2012 IEEE International Conference on Robotics and Automation
RiverCentre, Saint Paul, Minnesota, USA
May 14-18, 2012

Large-scale Semantic Mapping and Reasoning with Heterogeneous Modalities

Andrzej Pronobis and Patric Jensfelt
{pronobis, patric}@kth.se

Abstract— **This paper presents a probabilistic framework combining heterogeneous, uncertain, information such as object observations, shape, size, appearance of rooms and human input for semantic mapping. It abstracts multi-modal sensory information and integrates it with conceptual common-sense knowledge in a fully probabilistic fashion. It relies on the concept of spatial properties which make the semantic map more descriptive, and the system more scalable and better adapted for human interaction. A probabilistic graphical model, a chain-graph, is used to represent the conceptual information and perform spatial reasoning. Experimental results from online system tests in a large unstructured office environment highlight the system's ability to infer semantic room categories, predict existence of objects and values of other spatial properties as well as reason about unexplored space.**

I. INTRODUCTION

In this paper we deal with the problem of modeling space in order to understand it, reason about it and be able to act efficiently in it. We consider applications where the robot is operating in indoor office or domestic 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, objects and properties such as the size and shape of rooms are important, not only because of the interaction with humans but also for generating efficient robot behavior, knowledge representation and abstraction of spatial knowledge. This is what we mean by semantic mapping. The semantic mapping system we present will be used in the context of a mobile robot (see Fig. 1) but most of the system would remain unchanged if for example used as part of a wearable device.

This paper builds on our previous work [7], [16] and now focuses on semantic mapping presenting a complete semantic mapping system with several contributions also at a component level. The system makes use of multi-modal sensory information, including information gathered from humans where humans are attributed a ”sensor model” just like other sensors. It supports inference about unexplored concepts (e.g. objects, rooms) and allows for goal oriented exploration using a distribution of possible extensions to the known world. We present an extensive experimental evaluation, both offline and online where the whole system runs in real-time on an entire office floor.

A unique feature of our system is the ability to extract semantic information from multiple heterogeneous modalities and integrate it in a principled manner with conceptual common-sense knowledge in a fully probabilistic fashion. The system combines information about the existence of objects, landmarks, the appearance, geometry and topology of space as well as human asserted input. This is possible thanks to an architecture based on semantic properties of spatial entities. The properties correspond to human concepts of space and permit creation of a more descriptive spatial representation in which all entities have attributes as shown in Fig. 1 (e.g. large, square double office with multiple books).

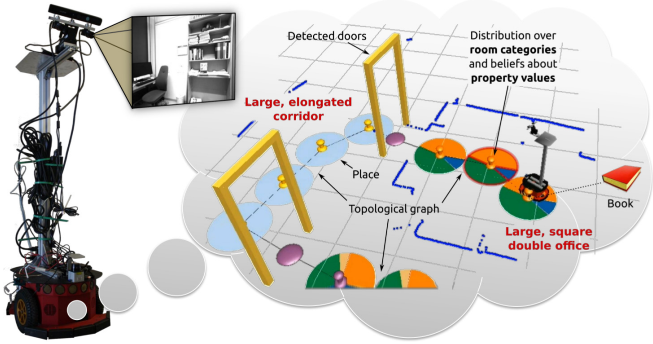

Fig. 1. Our robot platform and an illustration of a semantic map.

The presented approach is evaluated offline on a new comprehensive database, COLD-Stockholm, capturing appearance and geometry of almost 50 rooms belonging to different semantic categories as well as online in the same environment on a mobile robot. A video illustrating the system in action is available online at:

`http://www.semantic-maps.org`

The remaining sections first relate this work to other approaches in the literature and then discuss the problem of spatial understanding and present our framework from the representational and systems point of view. This is followed by a detailed presentation of our conceptual mapping and reasoning component and experimental evaluation.

II. RELATED WORK

The semantic mapping problem has only recently received significant attention. There exists a broad literature on mobile robot localization, mapping, navigation and place classification [3], [4], [20], [23], [19], [17]. Every such algorithm maintains a representation of space and performs spatial reasoning. However, this representation is usually specific to the particular problem and only captures a fraction of the broad spectrum of spatial knowledge. Other, more general frameworks, such as the Spatial Semantic Hierarchy [9] concentrate on lower levels of spatial knowledge abstraction

This work was supported by the SSF through its Centre for Autonomous Systems (CAS) and the EU FP7 project CogX. The help by Alper Aydemir, Moritz Göbelbecker and Kristoffer Sjöö is also gratefully acknowledged.

Authors are with Centre for Autonomous Systems, KTH Royal Institute of Technology, Stockholm, Sweden.

978-1-4673-1404-6/12/$31.00 ©2012 IEEE

	Place appearance	Place geometry	Object information	Topology	Human input	Segmen-tation	Conceptual map	Uncertain concepts	Inferring properties	Concepts acquired
[6]			✓			✓	✓		✓	
[25]		✓	✓		✓	✓	✓		✓	
[21]			✓			✓		✓		✓
[11]			✓							
[22]			✓			✓		✓	✓	✓
[15]		✓	✓							
[14]					✓	✓				
This work	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

TABLE I

PROPERTIES OF VARIOUS SEMANTIC MAPPING APPROACHES

and do not support higher-level conceptualization or representation of categorical information.

Table I compares properties of various semantic mapping approaches for indoor environments. None of the listed methods uses topology of the environment or general appearance of places as a source of semantic information. This is surprising given the large body of work on appearance-based place categorization [20], [23], [19], [17]. Two methods, [25] and [15] make use of geometric place information extracted from laser range sensors, and only [25] applies a previously developed place classification technique for this purpose. In [25], semantic cues can be obtained by a situated dialogue with a user and [14] build maps augmented with semantic symbols purely from human input. Almost every method is focused primarily on using objects for extracting spatial semantics [6], [25], [21], [11], [22], [15]. Objects clearly carry a lot of semantic information; however, they are also sparse and reliable object categorization in real-world environments is still a major open challenge. At the same time, valuable semantic cues are also encoded in geometry, general appearance and topology. The inability to fuse together all the sources of information is likely a result of the different character of the different inputs. In this work, we present a system able to combine all the aforementioned sources of semantic information: general appearance and geometry of places, object information, topological structure and human input.

The conceptual map in our system is also a unique feature. The most comprehensive related representations has been proposed in [6] and [25]. Both approaches encode an ontology of an indoor environment. However, those ontologies are built manually and use traditional AI reasoning techniques which are unable to incorporate uncertainty that is inherently connected with semantic information obtained through robot sensors in realistic environments. In contrast, we implement a probabilistic ontology and a probabilistic inference engine incorporating uncertainty in definitions of concepts and their links to instances of spatial entities. Moreover, the values of all properties for which direct evidence is not available can be inferred based on all the available semantic information. Additionally, as in case of [21] and [22] the concept definitions are acquired automatically from online databases and floor plans obtained from robotics datasets. Finally, we have shown [7], [1] that our approach can be combined with general planning components and is suitable for generating active robot behavior in a similar fashion to [22].

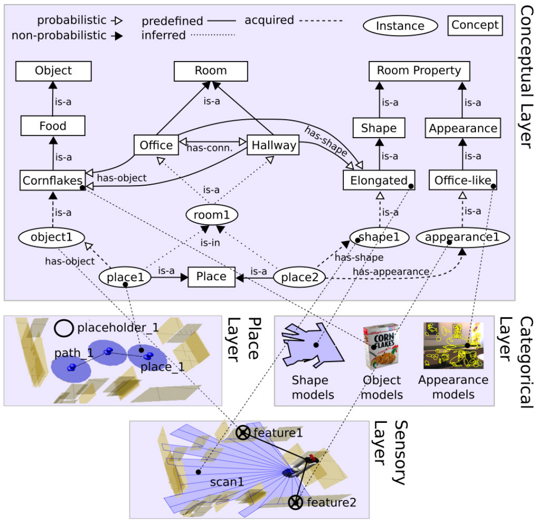

Fig. 2. The layered structure of the spatial representation and a visualization of an excerpt of the ontology of the conceptual layer. The conceptual layer comprises knowledge about concepts (rectangles), relations between those concepts and instances of spatial entities (ellipses).

III. SEMANTIC SPATIAL UNDERSTANDING

The functionality of our system is centred around the representation of complex, cross-modal, spatial knowledge that is inherently uncertain and dynamic. The representation employed here follows the principles presented in [18].

The primary assumption in our approach is that spatial knowledge should be abstracted to keep the representations compact, make knowledge more robust to dynamic changes, and allow the robot to infer additional knowledge about the environment based on combining background knowledge with observations. As one example of abstraction, we discretize the continuous space into discrete areas called *places*. Places connect to other places by *paths* which are generated as the robot travels between them forming a topological map. Hypothesized places, referred to as *placeholders*, are generated in the unexplored parts of space close to areas visited by the robot. This permits reasoning about unknown space [24]. An important concept employed by humans in order to group locations is a *room*. Rooms tend to share similar functionality and semantics and are typically assigned semantic categorical labels e.g. a double office. This make them appropriate units for knowledge integration over space.

A. Spatial Knowledge Representation

The structure of the spatial knowledge representation is presented in Fig. 2. The framework comprises four layers, each focusing on a different level of knowledge abstraction, from low-level sensory input to high-level conceptual symbols.

The lowest level of our representation is the sensory layer which maintains an accurate representation of the robot's environment corresponding to a metric map in our system. Above, the place layer contains the place, paths and placeholders. The categorical layer comprises universal categorical models (in our case static). These models describe objects and landmarks, as well as spatial properties such as a geometrical models of room shape or a visual models of appearance. On top is the conceptual layer, which is the primary focus of this paper. It is populated by instances of spatial concepts and creates a unified representation relating sensed instance knowledge from lower-level layers to general common-sense conceptual knowledge. Moreover, it includes a taxonomy of human-compatible spatial concepts. It is the conceptual layer which would contain the information that kitchens commonly contain cereal boxes and have a certain appearance and allows the robot to infer that the cornflakes box in front of the robot makes it more likely that the current room is a kitchen.

B. Conceptual Knowledge Representation

A visualization of the data representation of the conceptual layer is shown in Fig. 2. This representation is *relational*, describing common-sense knowledge as relations between concepts (e.g. kitchen has-object cornflakes), and describing instance knowledge as relations between either instances and concepts (e.g. object1 is-a cornflakes), or instances and other instances (e.g. place1 has-object object1). Relations in the conceptual map are either *predefined*, *acquired*, or *inferred*, and can either be deterministic or probabilistic. Probabilistic relations allow the expression of statistical dependencies and uncertainty as in the case of the “kitchen has-object cornflakes” or “room1 is-a hallway” relations which holds only with a certain probability. An acquired relation is one that is grounded in observations and generated as a result of a perceptual process. Predefined relations are given (and quantified in the case they are probabilistic) as part of a fixed ontology of common-sense knowledge. Inferred relations are the result of inference processes operating solely on the conceptual map.

The representation defines a taxonomy of concepts and associations between instances and concepts using hyponym relationships (is-a). Then, directed relations (has-a) are used to describe properties of room categories in terms of spatial properties, such as shape, size or appearance, and objects. Finally, we use undirected associative relations to represent connectivity between rooms.

IV. SEMANTIC MAPPING

A. Property-based Semantic Mapping

An important paradigm underpinning the design of our semantic mapping approach is the use of *properties of space*. Properties can be seen as attributes characterizing discrete spatial entities identified by the robot, such as places or placeholders. Additionally, properties can correspond to human concepts and thus provide another layer of spatial semantics shared between the robot and the user. The values of properties can be inferred from observations and other properties. Properties result from interpreting specific sensory information directly. They are modality specific and each property is connected to a model of sensory information. Higher level concepts, such as room categories, are defined based on the properties. As a result, to the conceptual reasoning, properties serve as connections between higher level concepts and low-level observations. Moreover, they permit building more specialized concepts that would be difficult to infer from uni-modal observations. The idea of using an intermediate level of properties in a feed-forward manner for place categorization has been evaluated previously as a proof of concept [16]. In this work, we generalize beyond a pure feed-forward strategy, so that both properties and room categories influence each other and provide a much more complete representation of space. Hence, we can define the problem of semantic mapping as that of estimating the joint probability distribution over categorical room labels and all values of properties of space for all places.

The current implementation of our system utilizes several types of properties assigned to places:

- *objects* - each object class results in one property associated with a place encoding the expected/observed number of such objects at a certain place
- *doorway* - determines if a place is located in a doorway
- *shape* - geometrical shape of a place extracted from laser data (e.g. elongated, square)
- *size* - size of a place extracted from laser data (e.g. large (compared to other typical rooms))
- *appearance* (e.g. office-like appearance) - visual appearance of a place

In addition to the properties of places, placeholders also have:

- *associated space* - the amount of visible free space around the placeholder not yet assigned to any place

For details about estimation of the placeholder property values, see [24]. We maintain a probability distribution over the property values in the system.

The property-based architecture has several advantages. First, it provides fine-grained and more descriptive representation of space. This can enhance the quality of human-robot interaction, increasing the robot’s ability to understand referring expressions and acquire spatial knowledge directly from humans as well as human’s understanding of the robot’s internal spatial knowledge. The additional semantic knowledge can also be used for generating a more efficient robot’s behavior, for example on the task of finding objects in large-scale environments [7], [1].

The approach has many of the advantages of high-level sensor fusion which was shown to outperform low-level feature integration for several problems (see [17] and references therein). It allows for integration of heterogeneous modalities and various types of models adapted to the characteristics of each modality (e.g. robust kernel-based discriminative models for high-dimensional data and probabilistic generative models for data of lower dimensionality or conceptual knowledge). Finally, it enhances the scalability of the approach in several ways. Instead of having to build a model from the level of sensor data for every new category, we can reuse the low level models. This saves memory (models of visual data can be hundreds of megabytes in size) and saves computations (calculations shared across categories).

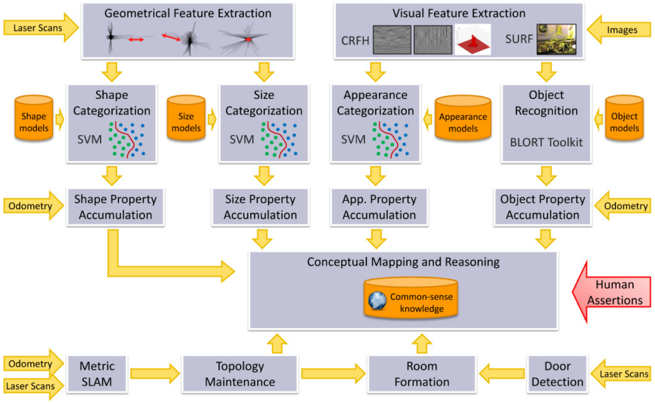

Fig. 3. Structure of the system and data flow between its main components.

The introduction of properties also facilitates training. Once models associated with properties are trained, training the system for a new category is decoupled from low-level sensor data.

B. *The Semantic Mapping System*

A visualization of the system components and data flow is presented in Fig. 3 and follows the principles outlined above. The layered structure of the spatial knowledge representation as well as the property-based architecture naturally permit the existence of data driven processes that abstract and integrate knowledge. In order to make those processes tractable, the updates of more abstract representations is performed only if a discrete value changes or a continuous values changes above a certain threshold (selected manually).

First, mapping and topology maintenance processes create the topological graph of places, paths and placeholders. A SLAM algorithm [5] builds a metric map of the environment. In our implementation the places are spread out over space like bread crumbs every one meter [25]. Unexplored space is covered with placeholders indicating location of potential places that can be discovered through exploration [24]. This approach to space discretization is limited and requires maintaining a global metric map of the environment. Vision-based topological mapping algorithms such as [4] could be used instead.

In the case of indoor environments, rooms are usually separated by doors or other narrow openings. Thus, we currently use the doorway place property in order to form rooms. A simple, template-based door detector [8] operates on laser range data and the doorway property of a place is set depending on whether the place is located inside a doorway. Then, based on the information about the connectivity of places and the doorway property value, 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 is using non-monotonic inference as described in [25]. We intend to involve all properties of space for room segmentation in the future.

The categorical sensory models of properties are continuously classifying the robot's observations obtained from the laser range finder and a camera. The estimated classification confidence information for each property value is then accumulated over each of the viewpoints observed by the robot while being in a certain place using a spatio-temporal accumulation algorithm presented in [17] and further normalized to form probabilities. The outcomes are then compared to previous observations in order to detect significant changes and fed into the conceptual mapping and reasoning component where they trigger probabilistic inference. If available, human asserted knowledge is provided to the conceptual mapping component where it is combined with the property values.

The resulting system operates in real-time on a standard laptop and is capable of semantic mapping of large scale environments. Since the probabilistic conceptual inference is computationally very efficient, it requires only a small fraction of the computational power. The system scales well not only with the number of room categories, but also with the size of the environment. The system dynamically segments space and integrates knowledge over time, space and multiple information sources. The next sections provide details about the sensory models as well as the the conceptual mapping component.

V. SENSORY MODELS OF PROPERTIES

To extract the semantic properties of spatial entities, the system employs a set of categorical models of sensory information. These models are implemented according to established object and scene modeling approaches.

a) Geometrical Property Models: Two independent models of shape and size properties are built. In both cases we use a set of simple geometrical features extracted from laser scans, as proposed in [17]. To provide sufficient robustness and tractability in the presence of noisy, high-dimensional information, we use kernel-based discriminative classifiers, namely Support Vector Machines (SVM) (see [17] for details). The models are trained from sequences of laser scans recorded in multiple instances of rooms of different shape and size. By including several different room instances into training, the acquired model can generalize sufficiently to provide categorization rather than instance recognition. We identified 3 room shapes (elongated, rectangular and square) as well as 3 room sizes (small, medium and large).

b) Appearance Property Models: We built two different models of general visual appearances of places, one for global and one for local image representation. The former was built from the Composed Receptive Field Histograms (CRFH) [17] calculated over the whole image, while the latter from local SURF features quantized into visual words [2]. The outputs of the two models were further integrated using the Generalized Discriminative Accumulation Scheme (G-DAS [17]). The models were trained on image sequences acquired in rooms belonging to various categories under different illumination conditions in order to generalize to new environments. We identified 7 different appearances: anteroom-like, bathroom-like, hallway-like, kitchen-like, lab-like, meetingroom-like, office-like.

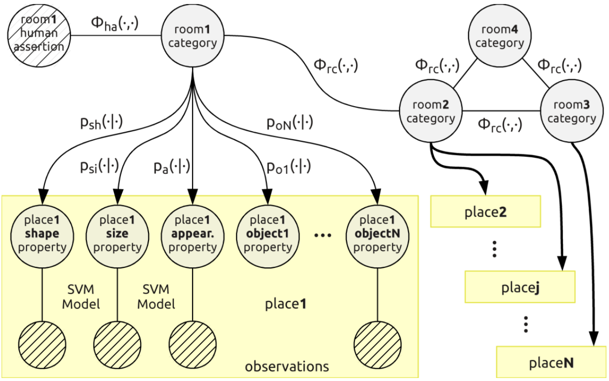

Fig. 4. Structure of the chain graph model of the conceptual map. The vertices represent random variables. The edges represent the directed and undirected probabilistic relationships between the random variables. The textured vertices indicate observations that correspond to sensed evidence.

c) Object Models: To model objects, we used the approach taken from the BLORT toolkit [13] based on SIFT recognition. We trained 6 object instance models for objects belonging to categories typically find in office environments: a book, a cereal box, a computer, a robot, a stapler, and a roll of toilet paper.

VI. PROBABILISTIC CONCEPTUAL MAPPING AND REASONING

To fully exploit the uncertainties provided by the sensory models of properties and permit uncertain spatial reasoning, the conceptual map is represented as a probabilistic *chain graph model* [10]. The structure is adapted at runtime according to the state of the underlying topological map. This is a unique feature of our approach compared to other semantic mapping systems (see Section II).

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 originating from possible loops in the topological graph. In order to perform inference on the chain graph, we first convert it into a factor graph representation and apply an approximate inference engine, namely Loopy Belief Propagation [12], to comply with time constraints imposed by the robotic applications.

A. Conceptual Map

The structure of the chain graph for the conceptual map is presented in Figure 4. Each discrete place instance is represented by a set of random variables, one for each property linked to that place. These are each connected to a random variable for the room category, representing the "is-a" relation between rooms and their categories in Figure 2. Moreover, the room category variables are connected by undirected links to one another according to the topological map. The doorway places are seen as transition areas between rooms and are not represented in the conceptual map. The potential functions $\phi_{rc}(\cdot,\cdot)$ describe knowledge about typical connectivity of rooms of certain categories (e.g. that kitchens are more likely to be connected to corridors than to other kitchens).

The remaining variables represent shape, size and appearance properties of space and the presence of objects. These are connected to observations of features extracted directly from the sensory input. These links are quantified by the categorical models of sensory information. Finally, the distributions $p_{sh}(\cdot|\cdot)$, $p_{si}(\cdot|\cdot)$, $p_a(\cdot|\cdot)$, $p_{o_i}(\cdot|\cdot)$ represent the common sense knowledge about shape, size, appearance, and object co-occurrence, respectively. It is assumed that the same object is never represented twice in the conceptual map and data association between object observations is performed while maintaining the sensory layer.

If human asserted input about room categories or other properties of the system is available, it can be seamlessly integrated with the other sources of information. Human assertions about semantic room categories are included by adding a new variable representing an observation of the human assertion and a potential $\phi_{ha}(\cdot,\cdot)$ representing the relation between the assertion and the room category. Identical procedure can be applied if the asserted knowledge is available about some other property of space, e.g. presence of an object.

B. Representing and Quantifying Relations

In our system, the "has-a" relations for room connectivity, shapes, sizes and appearances represented by the potential $\phi_{rc}(\cdot,\cdot)$ and distributions $p_{sh}(\cdot|\cdot)$, $p_{si}(\cdot|\cdot)$, $p_a(\cdot|\cdot)$, $p_{o_i}(\cdot|\cdot)$ were acquired by analyzing annotations in the database used in this paper. Co-occurrences between room categories of neighboring rooms as well as room categories and property values were counted and later normalized to form distributions.

The conditional probability distributions $p_{o_i}(\cdot|\cdot)$ are represented by Poisson distributions. The Poisson distribution was selected in order to easily model the expected number of object occurrences through its parameter λ as well as the ability to estimate λ from the probability of object existence obtained from common-sense knowledge databases. The probability of existence of an object of a certain category in a certain type of room was first bootstrapped using a part of the *Open Mind Indoor Common Sense* database[1]. Obtained object-location pairs were then used to generate '*obj* in the *loc*' queries to an online image search engine. The number of returned hits were used to obtain the probability value. More details about this approach can be found in [7]

The relations between human assertions and concepts (e.g. $\phi_{ha}(\cdot,\cdot)$) can be used to represent the uncertainty in perception of the human statements as well as a dependency between various assertions and concept values (e.g. both "kitchenette" and "kitchen" might be used to refer to a kitchen). In our system, we assign the potential value 0.8 when the assertion exactly matches the room category and we distribute the potential 0.2 evenly across all the remaining assertions.

[1] `http://openmind.hri-us.com/`

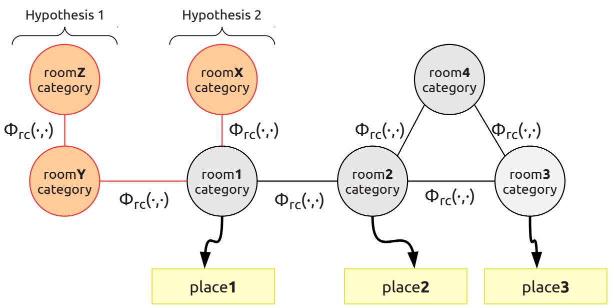

Fig. 5. Examples of extensions of the conceptual map permitting reasoning about unexplored space behind placeholder located in room 1.

C. Reasoning about Unexplored Space

The primary benefits of having a probabilistic relational conceptual representation is its capability to perform uncertain inference about some concepts based solely on their relations to other concepts rather than direct observations. This permits spatial reasoning about unexplored space and we will show two examples of that.

Consider the case of predicting the presence of objects of certain categories in a room with a known category. This can be easily performed in our model by adding variables and relations for object categories without providing the actual object observations. We will show through the experiments that the system is able to continuously predict the existence of objects based on other semantic cues.

Another way of using the predictive power of the conceptual map is to predict the existence of a room of a certain category in the unexplored space behind a placeholder. In such case, the conceptual map is extended from the room in which the placeholder exists with variables representing categories of hypothesized rooms for different possible room configurations in the unexplored space. For each configuration, the categories of the hypothesized rooms are calculated and the obtained probabilities of existence of rooms of certain categories are summed over all possible configurations.

In a simple case, we can consider only three hypotheses: (1) placeholder does not lead to a new room; (2) placeholder leads to a single new room; (3) placeholder leads to a new room connected to another new room. If we assign equal likelihood to the case (2) and (3), it is sufficient to calculate a probability of the placeholder leading to at least one room ($p(r)$). This can be estimated as follows: $p(r) = p(ph)(1 - p(d)) + p(d)$, where $p(ph)$ denotes the probability that the placeholder leads to another placeholder and thus potentially to another room and $p(d)$ is the probability associated with the placeholder doorway property. $p(ph)$ can be estimated from the associated space placeholder property.

VII. EXPERIMENTAL SCENARIO

All the categorical models used in the experiments were trained on the COLD-Stockholm database[2]. Several parts of the database were previously used during the RobotVision@ImageCLEF[3] contests and proved to be challenging in the context of room categorization. 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). Each data sample is labeled as belonging to one of the areas according to the position of the robot during acquisition. More detailed information about the database can be found online[2].

A. Experimental Setup

In order to guarantee that the system will never be tested in the same environment in which it was trained, we have divided 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 during our offline experiments and the online experiment was performed on the same floor.

For the purpose of the experiments, we have extended the annotation of the COLD-Stockholm database to include the 3 room shapes, 3 room sizes as well as 7 general appearances. The room size and shape, were decided based on the length ratio (elongated $(0, 0.4]$, rectangular $(0.4, 0.8)$, square $[0.8, 1])$ and maximum length of edges (small $[0m, 3m)$, medium $[3m, 5m)$, large $[5m, \infty))$ of a rectangle fitted to the room outline. These properties together with 6 object types defined 11 room categories used in our experiments: an anteroom, a bathroom, a computer lab, a robot lab, a conference hall, a hallway, a kitchen, a meeting room, and three types of offices, a double office, a single office and a professor's office. The three types of offices, the two types of labs as well as the meeting room and conference hall shared appearance properties (office-like, lab-like and meeting room-like respectively) and could only be discriminated by a using a combination of properties.

VIII. EXPERIMENTS

We first build and evaluate the performance of each of the sensory models of properties offline. To build the models, the rooms having the same values of properties were grouped to form the training and validation datasets. Then, parameters of the models were obtained by cross-validation. Finally, all training and validation data were collected together and used to train the final models. The evaluation was performed on test data acquired in previously unseen rooms.

The classification rates obtained for each of the properties and cues are presented in Tab. II. The rates represent the percentage of correct classifications obtained separately for each of the classes, and then averaged in order to exclude the influence of unbalanced testing set. We can see that all classifiers provided a recognition rate above 80%. Additionally,

[2] http://www.semantic-maps.org/db

[3] http://www.robotvision.info

Property	Cues	Classification rate
Shape	Geometric features	84.9%
Size	Geometric features	84.5%
Appearance	CRFH	80.5%
Appearance	BOW-SURF	79.9%
Appearance	CRFH + BOW-SURF	84.9%

TABLE II

CLASSIFICATION RATES OBTAINED FOR EACH PROPERTY AND CUE.

we see that integrating two visual cues (CRFH and BOW-SURF) increased the classification rate of the appearance property by almost 5 percentage points. For an additional analysis of results, we refer the reader to [16].

The obtained models were used in the semantic mapping system during the online experiments. The experiments were performed on the 6th floor of the building where the COLD-Stockholm database was acquired, i.e. in the part which was not used for training. The robot was manually driven through two parts of the environment consisting of 13 different rooms. It performed real-time semantic mapping without relying on any previous observations of the environment. The obtained maps of the two parts of the environment (A and B) as well as the robot trajectory are presented in Fig. 6.

The robot gathered observations of shapes, sizes, appearances and objects present in the environment and performed reasoning about the values of properties and room categories. If an observation of an object of a certain category was not available, the robot reasoned about its existence based on other available information. 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. 7. Each column in the plot corresponds to a single event, and the cells show the probabilities assigned to beliefs. For better analysis, compare the results in Fig. 6 and Fig. 7 using the room numbers as a reference.

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 objects of the category "computer". The next room was initially identified as a double office until the robot was given a human assertion 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. The remaining rooms were correctly identified as single offices (rooms 4 and 5) and a meeting room (room 6).

Looking at part B, we see that the system identified most of the room categories correctly with the exception of a single office (room 2), which due to a misclassification of size was incorrectly recognized as a double office. The robot was first driven to the robot lab, which was correctly categorized thanks to a combination of a appearance information (lab-like) and an object observation (a robot). Remaining rooms were mapped primarily based on general appearance information as well as geometric properties.

In several rooms, we did not provide any object observations (rooms 0, 1 in part A and 0, 2, 3, 5 in part B). Therefore all the object presence beliefs shown in Fig. 7 obtained for those rooms are predictions of unexplored concepts. The experiment showed that the system can deliver good performance by integrating multiple sources of semantic information. As previously mentioned, a video showcasing the system is available online.

IX. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a probabilistic framework combining heterogeneous, uncertain, information such as object observations, the shape, size, appearance of rooms and human input for semantic mapping. A graphical model, more specifically a chain-graph, is used to represent the semantic information and perform inference over it. We used the concept of spatial properties which allowed us to make the knowledge representation more descriptive and pave the way for better scalability. Finally, we showed how to use the representation in order to reason about unexplored concepts.

There are several ways in which the work presented in this report can be extended, however three are of particular importance. First, we intend to look at ways to make the segmentation of space part of the estimation process as is made in PLISS [19], and while doing so, rely on all available properties. Second, we plan to replace the current space discretization approach with a feature-based clustering technique such as in [4]. Finally, we will investigate the problem of detection and learning of novel properties and room categories to pave the way towards fully self-extendable semantic mapping.

REFERENCES

[1] A. Aydemir, M. Göbelbecker, A. Pronobis, K. Sjöö, and P. Jensfelt, "Plan-based object search and exploration using semantic spatial knowledge in the real world," in *Proc. of ECMR'11*.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. J. Van Gool, "Speeded-up robust features (SURF)," *CVIU*, vol. 110, no. 3, 2008.

[3] M. Cummins and P. M. Newman, "Highly scalable appearance-only SLAM - FAB-MAP 2.0," in *Proc. of RSS'09*.

[4] F. Dayoub, G. Cielniak, and T. Duckett, "A sparse hybrid map for vision-guided mobile robots," in *Proc. of ECMR'11*.

[5] J. Folkesson, P. Jensfelt, and H. I. Christensen, "The M-space feature representation for SLAM," *IEEE Tr. on Robotics*, vol. 23, no. 5, 2007.

[6] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J. A. Fernández-Madrigal, and J. González, "Multi-hierarchical semantic maps for mobile robotics," in *Proc. of IROS'05*.

[7] M. Hanheide, C. Gretton, R. W. Dearden, N. A. Hawes, J. L. 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*, Barcelona, Spain.

[8] P. Jensfelt, "Approaches to mobile robot localization in indoor environments," Ph.D. dissertation, Signal, Sensors and Systems (S3), Royal Institute of Technology, SE-100 44 Stockholm, Sweden, http://www.cas.kth.se/˜ patric/publications/phd.html, 2001.

[9] B. Kuipers, "Spatial Semantic Hierarchy," *AI*, vol. 119, no. 1-2, 2000.

[10] S. Lauritzen and T. Richardson, "Chain graph models and their causal interpretations," *J. of Royal Statistical Society*, vol. 64, no. 3, 2002.

[11] D. Meger, P.-E. Forssen, K. Lai, S. Helmer, S. McCann, T. Southey, M. Baumann, J. J. Little, and D. G. Lowe, "Curious George: An attentive semantic robot," *RAS*, vol. 56, no. 6, 2008.

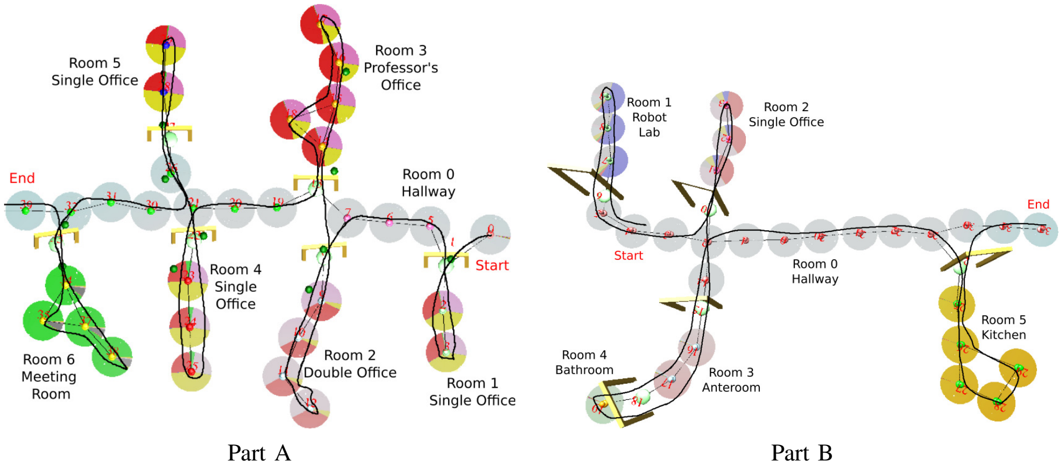

Fig. 6. Topological maps of the environment anchored to a metric map indicating the outcomes of room segmentation and categorization. The circles indicate the location of places in the environment and the black line shows the robot's trajectory. The pie charts indicate the probability distributions over the inferred room categories for each room (each fraction of the pie chart corresponds to a room category). In order to see the labels assigned to each fraction as well as a detailed view of the distributions across room categories, objects and values of spatial properties, see Fig. 7.

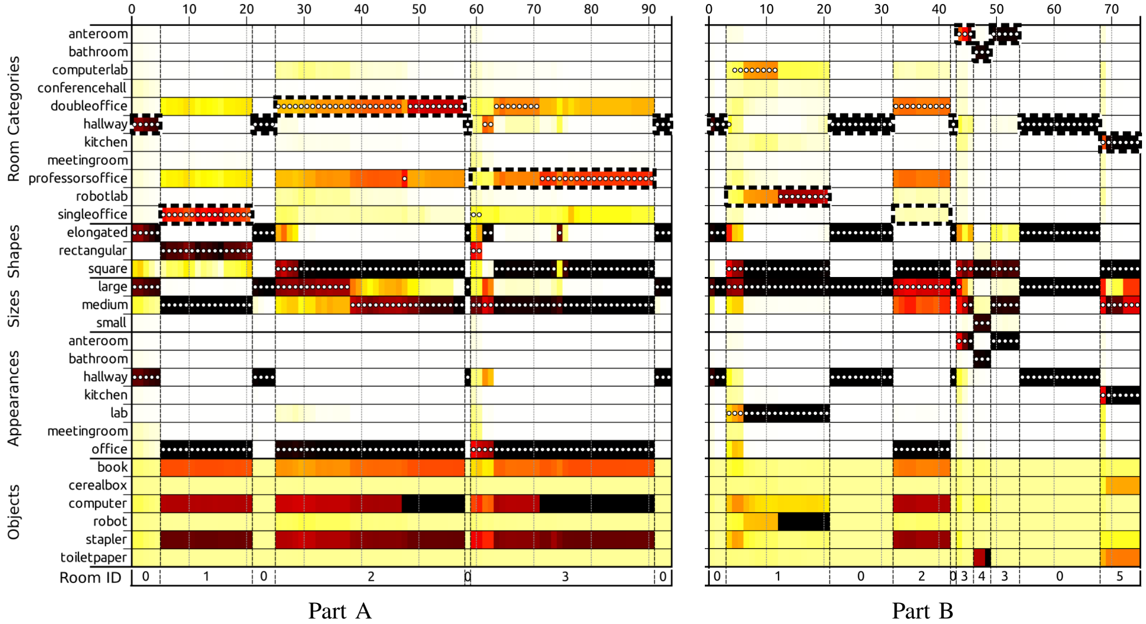

Fig. 7. Visualization of the events registered by the system during exploration and its beliefs about the categories of the rooms as well as values of the properties and object presence. Each row represents the development of probabilistic beliefs about a certain concept as the robot explored the environment (see the trajectory in Fig. 6). Darker colors indicate higher probability. The room category ground truth is marked with thick dashed lines. The MAP values are indicated with white dots.

[12] J. M. Mooij, "libDAI: A free and open source C++ library for discrete approximate inference in graphical models," *JMLR*, vol. 11, 2010.

[13] T. Mörwald, J. Prankl, A. Richtsfeld, M. Zillich, and M. Vincze, "BLORT - The blocks world robotic vision toolbox," in *ICRA Workshop Best Practice in 3D Perc. and Model. for Mobile Manipul.*, 2010.

[14] C. Nieto-Granda, J. G. Rogers, A. J. B. Trevor, and H. I. Christensen, "Semantic map partitioning in indoor environments using regional analysis," in *Proc. of IROS'10*, Taipei, Taiwan, 2010.

[15] A. Nüchter and J. Hertzberg, "Towards semantic maps for mobile robots," *RAS*, vol. 56, no. 11, 2008.

[16] A. Pronobis and P. Jensfelt, "Hierarchical multi-modal place categorization," in *Proc. of ECMR'11*.

[17] A. Pronobis, O. M. Mozos, B. Caputo, and P. Jensfelt, "Multi-modal semantic place classification," *IJRR*, vol. 29, no. 2-3, 2010.

[18] A. Pronobis, K. Sjöö, A. Aydemir, A. N. Bishop, and P. Jensfelt, "Representing spatial knowledge in mobile cognitive systems," in *Proc. of IAS'10*.

[19] A. Ranganathan, "PLISS: Detecting and labeling places using online change-point detection," in *Proc. of RSS'10*.

[20] A. Torralba, K. Murphy, W. Freeman, and M. Rubin, "Context-based vision system for place and object recognition," in *Proc. of ICCV'03*.

[21] S. Vasudevan and R. Siegwart, "Bayesian space conceptualization and place classification for semantic maps in mobile robotics," *RAS*, vol. 56, no. 6, 2008.

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

[23] J. Wu, H. I. Christensen, and J. M. Rehg, "Visual place categorization: problem, dataset, and algorithm," in *Proc. of IROS'09*.

[24] J. L. Wyatt, A. Aydemir, M. Brenner, M. Hanheide, N. Hawes, P. Jensfelt, M. Kristan, G.-J. M. Kruijff, P. Lison, A. Pronobis, K. Sjöö, A. Vrečko, H. Zender, M. Zillich, and D. Skočaj, "Self-understanding & self-extension:a systems and representational approach," *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 4, 2010.

[25] H. Zender, O. M. Mozos, P. Jensfelt, G.-J. M. Kruijff, and W. Burgard, "Conceptual spatial representations for indoor mobile robots," *RAS*, vol. 56, no. 6, 2008.