Representing Spatial Knowledge in Mobile Cognitive Systems

Andrzej PRONOBIS [a], Kristoffer SJÖÖ [a], Alper AYDEMIR [a], Adrian N. BISHOP [b], Patric JENSFELT [a]

[a] *Centre for Autonomous Systems, Royal Institute of Technology, Stockholm*
{pronobis, krsj, aydemir, patric}@kth.se

[b] *National ICT Australia (NICTA), Australian National University (ANU)*
adrian.bishop@nicta.com.au

Abstract. A cornerstone for cognitive mobile agents is to represent the vast body of knowledge about space in which they operate. In order to be robust and efficient, such representation must address requirements imposed on the integrated system as a whole, but also resulting from properties of its components. In this paper, we carefully analyze the problem and design a structure of a spatial knowledge representation for a cognitive mobile system. Our representation is layered and represents knowledge at different levels of abstraction. It deals with complex, cross-modal, spatial knowledge that is inherently uncertain and dynamic. Furthermore, it incorporates discrete symbols that facilitate communication with the user and components of a cognitive system. We present the structure of the representation and propose concrete instantiations.

1. Introduction

Many recent advances in the fields of robotics and artificial intelligence have been driven by the ultimate goal of creating artificial cognitive systems able to perform human-like tasks. Several attempts have been made to create integrated cognitive architectures and implement them on mobile robots [2,3,13,1,4]. There is an increasing interest in, and demand for, robots that are capable of dealing with complex and dynamic environments outside the traditional industrial workplaces. These next generation robots will not only have to track their position and navigate between points in space, but reason about space and their own knowledge, plan tasks and knowledge acquisition and interact with people in a natural way.

Spatial knowledge constitutes a fundamental component of the knowledge base of a cognitive agent providing a basis for navigation, reasoning, planning and episodic memories. Moreover, it is a common ground for communication between a robot and a human. In order for the process of acquisition, interpreting, storing and recalling of the spatial knowledge to be robust and efficient under limited resources and in realistic settings, the knowledge must be properly structured and represented. Such knowledge representation must address requirements imposed on the integrated system as a whole, but also resulting from properties of its components. Due to this central role, the design of a spatial knowledge representation should be one of the first steps in building a cognitive system.

In this work, we develop a structure of a spatial knowledge representation for a cognitive mobile system that we call COARSE (Cognitive lAyered Representation of Spatial knowledgE). We carefully analyze the role of a spatial representation and formulate design assumptions and requirements imposed by the functionality and components of an integrated system. Our representation is layered and represents knowledge at different levels of abstraction, from low-level sensory input to high level conceptual symbols. It is designed for representing complex, cross-modal, spatial knowledge that is inherently uncertain and dynamic and includes discrete symbols that facilitate communication with the user and components of the system. Moreover, we propose models and algorithms that could be used as instantiations of each layer of the representation.

This paper is motivated by the desire to create a framework that is powerful, robust and efficient, but most importantly suited for mobile agents performing typical human-like tasks. The literature contains many algorithms for spatial mapping and instantiations of mobile robotic systems. However, the existing representations are either designed for a very specific domain [7,12], they concentrate on a fraction of the spatial knowledge [20,23] or are designed to solve a single algorithmic task very efficiently rather than for use within a larger system [8,10,18]. The idea of this paper, is to take a step back, focus on structuring the whole body of spatial knowledge and see how an analysis of requirements can lead the way towards a powerful spatial representation for a cognitive mobile robot.

2. Related Work

There exists a broad literature on mobile robot localization, navigation and mapping and many algorithms relying on spatial knowledge have been proposed. These include solutions to such problems as Simultaneous Localization and Mapping (SLAM) [8,15,10,18] or place classification [16,20]. Every such algorithm maintains a representation of spatial knowledge. However, this representation is usually specific to the particular problem and designed to be efficient within the single mapping system detached from any other interacting components. Other, more general concepts, such as the Spatial Semantic Hierarchy [14] concentrate on lower levels of spatial knowledge abstraction and do not support higher-level conceptualization or representation of categorical information.

At the same time, we witness a growing interest in building artificial mobile cognitive systems [2,3,1,4]. These are complex, usually modular, systems that require a unified and integrated approach to spatial knowledge representation. The central role of spatial knowledge in those systems has been recognized and several authors proposed subsystems processing spatial knowledge integrated with other components such as dialogue systems [25,22]. However, neither of those provides a clear structure of the represented knowledge, perform a thorough analysis of the needs of different components of a mobile cognitive system or encapsulates all major aspects of spatial knowledge.

The most comprehensive relevant representation has been proposed in [25]. However, it has several major drawbacks that makes it unsuitable for systems that deal with dynamic and uncertain knowledge within large-scale, complex environments. First of all, the knowledge is never fully abstracted and is always grounded in an accurate global metric map. This makes the system less robust and scalable. Moreover, the categorical knowledge is not explicitly represented. The high-level conceptualization relies on rigid ontologies and ignores uncertainties associated with represented symbols. Finally, it is modality-specific and does not allow for knowledge fusion from multiple sources. In the rest of the paper, we propose an approach to spatial knowledge representation that addresses those problems.

3. Analysis of the Problem

Before designing a representation of spatial knowledge, it is important to review the aspects a representation should focus on. In this section, we analyze those aspects and propose our definition of a generic spatial knowledge representation. Then, we formulate the problem within the context of cognitive systems.

3.1. What is a Spatial Knowledge Representation?

Following the analysis by Davis [9], we formulate several points that characterize a general representation of spatial knowledge. A spatial representation can be seen as:

a) A substitution (surrogate) for the world that allows the agent to perform reasoning about the parts of the environment which are beyond its sensory horizon. Such a surrogate is naturally imperfect, and is incomplete (some aspects are not represented), inaccurate (captured with uncertainty), and will become invalid (e.g. due to dynamics of the world that cannot be observed and is too complex to be captured by the representation). Moreover, since the representation cannot be perfect, all the inferences based on that representation, such as the outcomes of the localization process, are uncertain. The only perfect representation of the world or the environment in which the agent operates is the environment itself.

b) A set of ontological commitments that determine the terms in which the agent thinks about space. The representation defines the aspects of the world that should be represented. Moreover, it defines the level of detail at which they should be represented as well as their persistence. The ontology should be understood in more general terms, from spatial concepts and their relations to categorical models or types of features extracted from the sensory input.

c) A set of definitions that determine the reasoning that can be (and that should be) performed within the framework and the possible inferences and their outcomes. The reasoning will typically correspond to determining the current location with respect to the internal map (topologically, semantically etc.), providing necessary knowledge for the navigation process, determining the properties of a location in space etc. Moreover, the representation defines how the location of the agent is represented and in what terms it is possible to refer to points in space (e.g. in terms of metric coordinates, semantic category of a place etc.).

d) A way of structuring the spatial information so that it is computationally feasible to perform all the necessary processing and inferences in a specified time (e.g. in real time) despite limited resources.

e) A medium of communication between the agent and human. If the agent is supposed to exchange information with humans, the representation must be designed in a way that allows the agent to interpret human expressions and generate expressions that are comprehensible to humans.

f) Similarly, a medium of communication between components of an integrated system.

3.2. Spatial Representation for Mobile Cognitive Systems

In this work, we narrow the focus to mobile cognitive systems. Based on the analysis of existing approaches [3,1,23] as well as ongoing research on artificial cognitive systems [2], we have identified several areas of functionality, usually realized through separate subsystems, that must be supported by the representation. These include localization, navigation, and autonomous exploration, but also understanding and exploiting semantics associated with space, human-like conceptualization and categorization of space, reasoning about spatial units and their relations, human-robot communication, action planning, object finding and visual servoing, and finally recording and recalling episodic memories.

Having in mind the aforementioned functionalities, aspects covered by a representation of spatial knowledge as well as limitations resulting from practical implementations, we have identified several desired properties and designed a representation reflecting those properties.

Complex, cross-modal, spatial knowledge in realistic environments is inherently uncertain and dynamic. Therefore, it is futile to represent the environment as accurately as possible. A very accurate representation must be complex, require a substantial effort to synchronize with the world and still cannot guarantee that sound inferences will lead to correct conclusions [9]. Our primary assumption is that the representation should instead be minimal and inherently coarse and the spatial knowledge should be represented only as accurately as it is required to support the functionality of the system. Furthermore, redundancy should be avoided and whenever possible and affordable, new knowledge should should be inferred from the existing information. It is important to note that uncertainties associated with represented symbols should be explicitly modeled.

Information should be abstracted as much as possible to make it robust to dynamic changes. Moreover, representations that are more abstract should be used for longer-term storage. At the same time, knowledge extracted from immediate observations can be much more accurate (e.g. for the purpose of visual servoing). In other words, the agent should use the world as an accurate representation whenever possible. It is important to mention that rich and detailed representations should not constitute a permanent base for more abstract ones (as is the case in [25]). Similarly, space should be represented on different spatial scales from single scenes to whole environments.

Space should be discretized into a finite number of spatial units. Discretization of continuous space is one of the most important abstracting steps as it allows to make the representation robust, compact and tractable. Discretization drastically reduces the number of states that have to be considered e.g. during the planning process [11] and serves as a basis for higher level conceptualization [25].

A representation should allow not only for representing instantiations of spatial segments visited by the robot. It is equally important to provide means for representing unexplored space. Furthermore, categorical knowledge should be represented that is not specific to any particular location and instead corresponds to general knowledge about the world. Typical examples would be categorical models of appearance of places [20] or objects [19].

Finally, we focus on the fundamental role of the representation in human-robot interaction. Spatial knowledge representation should model correspondence between the represented symbols and human concepts of space. Spatial properties (e.g. shape, size), semantic categories of rooms (e.g. kitchen, office) or spatial segments (e.g. rooms, floors, buildings) recognized by humans are examples of such concepts. This correspondence could be used to generate and resolve spatial referring expressions [24] or path descriptions.

4. Structure of the Representation

In this section, we propose a representation of spatial knowledge that adheres to the desired properties formulated above. Figure 1 gives a general overview of the structure of the representation. It is sub-divided into four layers which can be regarded as sub-representations focusing on different aspects of the world, abstraction levels of the spatial knowledge and different spatial scales. Moreover, each layer defines its own spatial entities and the way the agent's position in the world is represented. The properties of each layer are summarized in Table 1.

At the lowest abstraction level, we have the sensory layer which maintains an accurate representation of the robot's immediate environment extracted directly from the robot's sensory input. Higher, we have the place and categorical layers. The place layer provides fundamental discretisation of the continuous space into a set of distinct places. The categorical layer focuses on low-level, long-term categorical models of the robot's sensory information. Finally, at the top, we have the conceptual layer, which associates human concepts with the categorical models in the categorical layer and groups places into human-compatible spatial segments such as rooms. The following sections provide details about each of the layers.

4.1. Sensory Layer

In the sensory layer, a detailed robocentric model of the robot's immediate environment is represented based on direct sensory input as well as data fusion over space around the robot and short time intervals. The sensory layer stores low-level features and landmarks extracted from the sensory input together with their exact position with respect to the robot. Measures of uncertainty are also included in this representation. Landmarks that move beyond a certain distance are forgotten and replaced by new information. Thus, this representation is akin to a sliding window, with robocentric and up-to-date direct perceptual information. It is also essentially bottom-up only, though directives and criteria, such as guiding the attentional process, may be imposed from upper layers.

The representation in the sensory layer helps to maintain stable and accurate information about the relative movements of the robot. Moreover, it allows for maintaining and tracking the position of various features while they are nearby. This can be useful for providing "virtual sensing" such as 360° laser scans based on short-term temporal sensory integration as well as generation of features based on spatial constellations of landmarks located outside the field of view of the sensor. Additionally, it could be used for temporal filtering of sensory input or providing robustness to occlusions. Finally, the sensory layer can provide the low level robotic movement systems with data for deriving basic control laws such as for obstacle avoidance or visual servoing.

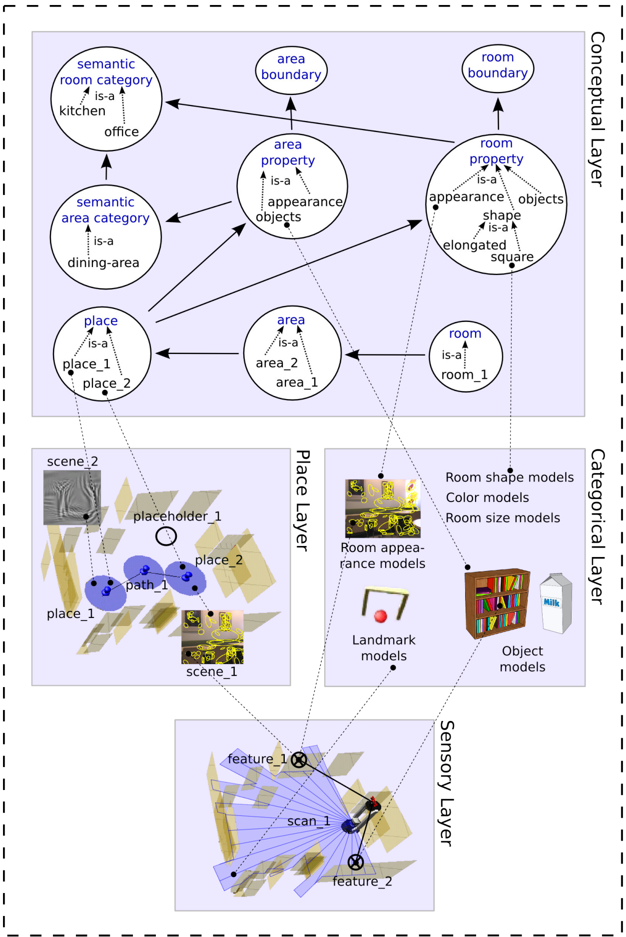

Figure 1. The layered structure of the spatial representation. The position of each layer within the representation corresponds to the level of abstraction of the spatial knowledge.

Property	Sensory Layer	Place Layer	Categorical Layer	Conceptual Layer
Aspects represented	Accurate geometry and appearance	Local spatial relations, coarse appearance, geometry	Perceptual categorical knowledge	High-level spatial concepts / Links concepts ↔ entities
Agent's position	Pose within the local map	Place ID	Relationship to the categorical models	Expressed in terms of high level spatial concepts
Spatial scope	Small-scale, local	Large-scale	Global	Global
Knowledge persistence	Short-term	Long-term	Very long-term	Life-long / Very long-term

Table 1. Comparison of properties of the four layers of the spatial representation.

4.2. Place Layer

The place layer is responsible for the fundamental, bottom-up discretization of continuous space. In the place layer, the world is represented as a collection of basic spatial entities called places as well as their spatial relations. Each place is defined in terms of features that are represented in the sensory layer, but also spatial relations to other places. The aim of this representation is not to represent the world as accurately as possible, but at the level of accuracy sufficient for performing required actions and robust localization despite uncertainty and dynamic variations. Similarly, the relations do not have to be globally consistent as long as they are preserved locally with sufficient accuracy. The representation of places in the place layer persists over long term.

Besides places, the place layer also defines paths between them. The semantic significance of a path between two places is the possibility of moving directly between one and the other. This does not necessarily imply that the robot has traveled this path previously. A link might be created for unexplored place e.g. based on top-down cues resulting from the dialogue with the user (e.g. when the user indicates part of the environment that should be of interest to the robot, but not immediately). In addition, the place layer explicitly represents unexplored space. Tentative places are represented which the robot would probably uncover if it moved in a certain direction.

The place layer operates on distinct places as well as their connectivity and spatial relations to neighboring places. No global representation of the whole environment is maintained. Still, since the local connectivity is available, global representation (e.g. a global metric map) can be derived when needed. This representation will not be accurate, but will preserve the connectivity and relaxed spatial relations between all the places.

4.3. Categorical Layer

The categorical layer contains long-term, low-level representations of categorical models of the robot's sensory information. The knowledge represented in this layer is not specific to any particular location in the environment. Instead, it represents a general long-term knowledge about the world at the sensory level. In this layer models of landmarks, objects or appearance-based room category or other properties of spatial segments such as shape, size or color are defined in terms of low-level features. The position of this layer in the spatial representation reflects the assumption that the ability to categorize and group sensory observations is the most fundamental one and can be performed in a feed-forward manner without any need for higher-level feedback from cognitive processes.

The categorical models stored in this layer give rise to properties that are utilized by conceptual layer. In many cases, the values of those properties will correspond to human spatial concepts, not to internal concepts of the robot (e.g. office-like appearance or elongated shape). The properties might require complicated models that can only be inferred from training data samples. In case of models that correspond to human concepts, they can be learned in a supervised fashion, using a top-down supervision signal.

4.4. Conceptual Layer

The conceptual layer provides an ontology that represents taxonomy of the spatial concepts and properties of spatial entities that are linked to the low-level categorical models stored in the categorical layer. This associates semantic interpretations with the low-level models and can be used to specify which properties are meaningful e.g. from the point of view of human-robot interaction. Moreover, the conceptual layer represents relations between the concepts and instances of those concepts linked to the spatial entities represented in the place layer. This makes the layer central for verbalization of spatial knowledge and interpreting and disambiguating verbal expressions referring to spatial entities.

The second important role of the conceptual layer is to provide definitions of the spatial concepts related to the semantic segmentation of space based on the properties of segments observed in the environment. A building, floor, room or area are examples of such concepts. The conceptual layer contains information that floors are usually separated by staircases or elevators and that rooms usually share the same general appearance and are separated by doorways. Those definitions can be either given or learned based on asserted knowledge about the structure of a training environment introduced to the system.

Finally, the conceptual layer provides definitions of semantic categories of segments of space (e.g. rooms) in terms of values of properties of those segments. The properties can reflect the general appearance of a segment as observed from a place, its geometrical features or objects that are likely to be found in that place.

5. Instantiations

This section indicates specific models and algorithms maintaining those models that we propose to use for representing knowledge stored in each layer.

We propose to realize the sensory layer using a robocentric, metric SLAM [6, 5]. Robocentric mapping reflects the properties of the sensory layer and allows for a straightforward treatment of forgetting knowledge that falls outside a certain horizon around the robot. The robocentric map can be seen as a sliding window centered on the robot and containing a detailed view of the world, which allows the robot to maintain a drift free estimate of the pose as long as it stays in a local region of space. The SLAM algorithm explicitly represents the uncertainty associated with the pose of the robot and the location of all landmarks in the local surrounding using a multivariate Gaussian distribution [6,5].

We propose to instantiate the place layer based on the mapping framework proposed in [21]. Central to the approach is the place map represented as a collection of places. A place is defined by a subset of values of arbitrary, possibly complex, distinctive features and spatial relations reflecting the structure of the environment. The features provide information about the world and can be perceived by an agent when at that place. In this sense, the places build on the perception of the agent and are based on its perceptual capabilities.

The categorical layer can be seen as an ensemble of categorical models of the robot's sensory information. The literature provides a broad range of models that could be used for this purpose. First, in order to represent visual and geometrical

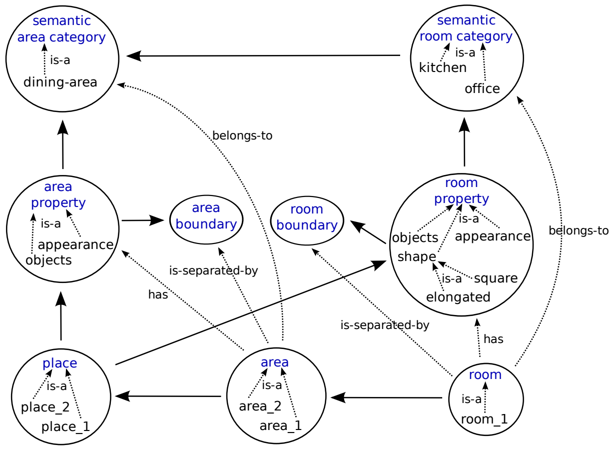

Figure 2. Overview of a possible instantiation of the conceptual layer. The solid arrows represent dependencies, while the dashed arrows illustrate the ontology that represents the taxonomy of spatial concepts and properties of spatial entities.

properties of areas in the environment, we suggest to use the multi-modal place classification algorithm presented in [20]. Other methods can be employed for representing landmarks (e.g. doors [17]) and object categories [19].

For the conceptual layer, we propose a possible instantiation presented in Figure 2. The conceptual layer provides an ontology that represents the taxonomy of the spatial concepts and properties as well as dependencies between the concepts, properties and instances of spatial entities. We use a fixed, handcrafted ontology for representing the taxonomy and a probabilistic model for representing the dependencies. In such an approach, the ontology is largely encoded in the structure of the probabilistic model. We represent the location of the robot within segments of space (e.g. a room or an area such as a dining area), the observed properties of areas and rooms as well as semantic categories of areas and rooms in terms of random variables. In the illustration in Figure 2, we can consider the circles as random variables and the solid arrows as dependencies within a graphical model. At the same time, the *is-a* relations link the random variables with their values. Further, the model represents the spatial hierarchy of segments of space. There is a dependency between the location of the robot at different levels of this hierarchy (e.g. a room and an area within the room). Moreover, the dependency between the instance of a place and the properties of areas and rooms observed from this place is represented. Those properties in turn influence the semantic categories of areas or rooms to which the place belongs. Finally, the proposed model represents the dependency between the area and room properties observed as the robot explores the environment and the probability that the robot crossed a boundary of a spatial segment. This link effectively defines the concepts of a room and an area and can be used to provide semantic segmentation of space.

6. Conclusions and Future Works

In this paper, we presented an analysis of the requirements for a spatial knowledge representation for cognitive systems and proposed a layered representation that conforms to those requirements. The representation provides a unified and coherent view on the structure of spatial knowledge and a basis for designing artificial cognitive systems. We further proposed specific models and algorithms as possible instantiations. Future work will focus on integrating those algorithms, which so far were only evaluated in separation, into a complete spatial subsystem providing spatial understanding capabilities for a mobile robot.

References

[1] COGNIRON: The Cognitive Robot Companion. Website: `www.cogniron.org`.

[2] CogX: Cognitive Systems that Self-Understand and Self-Extend. Website: `www.cogx.eu`.

[3] CoSy: Cognitive Systems for Cognitive Assistants. Website: `www.cognitivesystems.org`.

[4] RobotCub. Website: `www.robotcub.org`.

[5] A. N. Bishop and P. Jensfelt. A stochastically stable solution to the problem of robocentric mapping. In *Proc. of ICRA'09*.

[6] J. Castellanos, R. Martinez-Cantin, J. Tardos, and J. Neira. Robocentric map joining: Improving the consistency of EKF-SLAM. *Robotics and Autonomous Sys.*, 55(1), 2007.

[7] A. Chella and I. Macaluso. The perception loop in cicerobot, a museum guide robot. *Neurocomputing*, 72(4-6), 2009.

[8] M. Cummins and P. Newman. Highly scalable appearance-only SLAM - FAB-MAP 2.0. In *Proceedings of Robotics: Science and Systems V*, June 2009.

[9] R. Davis, H. Shrobe, and P. Szolovits. What is a knowledge representation? *AI Magazine*, 14(1):17–33, 1993.

[10] U. Frese and L. Schrder. Closing a million-landmarks loop. In *Proc. of IROS'06*.

[11] N. Hawes, H. Zender, K. Sjöö, M. Brenner, G.-J. M. Kruijff, and P. Jensfelt. Planning and acting with an integrated sense of space. In *Proc. of HYCAS'09*.

[12] P. Jensfelt, E. Förell, and P. Ljunggren. Automating the marking process for exhibitions and fairs. *Robotics and Autonomous Magazine*, 14(3):35–42, Sept. 2007.

[13] G.-J. Kruijff, H. Zender, P. Jensfelt, and H. Christensen. Situated dialogue and spatial organization: What, where and why? *Int. Journ. of Advanced Robotic Systems*, 4(1), 2007.

[14] B. Kuipers. The spatial semantic hierarchy. *Artificial Intelligence*, 119:191–233, 2000.

[15] M. Milford and G. Wyeth. Mapping a suburb with a single camera using a biologically inspired slam system. *IEEE Transactions on Robotics*, 24(5):1038–1053, Oct. 2008.

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

[17] A. Murillo, J. Kosecka, J. Guerrero, and C. Sagues. Visual door detection integrating appearance and shape cues. *Robotics and Autonomous Systems*, 56(6), 2008.

[18] L. Paz, P. Jensfelt, J. Tardós, and J. Neira. EKF SLAM updates in O(n) with Divide and Conquer SLAM. In *Proc. of ICRA'07*.

[19] A. Pinz. Object categorization. *Foundations and Trends in Computer Graphics and Vision*, 1(4):255–353, 2006.

[20] A. Pronobis, O. M. Mozos, B. Caputo, and P. Jensfelt. Multi-modal semantic place classification. *The International Journal of Robotics Research*, 29(2-3), 2010.

[21] A. Pronobis, K. Sjöö, A. Aydemir, A. N. Bishop, and P. Jensfelt. A framework for robust cognitive spatial mapping. In *Proc of ICAR'09*.

[22] T. Spexard, S. Li, B. Wrede, J. Fritsch, G. Sagerer, O. Booij, Z. Zivkovic, B. Terwijn, and B. Krose. BIRON, where are you? enabling a robot to learn new places in a real home environment by integrating spoken dialog and visual localization. In *Proc. of IROS'06*.

[23] S. Thrun, A. Bücken, W. Burgard, D. Fox, T. Fröhlinghaus, D. Henning, T. Hofmann, M. Krell, and T. Schmidt. Map learning and high-speed navigation in RHINO. In D. Kortenkamp, R. Bonasso, and R. Murphy, editors, *AI-based Mobile Robots: Case Studies of Successful Robot Systems*. MIT Press, 1998.

[24] H. Zender, G.-J. M. Kruijff, and I. Kruijff-Korbayová. Situated resolution and generation of spatial referring expressions for robotic assistants. In *Proc. of JCAI'09*.

[25] H. Zender, O. M. Mozos, P. Jensfelt, G.-J. M. Kruijff, and W. Burgard. Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems*, 56(6), 2008.