Multi-Layered Conceptual Spatial Mapping for Autonomous Mobile Robots

Hendrik Zender and **Geert-Jan M. Kruijff**
Language Technology Lab
German Research Center for Artificial Intelligence (DFKI GmbH)
Saarbrücken, Germany
{zender,gj}@dfki.de

Abstract

This paper presents an approach to spatial mapping for autonomous mobile robots that are to operate among, and interact with, non-expert human users. We argue that our approach of *conceptual spatial mapping* helps bridge the gap between the representations needed for low-level control of the robot, and the conceptual-topological representations of space humans have. Our approach maintains spatial knowledge on multiple interconnected layers. We show that a process for map acquisition, *human-augmented mapping*, which combines bottom-up and top-down influences from different modalities, will yield a rich multi-layered spatial representation. This representation enables the robot to perform complex actions in a human-populated environment. We show that our approach can be used to establish a notion of *situational* and *functional awareness*.

Introduction

Nowadays, robots have evolved from industrial assembly line machines to domestic service and entertainment robots. The next step ahead will take robots into our everyday lives, where they will no longer be operated by trained personnel but instead have to be of assistance to people from the general public. Thus a major challenge lies in facilitating the interaction between robots and humans. The modes of interaction must allow people to easily exercise precise control over their robots while preserving a high degree of flexibility for efficient cooperation of autonomous robots with their human users. The most intuitive way for humans to communicate with a robot, especially for non-expert users, is through spoken language. If such a dialogue is to be successful, the robot must make use of the same concepts to refer to things and phenomena in the world as a human would. For this, the robot needs to conceptually perceive the world similar to the way humans do. That is to say, it should be able to understand the spatial and functional properties of a human-designed and human-populated environment. At the same time, such understanding should allow it to safely navigate its environment and precisely control its actuators.

To recapitulate, A spatial knowledge representation for robotic assistants must thus address the issues involved with safe and reliable navigation control, with representing space in a way that resembles the way humans segment space into topological regions, and finally with the way linguistic reference to spatial entities is established in situated natural-language dialogue.

To meet these requirements, we propose a multi-layered spatial representation ranging from a metric map suitable for motion control to a conceptual-semantic view on a topological map that allows for conceptual reasoning. Our method has been integrated with a flexible, state-of-the-art linguistic framework that actively supports the map acquisition process and is used for situated dialogue about the environment. The aforementioned methods and techniques have been fully implemented in a *cognitive architecture* for a mobile robotic platform (Kruijff *et al.* 2007). We argue that our approach of adding a conceptual-semantic dimension to navigation and map acquisition presents an advance for communicative, cooperative service robots that are to collaborate with non-expert human users in domestic and office settings.

We will show how an extended notion of *human-augmented mapping*, i.e. the combination of a tutor-driven supervised map acquisition process with autonomous exploration and discovery of the environment by the robot, combines top-down and bottom-up approaches for mapping resulting in a spatial representation that is adequate for both robot control and human-robot interaction. In our architecture, qualitative and quantitative information about the environment stemming from different modalities is combined, thus yielding a more complete understanding of the world.

In our approach the distinction between *acquired*, *asserted*, *innate*, and *inferred* information plays a crucial role for the spatial knowledge processing mechanisms. We will also show how conceptual spatial knowledge primes the robot's behaviors, thus contributing to a notion of *situational awareness*.

Conceptual Spatial Mapping

Research in robotics has yielded robust and adaptive algorithms for autonomous robot navigation (Choset *et al.* 2005). These approaches often rely on a metrical, *quantitative*, representation of the environment. Research on human spatial cognition, however, indicates that humans adopt a partially hierarchical, conceptual-topological view on space (McNamara 1986; Hirtle & Jonides 1985) that is inherently vague and *qualitative* in nature. There are some approaches

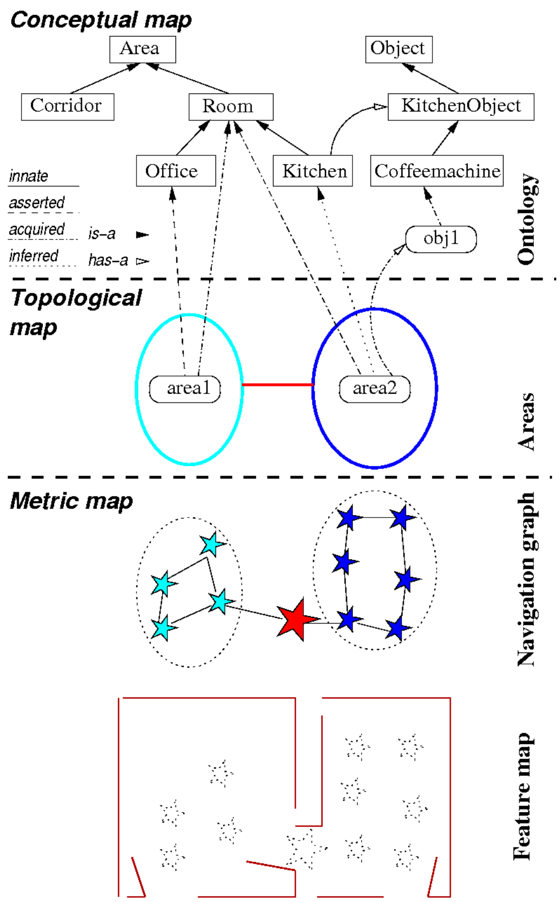

Figure 1: The multi-layered spatial representation

to endow robots with topological maps and adequate small-scale control strategies to move about within topological units, e.g. the *Spatial Semantic Hierarchy* (Kuipers 2000) and the *Route Graph* (Krieg-Brückner *et al.* 2005). These approaches are especially suitable for resolving verbal route descriptions. However, they differ from our approach in that they do not provide a conceptualization of indoor areas in terms of spatial and functional properties that would allow for resolution and generation of linguistic referring expressions.

Our notion of *conceptual spatial mapping* subsumes a *metric layer* that relies on the SLAM technique for exact, feature-based localization and mapping and is used for robot control and navigation, a *topological abstraction layer*, which partitions an indoor environment into topological areas, and a *conceptual layer*, which provides a semantic view on the spatial organization of the environment. The representations used in the individual layers have been chosen to address the requirements of reliable self-localization and exact mapping on the one hand and of providing a humanlike segmentation and categorization of spatial areas on the other. Figure 1 illustrates the layers of the map.

The exteroceptive sensors that are used in our implementation are a SICK laser range finder with a 180° field of view covering the front of the robot and a camera mounted on top of the robotic platform, an ActivMedia PeopleBot[1]. Wheel encoders provide odometry information. A microphone connected to a speech recognition software is used to record spoken input by the user.

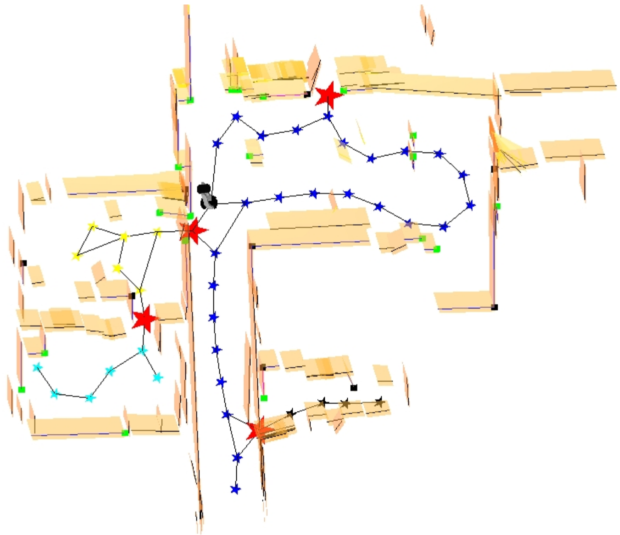

Figure 2: Metric and topological layers. Line features (extended to pseudo-3D walls) and navigation graph (colored stars and black lines) belong to the metric map. The partitioning of the navigation graph into topological areas is depicted by the different coloring of the place nodes.

Metric Layer

The metric map establishes an absolute frame of reference. Within this frame of reference, the SLAM module (Folkesson, Jensfelt, & Christensen 2005) stores the feature representations it needs to keep the robot localized. The features used here are lines that are extracted from laser range scans. Such lines typically correspond to walls and other flat, straight structures in the environment. While the robot moves around, a *navigation graph* representation of visited places (navigation nodes) and paths (edges), based on the notion of a ‘roadmap of virtual free-space markers’ as presented in (Newman *et al.* 2002), is constructed. This representation establishes a model of *free space* and its *connectivity*, i.e. reachability. In our implementation, it is used for path planning and navigation in known environments. We distinguish between two kinds of navigation nodes: place nodes and doorway nodes. Place nodes represent distinct places of free space on traveled routes. Whenever the robot passes through a narrow opening, the nearest node is placed in the center of the opening, converted to a doorway node. Furthermore, the set of navigation nodes provides a link for a topological map. Figure 2 gives an example of the visualization of the metric map including the line features (i.e. walls) and the navigation graph (i.e. traveled routes).

[1] http://www.mobilerobots.com

Topological Layer

The topological abstraction layer divides the set of navigation nodes into areas. An area consists of a set of (transitively) interconnected navigation nodes. In this view, the exact shape and boundaries of an area are irrelevant. The set of navigation nodes is partitioned into discrete areas on the basis of the door detection mechanism described in the previous paragraph. This approach complies with findings in cognitive psychology: humans segment space into regions that correspond to more or less clearly defined spatial areas (in the general sense of the word). The borders of these regions may be defined physically, perceptually, or may be purely subjective to the human. Walls in the robots environment are the physical boundaries of areas. Doors are a special case of physical boundaries that permit access to other areas. Figure 2 shows the topological partitioning by the color of the route nodes. Doorway nodes are depicted by large red stars. Note that in this figure, the hallway (top right) and the corridor (center) are treated as a single area because there is no separating doorway. In this example, a shortcoming of our approach is obvious: not all areas in a building are separated by doors. We hope to soon be able to cover these cases of perceptual boundaries by using the robot's perception of geometrical features of areas (Martínez Mozos & Burgard 2006) as another cue for segmenting space.

Conceptual Layer

In the conceptual map, knowledge stemming from vision and dialogue is anchored to the metric and topological maps. The conceptual map represents information about spatial areas and objects in the environment in an ontological reasoning module. The conceptual map contains a commonsense ontology of an indoor environment implemented as an OWL[2] ontology, cf. Figure 3. It describes taxonomies (*is-a* relations) of room types, and couples room types to typical objects found therein through *has-a* relations. These conceptual taxonomies have been handcrafted and cannot be changed online. However, instances of the concepts are added to the ontology during run-time. Using a description-logics based reasoning software[3], new knowledge can be inferred. For example, if the robot knows that it is in an area where there is a coffee machine and an oven, it can infer that it can categorize this area as a kitchen. Like this, linguistic references to areas can be generated and resolved.

Information processing

Depending on the origin of a piece of information, we distinguish between *acquired*, *asserted*, *innate*, and *inferred* knowledge. *Acquired knowledge* is derived from the robot's own sensors, including the spatial information encoded in the metric map and objects recognized by the vision-sensor. *Asserted knowledge* is provided by another agent, in our case the robot's tutor. It is typically given through verbal input (for example, the tutor might say "you are in the laboratory."). *Innate knowledge* is any kind of information that

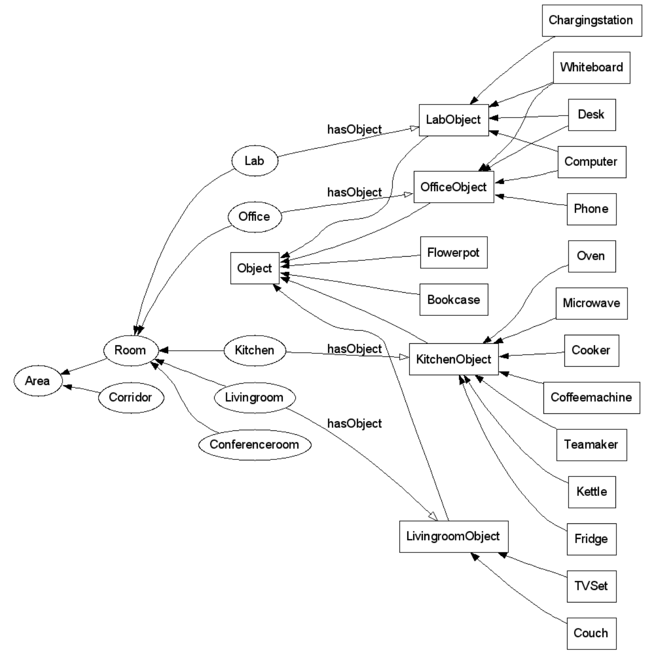

Figure 3: Commonsense ontology of an indoor environment. Depicted are only the parts that are necessary for automated reasoning. There are other roles and concepts in the taxonomy that are used only for spatial object memory.
The subconcept relation in the concept taxonomy (`is-a`) is expressed by arrows with black arrowheads. Hollow arrowheads denote the `hasObject` relation.

is incorporated into the architecture in a way that does not allow for on-line manipulation of the knowledge. In our architecture, the handcrafted commonsense conceptual ontology is an example of innate knowledge. Any piece of information that can be derived on the basis of the combination or evaluation of other information provides *inferred knowledge*, such as knowledge inferred by the description-logics based reasoning mechanisms in the conceptual map.

The individual layers of our multi-layered mapping approach are implemented in a distributed architecture. The information processing in our architecture follows the principles of *pushing* and *pulling* data. Whenever new information is received and processed and thus new knowledge is generated, relevant pieces of information are automatically pushed to any module that takes this kind of information as input. When the robot is executing behaviors, relevant context knowledge is pulled from those modules that provide the necessary information.

In the following sections, we will show how information is pushed between the modules involved in map acquisition, and how pulling information about the spatial context primes the robot's behavior, thus providing a degree of situational and functional awareness.

Interactive Map Acquisition

The map acquisition process exemplifies how information and knowledge is pushed between modules in our distributed

[2] `http://www.w3.org/TR/owl-guide/`
[3] `http://www.racer-systems.org`

architecture. It also relies on the combination of top-down and bottom-up influences on information processing.

Human-Augmented Mapping

The multi-layered representation is created using an enhanced method for concurrent semi-supervised map acquisition, i.e. the combination of a user-driven supervised map acquisition process with autonomous exploration discovery by the robot. This process is based on the notion of *Human-Augmented Mapping*, as introduced by Topp and Christensen (Topp & Christensen 2005). We additionally use a linguistic framework that actively supports the map acquisition process and is used for situated dialogue about the environment (Kruijff *et al.* 2007). The map can be acquired during a so-called *guided tour scenario* in which the human tutor shows the robot around and continuously teaches the robot new places and objects. During such a guided tour, the user can command the robot to follow him or to explore an area autonomously. Our system does not require a complete initial guided tour. It is as well possible to incrementally teach the robot new places and objects at any time the user wishes. With every new piece of information, the robot's internal representations become more complete. Still, the robot can always perform actions in, and conduct meaningful dialogue about, the aspects of its environment it already knows about.

Information Processing During Map Acquisition

Our approach to human-augmented mapping inherently combines several control strategies for map building. Tutor-given information primes the robot's expectations in a top-down manner. If the tutor tells the robot that "there is a printer," the image processing mechanisms will be primed to detect an instance of a printer in the robot's field of view. If the robot has been told that a specific area is called "kitchen", then—basing on ontology reasoning—it will expect there to be typical objects. By the same token, the acquisition of the metrical map and the topological abstraction level take place in a purely bottom-up manner. In (Kruijff *et al.* 2006), we show how the robot can initiate a clarification dialogue if it detects an inconsistency in its automatically acquired spatial representation. In principle this mechanism could be extended to handle contradictions or ambiguities in asserted and inferred knowledge.

The bottom-up acquisition of the spatial representation is done in a fix sequence of processing steps. The metric map constructed in the SLAM module works on input from the laser range finder and from the robot's wheel odometry. Within the multi-layered map, the SLAM module enables the robot to acquire knowledge about solid structures, as well as free and reachable space in its environment. Through use of a simple door detection mechanism, the free and reachable space is partitioned into topological areas. As soon as the robot acquires the knowledge about a new area in the environment, this information is pushed from the topological map to the conceptual map. There it is represented by creating a new instance of the concept `Area` that is anchored to the topological area by a unique identifier. A separate module for place classification (Martínez Mozos & Burgard 2006), which is also working on the laser range input, assigns semantic labels ('room' or 'corridor') to laser range readings. This classification is anchored to the pose at which it was obtained and pushed to the navigation map. There each navigation node is assigned a semantic label according to the majority vote of the classifications of the poses in its vicinity. As soon as the robot physically leaves an area, the majority vote of the node classifications is computed and pushed to the conceptual map, where it is used to further specify the area's ontological instance as `Corridor` or `Room`.

The visual object detection system uses SIFT features (Lowe 2004) to recognize previously learned objects, such as a television set, a couch, or a coffee machine. Whenever an image is matched to a training image, the pose of the robot is used to determine the position of the corresponding detected object. The positions of objects are stored in a local memory, but the acquired knowledge about the occurrence of an object is pushed to the conceptual map. There a new instance of the ontological concept of the object, e.g. `TVSet` is created and connected to the `Area` instance of the place in which it is located via the `hasObject` relation.

Whenever the user gives an assertion about areas in the environment or objects found therein, the dialogue subsystem pushes these assertions on to the conceptual map, where the ontology is updated with the new information.

Situational and Functional Awareness

While executing certain behaviors, such as autonomously navigating the environment or following a tutor, the modules responsible for the execution will pull relevant context information from the memories of the spatial mapping modules. This context knowledge determines certain parameters of the robot's behavior or leads even to the selection of a behavior that is specialized to fit the given situation.

We currently investigate how the information encoded in the robot's spatial representation can be used for a smarter, human- and situation-aware behavior. For one, the robot should exploit its knowledge about objects in the environment to move in a way that allows for successful interaction with these objects.

For instance, when following a person, the robot should make use of its knowledge about doors in the environment, such that it stops when it detects when the person it is following wants to leave the current area by opening a closed door and passing through it. In such a case, the robot should react accordingly, anticipating the tutor's holding open the door. As a first approach, we opt for the robot to increase the distance it keeps to its tutor when it detects that the tutor is approaching a door. A failure to understand the current situation would lead to a situation where the robot stubbornly moves behind the person until the person, trying to hold open the door, is trapped in the corner of the room by the robot. For this instance of *situational awareness*, we make use of a laser-range based people tracking (Schulz *et al.* 2003) and following algorithm combined with the information about doorway nodes present in the navigation graph. The people tracking module constantly pushes localization information of the tracked person, including the x-y-coordinates, an angular heading and the velocity, to the people following module. On the basis of this information, it pulls information about doors in the tutor's vicinity from the navigation map. In case the module determines that the tutor is approaching a door, the robot increases the distance it keeps to its tutor. Like that, the robot is prevented from blocking the doorway without interrupting its people following behavior in an unnatural way.

Apart from the bottom-up driven understanding of the robot's current situation, our system provides also a starting ground for establishing a notion of a top-down *functional awareness*. The commonsense knowledge encoded in the conceptual layer provides cues for where the robot can expect to find specific objects and thus, ultimately, where certain tasks can be accomplished. By knowing the general name (e.g. "kitchen") for a room, it is thus possible to know what actions can be performed there. Although most current robotic systems do not feature any sophisticated and reliable manipulators, robots will soon be used to perform fetch-and-carry tasks in domestic settings. A first step for a mobile robot to accomplish any task, no matter how complex, is to determine where to perform actions and how to get there. The conceptual spatial map presented in this paper provides the basis for the robot's awareness of functions of spatial areas based on the objects found therein.

Conclusions

We have presented an approach to multi-layered conceptual spatial mapping for autonomous robots. This approach addresses two important challenges of mobile robots that are to operate among and cooperate with humans. For one, it accounts for the need to have precise metric—quantitative—information about the environment that can be used for safe and reliable motion control and navigation of an autonomous robot. Secondly, our approach helps facilitating human-robot interaction by endowing the robot with a conceptual—more qualitative—understanding of its environment that resembles the way humans conceive of their environment.

We believe that our multi-layered representation of space, which includes geometrical information of the spatial areas, knowledge about objects in the environment, a topological abstraction layer, as well as a conceptual layer that links instance knowledge with commonsense world knowledge, serves as a good basis for integrating procedural-functional knowledge with situational awareness.

The information processing mechanisms underlying our approach have been described. We have shown how the spatial representation is constructed through fusion of autonomously acquired and tutor-asserted knowledge. In combination with innate conceptual knowledge, additional inferences can be drawn. We have also shown how the presence of specific context information influences the robot's behavior so that it is appropriate in a given situation.

Acknowledgements

This article presents work covered in Hendrik Zender's diploma thesis. More information can be found at `http://www.dfki.de/~zender`.

The research reported here was supported by the EU FP6 IST Cognitive Systems Integrated Project Cognitive Systems for Cognitive Assistants "CoSy" FP6-004250-IP.

References

Choset, H.; Lynch, K. M.; Hutchinson, S.; Kantor, G.; Burgard, W.; Kavraki, L. E.; and Thrun, S. 2005. *Principles of Robot Motion: Theory, Algorithms and Implementations*. Cambridge, MA, USA: The MIT Press.

Folkesson, J.; Jensfelt, P.; and Christensen, H. 2005. Vision SLAM in the measurement subspace. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation (ICRA 2005)*.

Hirtle, S. C., and Jonides, J. 1985. Evidence for hierarchies in cognitive maps. *Memory and Cognition* 13:208–217.

Krieg-Brückner, B.; Frese, U.; Lüttich, K.; Mandel, C.; Massokowski, T.; and Ross, R. 2005. Specification of an ontology for Route Graphs. In Freksa, C.; Knauff, M.; Krieg-Brückner, B.; Nebel, B.; and Barkowsky, T., eds., *Spatial Cognition IV*, volume 3343 of *Lecture Notes in Artificial Intelligence*. Heidelberg, Germany: Springer Verlag. 390–412.

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

Kruijff, G.-J. M.; Zender, H.; Jensfelt, P.; and Christensen, H. I. 2007. Situated dialogue and spatial organization: What, where…and why? *International Journal of Advanced Robotic Systems, special section on Human and Robot Interactive Communication* 4(2).

Kuipers, B. 2000. The Spatial Semantic Hierarchy. *Artificial Intelligence* 119:191–233.

Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2):91–110.

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

McNamara, T. 1986. Mental representations of spatial relations. *Cognitive Psychology* 18:87–121.

Newman, P.; Leonard, J.; Tardós, J.; and Neira, J. 2002. Explore and return: Experimental validation of real-time concurrent mapping and localization. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation (ICRA 2002)*, 1802–1809.

Schulz, D.; Burgard, W.; Fox, D.; and Cremers, A. B. 2003. People tracking with a mobile robot using sample-based joint probabilistic data association filters. *International Journal of Robotics Research* 22(2):99–116.

Topp, E., and Christensen, H. 2005. Tracking for following and passing persons. In *Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*, 70–76.