Cognitive Maps for Mobile Robots – An Object based Approach

Shrihari Vasudevan, Stefan Gächter, Marc Berger & Roland Siegwart

Autonomous Systems Laboratory (ASL)
Swiss Federal Institute of Technology Zurich (ETHZ)
8092 Zurich, Switzerland

{shrihari.vasudevan , stefan.gachter , r.siegwart}@ieee.org

***Abstract* - Robots are rapidly evolving from factory work-horses to robot-companions. The future of robots, as our companions, is highly dependent on their abilities to understand, interpret and represent the environment in an efficient and consistent fashion, in a way that is comprehensible to humans. This paper is oriented in this direction. It suggests a hierarchical probabilistic representation of space that is based on objects. A global topological representation of places with object graphs serving as local maps is suggested. Experiments on place classification and place recognition are also reported in order to demonstrate the applicability of such a representation in the context of understanding space and thereby performing spatial cognition. Further, relevant results from user studies validating the proposed representation are also reported. Thus the theme of the work is – representation for spatial cognition.**

Index Terms - Cognitive Spatial Representation, Robot Mapping, Conceptualization of spaces, Spatial Cognition

I. INTRODUCTION

Robotics today, is visibly and very rapidly moving beyond the realm of factory floors. Robots are working their way into our homes in an attempt to fulfill our needs for household servants, pets and other cognitive robot companions. If this "robotic-revolution" is to succeed, it is going to warrant a very powerful repertoire of skills on the part of the robot. Apart from navigation and manipulation, the robot will have to understand, interpret and represent the environment in an efficient and consistent fashion. It will also have to interact and communicate in human-compatible ways. Each of these is a very hard problem. These problems are made difficult by a multitude of reasons including the extensive amount of information, the huge number of types of data (multi-modality), the presence of entities in the environment which change with time, to name a few. Adding to all of these problems are the two simple facts that everything is uncertain and at any time, only partial knowledge of the environment is available.

The underlying representation of the robot is probably the single most critical component in that it constitutes the very foundation for all things we might expect the robot to do, these include the many complex tasks mentioned above. Thus, the extent to which robots will evolve from factory work-horses to robot-companions will in some ways (albeit indirectly) be decided by the way they represent their surroundings. This report is thus dedicated towards finding an appropriate representation that will make today's dream, tomorrow's reality.

II. RELATED WORK

Robot mapping is a relatively well researched problem, however, with many very interesting challenges yet to be solved. An excellent and fairly comprehensive survey of robot mapping has been presented in [1]. Robot mapping has traditionally been classified into two broad categories – metric and topological. Metric mapping [2] tries to map the environment using geometric features present in it. A related concept in this context is that of the relative map [3] – a map state with quantities invariant to rotation and translation of the robot. Topological mapping [4] usually involves encoding place related data and information on how to get from one place to another. More recently, a new scheme has become quite popular – the one of hybrid mapping [5, 6]. This kind of mapping typically uses both a metric map for precision navigation in a local space and a global topological map for moving between places.

The one similarity between all these representations is that all of them are navigation-oriented, i.e. all of them are built around the single application of robot-navigation. These maps are useful only in the navigation context and fail to encode the semantics of the environment. The focus of this work is to address this deficiency. Several other domains inspire our approach towards addressing this challenge – these include hierarchical representations of space, "high-level" [†] feature extraction, scene interpretation and the notion of a Cognitive Map.

The work presented here closely resembles those that suggest the notion of a hierarchical representation of space. Ref. [7] suggests one such hierarchy for environment modeling. In [8], Kuipers put forward a "Spatial Semantic Hierarchy" which models space in layers comprising respectively of sensorimotor, view-based, place-related and metric information. The work [9] probably bears the most similarity with the work presented in this paper. The authors use a naive technique to perform "object recognition" and add the detected objects to an occupancy grid map. The primary

[†] Objects, doors etc. are considered "high-level" features contrasting with lines, corners etc. which are considered "low-level" ones.

difference in the work presented here is that the proposed representation uses objects as the functional basis – i.e. the map is created and grown with the objects perceived.

Typically, humans seem to perceive space in terms of high-level information such as objects, states & descriptions, relationships etc. Thus, a human-compatible representation would have to encode similar information. The work reported here attempts to create such a representation using typical household objects and doors. It also attempts to validate the proposed representation in the context of spatial cognition. For object recognition, a very promising approach that has also been used in this work, is the one based on the SIFT [10]. In our experience, it was found to be a very effective tool for recognizing textured objects. Several works have attempted to model and detect doors. The explored techniques range from modeling/estimating door parameters [11] to those that model the door opening [12] and to those like [13], based on more sophisticated algorithms such as boosting. Ref. [13] also addresses the problem of scene interpretation in the context of spatial cognition. The authors use the AdaBoost algorithm and simple low-level scan features and vision together with hidden markov models to classify places.

This work takes inspiration from the way we believe humans represent space. The term "Cognitive Map" was first introduced by Tolman in a widely cited work, [14]. Since then, several works in cognitive psychology and AI / robotics have attempted to understand and conceptualize a cognitive map. Some of the more relevant theories are mentioned in this context. Kuipers, in [15], elicited a conceptual formulation of the cognitive map. He suggests the existence of five different kinds of information (topological, metric, routes, fixed features and observations) each with its own representation. More recently, Yeap et al. in their work [16] trace the theories that have been put forward to explain the phenomenon of early cognitive mapping. They classify representations as being space based and object based. The proposed approach in this work is primarily an object based one. Some of the most relevant object based approaches include the MERCATOR (Davis, 1986) and more recently RPLAN (Kortenkamp, 1993, [17]). The former bears the closest resemblance to some of the ideas put forward in this work. It should be emphasized that among most previously explored approaches classified as "object" based, either the works do not necessarily suggest a hierarchical representation or they do not use high-level features.

In summary, a single unified representation that is multi-resolution, multi-purpose, probabilistic and consistent is still a vision of the future; it is also the aspiration of this work. The approach can be understood as an engineering solution (as applicable to mobile robots) to the general Cognitive Mapping problem. Although being primarily object based, the proposed approach attempts to overcome some of the believed limitations of purely object based (i.e. no notion of the space) methods by incorporating some spatial elements (in this case doors). The kinds of elements that are incorporated will be gradually upgraded as the work is enhanced.

III. Approach

A. *Problem Definition*

This work is aimed at developing a generic representation of space for mobile robots. Towards this aim, in this particular work, two scientific questions are addressed - (1) How can a robot form a high-level probabilistic representation of space? (2) How can a robot understand and reason about a place?

The first question directly addresses the problems of high-level feature extraction, mapping and place formation. The second question may be considered as the problem of spatial cognition. Together, when appropriately fused, they give rise to the hierarchical representation being sought. This representation must consider and treat information uncertainty in an appropriate manner. Also, in order to understand places, the robot has to be able to conceptualize space; to be able to classify its surroundings and to recognize it, when possible.

B. *Overview*

Figures 1 & 2 respectively show the mapping process and the method used to demonstrate spatial cognition using the created map. In an integrated system, the mapping and reasoning processes cannot be totally separated, but it is done here so as to facilitate understanding of the individual processes. Subsection C elicits the details of the perception system – this includes the object recognition and door detection processes. Subsection D specifies the details on how the representation is created (process depicted in fig. 1) – both local probabilistic object graphs and individual places. It also addresses the issue of learning about place categories (kitchens, offices etc.). Subsections E explains how such a representation could be used for spatial cognition (process depicted in fig. 2) and the manner in which the representation is updated. All of these sections are briefly presented. For more details, the interested reader is referred to another recent report by the authors, [18]. The remaining parts of the papers discuss the experiments conducted, the user study and the conclusions drawn thereof. The main contribution of this paper is an enhancement of the previously reported results by the provision of relevant results from user studies, in support of our representation, as a cognitive validation of the theory.

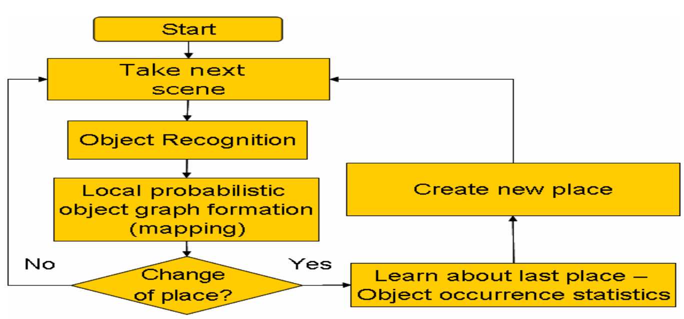

Fig. 1 The mapping process. High-level feature extraction is implemented as an object recognition system. Place formation is implemented using door detection. Beliefs are represented and appropriately treated. Together, these are encoded to form a hierarchical representation comprising of places, connected by doors and themselves represented by local probabilistic object graphs. Concepts about place categories are also learnt.

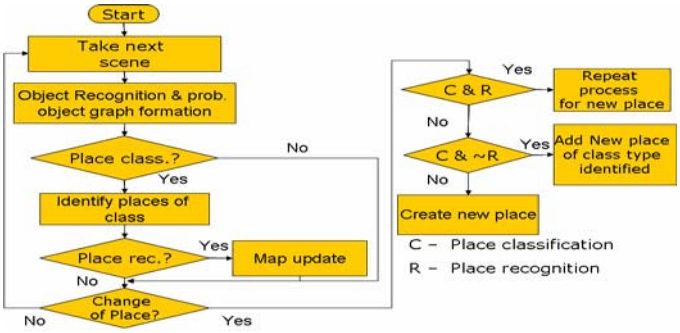

Fig. 2 The reasoning process for each place. First step is place classification – the robot uses the objects it perceives to classify the place into one of its known place categories (office, kitchen etc.). Next step is - recognizing specific instances of the place it is aware of – place recognition. Accordingly map update / adding of new place is done.

C. Perception

This work deals with representing space using high-level features. In particular, two kinds of features are used here – typical household objects and doors. Reliable and robust methods for high-level feature extraction are yet unavailable. It must be emphasized that the perception component of this work, is not the thrust of this work. Thus, established or simplified algorithms have been used.

For this work, a SIFT based object recognition system was developed (fig. 3) along the lines of [10]. The objects detected are used to represent places as explained in sub-section D. Doors are used in this work in the context of place formation. A method of door detection based on line extraction and the application of certain heuristics was used. The sensor of choice was the laser range finder. More details on the perception of objects and doors with regards to this work can be found in [18].

D. Representation

The representation put forward here is a hierarchical one that is composed of places which are connected to each other through doors and are themselves represented by local probabilistic object graphs (a probabilistic graphical representation composed of objects and relationships between them). Objects detected in a place are used to form a relative map for that local space. Doors are incorporated into the representation when they are crossed and link the different places together.

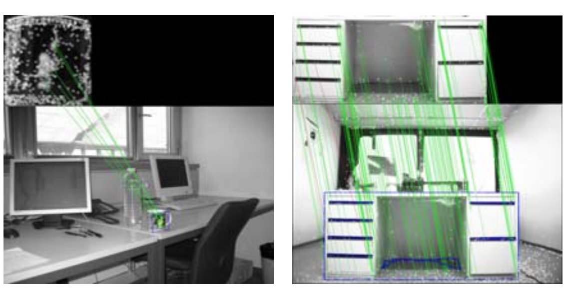

Fig. 3 Object recognition using SIFT features. Left image shows a mug being recognized, right image shows a table being recognized. Objects used in this work include cartons of different kinds, a table, a chair, a shelf & a mug.

Object graphs were used by the authors in [19]. The problem with this work is that the information encoded in the representation was purely semantic and not “persistent” i.e. not invariant and not re-computable based on current viewpoint. This work addresses this drawback by drawing on the relative mapping approach in robotics. It suggests the use of a probabilistic relative object graph as a means of local metric map representation of places. The metric information encoded between objects includes distance and angle measures in 3D space. These measures are invariant to robot translation and rotation in the local space. Such a representation not only encodes the inter-object semantics but also provides for a representation that could be used in the context of robot navigation.

The robot uses odometry to know the robot pose which is in turn used towards the creation of the relative object graph. A stereo camera is used to know the positions of various objects in 3D space. As mentioned before, the representation is probabilistic. “Existential” beliefs (discrete probability values) are obtained from the perception system for each object that is observed. Simultaneously, “precision” beliefs are maintained in the form of covariance matrices. By representing both kinds of beliefs, such a representation will serve in the context of high level reasoning / scene interpretation and yet be useful for lower level navigation related tasks. As mentioned earlier, the relative spatial information encoded, include distance and angle measures in 3D space. These also have associated existence and precision beliefs. Details on the sensor models used and the mathematical formulations for belief computation are mentioned in [18]. Concepts are learnt when creating the representation of various places. These encode the occurrence statistics (and thus likelihood values) of different objects in different place categories (office, kitchen, etc.). Thus, in a future exploration task, a robot could actually understand its environment and thereby classify its surroundings based on the objects it perceives.

E. Spatial Cognition (Place Classification / Recognition) and Map Update

Place classification is done in an online incremental fashion, with every perceived object contributing to one or more hypotheses of previously learnt place concepts. Place recognition is done by a graph matching procedure which matches both the nodes and its relationships to identify a node match. The aim is to find the maximal common set of identically configured objects between places the robot knows (previously mapped) and the one it currently perceives. A map update operation (internal graph representation is updated) is required both for handling the revisiting of places and the re-observation of objects while mapping a place. It involves the addition / deletion of nodes and the update of their beliefs. More details can be obtained from [18].

IV. Experiments

A. *System Overview and Scenario*

The robot platform shown in fig. 4 was used for this work. The robot is equipped with several sensors including encoders, stereo and two back-to-back laser range scanners. The robot was driven across 5 rooms covering about 20m in distance. The objects used (*and the way they are named*) for the representation comprised of different cartons (*carton, cartridge, xerox, logitech, elrob, tea*), a chair (*chair*), a mug (*mug*), a shelf (*shelf*), a table (*table*) & a book (*book*). The experiments were conducted in our lab – thus, the places visited included offices and corridors.

B. *Mapping*

Fig. 5 shows the path of the robot. The objects & doors recognized are shown in the object based map depicted in fig. 6. Finally, fig. 7 illustrates the complete probabilistic object-graph representation formed as a result of the process.

The robot performed the mapping process as per expectations. Objects and doors were recognized and the representation was formed as per the methods described in the previous sections. However, the robot often observed multiple doors at the same place (due to the presence of large cupboards) on either side of the door. Further, the robot created multiple occurrences of the corridor as, the topological information between places that is encoded was not used in the experiments in this work. Also, it did not see an identical set of objects through the corridor so as to be able to recognize the previously visited corridor. These two issues (fusing of doors and loop closing) would be addressed in subsequent works.

Fig. 4 The robot platform that was used for the experiments. The encoders, stereo vision system and laser scanners were used for this work.

C. *Spatial Cognition – Place classification/recognition*

The robot was made to traverse a previously visited place – *SV (office)* and *the corridor* (refer fig. 6). The locations of movable objects (all but the table, shelf and the door) were changed so that a significant configuration change of both places was observed. The robot was then made to interpret these places.

For the first place, the robot perceived the objects in the sequence *shelf – xerox – carton – table – logitech – cartridge*. Fig. 8 displays the object map for the “unknown” place. On seeing the first two objects, the robot successfully classified the place as an office. Subsequently the robot attempted to match this place with its knowledge of prior offices it has visited. When finally crossing the door, the robot found enough objects (including the door) that are located in a matching spatial configuration to a place that it has visited before. Thus, at this point, the “unknown” place was recognized as the place *SV (office)* and the internal map representation of the robot is updated to reflect the changes to the place that the robot had perceived. Figure 9 displays the updated internal representation of the robot. The corridor was also successfully classified. More details can be got from [18].

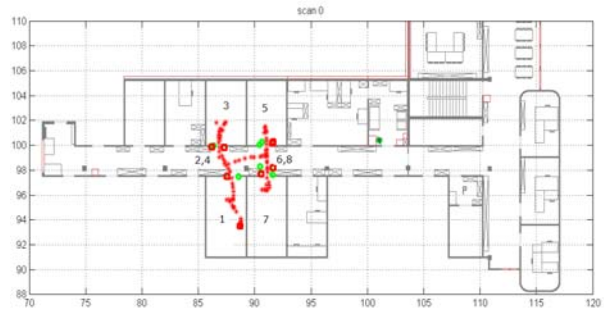

Fig. 5 Map displaying the robot path. The robot traverses through 4 rooms crossing a corridor each time it moves from one room to another. Green/red circles indicate the doors detected. The red circles also serve as the place references for the place explored on crossing the door. The numbers indicate the sequence in which the places were visited.

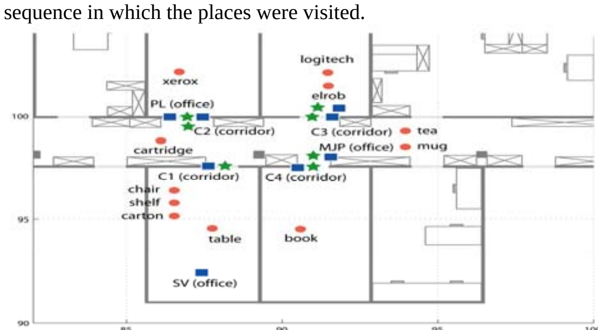

Fig. 6 Object based map produced as a result of exploring the test environment. Zoomed-in view of the above map. Blue squares are the place references, red circles are the objects and the green stars are the doors.

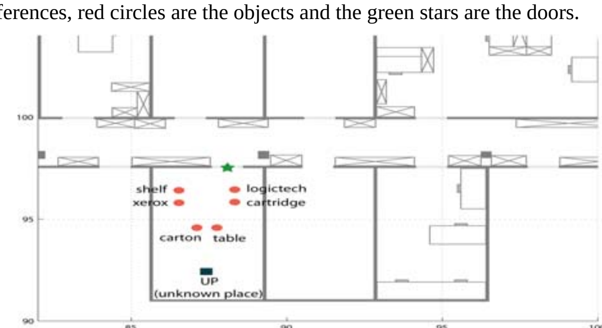

Fig. 8 First ‘unknown’ place at the time of place recognition. The configuration of the objects is different from that of the same place in Fig. 6. Note - The *carton* is above the table & *xerox* is above the shelf.

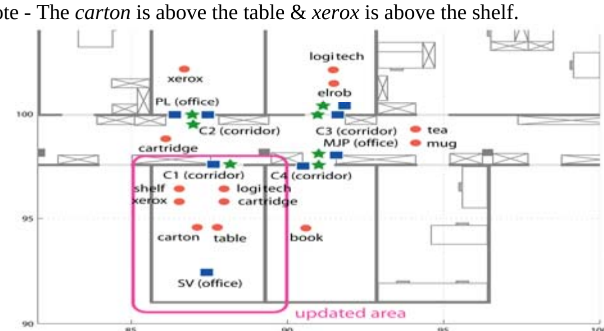

Fig. 9 Updated internal representation of the robot after place recognition

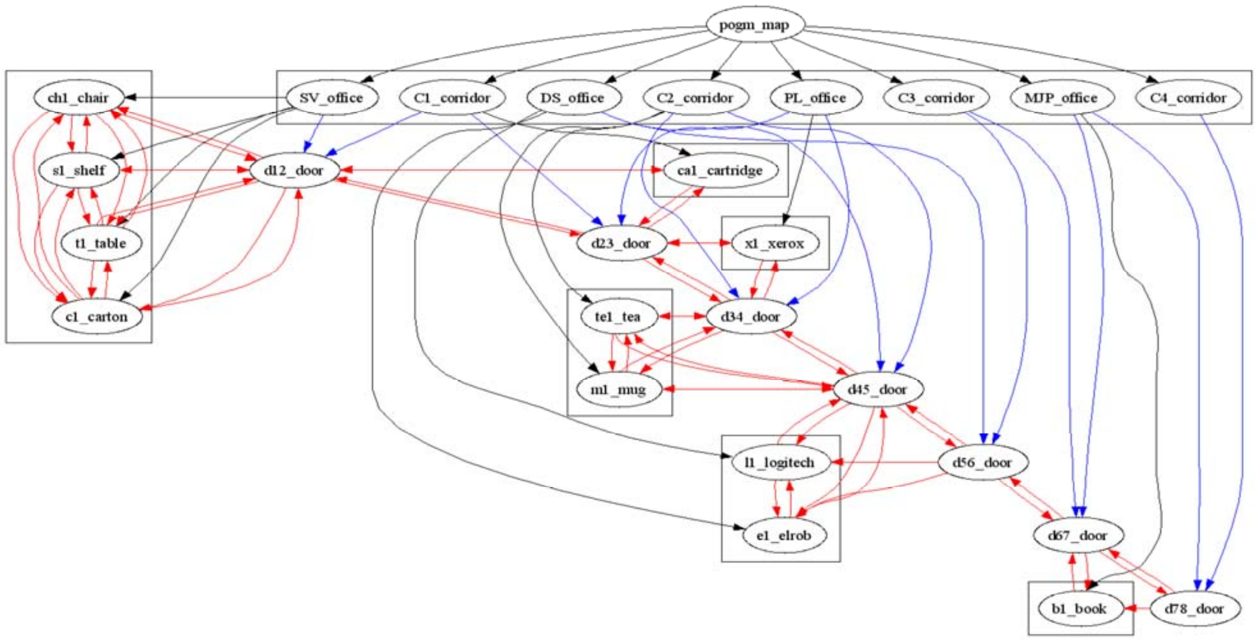

Fig. 7 Probabilistic object graph representation created as a result of exploring the path shown in fig. 5

V. User studies – A cognitive validation of the proposed representation

A. The study – objectives and methods

The broad aim of the study was to validate the proposed representation in a cognitive sense. The aim was to verify our approach and to find out what other details (kinds of features / data) the proposed representation could encode. As mentioned before, the complete representation is beyond the scope of this report. Thus, only results of the survey that are relevant to the aspects of the representation proposed here are quoted. The complete study will be reported in a more appropriate forum.

The study was performed with input from 52 people. The people were chosen from a diverse population spanning different nationalities, backgrounds and occupation. Both genders have been appropriately represented.

B. Relevant Results

In the tables that follow, most criteria correspond to their literal (dictionary) meanings. The "function" of a place refers to the typical functionality / purpose associated with a place. "Ground materials" refer to the floor material (wooden / carpeted /…). "Boundaries" refer to walls, doors, partitions etc. The percentages indicate the number of people, of the total number surveyed, that replied with information corresponding to the particular criteria for the place in consideration.

Survey takers were asked to imagine their presence in a living room, an office and a kitchen. They were then asked to describe what they understood / represented about that place in their minds. Table 1 shows the results obtained. The most common objects identified with an office were desks, chairs, computers etc. Living rooms were better understood in terms of the presence of sofas, armchairs, tables etc. and finally kitchens were typically identified with cooker, oven, sink, fridge, utensils etc.

TABLE 1
MEANS OF REPRESENTATION OF PLACES.

Criteria / Place	Office (%)	Living Room (%)	Kitchen (%)
Objects	98	96	98
Function	13	21	13
Boundaries	71	48	38
People	23	10	8
Size	17	25	35
Ambience	19	33	27
Luminosity	37	37	13
Ground Material	8	15	12
Smell	-	-	4

Next, users were taken to three places in our laboratory premises – a "standard" office, a refreshment room and lastly, a large electronics lab-office. Survey takers were asked to describe each place – what they saw in as much detail as possible. The typical ways in which survey takers tend to describe these places are conveyed through the following graphs shown in table 2.

TABLE 2
MEANS OF DESCRIPTIONS OF PLACES

Criteria / Place	Office (%)	Refreshment room (%)	Lab (%)
Objects	100	100	100
Function	52	90	63
Boundaries	40	10	4
Partitions	-	-	15

Finally, users were taken from one room to another and asked if they believed they were in a new place and the reason for their belief. The results obtained are shown in a graphical form below in fig. 10.

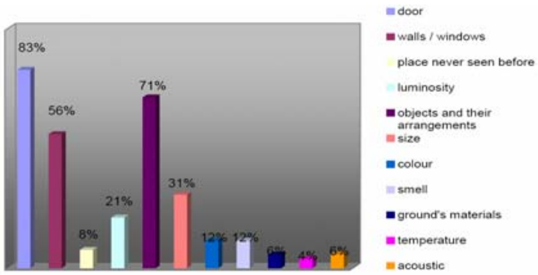

Fig.10 Criteria to ascertain a change of place

C. Analysis / Inference

The reason survey takers were first asked to imagine being in a place and then taken to such a place for questioning was to get both inputs – that of the accumulated (through experience) representation of the place and also that obtained from on-site scene interpretation. It was found that objects constituted a very critical component of both a representation and a description. People seem to understand places in terms of the high-level features (objects) that are present in it – the underlying philosophy of this work and the direction of our future works as well. It was also found that boundaries (walls / doors / windows) constituted an important component in describing the places and the "function" of the place (kitchen – cooking etc.) was an important descriptive element. The last graph seems to convey that boundary elements (such as doors and walls) and the arrangement of objects are critical to detecting a change of place. From an implementation perspective, this information seems to validate our choice for using the objects as the functional basis of the representation and doors as the links between places. Lastly, we believe that a transition between places occurs when there is a change of "visibility", a term we can now implement in terms of the other important factors that crop up in the graph shown in figure 10, including arrangement of objects, luminosity, size, color and ground materials. Thus these results not only validate the proposed representation but further provide ideas on the future enhancements (functionality of a place etc. need to be incorporated) to this representation and how it is formed.

VI. CONCLUSIONS & FUTURE WORK

A cognitive probabilistic representation of space based on high level features was proposed. The representation was experimentally demonstrated. Spatial cognition using such a representation was shown through experiments on place classification and place recognition. The uncertainty for all required aspects of such a representation were appropriately represented and treated. Relevant results from a user study conducted were also reported, thus validating the representation in a cognitive sense. They also suggest the next steps towards enhancing the proposed representation.

Fusing of doors and merging of places are both required to get a more appropriate representation of space. On the conceptual front, the suggested representation needs to be made richer but yet lighter and computationally efficient in applications. A more in-depth survey focusing on specific aspects of the representation is also warranted.

ACKNOWLEDGEMENTS

The work described in this paper was conducted within the EU Integrated Project COGNIRON ("The Cognitive Companion") and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.

REFERENCES

[1] S. Thrun, "Robotic mapping: A survey", In G. Lakemeyer and B. Nebel, editors, Exploring Artificial Intelligence in the New Millenium. Morgan Kaufmann, 2002.

[2] R.Chatila and J.-P. Laumond. Position referencing and consistent world modeling for mobile robots. In the Proceedings of the IEEE International Conference on Robotics and Automation, 1985.

[3] A. Martinelli, A. Svensson, N. Tomatis and R. Siegwart, "SLAM Based on Quantities Invariant of the Robot's Configuration", IFAC Symposium on Intelligent Autonomous Vehicles, 2004.

[4] H. Choset and K. Nagatani, "Topological simultaneous localization and mapping (SLAM): toward exact localization without explicit localization", IEEE Transactions on Robotics and Automation, Volume: 17 Issue: 2, Apr 2001.

[5] S. Thrun, "Learning metric-topological maps for indoor mobile robot navigation", Artificial Intelligence, Volume 99, Issue 1, February 1998, Pages 21-71.

[6] N. Tomatis, I. Nourbakhsh and R. Siegwart, "Hybrid Simultaneous Localization and Map Building: A Natural Integration of Topological and Metric", Robotics and Autonomous Systems, 44, 3-14., July 2003.

[7] A. Martinelli, A. Tapus, K.O. Arras and R. Siegwart, "Multi-resolution SLAM for Real World Navigation", In Proceedings of the 11th International Symposium of Robotics Research, Siena, Italy, 2003.

[8] B. Kuipers, "The Spatial Semantic Hierarchy", Artificial Intelligence, 119: 191-233, May 2000.

[9] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J.A. Fernández-Madrigal and J. González, "Multi-Hierarchical Semantic Maps for Mobile Robotics", Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS), pp. 3492-3497. Edmonton, CA, 2005.

[10] D. G. Lowe, "Distinctive image features from scale-invariant key-points", International Journal of Computer Vision, vol. 60, no. 2, 2004.

[11] D. Anguelov, D. Koller, E. Parker, S. Thrun, "Detecting and Modeling Doors with Mobile Robots", Proceedings of the International Conference on Robotics and Automation (ICRA), 2004.

[12] D. Kortenkamp, L. D. Baker, and T. Weymouth, "Using Gateways to Build a Route Map," Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems, 1992.

[13] C. Stachniss, Ó. Martínez-Mozos, A. Rottmann, and W. Burgard, "Semantic Labeling of Places", In Proc. of the Int. Symposium of Robotics Research (ISRR).San Francisco, CA, USA, 2005.

[14] E. C. Tolman. "Cognitive maps in rats and men." Psychological Review, 55:189-208, 1948.

[15] B. J. Kuipers, "The cognitive map: Could it have been any other way?", Spatial Orientation: Theory, Research, and Application. New York: Plenum Press, 1983, pages 345-359.

[16] W.K. Yeap and M.E. Jefferies, "On early cognitive mapping", Spatial Cognition and Computation 2(2) 85-116, 2001.

[17] D. Kortenkamp. "Cognitive maps for mobile robots: A representation for mapping and navigation". PhD Thesis, University of Michigan, 1993.

[18] S. Vasudevan, V. Nguyen and R. Siegwart, "Towards a Cognitive Probabilistic Representation of Space for Mobile Robots", In the Proceedings of the IEEE International Conference on Information Acquisition (ICIA), 20-23 August 2006, Shandong, China.

[19] A. Tapus, S. Vasudevan, and R. Siegwart, "Towards a Multilevel Cognitive Probabilistic Representation of Space", In Proceedings of the International Conference on Human Vision and Electronic Imaging X, part of the IS&T/SPIE Symposium on Electronic Imaging 2005, 16-20 January 2005, CA, USA.