as a conceptual knowledge base representing a hierarchy of concepts, properties and relations. In particular, this type of knowledge base is usually represented as a taxonomy of concepts considered to be representative of any environment, linked by an *is-a* relation (Galindo et al. 2005) and (Martínez Mozos and Burgard 2006). The term $\mathcal{C}oncepts$, in fact, refers to a set of symbols denoting the abstraction of a certain number of elements, which can be used in the World Knowledge to characterize the specific instances of the environment. In the representation used in this system, three classes have been considered:

1. *Areas*: Places in the environment (corridors, rooms, etc.);
2. *Structural elements*: Static entities that form the environment and that topologically connect areas (windows, doors, etc.);
3. *Objects*: Elements in the environment not related to its structure and located within areas (printers, tables, etc.).

The knowledge base, therefore, contains environmental properties of the defined concepts, like size, functionalities of objects and connections among places. Those, in fact, are useful to describe the general knowledge about an environment and to support the robot's task execution. For example, in a home environment, the information that a fridge can be usually found in a kitchen is useful to the robot when performing some tasks related to that object, even if its exact location is not known. For this reason, symbols like $Fridge$ and $Kitchen$[1] are part of the general Domain Knowledge which is commonly used in this system. Also synonyms are stored in the KB, thus providing to the user the possibility to refer to the same object with different natural language expressions with the same meaning. The structure of the taxonomy, during interactions, is therefore explored, and more specific or more generic concepts are extracted, together with spatial relations useful for the disambiguation of the targets to be reached by the robot. An important difference with respect to previous work is the role of such component in the system: indeed, it is not related to the classification of spaces and objects, but to the inheritance of properties which may support both map acquisition and different actions of the robot in the environment. For example, spatial properties can be used to build a metric representations of the objects; functional properties can be used to determine their location and possible uses (e.g., for the developement of behavioral plans); physical properties can be used to check preconditions for executing some action.

The process of enriching a low-level representation of the world is usually realized applying a set of labels to each location of the map. The term $\mathcal{L}abel$ is used to refer to a set of symbols indicating specific instances of objects or locations. For example, $fridge1$ is a label denoting a particular fridge and $kitchen1$ is a label denoting a particular kitchen[2]. In this article, the associations between labels and concepts are denoted as $label \mapsto Concept$ (e.g., $fridge1 \mapsto Fridge$), meaning that the label $label$ is related to the concept $Concept$ (i.e., $fridge1$ is a fridge). It is important to notice, however, that the meaning of the labels is simply that of a pointer to the concept, rather than an instance of it. In this way, a labeled object can be enriched with general domain information, but it is not required to be consistent with the domain KB.

In this work, labels have been used to denote functional areas, and to represent the presence of structural elements of the environment or objects. The representation which has been adopted is built, in particular, with the support for a much more detailed description as an added value. This goal is achieved thanks to the exploitation of the hints provided by the user through human-robot interaction. Intuitively, such an approach is helpful when the results provided by state-of-the-art techniques are still too far from the goal and, therefore, the problem is still challenging (e.g., classifying a chair from images). Instructing the system to recognize the specific instances of chairs that the user has in his/her home, indeed, is a practical and effective solution.

In detail, looking for a rich representation, the representation formalism of the World Knowledge has been structured in four different knowledge structures. The first structure is the *Metric Map*, which is represented as an occupancy grid generated by a SLAM method. This map has usually a fine discretization (e.g., 5 cm) and is used for low-level robot tasks, such as localization and navigation. The second representation are *Instance Signatures*, which are represented as a data base of structured data, where each instance has a unique label ($l \in \mathcal{L}abel$), an associated con-

[1] Concepts are written with the first letter capitalized.

[2] In this article labels are denoted with all lowercase letters, typically followed by a digit.

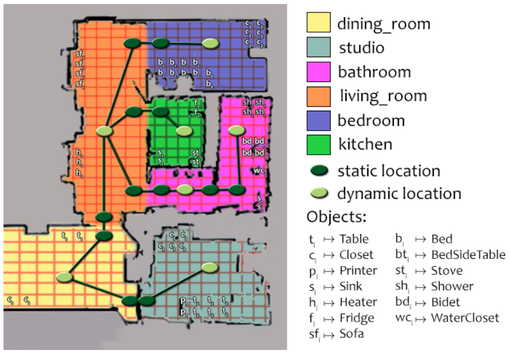

Figure 1: Example of World Knowledge.

cept ($C \in \mathcal{C}oncept$) such that $l \mapsto C$, and a set of properties (including, for example, the position in the environment) expressed as attribute-value pairs. The third representation is the *Cell Map*, represented as a discretization of the environment in cells of variable size. Each cell represents a portion of a physical area and is an abstraction of locations that are not distinguishable from the point of view of robot high-level behaviors. The Cell Map also includes a function $f : Cell \rightarrow 2^{\mathcal{L}abel}$, that maps each cell to a set of labels associated to concepts in the conceptual KB of the Domain Knowledge, and a connectivity relation $Connect \subseteq Cell \times Cell$, that describes the connectivity between adjacent cells. The fourth and final structure is the *Topological Graph*, which is a graph where nodes are locations associated to cells in the Cell Map and edges are connections between these locations. Locations are distinguished in two types: static and dynamic. For the static locations, the corresponding positions (i.e., the correspondences with the metric map) are fixed, while in the dynamic locations, the corresponding positions are variable within a given area of the environment. Since the Topological Graph is used by the robot for navigation purposes, the edges also contains the specific navigation behavior that is required for the robot to move from one location to another.

Figure 1 shows a graphical representation of the World Knowledge as described above. The Metric Map, generated by the SLAM method with a resolution of 5cm, is shown as background of the image, where black pixels represent occupied cells. Every object and location reported in the figure has an entry in the data base of instance signatures. In particular, each instance is determined by a unique label representing an entity within the mapped environment, (e.g., $fridge1$), the corresponding concept according to the conceptual KB (e.g., $f_1 \mapsto Fridge$), and a set of properties (e.g., position = $< x, y, \theta >$, color = $white$, open = $false$, ...). The Cell Map, obtained with the methodologies described in the next section, is shown in the same image: cells are delimited by solid borders while colors and text in the cell specify the labels associated to each cell. The labels refer to a typical area of a home (kitchen, living room, bathroom, bedroom, etc.) and to typical objects (tables, chairs, heaters, doors, etc.), while each cell can c more than one label, thus representing that the same area can be associated with more concepts. For example, the green cells labeled with f_1 (top-right corner of the kitchen area) are mapped to the labels $\{f_1, kitchen1\}$, respectively associated with the concepts $Fridge$ and $Kitchen$, i.e. $f_1 \mapsto Fridge$ and $kitchen1 \mapsto Kitchen$. Thus, the corresponding area is characterized as being occupied by a fridge and as belonging to the kitchen. Connectivity relations can be derived by adjacent cells. Finally, the Topological Graph, is shown in the figure as a graph connecting oval nodes: dark nodes are static locations, while light nodes are dynamic locations, that are associated to the main areas (labels) of the environment. The static locations denote specific positions that are of interest for the robot tasks (e.g., the position to enter in a room), while the dynamic locations are used to denote areas, where instantiation of the position (i.e. the mapping to the metric map) for a navigation behavior is executed at run-time, depending on the current status of the robot and on its goals.

Semantic Map Building and Reasoning

The building process of the proposed spatial and environmental knowledge representation is divided into two phases. The first step automatically generates a *Grid Map*, which is a discretization of the input 2D Metric Map. The output of such procedure, is later processed to produce the Cell Map and the Topological Graph, on which reasoning is performed.

Automatic Building of the Grid Map

The Grid Map represents a suitable discretization of the environment, enabling the system to later associate a position in the real world (i.e., a portion of the Metric Map) to a particular set of labels (i.e., entries of the instance signature database), thus realizing an effective qualitative representation of the environment. In order to produce such an abstraction, first the algorithm shown in Algorithm 1 is applied, taking as input the Metric Map and outputting a grid layer that reflects the structure of the environment. Indeed, each cell of the grid represents a portion of a physical area of the environment and, from a functional point of view of the robot's high-level behaviors, it is an abstraction of locations that are not distinguishable. Since the system is focused on knowledge representation for indoor environments, the production of the grid relies on the extraction of the lines corresponding to the walls. Those are highly informative, being commonly used for the separation of different functional areas and usually providing, from a top-view, a regular structure to the building. Indeed, walls can usually be represented as straight horizontal or vertical lines, thus describing the map through the use of simple elements. However, even considering those regular patterns, the lines corresponding to the walls cannot be extrapolated directly looking at the metric map. In fact, three problems arise when considering the occupancy grid in a straightforward manner: *(i)* noise, due to positions occupied by humans moving in the environment during the map acquisition or by very small objects, such as table or chair legs; *(ii)* incomplete laser-scans, which compromise the regular structure of the building (e.g., caused by closed doors); *(iii)* big objects (e.g., a closet), at the border

Algorithm 1: Pseudocode for the Grid Map Building

Data: $\mathcal{G}$: grid lines set, Q: processing queue, $\mathcal{W}$: wall lines set, $\mathcal{D}$: detected lines set

Input: *map*: metric map

Output: $\mathcal{G}$

1 $\mathcal{G} \leftarrow \varnothing$;
2 $Q \leftarrow \varnothing$;
3 $\mathcal{W} \leftarrow \varnothing$;
4 $\mathcal{D} \leftarrow \varnothing$;
5 $edge\_map \leftarrow CannyEdgeDet(metric\_map, 1, 30)$;
6 *threshold* $\leftarrow$ max_threshold;
7 **while** *threshold* $>$ *min_threshold* **do**
8   *min_distance* = $\max(\frac{100}{threshold}, 5)$;
9     $\mathcal{D} \leftarrow$ *HoughTransform*(map, distance_resolution
10   = 2, angle_resolution = $\pi/2$, threshold);
11   *SelectParallelLinesWithDist*(min_distance, $\mathcal{W}$, $\mathcal{D}$);
12 *extendLines*($\mathcal{W}$);
13 $\mathcal{Q} \leftarrow$ *SelectPairOfParallelAndConsecutiveLines*($\mathcal{W}$);
14 **while** $Q \neq \varnothing$ **do**
15     $< first\_line, second\_line > \leftarrow$ *last_element*(Q);
16   **if** *areVertical(*first_line, second_line*)* **then**
17   *max_dist* $\leftarrow x_{min}$;
18   **else**
19   *max_dist* $\leftarrow y_{min}$;
20   *val* $\leftarrow \lfloor \frac{|dist(first\_line) - dist(second\_line)|}{max\_dist} \rfloor$;
21   **if** *val* $\geqslant 2$ **then**
22   dist2 $\leftarrow \frac{dist(first\_line) + dist(second\_line)}{2}$;
23   new_line $\leftarrow$ *line*(dist2,*angle*(first_line));
24         $\mathcal{G} \leftarrow$ new_line;
25   Q $\leftarrow < first\_line, new\_line >$;
26   Q $\leftarrow < new\_line, second\_line >$;

or within the building, which produce occupied locations in the metric map.

In order to reduce the noise and effectively detect the edges of the building a pre-processing step is required, where the Canny Edge Detector is applied (line 5). The determination of the threshold values has been realized through a continuous, multi-map tuning procedure, considering a trade-off between overall performance and noise reduction. In detail, these values have been selected to be 1 for the lower threshold and 30 for the upper one, obtaining a low reduction of the noise which, however, ensures that good lines are never discarded from the map.

Recognizing walls requires an additional step to resolve two problems: noise and incomplete scans in the metric map. The common sense heuristic behind wall detection that has been used in this work is that of considering, in a first moment, only long and continuous lines. Such lines are detected through the use of the *Hough Transform* (line 10). Furthermore, due to the regular structure which usually characterizes our buildings, only horizontal and vertical lines are considered as candidates in this process. Then, gradually reducing the threshold value and, consequently, the length constraints (line 7), all the other lines are accepted only if their distance from the previously acquired lines is greater than a particular value; such value is determined adaptively as a function of the accumulator threshold. In this way, a big portion of false-positives corresponding to irregularities in the laser scans and to noise are discarded, being evaluated in the last cycles of execution of the algorithm, when almost all the walls have been already detected.

Having automatically detected the walls of the metric map, the Grid Map can be computed. In order to achieve this result, two steps are required. First of all, the lines generated through the wall detection are extended to the whole image (line 12), by computing the minimum horizontal and vertical distances (x_{min}, y_{min}). Then, considering each pair of parallel lines, a new one is inserted between them (lines 21-26) only if their distance is at least twice x_{min} or y_{min}, depending on their angle. This adaptive procedure, creates a grid whose cells have a size between $x_{min} \cdot y_{min}$ and $2x_{min} \cdot 2y_{min}$. The obtained discretization provides, without any loss of information, a qualitative representation that is consistent both with the metric and symbolic layers. On one hand, the grid is built on the basis of the walls and, therefore, it is consistent with the structure of the environment; on the other hand, the granularity is high enough to provide to the robot the basis for a full symbolic and semantic representation of the building.

Cell Map and Topological Graph

The Cell Map is the linking layer between metric and symbolic information, in which each label of the Instance Signature Data Base is associated to a cell of the previously described Grid Map. This abstraction level is represented by a matrix, in which each element of the data structure represents a cell of the grid map.

The cell map abstraction endows the robot with the capability to acquire the environmental and spatial knowledge needed to effectively perform tasks. In particular, the following knowledge categories has been selected to aid the robot in performing its tasks: *(i) Areas*, represented by a string with the tag of the functional area; *(ii) Objects*, defined with a vector containing the tag of each object located in the cell; *(iii) Doors*, described by a vector of boolean values representing the presence of a door between that cell and a surrounding one; *(iv) Properties*, expressed by a vector containing an list of properties for each tagged object.

On the basis of such knowledge, the environment is segmented in different functional areas. In particular, all the cells of the grid within the building are assigned to an area-label, while the external elements are considered to be out of any functional area. This task is performed through the creation of a *Room Map* on the basis of the contours extracted from the metric map, together with the off-line acquired knowledge about areas and doors.

The segmentation process is based on a simple heuristic: each room is distinguished from the others thanks to,

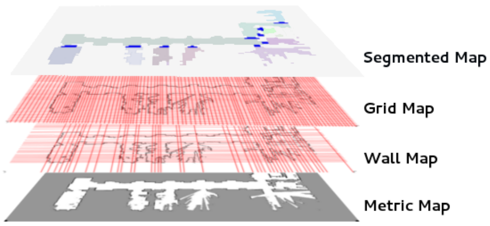

Figure 2: Sequential building steps for achieving area segmentation of the environment.

at least, a door. Therefore, evaluating all the acquired tags, a set of lines of appropriate size, corresponding to the doors, is added to the detected contours of the metric map, closing each functional area. Each area-tag is, then, used as a seed for a Watershed-based region growing algorithm. Finally, the opening morphological operator (which is an erosion followed by dilation) is applied to the obtained room map. Each room can then be extracted applying a color based segmentation starting from the tag of the desired area. The results of the steps described so far in this section are depicted in Figure 2.

The other two categories of knowledge (i.e., objects and their properties) are acquired on-line through human-robot interaction. For such purpose, a specific multi-modal interaction has been devised, where the user is asked to point objects in the environment through the aid of a commercial laser pointer, while uttering their name. The properties are later extracted by analyzing the visual perceptions of the robots and by querying the Domain Knowledge. Since objects and their properties are key components to operate in the environment, an intermediate layer of abstraction is required between such tagging procedure and the integration of the off-line knowledge stored. For such reason, a module devoted to the memorization and processing of the objects and their properties has been developed.

The object handler is responsible for the positioning and computation of the vertices corresponding, in the map, to the tagged elements. Subsequently, all the cells of the grid included between those corresponding to the vertices are assigned to the label of the acquired element. Finally, in order to fully reproduce the object in the cell map on a qualitative level, each label is enriched with the properties memorized by the object handler. In this way, the robot can perform reasoning and disambiguate between objects of the same type, in order to interact both with the environment and the user.

The final knowledge representation that is built is the Topological Graph. A dedicated module has been developed to construct such representation needed for navigating and acting in the environment. Thus, for each room mapped in the cell map, a dynamic node is created while, for each doorway that connects such a room to another, a static node is created with its fixed position set in front of the considered door. Finally, each static node is connected to the dynamic node of the room it belongs to and to the static node that represents the location in front of the other side of its related door. The construction of the topological graph highlights a twofold aspect of our system: both the knowledge provided through user interaction and a-priori high-level knowledge contribute to refine the overall mapping process. For a graphical representation of the final result obtained with this procedure, we refer to Figure 1.

Reasoning on the Semantic Map

Given the various kinds of knowledge built in the Semantic Map, various forms of reasoning can be performed. To this end, all the knowledge included in the Cell Map and in the Topological Graph is automatically translated to Prolog assertion predicates. In particular, each cell of the cell map is represented with a predicate `cellIsPartOf(XCoord, YCoord, AreaTag)` and each object is represented with the two predicates `object(Id, XCoord, YCoord, Properties)` and `Type(Id, Type)`. For example, the knowledge of a white plug located in the cell with grid map coordinates 45, 67 belonging to the corridor area will be represented with the three predicates `cellIsPartOf(45, 67, corridor)`, `object(plug1, 45, 67, color-white)` and `objectType(plug1, plug)`. The knowledge stored in the topological graph is instead represented as an acyclic graph in prolog with the predicates `dynamicNode(Id, XCoord, YCoord)`, `staticNode(Id, XCoord, YCoord)` and `arc(Id1, Id2)`, respectively for the dynamic nodes, static nodes and arcs of the graph. Having translated in Prolog all the knowledge stored in the representation built with the previously described method, performing certain kind of inference on it becomes straight forward. In particular, a module for the resolution of spatial relation references has been written in Prolog. In this module, one predicate for each of the relations *next to, nearest, near, in, left of, right of, belonging to, in front of* and *into* has been written for effectively resolving such references by exploiting the connectivity relations between cells. Moreover, inheritance of properties based on the domain ontology, as well as the possibility of searching for objects that specify a given property (e.g., a specific color) have also been implemented.

Experimental Validation

A first experimental validation was carried out in order to show the feasibility, the effectiveness, and the robustness of the proposed representation. Several experiments were therefore conducted by using the publicly available Radish Data Set[3]. Specifically, six 2D metric maps[4], obtained from different SLAM methods, have been processed in order to validate the area segmentation approach on a number of different occupancy grids. In addition, four maps generated by our robots have been processed, obtaining a total of ten completely different environments. Since only the environments

[3] `http://radish.sourceforge.net/index.php`

[4] The algorithm was applied to the following maps: *albert-b-laser*, *ap_hill_07b*, *ubremen-cartesium*, *intel_lab*, *belgioioso* and a portion of *hospital_floorplan_fort_sam_houston*

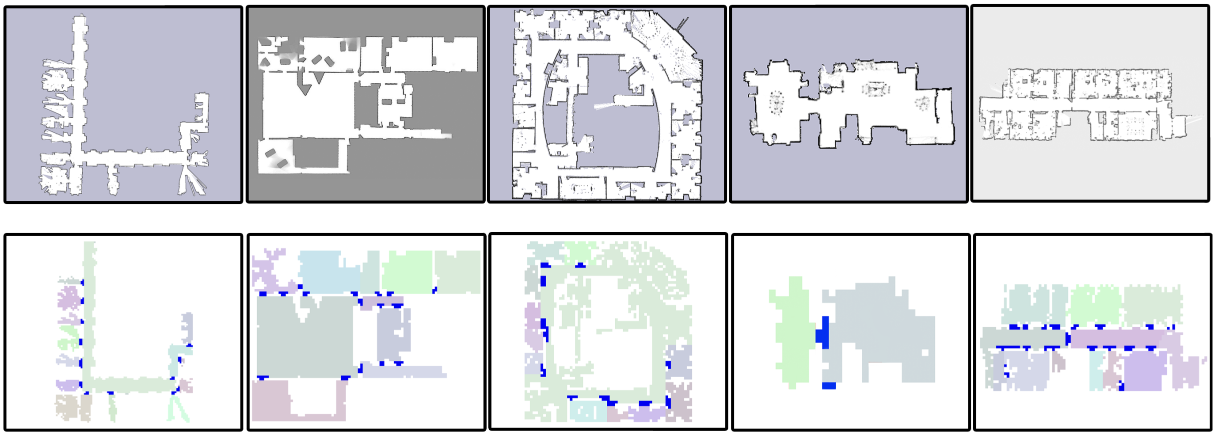

Figure 3: 2D metric maps (top images) and results obtained by applying the area segmentation method (bottom images).

represented in the maps produced by our robots were accessible, only such maps have been augmented with the knowledge about objects, acquired on-line through a specific developed human-robot interaction.

Qualitative results obtained by applying our automatic area segmentation algorithm are shown in Figure 3, where both the initial input metric maps and the final segmented maps are depicted. For all the considered environments, our representation allows to correctly distinguish between different functional areas, as well as doors. It can be observed that a very good correspondence can be obtained between the real environment and the symbolic representation resulting from the application of our algorithm. Finally, it is worth noticing that, even if the system has been developed for ordinary buildings, it can be applied, without loss of generality, to environments with irregular edges, as it can be seen in the central map shown.

After evaluating the method for building the KB, we are currently addressing the improvement of the performance of the robot in the interaction with the user. To this end, we are carrying out two different quantitative sets of experiments in order to measure: 1) The time and the number of interactions needed by the robot to disambiguate a particular target, keeping constant its knowledge; 2) The effectiveness of the system with respect to different amount of knowledge available to the robot. The setup for the evaluation of these two experimental settings is the following. A set of commands $\mathcal{S}$ has been collected through a web interface by asking non-expert users to provide directions to move the robot in front of a target, by exploiting spatial relations between objects present in the scenes. Then, the robot is located in a random position of the environment and it is told, through natural language interaction with multiple users, to reach one of the targets in the environment that have been pointed out by the non-expert volunteers.

The first set of experiments consists in comparing the proposed approach with a basic version of the system (Bastianelli et al. 2013), used as a baseline. A rich knowledge about the environment (including the 2D metric map, as well as all the objects of interest in the environment) is made available to the robot. In the basic version the robot can understand a particular referred target only if its specific natural language tag is used by the user. For example, given the command *"go to the closet"* the robot moves to the specified target only if *"closet"* is in the knowledge base. Furthermore, if multiple instances of *"closet"* are present in the knowledge base the robot, when ordered to reach a generic closet, moves in front of the first temporally tagged one, without engaging in a clarifying dialog with the user. In the current system, instead, synonyms, generalizations, and specifications of objects, as well as multiple spatial relations can be exploited to refer to targets and to disambiguate between objects, that have been tagged with the same label, trough clarifying dialogs.

The second set of quantitative experiments aims instead at measuring the effectiveness of the system with respect to different amount of knowledge available to the robot. The same setup of the first set of experiments is used, given the set of commands $\mathcal{S}$ to the robot. This time, the number of known objects by the robotic system is gradually varied.

Visual evidence of our experiments as well as all the maps processed by our automatic segmentation algorithm are available at `www.dis.uniroma1.it/~gemignani/Articles/ResSemMaps.html`.

Conclusion

In this paper we have presented a novel method for building an effective representation of the environment. By taking as input a metric map produced by a SLAM algorithm and a set of area tags, we have shown how to represent the environmental and spatial knowledge through the automatic construction of a Cell Map and a Topological Graph. Exploiting these two knowledge representations the resolution of spatial reference has been addressed.

In order to test the proposed approach, such an automatic method has been applied to ten different maps, both downloaded from the internet and created from scratch. The developed algorithms have also been embedded in a more complex system that has been installed on two different robots able to interact with multiple users. The performed experiments have underlined the effectiveness of the proposed approach, by testing it in two very different kind of environments.

For future works, a number of additional features need to be addressed. In particular, we are investigating the issues that arise during the update and maintenance of the knowledge base, when considering dynamic objects that change location.

References

Anguelov, D.; Koller, D.; Parker, E.; and Thrun, S. 2004. Detecting and modeling doors with mobile robots. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*.

Bastianelli, E.; Bloisi, D.; Capobianco, R.; Gemignani, G.; Iocchi, L.; and Nardi, D. 2013. Knowledge representation for robots through human-robot interaction. *arXiv:1307.7351*.

Brunskill, E.; Kollar, T.; and Roy, N. 2007. Topological mapping using spectral clustering and classification. In *Proceedings of IEEE/RSJ Conference on Robots and Systems (IROS)*.

Buschka, P., and Saffiotti, A. 2002. A virtual sensor for room detection. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 637–642.

Diosi, A.; Taylor, G.; and Kleeman, L. 2005. Interactive slam using laser and advanced sonar. In *Proceedings of the IEEE International Conference on Robotics and Automation*, 1103–1108.

Fabrizi, E., and Saffiotti, A. 2002. Augmenting topology-based maps with geometric information. *Robotics and Autonomous Systems* 40(2):91–97. Online at http://www.aass.oru.se/˜asaffio/.

Friedman, S.; Pasula, H.; and Fox, D. 2007. Voronoi random fields: Extracting the topological structure of indoor environments via place labeling. In *Proceedings of 19th International Joint Conference on Artificial Intelligence (IJCAI)*.

Galindo, C.; Saffiotti, A.; Coradeschi, S.; Buschka, P.; Fernández-Madrigal, J.; and González, J. 2005. Multi-hierarchical semantic maps for mobile robotics. In *Proceedings of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 3492–3497.

Galindo, C.; Fernández-Madrigal, J.-A.; González, J.; and Saffiotti, A. 2008. Robot task planning using semantic maps. *Robotics and Autonomous Systems* 56(11):955–966.

Goerke, N., and Braun, S. 2009. Building semantic annotated maps by mobile robots. In *Proceedings of the Conference Towards Autonomous Robotic Systems, Londonderry, UK*.

International Federation of Robots. 2013. Service Robots. `http://www.ifr.org/service-robots/`.

Kruijff, G.; Zender, H.; Jensfelt, P.; and Christensen, H. 2006. Clarification dialogues in human-augmented mapping. In *Proceedings of the 1st Annual Conference on Human-Robot Interaction (HRI'06)*.

Martínez Mozos, O., and Burgard, W. 2006. Supervised learning of topological maps using semantic information extracted from range data. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2772–2777.

Martinez Mozos, O.; Triebel, R.; Jensfelt, P.; Rottmann, A.; and Burgard, W. 2007. Supervised semantic labeling of places using information extracted from sensor data. *Robotics and Autonomous Systems* 55(5):391–402.

Mozos, O. M.; Mizutani, H.; Kurazume, R.; and Hasegawa, T. 2012. Categorization of indoor places using the kinect sensor. *Sensors* 12(5):6695–6711.

Mozos, O. M.; Stachniss, C.; and Burgard, W. 2005. Supervised learning of places from range data using adaboost. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, 1730–1735. IEEE.

Nieto-Granda, C.; III, J. G. R.; Trevor, A. J. B.; and Christensen, H. I. 2010. Semantic map partitioning in indoor environments using regional analysis. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1451–1456.

Nüchter, A.; Wulf, O.; Lingemann, K.; Hertzberg, J.; Wagner, B.; and Surmann, H. 2005. 3D Mapping with Semantic Knowledge. In *RoboCup 2005: Robot Soccer World Cup IX*.

Pangercic, D.; Pitzer, B.; Tenorth, M.; and Beetz, M. 2012. Semantic object maps for robotic housework-representation, acquisition and use. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, 4644–4651. IEEE.

Pronobis, A., and Jensfelt, P. 2012. Large-scale semantic mapping and reasoning with heterogeneous modalities. In *Proceedings of the 2012 IEEE International Conference on Robotics and Automation (ICRA'12)*.

Randelli, G.; Bonanni, T. M.; Iocchi, L.; and Nardi, D. 2013. Knowledge acquisition through humanrobot multimodal interaction. *Intelligent Service Robotics* 6:19–31.

Thrun, S. 1998. Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence* 99(1):21–71.

Wu, J.; Christenseny, H. I.; and Rehg, J. M. 2009. Visual place categorization: Problem, dataset, and algorithm. In *Proceedings of IEEE/RSJ Conference on Robots and Systems (IROS)*.

Zender, H.; Martínez Mozos, O.; Jensfelt, P.; Kruijff, G.; and Burgard, W. 2008. Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems* 56(6):493–502.