Multi-modal Semantic Mapping

Andrzej Pronobis and Patric Jensfelt

***Abstract*—A cornerstone for mobile robots operating in man-made environments and interacting with humans is representing and understanding the human semantic concepts of space. In this paper, we present a multi-layered semantic mapping algorithm able to combine information about the existence of objects in the environment with knowledge about the topology and semantic properties of space such as room size, shape and general appearance. We use it to infer semantic categories of rooms and predict existence of objects and values of other spatial properties. We perform experiments on a mobile robot showing the efficiency and usefulness of our system.**

I. Introduction

In this paper we focus on the understanding of space to facilitate interaction between humans and robots and increase the efficiency of the robot performing tasks in man-made environments. We consider applications where the robot is operating in indoor environments, i.e. environments which have been made for and are, up until now, almost exclusively inhabited by humans. In such an environment human concepts such as rooms and objects and properties such as the size and shape of rooms are important, not only because of the interaction with humans but also for knowledge representation and abstraction of spatial knowledge.

The main contribution of this work is a way of combining information about the existence of objects, the appearance, geometry and topology of space for semantic mapping and room categorization in a principled manner. An important characteristic of our approach is that the lowest levels of information are decoupled from high-level room categorization by introducing the so called spatial properties. Those properties can be appearance-based e.g. general visual appearance of a room, geometry-based e.g. room size or shape obtained from laser range data, or object-based e.g. the presence of an object of a specific type. Furthermore, by incorporating information about the topology of space we can infer properties of space even without having made any observations there. For example, starting in an office the system would be able to say that it is very likely that the neighboring room is elongated and is a corridor because that is the typical topology.

The property-based architecture has several advantages. It paves the way for better scalability. It makes training of new categories easier. It permits describing space at much finer level of granularity. The properties can correspond to human concepts of space. The use of such human understandable properties provides better support for verbalization of knowledge, e.g. the corridor is large (size property) and elongated (shape property) as well as the dual, i.e. interpreting what a human says and ultimately learning models for new categorizes based on human input. Additional spatial properties such as based on actions observed in the environment can be easily incorporated. Finally, human input can be treated in the same principled way as the information from a camera or a laser scanner. That is, if a human tells us that there is a certain object nearby or that we are in the room next to the kitchen this type of information can be incorporated.

Most previously published semantic mapping approaches rely purely on object information [2, 9, 10, 5]. In [11], Zender *et al.* combines a laser-based place classification method with object recognition for semantic environment descriptions. However, in this case, the modalities are integrated in an ad-hoc way through a manually built OWL-DL ontology, and the reasoner fails to include all the uncertainty associated with place classification or object detection. In contrast, our method permits integration of multiple sources of knowledge and performs all the reasoning in a fully probabilistic fashion using automatically gathered conceptual information.

The proposed semantic mapping system is implemented on a mobile robot platform and successfully evaluated in a typical real-world office environment.

II. Semantic Spatial Concepts

We begin with an outline of some of the important spatial concepts employed in our approach. Our primary assumption is that spatial knowledge should be abstracted. This keeps the complexity under control, makes the knowledge more robust to dynamic changes, and allows to infer additional knowledge about the environment. One of the most important steps in abstraction of spatial knowledge is discretization of continuous space. In our view, the environment is decomposed into discrete areas called places. Places connect to other places using paths which are generated as the robot travels the distance between them. Thus, places and paths constitute the fundamental topological graph of the environment.

An important concept employed by humans in order to group locations is a room. Rooms tend to share similar functionality and semantics which make them a good candidate for integrating semantic knowledge over space. In the case of indoor environments, rooms are usually separated by doors or other narrow openings. Thus, we propose to use a door detector and perform reasoning about the segmentation of space into rooms based on the doorway hypotheses.

Many other concepts than simply related to the topology are being used by humans to describe space. In this work, we focus on the combination of objects, which we believe are strongly related to the semantic category of a place where they are typically located, with other spatial properties. As properties, we identify shape of a room (e.g. elongated), size of a room (e.g. large, compared to other typical rooms) as well as the general appearance of a room (e.g. office-like appearance).

III. The Conceptual Map

The key component of our semantic mapping approach is the probabilistic conceptual map. In order to fully exploit

the uncertainties provided by the multi-modal lower-level models, the map encodes an uncertain ontology and employs a probabilistic inference engine.

A. Uncertain Ontology

The ontology of spatial concepts and instances of those concepts implemented in the conceptual map is presented in Fig. 2. In order to represent the uncertainty associated with some of the relationships, we extended the standard ontology notation by annotating relations as either probabilistic or non-probabilistic. The resulting ontology defines a taxonomy of concepts through hyponym relationships (is-a) as well as relations between concepts (has-a relationships). As in [11], the ontology distinguishes three primary sources of knowledge: *predefined* (taxonomy and conceptual common-sense knowledge, e.g. the likelihood that cornflakes occur in kitchens), *acquired* (knowledge acquired using the robot's sensors), and finally *inferred* (knowledge generated internally, e.g. that the room is likely to be a kitchen, because you are likely to have observed cornflakes in it). We could further differentiate between acquired knowledge and *asserted* knowledge which can be obtained by interaction with a human.

The ontology ties the concepts to instance symbols derived from the lower level representations. The instance knowledge includes the presence of objects and sensed spatial properties such as shape, size, appearance and topology. The conceptual knowledge comprises common-sense knowledge about the occurrence of objects in rooms of different semantic categories, and the relations between these categories and the aforementioned spatial properties. In our system, the "has-a" relations for rooms, objects, shapes, sizes and appearances were acquired by analyzing common-sense knowledge available through the world wide web (for details see [6]) as well as annotations available together with the database described in this paper.

B. Probabilistic Inference

The conceptual map is implemented using a chain graph probabilistic model [4] for reasoning. Chain graphs are a natural generalization of directed (Bayesian Networks) and undirected (Markov Random Fields) graphical models. As such, they allow for modeling both "directed" causal as well as "undirected" symmetric or associative relationships, including circular dependencies.

The structure of the chain graph model is presented in Fig. 1. The structure of the model depends on the topology of the environment. Each discrete place is represented by a set of random variables connected to variables representing semantic category of a room. Moreover, the room category variables are connected by undirected links to one another according to the topology of the environment. The potential functions $\phi_{rc}(\cdot,\cdot)$ represent the type knowledge about the connectivity of rooms of certain semantic categories.

The remaining variables represent shape, size and appearance properties of space and presence of a certain number of instances of objects as observed from each place. These can be

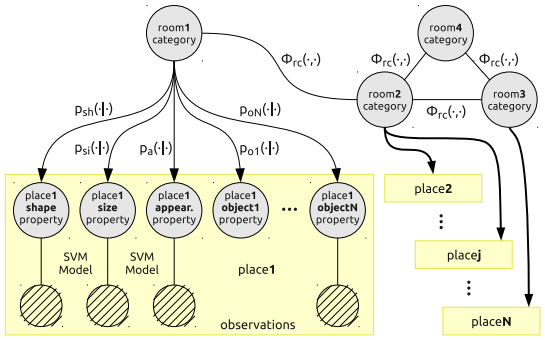


Fig. 1. Structure of the chain graph model compiled from the conceptual map. The vertices represent random variables. The edges represent the directed and undirected probabilistic relationships between the random variables. The textured vertices indicate observations that correspond to sensed evidence.

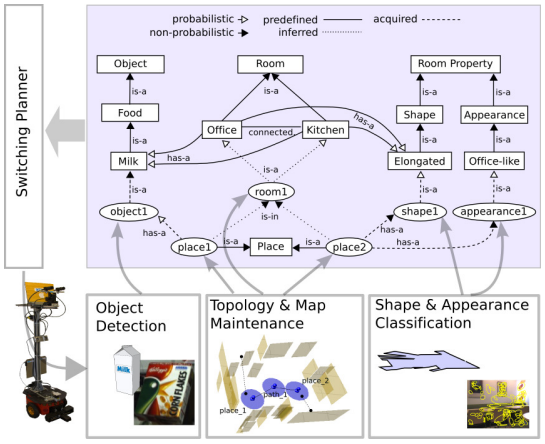


Fig. 2. View of the processes and representations of the system. Sensing processes (at the bottom) discretize and categorize sensor input into instances (shown as ellipses) and acquired relations. The conceptual map also comprises knowledge about concepts (rectangles) of which only an excerpt in shown.

connected to observations of features extracted directly from the sensory input and quantified by the categorical models of sensory information. Finally, the functions $p_{sh}(\cdot|\cdot)$, $p_{si}(\cdot|\cdot)$, $p_a(\cdot|\cdot)$, $p_{o_i}(\cdot|\cdot)$ utilize the common sense knowledge about object, spatial property and room category co-occurrence to allow for reasoning about other properties and room categories. The conditional probability distributions $p_{o_i}(\cdot|\cdot)$ are represented by Poisson distributions. The parameter λ of the distribution allows to set the expected number of object occurrences. In our experiments the parameter was calculated to match the probability of there being no objects of a certain category in a room of a certain category as provided by the common sense knowledge databases.

IV. System Overview

We now explain the other processes in the semantic mapping system the provide input to the conceptual map as shown in Fig. 2. First, mapping and topology maintenance processes create a topological place map. A SLAM algorithm [1] builds a metric map of the environment. The metric map is further discretized into places distributed spatially in the metric map. The places together with paths obtained by traversing from one place to another constitute a topological graph. Then, based on

Fig. 3. Examples of images from the COLD-Stockholm database acquired in 9 different rooms. A video illustrating the acquisition process is available on the website of the database.

the information about the connectivity of places and the output of a template-based laser door detector, a process forms rooms by clustering places that are transitively interconnected without passing a doorway. Since the door detection algorithm can produce false positives and false negatives, room formation must be a non-monotonic process to allow for knowledge revision. It is handled by a general purpose rule engine able to make non-monotonic inferences in its symbolic knowledge. The approach is an adaptation of the one by [3].

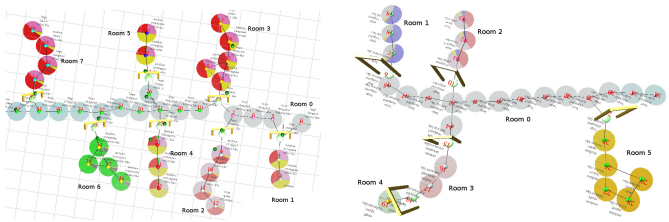
Geometry and appearance classification is based on categorical place models [8] and provides information about the shape, size and general appearance of rooms. The categorical models are provided with sensory information from the laser scanner and a camera. This information is classified and confidence estimates are provided indicating the similarity of the sensory input to each of the categorical models. The estimated confidence information is then accumulated over each of the viewpoints observed by the robot while being in a certain place [8] and further normalized to form potentials. Similar independent process performs object detection and recognition based on visual object models [7]. The results are fed back into the chain graph triggering an inference in the probabilistic model. Accordingly, room categorization is performed as a result of the reasoning process in the conceptual map.

V. Experiments

A. Experimental Scenario

All the categorical models used in the experiments were trained on the COLD-Stockholm database (http://www.cas.kth.se/cold-stockholm). The database consists of multiple sequences of image, laser range and odometry data. The acquisition was performed on four different floors (4th to 7th) of an office environment, consisting of 47 areas (usually corresponding to separate rooms) belonging to 15 different semantic and functional categories and under several different illumination settings (cloudy weather, sunny weather and at night). Examples of images from the COLD-Stockholm database are shown in Fig. 3.

In order to guarantee that the system will never be tested in the same environment in which it was trained, we have divided



Part A Part B

Fig. 4. Topological maps of the environment anchored to a metric map indicating the outcomes of room segmentation and categorization. The circles indicate the location of places in the environment and the colors indicate the inferred room categories. For the detailed information about the inferred categories, see Fig. 5.

the COLD-Stockholm database into two subsets. For training and validation, we used the data acquired on floors 4, 5 and 7. The data acquired on floor 6 were used for testing.

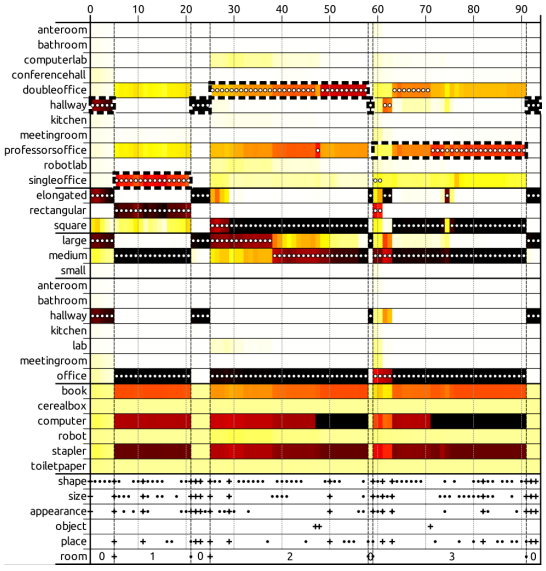
For the purpose of the experiments presented in this paper, we have extended the annotation of the COLD-Stockholm database to include 3 room shapes, 3 room sizes as well as 7 general appearances. The room size and shape, were decided based on the length ratio and maximum length of edges of a rectangle fitted to the room outline. These properties together with 6 object types defined 11 room categories used in our experiments. The values of the properties as well as the room categories are listed in Fig. 5.

B. Experimental Results

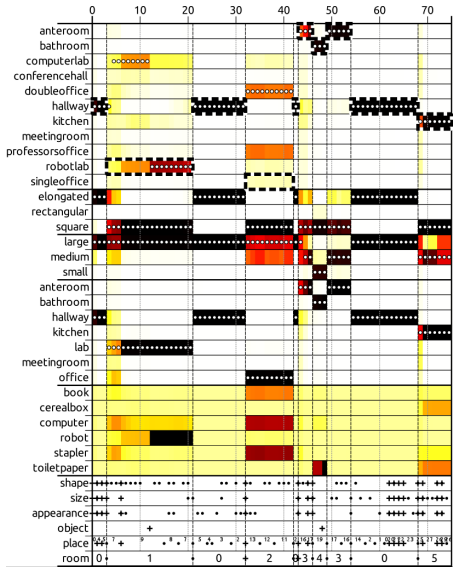
The models trained on the COLD database were used in the semantic mapping system. The experiments were performed on the 6th floor of the building, i.e. in the part which was not used for training. The robot was manually driven through 15 different rooms while performing real-time semantic mapping without relying on any previous observations of the environment. The obtained maps of parts of the environment (A and B) are presented in Fig. 4.

The robot recorded beliefs about the shapes, sizes, appearances, objects found and the room categories for every significant change event in the conceptual map. The results for the two parts of the environment are presented in Fig. 5. Each column in the plot corresponds to a single event and the source of that event is indicated using dots (changes) and crosses (additions) at the bottom. At certain points in time, the robot was provided with asserted human knowledge about the presence of objects in the environment.

By analyzing the events and beliefs for part A, we see that the system correctly identified the first two rooms as a hallway and a single office using purely shape, size and general appearance (there are no object related events for those rooms). The next room was properly classified as a double office, and that belief was further enhanced by the presence of two computers. The next room was initially identified as a double office until the robot was given information that there is a single computer in this room. This was an indication that the room is a single person office that due to its dimensions is likely to belong to a professor.



Part A



Part B

Fig. 5. Visualization of the events registered by the system during exploration and its beliefs about the categories of the rooms as well as the values of the properties. The room category ground truth is marked with thick dashed lines while the MAP value is indicated with white dots. The colors indicate the strength of the beliefs after each event (the darker the stronger). Source of each event is indicated at the bottom of the plot (e.g. sensed appearance or detected object etc.). A video showcasing the system is available at: http://www.pronobis.pro/research/semantic-mapping.

Looking at part B, we see that the system identified most of the room categories correctly with the exception of a single office which due to a misclassification of size was incorrectly recognized as a double office. The experiment proved that the system can deliver an almost perfect performance by integrating multiple sources of semantic information.

VI. Conclusions

In this paper we have presented a probabilistic framework combining heterogenous, uncertain, information such as object observations, the shape, size and appearance of rooms for semantic mapping. A graphical model, more specifically a chain-graph, is used to represented the semantic information and perform the inference over it. We introduced the concept of properties between the low level sensory data and the high level concepts such as room categorizes. The properties allowed us to decouple the learning processes at the different levels and describe space at much finer level of granularity. By making the properties understandable to humans, possibilities open in terms of spatial knowledge verbalization and interpretation of human input.

Acknowledgement

This work was supported by the SSF through its Centre for Autonomous Systems (CAS) and the EU FP7 project CogX.

References

[1] J. Folkesson, P. Jensfelt, and H. I. Christensen. The m-space feature representation for SLAM. *IEEE Trans. Robotics*, 23(5):1024–1035, October 2007.

[2] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J. A. Fernandez-Madrigal, and J. Gonzalez. Multi-hierarchical semantic maps for mobile robotics. In *Proc. of IROS'05*.

[3] N. Hawes, M. Hanheide, J. Hargreaves, B. Page, and H. Zender. Home alone: Autonomous extension and correction of spatial representations. In *Proc of ICRA'11*.

[4] S. L. Lauritzen and T. S. Richardson. Chain graph models and their causal interpretations. *J. Roy. Statistical Society, Series B*, 64(3):321–348, 2002.

[5] D. Meger, M. Muja, S. Helmer, A. Gupta, C. Gamroth, T. Hoffman, M. Baumann, T. Southey, P. Fazli, W. Wohlkinger, P. Viswanathan, J. Little, D. Lowe, and J. Orwell. Curious George: An Integrated Visual Search Platform. In *Proc. of CRV'10*, 2010.

[6] M.Hanheide, C.Gretton, R. Dearden, N. Hawes, J. Wyatt, A.Pronobis, A. Aydemir, M. Göbelbecker, and H. Zender. Exploiting probabilistic knowledge under uncertain sensing for efficient robot behaviour. In *Proc. of IJCAI'11*.

[7] T. Mörwald, J. Prankl, A. Richtsfeld, M. Zillich, and M. Vincze. BLORT - The blocks world robotic vision toolbox. In *Proc of ICRA BRICS Workshop*, 2010.

[8] A. Pronobis, O. Martinez Mozos, B. Caputo, and P. Jensfelt. Multi-modal semantic place classification. *IJRR*, 29 (2-3), February 2010.

[9] S. Vasudevan and R. Siegwart. Bayesian space conceptualization and place classification for semantic maps in mobile robotics. *RAS*, 56:522–537, 2008.

[10] P. Viswanathan, D. Meger, T. Southey, J. Little, and A. Mackworth. Automated spatial-semantic modeling with applications to place labeling and informed search. In *Proc of CRV'09*.

[11] H. Zender, O. M. Mozos, P. Jensfelt, G.-J. M. Kruijff, and W. Burgard. Conceptual spatial representations for indoor mobile robots. *RAS*, 56(6):493–502, 2008.