Multi-Hierarchical Semantic Maps for Mobile Robotics

C. Galindo[1], A. Saffiotti, S. Coradeschi, P. Buschka
Center for Applied Autonomous Sensor Systems
Dept. of Technology, Örebro University
S-70182 Örebro, Sweden
{*alessandro.saffiotti,silvia.coradeschi,par.buschka*}*@aass.oru.se*

J.A. Fernández-Madrigal, J. González
System Engineering and Automation Dept.
University of Málaga
Campus Teatinos, 29071 Málaga, Spain
{*jafma,jgonzalez*}*@ctima.uma.es*

Abstract—The success of mobile robots, and particularly of those interfacing with humans in daily environments (e.g., assistant robots), relies on the ability to manipulate information beyond simple spatial relations. We are interested in semantic information, which gives meaning to spatial information like images or geometric maps. We present a multi-hierarchical approach to enable a mobile robot to acquire semantic information from its sensors, and to use it for navigation tasks. In our approach, the link between *spatial* and *semantic* information is established via *anchoring*. We show experiments on a real mobile robot that demonstrate its ability to use and infer new semantic information from its environment, improving its operation.

Index Terms—Semantic maps, Mobile robots, Anchoring, Knowledge representation, Abstraction, Symbol grounding.

I. INTRODUCTION

The study of robot maps is one of the most active areas in mobile robotics. This area has witnessed enormous progress in the last ten years mostly based on metric and/or topological representations (see [1] for a comprehensive survey). There are, however, other types of information which would be needed in order to autonomously perform a variety of tasks: for instance, the robot may need to know that a given area in a map is a kitchen, or that corridors in a public building are usually crowded during daytime but empty at night. In other words, the robot needs to have some *semantic information* about the entities in the environment. Semantic information can be used to reason about the functionalities of objects and environments, or to provide additional input to the navigation and localization subsystems. Semantic information is also pivotal to the ability of the robot to communicate with humans using a common set of terms and concepts [2].

The need to include semantic information in robot maps has been recognized for a long time [3], [4]. In fact, most robots that incorporate provisions for task planning and/or for communicating with humans, store some semantic information in their maps (e.g., [5], [6]). Common information includes the classification of spaces (rooms, corridors, halls) and the names of places and objects. This information can be used to decide the navigation mode to use, or for task planning. However, while geometric and topological information is usually acquired automatically by the robot through its

sensors, semantic information is most often hand-coded into the system.

Recently, a few authors have reported systems in which the robot can acquire and use semantic information [7], [8]. In most cases, however, the acquisition is done via a linguistic interaction with a human and not using the robot's own sensors. An interesting exception is [9], in which the robot extracts semantic information from 3D models built from a laser scanner. This work, inspired by work on 3D scene analysis in vision [10], is similar in spirit to the one proposed here, but its scope is narrower, being limited to the classification of surface elements (ceilings, floors, door, etc.).

In this paper, we propose an approach to allow a mobile robot to build a semantic map from sensor data, and to use this semantic information in the performance of navigation tasks. In our approach, we maintain two parallel hierarchical representations: a spatial representation, and a semantic one. These representations are based on off-the-shelf components from the field of robot mapping and from the field of AI and knowledge representation, respectively. The link between these components is provided by the concept of *anchoring*, which connects symbolic representations and sensor-based representations [11]. The semantic information can be used to perform complex types of reasoning. For instance, it can be used by a symbolic planner to devise contingency plans that allow the robot to recover from exceptional situations.

In the rest of this paper, we describe our approach in some detail, and present several experiments performed on an iRobot Magellan robot equipped with a sonar ring, a laser, and a color camera. In these experiments, we show how the robot can: (1) acquire semantic information from sensor data, and link this information to those data; (2) use generic knowledge to infer additional information about the environment and the objects in it; and (3) use this information to plan the execution of tasks and to detect possible problems during execution.

II. OVERALL APPROACH

In our approach we endow a mobile robot with an internal representation of its environment from two different perspectives: (i) a spatial perspective, that enables it to reliably plan and execute its tasks (e.g., navigation); and (ii) a semantic perspective, that provides it with a human-like interface and inference capabilities on symbolic data (e.g., a *bedroom* is a *room* that contains a *bed*). These two sources of knowledge,

---

[1]Cipriano Galindo was with the Center for Applied Autonomous Sensor Systems (Örebro University) as a Marie Curie fellow during the preparation of this manuscript. His home affiliation is the System Engineering and Automation Dept. (University of Málaga) (email: cipriano@ctima.uma.es).
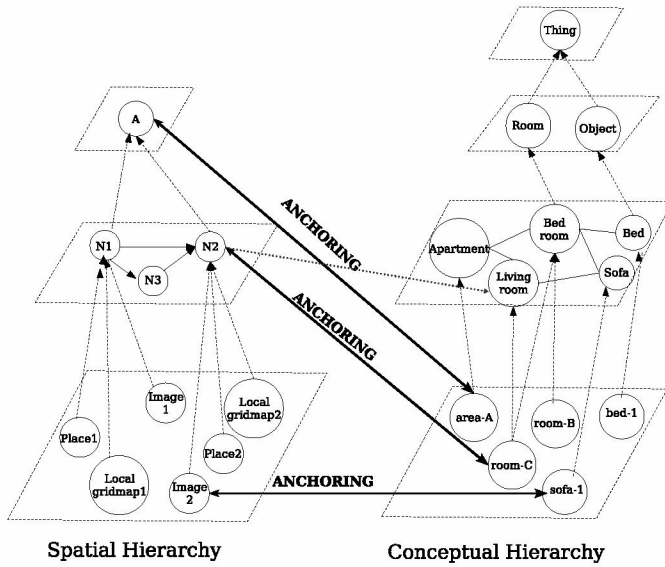
Fig. 1. The spatial and semantic information hierarchies. On the left, spatial information gathered by the robot sensors. On the right, semantic information that models concepts in the domain and relations between them. Anchoring is used to establish the basic links between the two hierarchies (solid lines). Additional links can then be inferred by symbolic reasoning (dotted line).
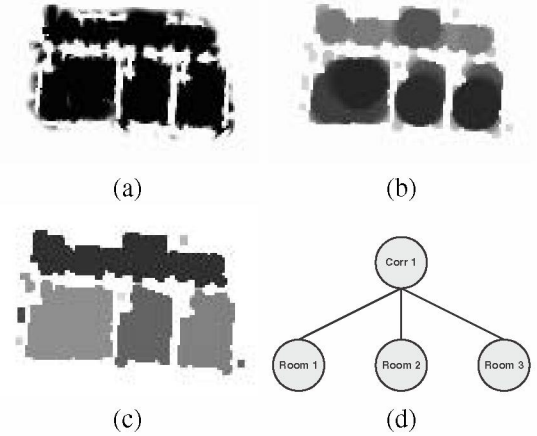

Fig. 2. Topological extraction. (a) Original gridmap; (b) Fuzzy morphological opening; (c) Watershed segmentation; (d) Extracted topology.

spatial and semantic, are interrelated through the concept of *anchoring* [11], that connects internal symbols (e.g., bed-1) to sensor data that refers to the same physical entities in the environment (e.g., an image of a bed).

Fig. 1 depicts our approach. It includes two hierarchical structures, the spatial and the conceptual hierarchies. The *Spatial Hierarchy* arranges its information in different levels of detail: (i) simple sensorial data like camera images or local gridmaps, (ii) the topology of the robot environment, and (iii) the whole environment represented by an abstract node. Additional intermediate levels could also be included.

The *Conceptual Hierarchy* represents concepts (categories and instances) and their relations, modeling the knowledge about the robot environment. This permits the robot to do inferences about symbols, that is instances of given categories.

The integration of both sources of knowledge enables the robot to carry out complex navigational tasks involving semantic information. For example, the robot can execute a task like *go to the living-room* by inferring that the spatial element $N2$ is identified as a living-room[1] since it includes a percept (sensorial data) anchored to the symbol `sofa-1`, which is an instance of the general class *sofa*. This inference is graphically represented by a dotted line in Fig. 1.

In our work we manage the Spatial Hierarchy by using a mathematical model called *AH-graph* model [12], which has proved its suitability in reducing the computational effort of robot operations such as path-search [13] or symbolic task planning [14]. On the other hand, the conceptual hierarchy has been modeled by employing standard AI languages, like the *NeoClassic* language [15], in order to provide the robot with

inference capabilities. Overall, our multi-hierarchical map can be classified as a hybrid metric-topological-semantic map according to the taxonomy proposed in [16].

III. THE TWO HIERARCHIES

A. *The Spatial Hierarchy*

The Spatial Hierarchy contains spatial and metric information from the robot environment

This model is based on *abstraction*, that helps to minimize the information required to plan tasks by grouping/abstracting symbols within complex and high-detailed environments.

The basic sensorial information held by the Spatial Hierarchy (at the lowest level) are images of objects and local gridmaps, which are a representation of the local workspace of the robot. Such percepts are abstracted into the upper levels of the hierarchy to represent the topology of the space. Nodes represent open areas, such as rooms, corridors, and arcs the possibility of navigation from one to another. Finally, the whole spatial environment of the robot is represented at the highest level as a single node. The use of small local metric maps connected into a global topological map allows us to preserve the accuracy provided by metric maps, while covering the large extent that can be afforded by a topological map with limited computational effort [16].

The construction of this hierarchy is based on the techniques presented in [17], [18]. It uses the data gathered from the robot sensors, i.e., ultrasonic sensors, to build an occupancy grid map of the surroundings of the robot (see Fig. 2a), that can be seen as an image where values of emptiness (and occupancy) correspond to gray-scale values. This grid map is then segmented, using image processing techniques, into *large open spaces*, that can be anchored to room names. To do so, the gray-scale image is filtered using fuzzy mathematical morphology, which yields a new gridmap where values of cells represent a membership degree to circular open spaces (Fig. 2b). The resulting grid is then segmented into connected components using a technique called watersheding in order to extract the topology of the open space in the environment.
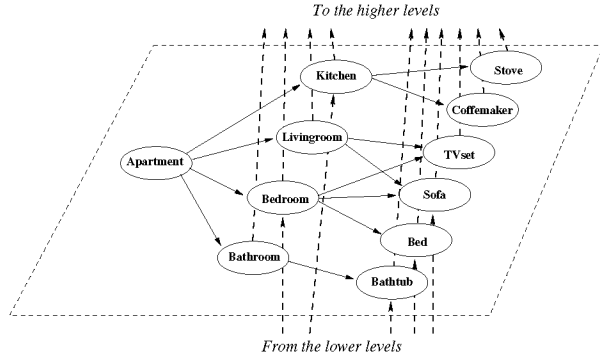
---

[1]Note that $N2$ could also be a *bedroom* in this example. Multiple possibilities when deducing the type of a room are treated in Sec. V.

To the higher levels

From the lower levels

Fig. 3. Detail of Level 1 of our conceptual hierarchy. Horizontal links are *"has"* links, like the ones in the description above. Vertical links are *"is-a"* links, and go to the other levels of the hierarchy.

Fig. 5 shows an example of the Spatial Hierarchy constructed by the robot in one of our experiments.

B. The Conceptual Hierarchy

The Conceptual Hierarchy models semantic knowledge about the robot environment. All concepts derive from a common ancestor called *Thing*, at the top level of the hierarchy. At the next level (level 2) there are the two general categories *Objects* and *Rooms* of interest for our domain. At level 1 we find specific concepts (kitchen, bedroom, bed, sofa, etc.) derived from these categories. These concepts may incorporate constraints, like the fact that a bedroom must have at least one bed. Finally, at level 0 we have individual instances of these concepts, denoted by symbols like `room-C` or `sofa-1`.

In our work, we use a well-known system for knowledge representation and reasoning developed within the AI community, called *NeoClassic* [15]. The following is an example of how the concept of a kitchen can be defined in the NeoClassic language. Intuitively, a kitchen is a room that has a stove and a coffee machine, but does not have a bed, bathtub, sofa or TV set.

```
(createConcept Kitchen
    (and Room
        (atLeast 1 stove)
        (atLeast 1 coffee-machine)
        (and (atMost 0 bathtub) (atMost 0 sofa)
            (atMost 0 bed) (atMost 0 tvset))))
```

Fig. 3 shows the full Level 1 of the conceptual hierarchy outlined in Fig. 1 above.[2] The inference mechanisms in Neo-Classic (and in most other knowledge representation systems) allow the robot to use this knowledge to perform several types of inferences. For instance, if we know that `room-D` is a room and that it contains `obj-1` which is a bathtub, then we can infer that `room-D` is a bathroom.

---

[2]This semantic knowledge is admittedly overly simplistic, and it is used here mainly for illustration purposes.

IV. LINKING THE HIERARCHIES

A. Anchoring

According to [11], anchoring is the process of creating and maintaining the correspondence between symbols and sensor data that refer to the same physical objects.

In our framework, the anchoring process is needed to connect the symbols at the ground level of the Conceptual Hierarchy to the sensor data acquired from the sonars and the video camera. More specifically, the anchoring process creates the connection between sensor data representing rooms and corridors in the Spatial Hierarchy (e.g., a local gridmap) to the corresponding symbol in the Conceptual Hierarchy (e.g., `room-C`); and sensor data representing objects in the Spatial Hierarchy (e.g., the segmented image of a sofa) to the corresponding symbol (e.g., `sofa-1`). The anchoring process is also needed to maintain this information over time, even when the objects and places are not in view, so that they can be recognized when they come back into view.

In our work, we use the computational framework for anchoring defined in [19]. In that framework, the symbol-data correspondence for a specific object is represented by a data structure called an *anchor*. An anchor includes pointers to the symbol and sensor data being connected, together with a set of properties useful to re-identify the object, e.g., its color and position. These properties are also used as input by the control routines.

B. Inference using anchoring

The Conceptual Hierarchy has at its lowest level symbols denoting individual objects like `room-A` and `sofa-1`. These symbols are instances of classes that in turn are part of more abstract classes. The Spatial Hierarchy has at its lowest level gridmaps and at its higher levels increasingly abstract spatial entities like rooms and corridors. The entities recognized at different levels of the Spatial Hierarchy are connected by the anchoring process to the symbols that denote these entities in the Conceptual Hierarchy. This allows the application of inferences on spatial objects. For instance, by connecting a spatially recognized room, `R1`, to the symbol `room-A` which is an instance of the class *kitchen*, the robot is able to fulfill the user high-level command *go to the kitchen*.

The anchoring process also connects objects acquired by vision to individual symbols. Objects are recognized by shape and color and an anchor is created containing the symbol, the symbolic description and the perceptual properties of the object including its position. The position allows the system to determine in which room or corridor the object is.

The anchoring connection between both hierarchies (*Spatial* and *Conceptual*) allows the robot use semantic information in several ways. By recognizing objects of a particular type inside a room, the robot may be able to classify that room. For example if a *stove* is detected inside `room-A`, the inference system will deduct that the room is a *kitchen*. Another way to use semantic information is to infer the probable location of an object which has not been previously
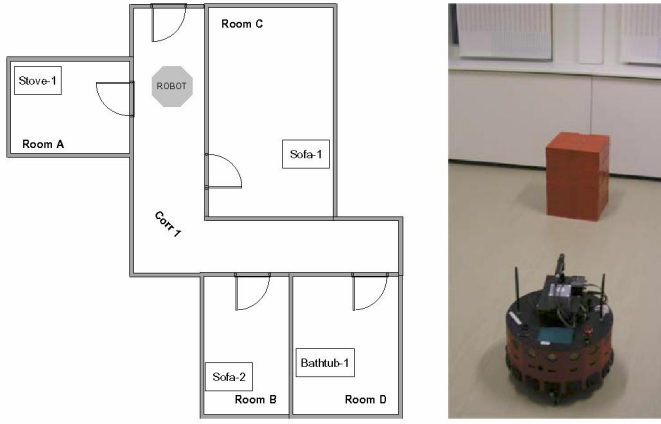
Fig. 4. Experimental setup. (Left) Plan of our home-like scenario. (Right) Our mobile robot identifying a red box (a sofa).
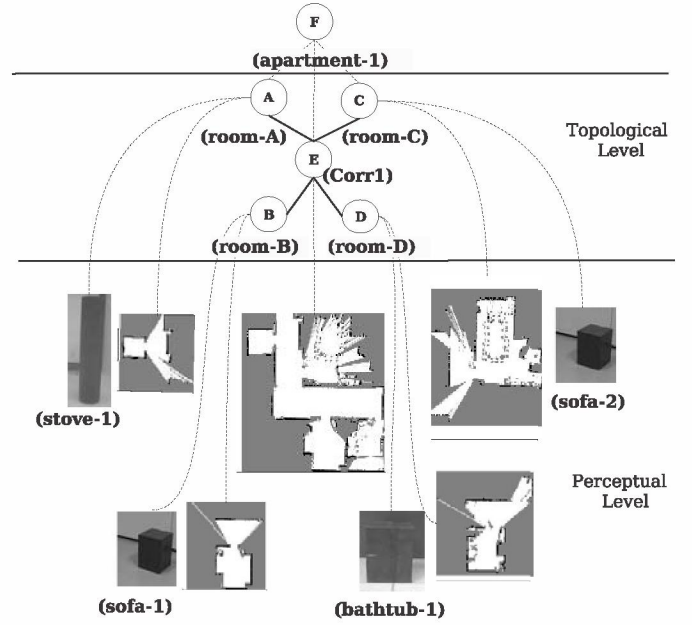

Fig. 5. Constructed Spatial Hierarchy. Spatial information is anchored to symbols at the lowest level of the Conceptual Hierarchy. For the sake of clarity, connected symbols are shown in brackets below each spatial entity.

observed. For instance, if the robot is looking for a *stove*, it can use the semantic net shown in Fig. 3 to decide that only the kitchen needs to be explored.

V. EXPERIMENTS

In order to test our approach, a variety of experiments have been conducted on a Magellan Pro robot using an implementation of the *Thinking Cap* robotic architecture [20]. We have reproduced in our laboratory a home-like environment as the one shown in Fig. 4 embracing four rooms (a kitchen, a living-room, a bathroom, and a bedroom) connected by a corridor. The robot used in our experiments incorporates a laser range finder, 16 sonars, and a color camera for object recognition. Since reliable and robust object recognition is out of the scope of this paper, in our experiments the vision system has been simplified to recognize only combinations of colors and simple shapes like boxes and cylinders, identifying them as furniture (i.e., a red box represents a sofa, a green box a bathtub, a green cylinder a stove, etc.).

Using this setup, we have carried out the following three different types of experiments.

A. Model Construction

The creation of the Spatial Hierarchy and its connection to the Conceptual Hierarchy enables the robot to infer the type of rooms according to their objects. This experiment was performed by tele-operating the robot within its environment while it stores spatial information of detected rooms and objects and connects (anchors) them to symbols at the lowest level of the Conceptual Hierarchy. Fig. 5 shows the Spatial Hierarchy constructed in our experiments which holds local gridmaps of rooms and images of recognized objects at the lowest level, the topology of the environment at the first level, and the whole robot workspace at the highest level.

In this experiment, the symbols created at the lowest level of the Conceptual Hierarchy (shown in the figure in brackets) are `sofa-1`, `sofa-2`, `bathtub-1`, and `stove-1` for objects, and `room-A`, `room-B`, `room-C`, and `room-D` for rooms. Object's symbols are classified in their correspondent

type by the vision system while the type of rooms is inferred following the semantic description shown in Fig. 3, through the type of objects that they contain. Thus, `room-D` is unequivocally classified as a *bathroom* and `room-A` as a *kitchen* since both rooms contain objects (`bathtub-1` and `stove-1` respectively) that unequivocally identify these types. However, in some cases, the type of a room can not be unequivocally determined by the observed objects. This is the case for `room-B` and `room-C`, where only a sofa has been observed: given this incomplete information, the robot describes the type of this room by a disjunction: *livingroom* OR *bedroom*.

B. Navigation

These experiments were intended to prove the utility of our approach as a mechanism for human-robot communication, allowing users to give symbolic instructions to the robot. In the first experiment, we ask the robot to solve the task *go to the bathroom*. To do so, the inference system needs to find an instance of the general category *bathroom* to be used as the destination for the robot, i.e., `room-D`. This symbolic information, however, cannot be directly handled by the navigation system, which requires instead the spatial information related to the destination. Such spatial information is retrieved by following the anchoring link that connects the desired destination to the topological element D in the Spatial Hierarchy. In this way, the initial symbolic task is translated to the executable task *go to D*. This task is then performed using the topological and metric information for navigation stored in the Spatial Hierarchy — see Fig. 6 (top).

The use of knowledge representation techniques allows us to represent and reason about situations of ambiguity, which
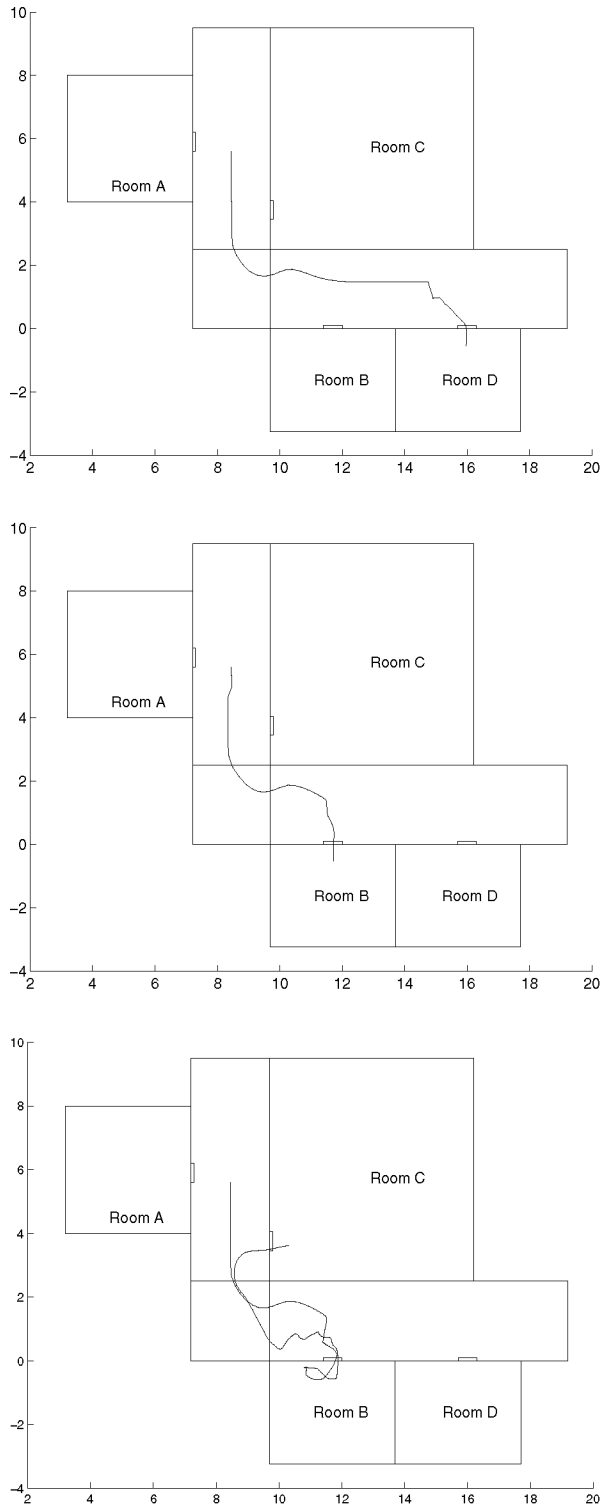
Fig. 6. Experimental runs. Path followed by the robot when performing the tasks: (top) *go to the bathroom*; (middle) *go to the bedroom*; (bottom) *approach the TV set*.

for instance, by using an AI planner.

In order to test this possibility, we have performed an experiment in which we gave the robot the task to *go to the bedroom*. Based on the available environmental information, both `room-B` and `room-C` could be classified as bedroom by the semantic inference system. To cope with this type of situations, our system is equipped with a state of the art AI planner, called *PTLplan* [21]. PTLplan is a conditional possibilistic/probabilistic planner which is able to reason about uncertainty and about perceptual actions. PTLplan searches in a space of epistemic states, representing a set of hypotheses about the actual state of the world. For instance, in one epistemic state `room-C` is a a bedroom, while in another one it is a living-room. PTLplan can also reason about perceptual actions, which make observations that may discriminate between different epistemic states, thus reducing the uncertainty. In our example, PTLplan has produced the following conditional plan, where the perceptual action *check-for-bedroom* looks for objects that unequivocally identifies a bedroom (i.e., a bed).

```
((MOVE CORR1 B)
 (CHECK-FOR-BEDROOM)
 (COND
    ((IS-A-BEDROOM ROOM-B = F)
     (MOVE ROOM-B CORR1)
     (MOVE CORR1 ROOM-C)
     (CHECK-FOR-BEDROOM)
     (COND
        ((IS-A-BEDROOM ROOM-C = T) :SUCCESS)
        ((IS-A-BEDROOM ROOM-C = F) :FAIL)))
  ((IS-A-BEDROOM ROOM-B = T) :SUCCESS)))
```

This plan can be read as follows: go to one of the candidate rooms (`Room-C`); observe the furniture in order to classify it; if the room is classified as a *bedroom*, then the goal is achieved; else go to the other candidate room (`Room-B`), observe the furniture, and succeed or fail depending of the result of the classification. Fig. 6 (middle) shows a sample execution of this task, in which the robot finds a bed in the first room visited (`Room-B`).

The semantic information stored in the Conceptual Hierarchy also enables the robot to reason about the location of objects non-previously observed. To test this ability, a different type of navigation experiment was devised in which we asked the robot to *approach the TV set*. Since there are no instances of the category *TVset* in the Conceptual Hierarchy, the inference system generates their probable location according to the available semantic knowledge: a bedroom or a living-room, that is, `room-B` or `room-C`. The robot then uses PTLplan to generate a conditional plan similar to the one above, in which the robot visits each room and performs the perceptual action *look-for-tv* in each one. Fig. 6 (bottom) shows a sample execution of this task, in which the robot does not find a TV set in the first room visited (`Room-B`) but it finds it in the second one (`Room-C`).

may arise when the robot's information about the environment is incomplete. Semantic information can be exploited to autonomously devise strategies to resolve these ambiguities,

C. Detecting localization errors

The navigation system of the robot can use the semantic information to detect localization errors by reasoning about the expected location of objects. We have tested this feature through the following experiment. The robot was placed at the entrance of `room-C` previously identified as a living-room. The odometric position was approximatively ($x = 10$, $y = 4$). An error in the odometric system was artificially introduced by lifting the robot and placing it in front of `room-A`. Inside `room-A`, the robot recognized a stove. The inference system then signaled an exception, since a living room should not contain a stove according to the given semantic knowledge. Assuming a reliable vision system, this exception was attributed to a self-localization error and reported to the navigation system. In our experiment, this system corrected the internal position of the robot using to the new observed object as a landmark (i.e., $x = 6$ and $y = 6$).

VI. CONCLUSIONS

This paper has presented a multi-hierarchical map for mobile robots which includes spatial and semantic information. The spatial component of the map is used to plan and execute robot tasks, while the semantic component enables the robot to perform symbolic reasoning. These components are tightly connected by an anchoring process.

Our approach has been successfully tested on a real mobile robot demonstrating the following abilities: (i) interface with humans using a common set of concepts, (ii) classify a room according to the objects in it, (iii) deduce the probable localization of an object, (iv) deal with ambiguities, and (v) detect localization errors based on the typical locations of objects. Although in our experiments we have chosen specific techniques for the spatial and the semantic components, it should be emphasized that our approach does not depend on this choice: one could reproduce our results using a different representation for the semantic component, for the spatial component, or for both. An additional interesting possibility would be to use learning techniques to automatically acquire the semantic structure of the domain.

ACKNOWLEDGMENTS

REFERENCES

[1] S. Thrun. Robotic mapping: A survey. In G. Lakemeyer and B. Nebel, editors, *Exploring Artificial Intelligence in the New Millenium.* Morgan Kaufmann, 2002.

[2] J.A. Fernández and C. Galindo and J. González. Assistive Navigation of a Robotic Wheelchair using a Multihierarchical Model of the Environment. *Integrated Computer-Aided Eng.* 11(11):309–322, 2004.

[3] B. Kuipers. Modeling spatial knowledge. *Cognitive Science*, 2, 1978.

[4] R. Chatila and J.P. Laumond. Position referencing and consistent world modeling for mobile robots. In *Proc of the IEEE Int Conf on Robotics and Automation*, pages 138–145, 1985.

[5] S. Thrun, M. Bennewitz, W. Burgard, A.B. Cremers, F. Dellaert, D. Fox, D. Hähnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz. MINERVA: A second generation mobile tour-guide robot. In *Proc of the IEEE Int Conf on Robotics and Automation*, pages 1999–2005, 1999.

[6] M. Beetz, T. Arbuckle, T. Belker, A.B. Cremers, D. Schulz, M. Bennewitz, W. Burgard, D. Hähnel, D. Fox, and H. Grosskreutz. Integrated Plan-Based Control of Autonomous Robots in Human Environments. *IEEE Intelligent Systems* 16(5):56–65, 2001.

[7] C. Theobalt, J. Bos, T. Chapman, A. Espinosa, M. Fraser, G. Hayes, E. Klein, T. Oka, and R. Reeve. Talking to Godot: Dialogue with a mobile robot. In *Proc of IROS*, pages 1338–1343, Lausanne, CH, 2002.

[8] M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, and D. Brock. Spatial language for human-robot dialogs. *IEEE Trans. on Systems, Man, and Cybernetics*, 2(C-34):154–167, 2004.

[9] A. Nüchter, H. Surmann, K. Lingemann, and J. Hertzberg. Semantic scene analysis of scanned 3D indoor environments. In *Proc of the VMV Conference*, Munich, DE, 2003.

[10] O. Grau. A scene analysis system for the generation of 3D models. In *Prof of the IEEE Int Conf on Recent Advances in 3d Digital Imaging and Modeling*, pages 221–228, Canada, 1997.

[11] S. Coradeschi and A. Saffiotti. An Introduction to the Anchoring Problem. *Robotics and Autonomous Systems* 43(2-3):85-96, 2003.

[12] J.A. Fernández and J. González. Multi-Hierarchical Representation of Large-Scale Space. *Kluwer Academic Publishers*, 2001.

[13] J.A. Fernández and J. González. Multihierarchical Graph Search. *IEEE T. on Pattern Analysis and Machine Inteligence* 24(1):103-113, 2002.

[14] C. Galindo, J.A. Fernández and J. González. Hierarchical Task Planning through World Abstraction. *IEEE T. on Robotics* 20(4):667-690, 2004.

[15] P.F. Patel-Schneider, M. Abrtahams, L. Alperin, D. McGuinness and A. Borgida. NeoClassic Reference Manual: Version 1.0. AT&T Labs Research, Artificial Intelligence Principles Research Department, 1996.

[16] P. Buschka and A. Saffiotti. Some Notes on the Use of Hybrid Maps for Mobile Robots. In *Proc. of the 8th Int. Conf. on Intelligent Autonomous Systems* (IAS) 547–556. Amsterdam, NL, 2004.

[17] E. Fabrizi and A. Saffiotti. Extracting Topology-Based Maps from Gridmaps. In *Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA)*, San Francisco, CA, pp. 2972–2978, 2000.

[18] P. Buschka and A. Saffiotti. A Virtual Sensor for Room Detection. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, Lausanne, CH, pp. 637-642, 2002.

[19] S. Coradeschi and A. Saffiotti, Anchoring symbols to sensor data: preliminary report. In , *Proc. of the Seventeenth National Conference on Artificial Intelligence (AAAI-2000)*, Austin, pp. 129-135, 2000.

[20] A. Saffiotti and K. Konolige and E.H. Ruspini, A multivalued-logic approach to integrating planning and control. *Artificial Intelligence* 76(1-2):481-526, 1995.

[21] L. Karlsson, Conditional progressive planning under uncertainty. In *Proc. of the 17th IJCAI Conf.*, pages 431-438. AAAI Press, 2001.