Chapter 7
Conceptual Spatial Representation of Indoor Environments*

7.1 Introduction

In the last years, there has been an increasing interest in service robots, such as domestic or elderly care robots, whose purpose is to assist people in human-like environments. These service robots have to interact with people having little or no formal training in robotics. In such situations, the communication and interaction between robots and humans become key issues for these systems.

One of the most intuitive and powerful ways for humans to communicate is spoken language. It is therefore interesting to design robots that are able to speak with people and understand their words and expressions. For this, the robot needs to perceive the world in a similar way humans do. However, when comparing the way robots typically perceive and represent the world with the findings from cognitive psychology on how humans do it, it is evident that there is a large discrepancy. Bridging the gap between human and robot spatial representations is thus of great importance.

In this chapter we give an overview of an integrated approach for creating conceptual representations of human-made environments using a mobile robot. In this representation, the concepts refer to spatial and functional properties of typical places found in indoor environments. The complete model is composed of layers containing maps at different levels of abstraction. The lower layers contain a metric map, a navigation map and a topological map, each of which plays a role in navigation and self-localization of the robot. On the topmost level of abstraction, the conceptual map provides a richer semantic view of the spatial organization.

The complete multi-layered representation is created using a combination of user-driven map acquisition process together with autonomous exploration and discovery of the environment. This process is actively supported by a linguistic framework, which allows the user to communicate with the robot using natural language only. The complete system in shown in Fig.7.1.

The rest of the chapter is organized as follows. In the next section, we present the multi-layered conceptual spatial representation. In Sect. 7.3, we describe

* This chapter originated from a joint work with Hendrik Zender and Patric Jensfelt.

Ó.M. Mozos: Semantic Labeling of Places with Mobile Robots, STAR 61, pp. 83–97.
springerlink.com

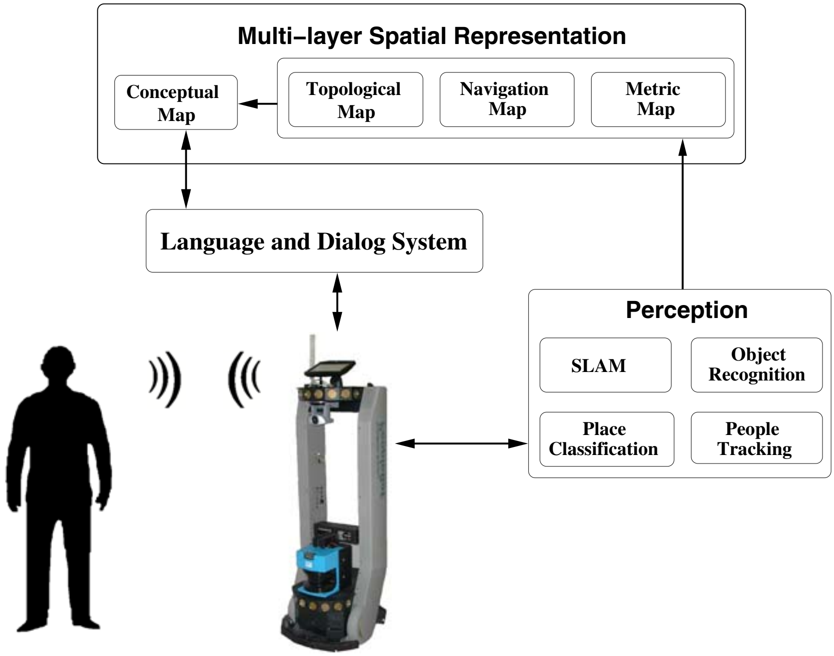


Fig. 7.1 The integrated system used for conceptual spatial representations of indoor environments. The arrows indicate the direction in which the information flows along the different subsystems. The communication between the user and the robot is done using natural language only.

implementation details of the complete system. Section 7.4 presents a demo which shows the capabilities of the service robot. We discuss related work in Sect. 7.5. Finally, we conclude in Sect. 7.6.

7.2 Multi-layered Conceptual Mapping

The goal of the multi-layered conceptual mapping is to generate spatial representations that enable a mobile robot to create models of human-made environments in a similar way as people do. The concepts represented in this model correspond to spatial and functional properties of typical indoor environments.

Our representation is based on findings in cognitive psychology [15] that assume that topological areas are the basic spatial units suitable for situated interaction between humans and robots. In addition, people usually refer to places according to the functions ascribed to them.

Taking these ideas into account, the final representation model is divided into layers, each representing a different level of abstraction. At the lowest level the system uses a laser scanner and the odometry of the mobile robot to create a metric map of the environment. On top of the metric representation, a navigation map is constructed. The navigation map is used by the mobile robot to travel along routes. In an upper layer we find the topological map, which represents the different areas found in the environment. This layer uses the detected doorways as criteria for splitting areas in the environment. On top of these layers the conceptual map contains the spatial information and the knowledge about the objects found in the environment, as well as the relations between them. This layer gathers information coming from lower maps together with information from different modalities such as proximity information, vision or dialogue, to allow symbolic reasoning and situated dialogue. Fig. 7.2 depicts the four layers of the conceptual spatial representation.

7.2.1 Metric Map

The first layer of the model contains a metric representation of the environment in an absolute frame of reference (cf. Fig. 7.2, bottom). The geometric primitives of this metric map consist of lines extracted from laser range scans. Such lines typically correspond to walls and other flat structures in the environment. The complete metric map is created by a mobile robot using simultaneous localization and mapping (SLAM) techniques. In particular, we apply the framework introduced in [5], which uses general representations for features that address symmetries and constraints in the feature coordinates to be added to the map with partial initialization. The number of dimensions for a feature can grow with time as more information is acquired. The basis for integrating the feature observations is the extended Kalman filter (EKF). An example metric map created using this method is shown in Fig. 7.3.

7.2.2 Navigation Map

The second layer contains the navigation map, which is represented by a graph (cf. Fig. 7.2). This representation is based on the notion of a roadmap of virtual free-space markers [13, 16]. In this approach, navigation nodes are inserted in the map as the robot moves through the environment. A new node is added whenever the robot has traveled a certain distance from the closest existing one. The graph serves for planning and autonomous navigation in the known part of the environment.

The navigation nodes are classified into *doorway* and *place* nodes. Doorway nodes indicate the transition between different places and represent possible doors. They are detected and added whenever the robot passes through a narrow opening. Later, the status (open/closed) of a known door can be monitored using the laser scanner. Additionally, doorway nodes are assigned information about the door opening such as its width and orientation.

Place nodes are in turn classified into two classes, namely *corridor* and *room*. The classification of place nodes is done following the approach introduced in Chap. 3. Laser observations are constantly obtained by the robot and classified into one of the two previous classes. When the robot is located at a position in the map corresponding to a place node, it assigns this node the label resulting from the classification of the current observation. To increase the robustness of this method the robot constantly

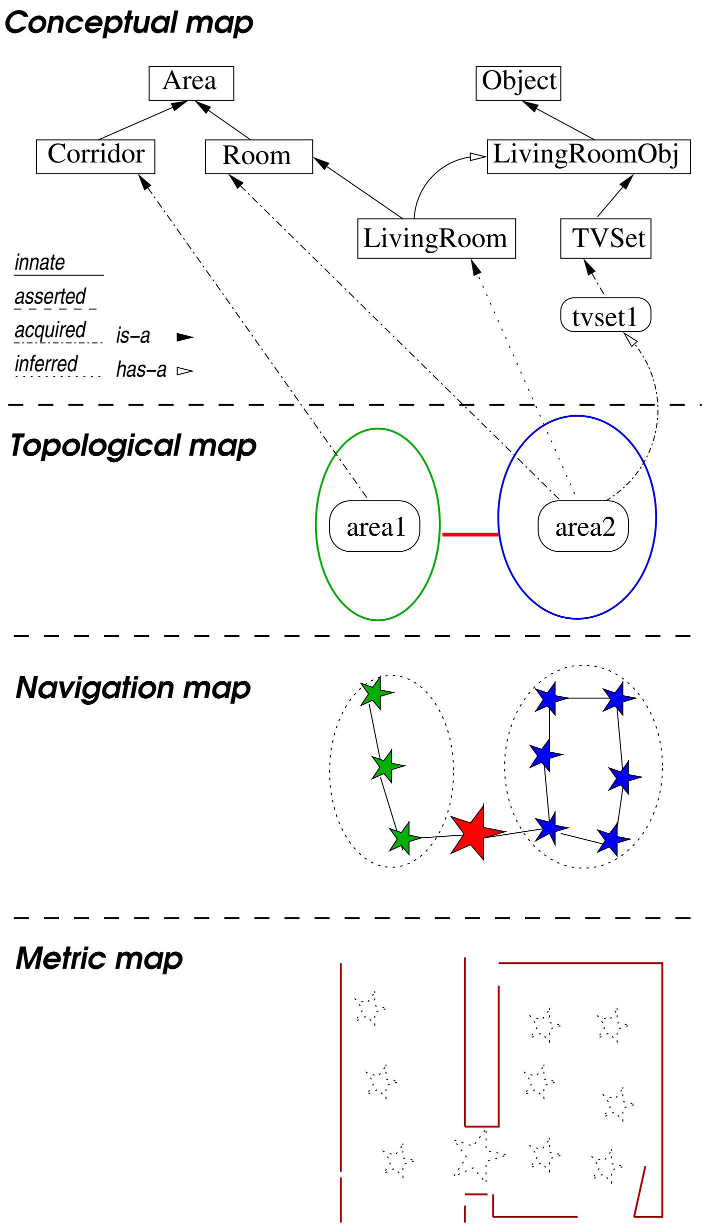


Fig. 7.2 Layered spatial representation for an indoor environment.

stores the last N classifications. Whenever a place node is reached, a majority vote is obtained with these last N labels and a final classification is assigned to the node.

The place classification of Chap. 3 is based on a 360° field of view. However the robot used in our system is equipped only with one laser covering a restricted 180°

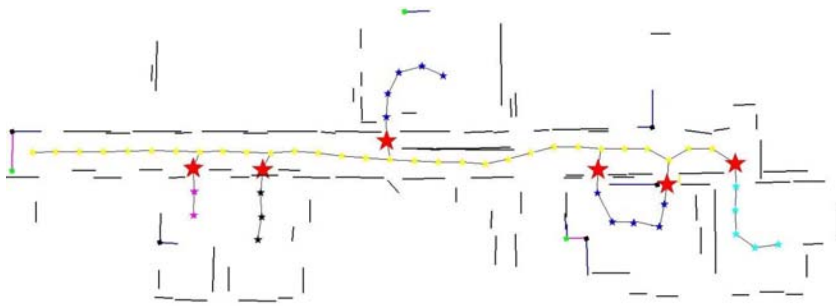


Fig. 7.3 Example of the two first layers in the spatial representation. The metric map is composed of lines. The navigation map is represented by stars. Different areas are represented by different colors/gray levels. Big stars indicate the doorways separating the places.

field of view at the front of the robot (see the robot in Fig 7.1). To solve this problem we follow the approach described in Sect. 3.5 and maintain a local map around the robot which permits us to simulate the rest of the beams covering the rear part of the robot. The classification of places and doorways forms the basis of the conceptual system.

7.2.3 Topological Map

The topological map divides the set of nodes in the navigation graph into different areas. An area consists of a set of interconnected nodes with the same place classification. The nodes are partitioned on the basis of the detection of doorways. This process is shown in Fig. 7.2.

In the topological layer, the exact shape and boundaries of an area are irrelevant. This approach complies with previous studies [8, 15], which state that humans segment space into regions that correspond to more or less clearly defined spatial areas.

Note that this method for topological map extraction is an alternative to the one presented in Chap. 4, where we applied an offline approach using simulated range data for the classification of the free poses in the map. Moreover, that method needed the complete map before extracting the topology. In contrast, the approach of this chapter is based mainly on the detection of doorways as the boundaries between different regions. Then the nodes in the different regions are labeled according to their semantic classification. This procedure is more appropriate for an online creation of the topological map during a guided tour.

7.2.4 Conceptual Map

The conceptual map provides the interface between the lower levels and the communication system, which is based in natural language. This layer contains the knowledge about the space in the indoor environment together with knowledge about the

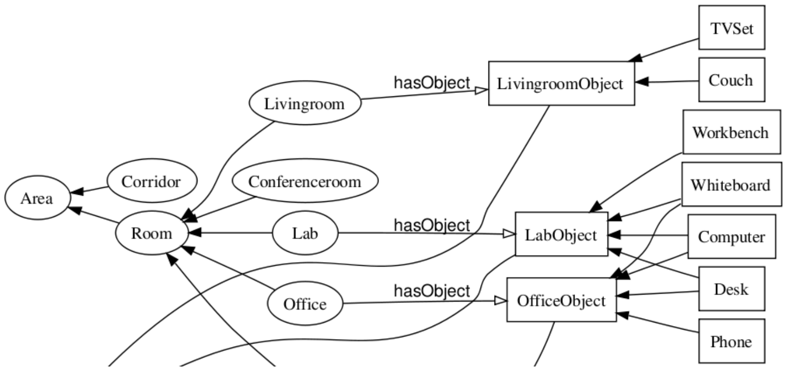


Fig. 7.4 Illustration of a part of the commonsense ontology of an indoor office environment.

entities found therein. In addition, the layer includes the relations between the different components, allowing symbolic reasoning.

Based on the work by Zender [24], the conceptual level is provided with a commonsense OWL ontology [20] of a typical office environment. Part of this ontology is shown in Fig. 7.4. The ontology describes *is-a* relations of room types. Moreover, it shows typical objects found in different places with *has-a* relations. The ontology is created by hand using a priori knowledge about typical office environments. In this ontology, the knowledge about classes and their relations are fixed during the operation of the robot. However, new instances can be added to the representation in real-time by the robot. Using this representation, a reasoner [7] can infer information about the world that is neither given verbally nor actively perceived. In this way, linguistic references to spatial areas can be generated.

The information represented in the ontology is the result of the fusion from *acquired*, *asserted*, and *innate conceptual* knowledge.

7.2.4.1 Acquired Knowledge

The acquired knowledge comprises the information about objects and places that the mobile robot is able to detect in an autonomous manner while moving around. In addition, this knowledge spreads in a bottom-up manner until reaching the conceptual layer. For example, each topological area is represented in the conceptual map as an ontological instance of the type `Area`. As soon as the area is classified as a room or corridor, the instance change its type to a more specific one such as `Room`.

In addition, whenever a new object in the environment is recognized, a new instance of the corresponding type, e.g. `Couch`, is added to the ontology. Moreover, the instance representing the object and the instance of the area where the object is located are related via the *has-a* relation. This process is shown in Fig. 7.2.

7.2.4.2 Asserted Knowledge

The process of acquiring the knowledge about the environment is carried out during a guided tour with the robot [22]. In this tour, a user shows the environment to the robot and names areas and certain objects that should be relevant to it. Typical assertions in a guided tour include "You are in the corridor," or "This is the charging station." These assertions are stored in the conceptual map, either by specifying the type of the current area or by creating a new object instance of the asserted type and linking it to the area instance with the *has-a* relation (cf. Fig. 7.2).

7.2.4.3 Innate Conceptual Knowledge

The innate conceptual knowledge is represented by an ontology that models conceptual commonsense knowledge about an indoor office environment as shown in Fig. 7.4. In the top level we found the two base concepts `Area` and `Object`. `Area` can be further divided into `Room` or `Corridor`. The basic-level subconcepts of `Room` are characterized by the instances of `Object` that are found there with the *has-a* relation. For example, a room with a TV set is represented by the subconcept `LivingRoom` (cf. Fig. 7.2).

7.2.4.4 Inferred Knowledge

Applying a reasoner software [7] to the ontology, a service robot can infer more specific categories for known areas. For example, combining the acquired information that a given topological area is classified as a room and contains a couch, together with the innate conceptual knowledge given in our commonsense ontology, it can be inferred that this area is an instance of `LivingRoom`. On the contrary, if an area is classified as a corridor and the user shows the robot a charging station in that area, the robot can not infer further categories for this area, since according to Fig. 7.2 there are no more subcategories in the original ontology.

7.2.4.5 Ambiguities

The presented conceptual model supports ambiguous classification of areas. That means that the same room can be referred to using different terms. This capability facilitates the interaction with many people simultaneously, since the way people refer to the same room can differ from situation to situation and from speaker to speaker [22]. As an example, a room described as a kitchen by one person can be seen as a recreation room by another person.

7.3 System Integration

The previous approach was implemented in a real robot as part of the explorer scenario in the CoSy project [3]. The robot acquired information about the environment

(a) Couch (b) Bookcase

Fig. 7.5 Two example objects used in the CoSy explorer scenario. (**a**) A couch. (**b**) A bookcase.

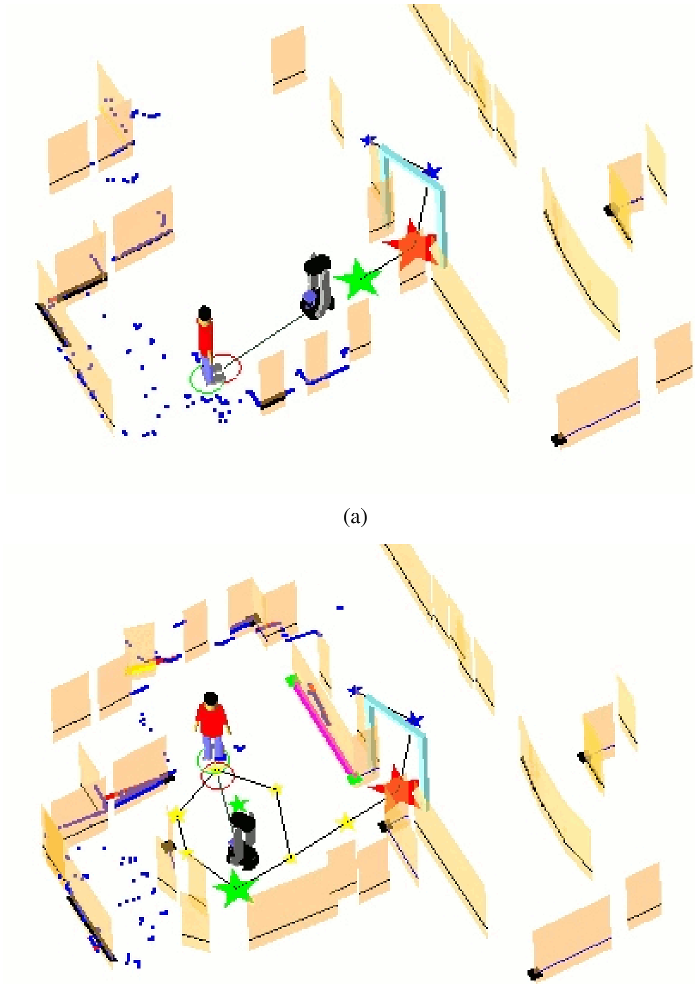
using different sensors, namely a laser range finder and a camera. This information was used for object recognition, place classification, and people tracking. All these perception components are also part of the navigation subsystem, which used the sensors for SLAM and motion planning. The complete system is shown in Fig. 7.1.

The approach was implemented and integrated in an ActivMedia PeopleBot mobile platform (robot in Fig. 7.1). The robot was equipped with a SICK laser range finder, which was used for the metric map creation, people following, and for the semantic classification of places. Additionally, a camera was used only for object detection. The detection systems used SIFT features for finding typical objects like a television set, a couch or a bookcase. We recognized instances of objects and not categories [14]. The objects should be previously shown to the robot and learned by it. Examples of objects used for recognition are shown in Fig. 7.5.

The communication with people was done using spoken language only (cf. Fig. 7.1). The user could talk to the robot using a bluetooth headset and the robot replied using a set of speakers mounted on the mobile platform. Voice commands were processed using a real time speech recognition system [17].

The information coming from the sensors was used to create a multi-layered conceptual and spatial representation of the indoor environment. Some of the information needed at the conceptual level to complete this representation was given by the user through spoken dialogues. The system additionally used a linguistic framework that actively supported the map acquisition process and was used for situated dialogue about the environment. The robot could also initiate a clarification dialogue if it detected an inconsistency in its spatial representation, illustrating the mixed-initiative capabilities of the dialogue system [10, 11].

As an additional tool, we used an online viewer for the metric and navigation maps. Examples of snapshots are shown in Fig. 7.6. The output of this program was composed of the lines extracted by our SLAM implementation extended to 3D planes to facilitate the visualization. The viewer showed the different nodes and edges used to construct the navigation map. Nodes corresponding to doorways were drawn bigger and with red color and with an associated doorframe as shown in Fig. 7.6. Finally, the robot and the user were constantly shown in the positions where



(a)

(b)

Fig. 7.6 Two snapshots of the online viewer used during the experiment. The stars indicate the nodes in the navigation map. Small and blue for corridor, small and yellow for room, big and red for doorways and medium and green for the actual position of the robot. Additionally, lines are extended to 3D planes and simulated doorways are drawn for facilitating the visualization. The person is drawn in the position detected by the people following software. (**a**) The robot enters a room after detecting a doorway. (**b**) The complete map of the room is created using lines.

they were localized. The localization of the robot was calculated using the SLAM method introduced in Sect. 7.2.1, while the pose of the person was estimated using people tracking methods based on laser readings only [18].

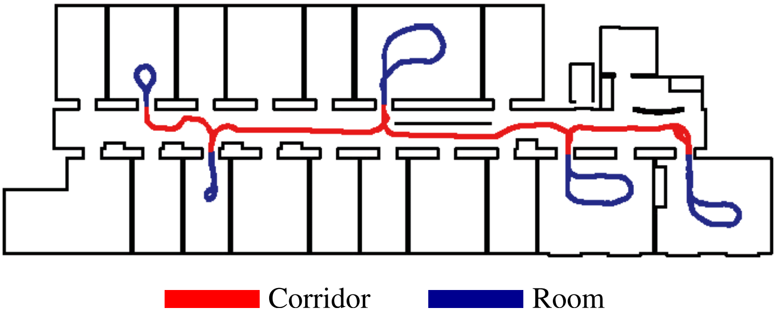


Fig. 7.7 Trajectory followed by the robot to train the classifier for distinguishing between corridor and room. The different places are depicted with distinct colors.

7.4 Demo

In order to show all the functionalities explained in the previous sections, we carried out a demo on the 7th floor of the CAS building at the Royal Institute of Technology in Stockholm. In this demo the robot, together with a user, went through different situations along the environment. The complete demo was carried out non-stop, i.e. we did not stop the robot or restart the system at any moment. The demo was thought of as a test, and for this reason we forced some artificial situations to simulate possible real ones, e.g. a false doorway. The aim of the demo was to show how the robot was able to learn its environment while interacting with a tutor. A video showing the complete demo is available at the CoSy project website [3] under the explorer scenario.

Before running the experiment, some previous knowledge was needed. First, the robot was provided with an ontology representing the general knowledge about the environment. We used the ontology (partially) depicted in Fig. 7.4. Second, the classification of places was based on previous general knowledge about the geometry of rooms and corridors in typical office environments. This knowledge was encoded in a classifier based on laser readings, as explained in Sect. 7.2.2. This classifier was trained using examples of corridors and rooms from real environments, as the one shown in Fig. 7.7. These two kinds of knowledges are independent of the environment used for testing, in the sense that the robot does not need to be physically present in the test environment to acquire the information.

Finally, the robot had to recognize different objects, such as couches or TV sets, using vision (see Fig. 7.5 for some examples). Due to the fact that we performed instance recognition rather than categorization, the objects we wanted to recognize had to be presented to the robot before running the experiment. For this purpose, we positioned the robot in front of these objects, acquired a training image and labeled it with the corresponding term. This term was then added to a small database of objects and also included in the language systems for its posterior use.

The demo started in the corridor, where the robot was positioned close to the charging station. The user activated the robot and told it that it was located at the

Fig. 7.8 The user wakes up the robot and the demos starts.

charging station. This situation is depicted in Fig. 7.8. The user then asked the robot to follow him. From this point the robot started dropping markers (navigation nodes), which were correctly classified as corridor. Then the person followed by the robot entered a room through a doorway. This doorway was recognized by the robot and the corresponding node was included in the map. From this point, the next nodes were classified as a new area and thus correctly labeled as room.

To show the utility of the clarification dialogues, we forced the robot to detect a false doorway inside a room by putting a bucket close to a table. This created an illusion of a doorway when using only the front laser as sensor. The robot passed through this false doorway and came back to a previously visited node. At this point the robot inferred that there was an inconsistency in the map, and initialized a clarification dialogue asking if there was a door there previously. The user denied this fact and the map was updated accordingly.

The inference of new subconcepts was demonstrated in the following situation. After the false doorway, and while staying inside the room, the robot was asked for the current place and it answered with the indefinite description "a room". The term `Room` was obtained from the classification of the navigation nodes belonging to the current area. A majority vote among these nodes was used in case the node classification was not unanimous. Then the robot was asked to look around. This command activated the vision-based object detection capabilities of the robot. The robot moved and detected a couch, and then a television set. After that, the user asked the robot for the name of the place. Due to the inference over the detected objects and places, the robot categorized the place as a `Livingroom`. Note that previous to the detection of objects the same place was categorized as a `Room`.

Fig. 7.9 Following the order "go to the television", the robot approaches the navigation node from where it saw the television set the last time.

Finally, we showed how the navigation map was used by the robot to come back to previously visited places. After the door opening situation, the robot was asked to go to the television set. The robot then navigated to the node where the television set was last detected (cf. Fig. 7.9). This functionality permitted the user to command the robot to places without the need of giving concrete coordinates. It is also more powerful in the sense that the user may not know the concrete name of the place, but he can remember it as 'the room with a television".

Finally, the robot was commanded to go to the charging station. Again the robot followed the navigation map until it positioned itself on the station, thus finishing the experiment.

7.5 Related Work

Research in spatial representations has yielded different multi-layered environment models. Vasudevan et al. [23] suggest a hierarchical probabilistic representation of space based on objects. The work by Galindo et al. [6] presents an approach containing two parallel hierarchies, spatial and conceptual, connected through anchoring. Inference about places is based on objects found in them. Furthermore, the Hybrid Spatial Semantic Hierarchy (HSSH) was introduced by Beeson et al. [1]. This representation allows a mobile robot to describe the world using different models each containing its own ontology. In contrast to these approaches our implementation uses human augmented mapping to collect information. Furthermore, the communication with the robot is made entirely using natural language and dialogues.

Moreover, our conceptual representation comes from the fusion of acquired, asserted, inferred, and innate knowledge.

There are more cognitively inspired approaches to describe indoor environments. These approaches do not necessarily rely on an exact global self-localization, but rather require the execution of a sequence of local behaviors. Kuipers [12] presented the Spatial Semantic Hierarchy (SSH). Alternatively, the Route Graph model was introduced by Krieg-Brückner et al. [9]. Both theories propose a cognitively inspired multi-layered representation of a map, which is at the same time suitable for robot navigation. Their central layer of abstraction is a topological representation. Our approach differs in that it provides an abstraction layer that can be used for reference resolution of topological entities.

A number of systems have been implemented that permit a robot to interact with humans in their environment. Rhino [2] and Robox [19] are robots that work as tour-guides in museums. Both robots rely on an accurate metric representation of the environment and use limited dialogue to communicate with people. The robot BIRON [21] is endowed with a system that integrates spoken dialogue and visual localization capabilities on a robotic platform. However, these systems differ from ours in the degree to which conceptual spatial knowledge and linguistic meaning are grounded in, and contribute to, situational awareness.

Acquiring a map of the environment with the help of a tutor has been considered in different works. For example, Diosi et al. [4] creates a metric map through a guided tour. The map is then segmented according to the labels given by the instructor. Finally, in the work by Topp et al. [22], a graph based model incorporates information from a user that presents the environment. In contrast to these works, the approach presented in this chapter provides the model with the autonomous acquisition by the robot of knowledge about space and entities.

7.6 Conclusion

In this chapter we presented the semantic classification of places as part of a integrated approach for creating conceptual representations of human-made environments. In this model, the concepts represent spatial and functional properties of typical office indoor environments. This representation is based on multiple maps at different levels of abstraction. The complete system was integrated and tested in a service robot which included a linguistic framework with capabilities for situated dialogue and map acquisition. The presented demo showed that the system was able to provide a high level of human-robot communication and a certain degree of social behavior.

References

1. Beeson, P., MacMahon, M., Modayil, J., Murarka, A., Kuipers, B., Stankiewicz, B.: Integrating multiple representations of spatial knowledge for mapping, navigation, and communication. In: Interaction Challenges for Intelligent Assistants, AAAI Spring Symposium, Stanford, CA, USA (2007)

2. Burgard, W., Cremers, A.B., Fox, D., Hähnel, D., Lakemeyer, G., Schulz, D., Steiner, W., Thrun, S.: Experiences with an interactive museum tour-guide robot. Artificial Intelligence 114(1-2) (2000)
3. CoSy. Cognitive systems for cognitive assistants (2004), http://www.cognitivesystems.org/
4. Diosi, A., Taylor, G., Kleeman, L.: Interactive SLAM using laser and advanced sonar. In: Proceedings of the IEEE International Conference on Robotics and Automation, Barcelona, Spain (April 2005)
5. Folkesson, J., Jensfelt, P., Christensen, H.I.: Vision SLAM in the measurement subspace. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 30–35 (2005)
6. Galindo, C., Saffiotti, A., Coradeschi, S., Buschka, P., Fernández-Madrigal, J.A., González, J.: Multi-hierarchical semantic maps for mobile robotics. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Alberta, Canada (2005)
7. Haarslev, V., Mölle, R.: Racer: A core inference engine for the semantic web. In: Proceedings of the International Workshop on Evaluation of Ontology-based Tools, Florida, USA, October 2003, pp. 27–36 (2003)
8. Hirtle, S.C., Jonides, J.: Evidence for hierarchies in cognitive maps. Memory and Cognition 13, 208–217 (1985)
9. Krieg-Brückner, B., Röfer, T., Carmesin, H.-O., Müller, R.: A taxonomy of spatial knowledge for navigation and its application to the Bremen autonomous wheelchair. In: Freksa, C., Habel, C., Wender, K.F. (eds.) Spatial Cognition 1998. LNCS (LNAI), vol. 1404, pp. 373–397. Springer, Heidelberg (1998)
10. Kruijff, G.-J.M., Zender, H., Jensfelt, P., Christensen, H.I.: Clarification dialogues in human-augmented mapping. In: Proceedings of the 1st ACM Conference on Human-Robot Interaction, Salt Lake City, UT, USA (2006)
11. Kruijff, G.-J.M., Zender, H., Jensfelt, P., Christensen, H.I.: Situated dialogue and spatial organization: What, where... and why? International Journal of Advanced Robotic Systems 4(2) (March 2007)
12. Kuipers, B.: The Spatial Semantic Hierarchy. Artificial Intelligence 119, 191–233 (2000)
13. Latombe, J.C.: Robot Motion Planning. Academic Publishers, Boston (1991)
14. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
15. McNamara, T.P.: Mental representations of spatial relations. Cognitive Psychology 18, 87–121 (1986)
16. Newman, P., Leonard, J., Tardos, J.D., Neira, J.: Explore and return: Experimental validation of real-time concurrent mapping and localization. In: Proceedings of the IEEE International Conference on Robotics and Automation, Washington, D.C., USA, pp. 1802–1809 (2002)
17. Nuance. Nuance speech recognition system developer's manual version 6.2 (1999), http://www.nuance.com/
18. Schulz, D., Burgard, W., Fox, D., Cremers, A.B.: People tracking with a mobile robot using sample-based joint probabilistic data association filters. International Journal of Robotics Research 22(2), 99–116 (2003)
19. Siegwart, R., Arras, K.O., Bouabdallah, S., Burnier, D., Froidevaux, G., Greppin, X., Jensen, B., Lorotte, A., Mayor, L., Meisser, M., Philippsen, R., Piguet, R., Ramel, G., Terrien, G., Tomatis, N.: Robox at expo.02: A large scale installation of personal robots. Robotics and Autonomous Systems 42, 203–222 (2003)
20. Smith, M.K., Welty, C., McGuinness, D.L.: OWL web ontology language guide (2004)

21. Spexard, T., Li, S., Wrede, B., Fritsch, J., Sagerer, G., Booij, O., Zivkovic, Z., Terwijn, B., Kröse, B.: BIRON, where are you? - enabling a robot to learn new places in a real home environment by integrating spoken dialog and visual localization. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (2006)
22. Topp, E.A., Huettenrauch, H., Christensen, H.I., Eklundh, K.S.: Bringing together human and robotic environment representations – a pilot study. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China (October 2006)
23. Vasudevan, S., Gächter, S., Berger, M., Siegwart, R.: Cognitive maps for mobile robots– an object based approach. In: Proceedings of the IEEE/RSJ IROS 2006 Workshop: From Sensors to Human Spatial Concepts, Beijing, China (2006)
24. Zender, H.: Learning spatial organization through situated dialogue. Master's thesis, Department of Computational Linguistics, Saarland University, Saarbruecken, Germany (2006)