Representing Spatial Knowledge in Mobile Cognitive Systems

Andrzej Pronobis, Kristoffer Sjöö, Alper Aydemir, Adrian N. Bishop and Patric Jensfelt
Centre for Autonomous Systems, Royal Institute of Technology
SE-100 44 Stockholm, Sweden
{pronobis, krsj, aydemir, adrianb, patric}@kth.se

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 possible instantiations for each layer of the representation and provide a proof of concept realized within an integrated cognitive system.**

I. 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], [19], [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 do so robustly, reason about space and their own knowledge, plan tasks including acquiring missing knowledge and interact with people in a natural way.

A cornerstone for cognitive mobile agents is to understand the space in which they operate. Spatial knowledge constitutes a fundamental component of the knowledge base of a cognitive agent providing a basis not only for reliable navigation but also abstract and cross-modal reasoning, planning and episodic memories as well as 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 within the system. Due to its central role, the design of spatial knowledge representation should be one of the first steps in building a cognitive system and result from the requirements placed on the system as a whole as well as resulting from the limitations of all interacting components.

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 a cognitive system. Then, we present possible instantiations for each layer of the representation and propose models and algorithms that could be used to store and maintain information. Finally, we briefly describe a proof of concept for most of the elements of the representation realized within the integrated system of the EU FP7 project CogX.

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 within complex, dynamic environments. Although the literature contains many algorithms for spatial mapping and instantiations of mobile robotic systems, there is little work on structuring the whole body of spatial knowledge within an integrated cognitive system. Moreover, little emphasis is placed on rigorous analysis of fundamental requirements and properties of a spatial representation. The existing representations are either designed for a very specific domain [9], [16], they concentrate on a fraction of the spatial knowledge [28], [31] or are designed to solve a single algorithmic task very efficiently rather than for use within a larger system [10], [12], [26]. The idea of this paper, is to take a step back and see how an exhaustive analysis of requirements can lead the way towards a powerful spatial representation for cognitive mobile robot and provide a guidebook for selecting the most suitable instantiations.

II. 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

This work was supported by SSF through the Centre for Autonomous Systems (CAS) and the EU FP7 project CogX and the Swedish Research Council, contract 621-2006-4520 (K. Sjöö) and 2005-3600-Complex (A. Pronobis).

(SLAM) [10], [24], [12], [26] or place classification [25], [28]. 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 [21] 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 [34], [30]. 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 [34]. However, it has several major drawbacks that makes it unsuitable for systems that deal with inherently 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 the aforementioned problems.

III. Analysis of the Problem

In this section, we first propose our definition and analyze the roles of a generic spatial representation. Then, we formulate the problem within the context of cognitive systems in terms of requirements and desired properties.

A. What is a Spatial Knowledge Representation?

Following the analysis by Davis [11], 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, e.g. a kitchen, spatial relations 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. Apart from a middleware that can provide exchange of information, the information must be structured in a way that it can be efficiently accessed and interpreted by each of the communicating components.

In view of the description given above, a cognitive spatial representation is much broader than a cognitive map as defined by Kuipers [20].

B. 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], [31] as well as ongoing research [2] on artificial cognitive systems, we have identified several areas of functionality, usually realized through separate subsystems, that must be supported by the representation. These obviously include spatial localization, navigation, route finding 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, characteristics of a spatial representation as well as limitations resulting from practical implementations, we have identified several desired properties of a spatial representation for mobile cognitive systems.

First, since the representation is unavoidably uncertain, it is futile to represent the world as accurately as possible. A very accurate representation must be complex, require a substantial effort to synchronize with the dynamic world and still cannot guarantee that sound inferences will lead to correct conclusions [11]. Here, we assume that the representation should instead be minimal and inherently coarse. Only as much knowledge should be represented as it is required to provide all the necessary functionality for the system. Furthermore, redundancy should be avoided and whenever possible and affordable, new knowledge 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 in order to make it robust to the dynamic changes in the world and representations that are more abstract should be used for longer-term storage. At the same time, knowledge extracted from immediate observations might 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 [34]). Additionally, it should be possible to determine the location of the agent as well as interpret references to points in space at each level of abstraction.

Similarly to abstraction levels, space should be represented on different spatial scales from single scenes to whole environments. Moreover, space should be discretized into a finite number of spatial units. Discretization of continuous space is one of the most important abstracting steps in representing spatial knowledge 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 [13] and serves as a basis for higher level conceptualization [34].

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 which relate to spatial entities already explored. 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 [28], objects [27] or properties of space recognized by humans (e.g. shape, size etc.).

Another important group of properties is related to the fundamental role of the representation in human-robot interaction. The representation should allow for modelling correspondence between the represented symbols and human concepts of space. A typical example would be human-like segmentation of space (e.g. into rooms, floors, buildings) or human spatial properties. This correspondence should allow for generating and resolving spatial referring expressions [33] as well as path descriptions.

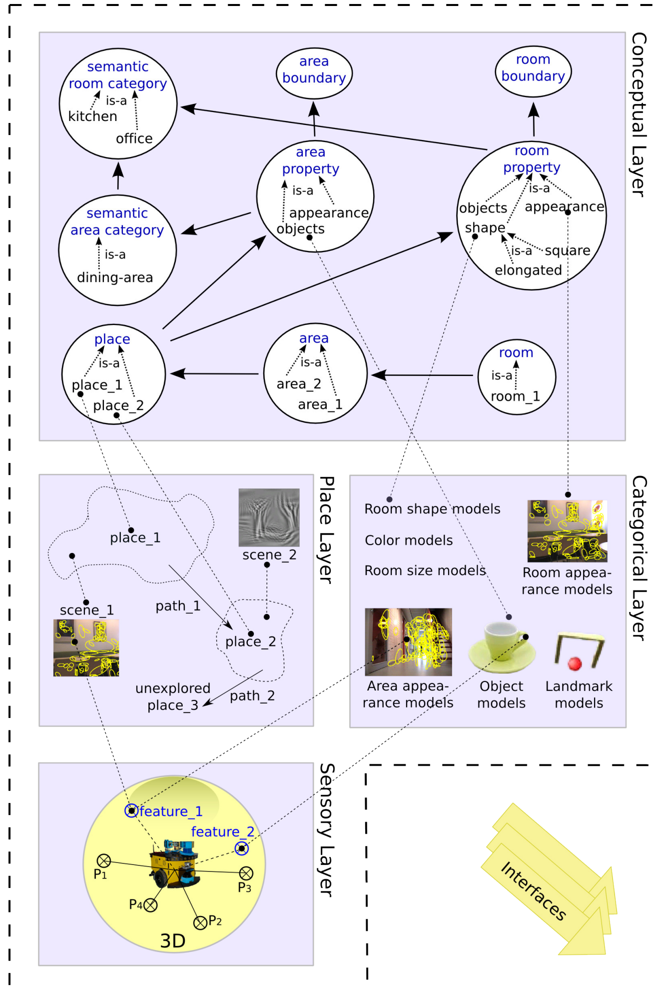


Fig. 1. The layered structure of COARSE. The position of each layer within the representation corresponds to the level of abstraction of the spatial knowledge.

Finally, the representation should be adaptable and should not grow unbounded despite the continuous changes in the environment. In the next sections, we propose a representation of spatial knowledge that adheres to the desired properties formulated above.

IV. Structure of the Representation

The primary contribution of this paper is the cognitive layered representation of spatial knowledge (COARSE). A general overview of the structure of the representation is presented in Figure 1. The fundamental building block of the representation is a layer which can be regarded as a sub-representation focusing on a different abstraction of the spatial knowledge. We distinguish between 4 layers: sensory layer, place layer, categorical layer and conceptual layer. Each layer focuses on different aspects of the world and represents those aspects using specific symbols. The layers focus on different spatial scales, define their own spatial entities and the way the agent's location in the world is represented. Finally, the layers define the persistence of the

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 $\leftrightarrow$ 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
Knowledge decay	Replacement	Generalization, forgetting	Generalization	None / Forgetting
Information flow	Bottom-up	Primarily bottom-up	Primarily bottom-up	Top-down and bottom-up

TABLE I

COMPARISON OF PROPERTIES OF THE FOUR LAYERS OF COARSE.

information they represent. A comparison of properties of the four layers of COARSE is presented in Table I.

As previously mentioned, redundancy should be avoided; however, information might be abstracted and thus moved between the layers. As a result, each layer also corresponds to a map and an actual storage of information. Apart from layers, the representation provides interfaces that can be seen as sets of inferences used to derive new symbols for specific purposes. However, those symbols are outside the scope of the layers and thus will not be updated.

The following sections provide details about each of the layers as well as a general overview of possible interfaces.

A. 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, the representation in the sensory layer is akin to a sliding window, with robot-centric 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. It can contain data of both a 2D and 3D nature.

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.

B. 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 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; however, knowledge that is not accessed or updated can be compressed, generalized and finally forgotten.

Besides places, the place layer also needs to define 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 robot is guided and the user indicates part of the environment that should be of interest to the robot, but not immediately).

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.

C. 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. For instance, this is the layer where 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 higher-level layers. In order to enhance communication with the user, 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. Due to the high complexity of the models, unused knowledge might be compressed and generalized.

D. 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 those concepts and the spatial entities. 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 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.

Another role of the conceptual layer is to provide definitions of semantic categories of segments of space (e.g. areas or rooms) in terms of the values of properties of those segments. These properties can reflect the general appearance of a segment as observed from a place, its geometrical features or objects found in the place.

E. Interfaces

Apart from the layers, we propose to introduce a concept of an interface, i.e. a set of inferences that might be used to derive new symbols from the knowledge represented within the layers. Interfaces usually support specific functionalities within the system, which for the sake of robustness and efficiency are not an inherent part of any layer and thus are not updated or maintained. A typical example would be visualization of information represented by the system. Here, we give examples of two interfaces that could be realized within the proposed spatial representation:

- Metric interface providing approximate global metric maps or metric maps for a selected region of space. Metric maps could be generated based on the local connectivity information preserved in the place layer, relaxed e.g. using a method proposed in [24].
- Topological interface providing global topological graph for the environment as well as approximate costs of traveling between the topological nodes.

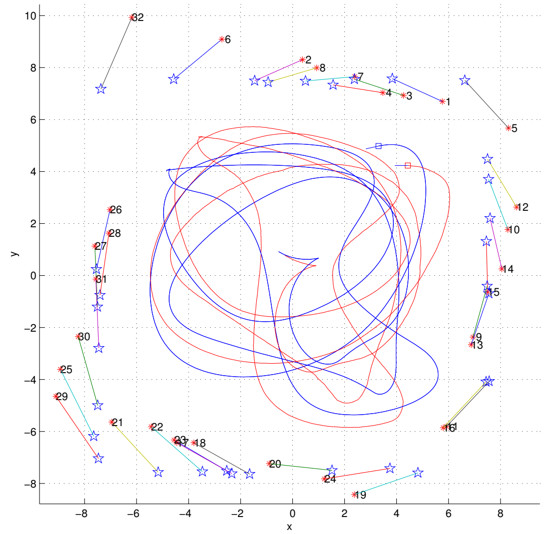


Fig. 2. Global SLAM map and estimated robot trajectory is rotated and displaced quite significantly.

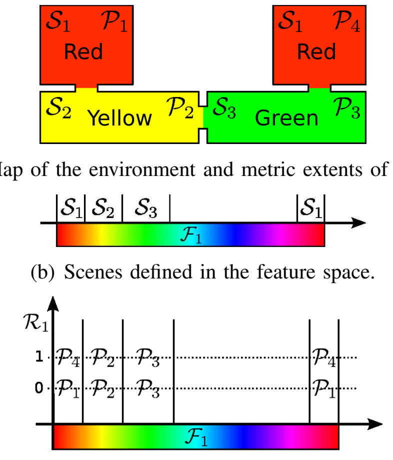
V. Possible Instantiations

This section proposes and briefly describes specific models that could be used to represent knowledge stored in each layer and algorithms maintaining those models according to the principles described in the previous section.

A. Sensory Layer

The sensory layer can be realized using a robocentric SLAM algorithm. Robocentric mapping exhibits a number of advantages over traditional SLAM methods. For example, the traditional SLAM state vector in a global coordinate system is not observable as discussed in [14] given only relative landmark-robot measurements such range and/or bearing. Another problem with global mapping algorithms is that of estimator inconsistencies caused by accumulated linearization errors [17], [5], [15]. Commonly, a global map and robot pose estimate can rotate and/or translate a significant amount. Figure 2 shows a birds-eye view of such a case and highlights the flaws of world-centric mapping.

This distortion of the global map shown in Figure 2 is caused by an unobservable state in global SLAM, as was shown in [22]. In [8], [7], [6] the concept of robocentric mapping is introduced and this concept is shown to better deal with linearization errors than the traditional SLAM formulation. In general, robocentric mapping constitutes a well-posed problem with an observable state for many classes of sensor measurement. From the point of view of a cognitive agent, we are using a robocentric mapping algorithm in order to emphasize the mapping function in a vicinity around



(a) Map of the environment and metric extents of places.

(b) Scenes defined in the feature space.

(c) Places defined in the descriptor space.

Fig. 3. Illustrative example of an environment and definitions of places in the descriptor space.

the robot. This has practical importance for certain actions likely to be performed by any cognitive agent such as object manipulation.

B. Place Layer

We propose to instantiate the place layer based on the mapping framework recently proposed in [29]. 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 process of defining places consists of two steps. First, the space of all features is divided into disjoint sets termed *scenes*. This provides an initial segmentation of space corresponding to groupings of similar feature values. Then, spatial relations are formed (e.g. adjacency relation) for each point in space with respect to the scenes e.g. by integrating features temporally. This allows for disambiguating between locations that have the same appearance but are separate in space.

To illustrate the idea, let us consider a simple example of a small environment presented in Figure 3(a) consisting of 4 rooms characterized by the color of the floor. We define a single feature $f_1 \in \mathcal{F}_1$ that corresponds to the hue of the floor color at each location. Then, we can define a feature space by the range of the hue values. If we divide the feature space into regions as presented in Figure 3(b), we can differentiate between three scenes: red ($\mathcal{S}_1$), yellow ($\mathcal{S}_2$) and green ($\mathcal{S}_3$). We can clearly see that the scene $\mathcal{S}_1$ corresponds to two separate rooms which could be distinguished if we consider their relations to other scenes. Let us define an adjacency relation with respect to the scene $\mathcal{S}_3$, $r_1 \in \mathcal{R}_1 = \{1, 0\}$, and create a descriptor space $\mathcal{D} = \mathcal{F}_1 \times \mathcal{R}_1$. In that descriptor space, we can create four non-overlapping places $\mathcal{P}_1$-$\mathcal{P}_4$ by dividing the scene $\mathcal{S}_1$ into two places, one of which is adjacent to the scene $\mathcal{S}_3$ and the other is not. This division is reflected in the clustering of space presented in Figure 3(c).

The place map is defined in terms of the agent's perception of space and adapts to its perceptual capabilities. Moreover, the perceived features can be abstract and non-metric and describe for instance visual properties of the world. In this sense, the map is subjective and robocentric as the robot's observations do not have to be expressed in terms of any objectively defined quantities or any global coordinate system. The map is fragmented (consists of a set of independent places), topological and does not require maintaining global spatial consistency.

C. Categorical Layer

The categorical layer can be seen as an ensemble of categorical models of the robot's sensory information. The literature provides several models that could be used for this purpose. Let's concentrate on the available approaches to place classification that could be used to represent appearance-based or geometry-based properties of spatial segments. These are either purely geometrical and rely on cues extracted from laser range data, employ visual cues or use a combination of both modalities. An approach employing the AdaBoost classifier and a set of simple geometrical features extracted from laser range scans was proposed in [25]. Torralba *et al.* [32] employs a global visual descriptor and a Hidden Markov Model in order to classify places. Finally, combination of different sensory modalities was used in [28], where discriminative models of places were built using Support Vector Machines from geometric laser-based features and a combination of global and local visual cues. Apart from representing appearance or geometry-based classes of rooms or areas, similar models could be used for other properties of spatial entities such as shape, size or color.

Other methods could be employed for landmark and object modelling, detection and categorization. Murillo *et al.* describe a method for visual door detection using appearance and shape cues. A system that uses information signs as landmarks, and interprets them through its ability to read text and recognize icons is proposed in [23]. Finally, the vision literature provides a broad range of approaches to the problem of object categorization. An extensive review of the state-of-the-art can be found in [27].

D. Conceptual Layer

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. Here, we propose a possible instantiation of the layer presented in Figure 4. In order to address the needs of the conceptual layer that represents conceptual knowledge of different type, we propose to 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 propose to represent

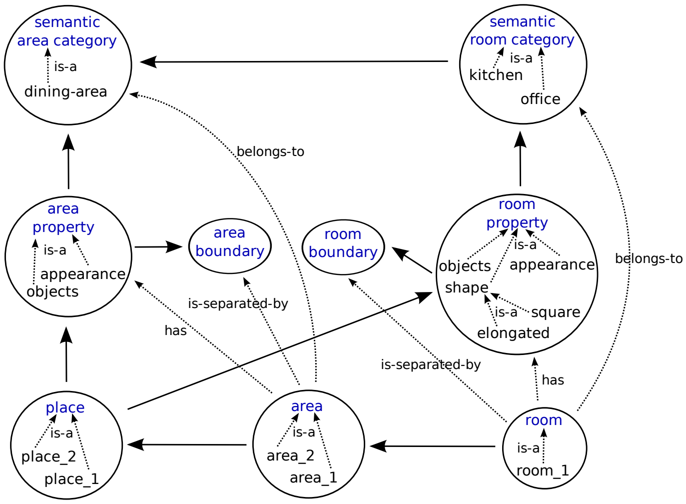


Fig. 4. Overview of a possible instantiation of the conceptual layer. The solid arrows represent dependencies, while the dashed arrows illustrate the ontology that represents the taxonomy of spatial concepts and properties of spatial entities.

the location of the robot within segments of space (e.g. a place, 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.

Let's focus on the example presented in Figure 4. 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. The model represents the spatial hierarchy of segments of space. There is clearly 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 model represents the dependency between the instance of a place and the properties of areas and rooms observed from this place. Those in turn influence the semantic categories of areas or rooms to which the place belongs. For example, this link represents the strong connection between the semantic category of a room and typical objects found in that room.

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 in terms of the properties and can be used to provide semantic segmentation of space.

VI. Proof of Concept

In this section, we briefly describe a proof of concept for most of the elements of the representation realized within the integrated system of the EU FP7 project CogX. In this initial implementation, we simulated the general knowledge structure, still on top of algorithms that do not adhere fully to the principles behind the representation. This way, the whole system violates some of the assumptions behind the proposed representation; however, it allows validating correctness and usefulness of the general knowledge structure within an integrated cognitive system.

We assembled a system consisting of several subarchitectures (groupings of large functional portions) within the software framework based on the CoSy Architecture Schema (CAS, [18]). The following subarchitectures were included: *the spatial subarchitecture* responsible for maintaining the spatial knowledge and relating the robot to that knowledge; *the planning subarchitecture*, which given goals, dynamically decides which subarchitectures need to provide which data in order to achieve them; *the motivation subarchitecture* which monitors the system state and decides which goals should be concatenated into a single one for the planner; and finally *the binding subarchitecture* which unifies cross-modal information.

Within the spatial subarchitecture, we implemented the place layer as a set of independent places connected with neighboring places using paths. These places were generated on top of a global metric map, which effectively replaced most of the functionality of the sensory layer; however, other components of the system did not have access to global metric coordinates. Moreover, according to the requirements for the spatial representation, we introduced the concept of placeholders in order to represent hypotheses for places in the yet unexplored part of the environment. The categorical layer was realized through categorical models of room appearance and geometry based on the algorithm proposed in [28]. Additionally, a simple door detection algorithm was used as a replacement for a more sophisticated landmark model. Finally, the conceptual layer was instantiated using a description-logic based deductive reasoning that relied on the the discretization of space and properties assigned to places, maintained a *taxonomy* of concepts and the knowledge about individuals and performed reasoning resulting in semantic segmentation of space.

The system was successfully tested in an autonomous exploration scenario on a Pioneer differential-drive robot base, equipped with a Hokuyo laser scanner, and surmounted by a custom superstructure bearing a stereo camera and a laptop that performs all information processing. Based on the outcomes of the initial experiment as well as the design process of the integrated system, we can conclude that the framework facilitates integration between the subarchitectures and greatly enhances the scalability of the whole system. Additionally, the experiments demonstrated the feasibility of the representation in carrying out spatial abstraction from continuous sensory information to amodal symbolic planning.

VII. Conclusions and Future Works

In this paper, we presented a thorough analysis of the requirements for a spatial knowledge representation for mobile 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. Moreover, it facilitates integration between components improving the scalability of the system. We further proposed specific models and algorithms as possible instantiations and provided a proof of concept based on an initial implementation within an integrated system.

The future work will focus on creating components providing instantiations for each of the layers of the representation that fully adhere to presented principles and constitute a part of an integrated system. Currently, we pursue research on robocentric SLAM algorithms, semantic categorization of space, high-level probabilistic conceptual reasoning and efficient instantiations of the place map. From the point of view of the structure of the representation, we plan to investigate how small-scale and large-scale space could be represented within the same, unified framework.

REFERENCES

[1] "COGNIRON: The Cognitive Robot Companion," website: http://www.cogniron.org.

[2] "CogX: Cognitive Systems that Self-Understand and Self-Extend," website: http://www.cogx.eu/.

[3] "CoSy: Cognitive Systems for Cognitive Assistants," website: http://www.cognitivesystems.org/.

[4] "RobotCub," website: http://www.robotcub.org/.

[5] T. Bailey, J. Nieto, J. Guivant, M. Stevens, and E. Nebot, "Consistency of the EKF-SLAM algorithm," in *Proc. of the IROS'06*.

[6] A. N. Bishop and P. Jensfelt, "Stochastically convergent localization of objects and actively controllable sensor-object pose," in *Proc. of the 10th European Control Conference (ECC'09)*.

[7] ——, "A stochastically stable solution to the problem of robocentric mapping," in *Proc. of ICRA'09*.

[8] J. Castellanos, R. Martinez-Cantin, J. Tardos, and J. Neira, "Robocentric map joining: Improving the consistency of EKF-SLAM," *Robotics and Autonomous Systems*, vol. 55, no. 1, January 2007.

[9] A. Chella and I. Macaluso, "The perception loop in cicerobot, a museum guide robot," *Neurocomputing*, vol. 72, no. 4-6, 2009.

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

[11] R. Davis, H. Shrobe, and P. Szolovits, "What is a knowledge representation?" *AI Magazine*, vol. 14, no. 1, pp. 17–33, 1993.

[12] U. Frese and L. Schrder, "Closing a million-landmarks loop," in *Proc. of IROS'06*.

[13] 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 *Proceedings of the 1st International Workshop on Hybrid Control of Autonomous Systems – Integrating Learning, Deliberation and Reactive Control (HYCAS)*, Pasadena, CA, USA, July 2009.

[14] G. Huang, A. Mourikis, and S. Roumeliotis, "Analysis and improvement of the consistency of extended Kalman filter based SLAM," in *Proc. of the ICRA'08*.

[15] S. Huang and G. Dissanayake, "Convergence and consistency analysis for extended kalman filter based SLAM," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 1036–1049, October 2007.

[16] P. Jensfelt, E. Förell, and P. Ljunggren, "Automating the marking process for exhibitions and fairs," *Robotics and Autonomous Magazine*, vol. 14, no. 3, pp. 35–42, Sept. 2007.

[17] S. Julier and J. Uhlmann, "A counter example to the theory of simultaneous localization and map building," in *Proc. of ICRA'01*.

[18] G.-J. Kruijff, M. Brenner, and N. Hawes, "Continual planning for cross-modal situated clarification in human-robot interaction," in *Proc. of RO-MAN'08*.

[19] G.-J. M. Kruijff, H. Zender, P. Jensfelt, and H. I. Christensen, "Situated dialogue and spatial organization: What, where... and why?" *International Journal of Advanced Robotic Systems (ARS), Special Issue on Human-Robot Interaction*, vol. 4, no. 1, March 2007.

[20] B. Kuipers, "The cognitive map: Could it have been any other way," in *Spatial Orientation: Theory, Research, and Application*. Plenum Press, 1983, pp. 345–359.

[21] ——, "The spatial semantic hierarchy," *Artificial Intelligence*, vol. 119, pp. 191–233, 2000.

[22] K. W. Lee, W. Wijesoma, and I. Javier, "On the observability and observability analysis of SLAM," in *Proc. of IROS'06*.

[23] M. Mata, J. M. Armingol, A. de la Escalera, and S. M. A., "A visual landmark recognition system for topological navigation of mobile robots," in *Proc. of ICRA'01*.

[24] M. Milford and G. Wyeth, "Mapping a suburb with a single camera using a biologically inspired slam system," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1038–1053, Oct. 2008.

[25] 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 Journal*, vol. 55, no. 5, pp. 391–402, May 2007.

[26] 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*.

[27] A. Pinz, "Object categorization," *Foundations and Trends in Computer Graphics and Vision*, vol. 1, no. 4, pp. 255–353, 2006.

[28] A. Pronobis, O. Martínez Mozos, and B. Caputo, "SVM-based discriminative accumulation scheme for place recognition," in *Proc. of ICRA'08*.

[29] A. Pronobis, K. Sjöö, A. Aydemir, A. N. Bishop, and P. Jensfelt, "A framework for robust cognitive spatial mapping," in *Proceedings of the 14th International Conference on Advanced Robotics*, Munich, Germany, June 2009.

[30] 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*.

[31] 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 *AI-based Mobile Robots: Case Studies of Successful Robot Systems*, D. Kortenkamp, R. Bonasso, and R. Murphy, Eds. MIT Press, 1998.

[32] A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin, "Context-based vision system for place and object recognition," in *Proc. of ICCV'03*.

[33] 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*.

[34] H. Zender, Óscar Martínez Mozos, P. Jensfelt, G.-J. M. Kruijff, and W. Burgard, "Conceptual spatial representations for indoor mobile robots," *Robotics and Autonomous Systems*, vol. 56, no. 6, pp. 493–502, June 2008.