Available online at www.sciencedirect.com
ScienceDirect
Robotics and Autonomous Systems 55 (2007) 359–371
Robotics and Autonomous Systems
www.elsevier.com/locate/robot

Cognitive maps for mobile robots—an object based approach

Shrihari Vasudevan*, Stefan Gächter, Viet Nguyen, Roland Siegwart

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

Available online 30 January 2007

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. The work presented here 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 proposed. The work also details the first efforts towards conceptualizing space on the basis of the human compatible representation so formed. Such a representation and the resulting conceptualization would be useful for enabling robots to be cognizant of their surroundings. Experiments on place classification and place recognition are reported in order to demonstrate the applicability of such a representation towards 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.

Keywords: Robot mapping; Cognitive spatial representation; Conceptualization of space; Spatial cognition

1. 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 — 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.

2. Related work

Robot mapping is a 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 [2,3] tries to map the environment using geometric features present in it. A related concept in this context is that of the relative map [4] — a map state with quantities invariant to rotation and translation of the robot. Topological mapping [5,6] usually involves encoding place related data and information on how to get from one place to

* Corresponding author. Fax: +41 44 63211 81.
E-mail address: shrihari.vasudevan@ieee.org (S. Vasudevan).

0921-8890/$ - see front matter
doi:10.1016/j.robot.2006.12.008

another. More recently, a new scheme has become quite popular — the one of hybrid mapping [7,8]. 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[1] feature extraction, scene interpretation, the notion of a cognitive map and finally the field of human robot interaction (HRI).

The work presented here closely resembles those that suggest the notion of a hierarchical representation of space. Ref. [9] suggests one such hierarchy for environment modelling. In [10], Kuipers put forward a *Spatial Semantic Hierarchy* which models space in layers comprising sensorimotor, view-based, place-related and metric information. The work [11] probably bears the most similarity with that 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 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 and descriptions, relationships etc. This seems both intuitive and is also subsequently validated through user studies that were conducted as a part of this work (detailed in Section 5). Thus, a *cognitive* spatial representation could be expected 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 SIFT [12]. 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 modelling/estimating door parameters [13] to those that model the door opening [14] and to those like [15], based on more sophisticated algorithms such as boosting. Ref. [15] 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, [16]. 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 [17], 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 [18] 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 [19] and more recently RPLAN [20]. 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 being 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 and is the aspiration of this work. The proposed 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 proposed representation has the potential to enable robots to not only navigate its surroundings but also to conceptualize space and perform spatial cognition. The work is thus part of an endeavour to make robots more compatible and acceptable to us.

As robots are becoming more intelligent, they are also tending to be increasingly socially interactive. Ref. [21] gives a nice survey of recent advances in socially interactive robots. Dialogue and Natural Language Processing (NLP) will form a critical component of the interaction between a human and a robot. This in turn will be the deciding factor towards the compatibility and acceptability of robots in our homes. The works [22,23] are examples of some recent efforts towards integrating dialogue and NLP in robotics. Most works in mobile robotics, however, have until now restricted themselves to navigation related problems. Thus, few works in mobile robotics evaluate their concepts in human centred experiments. A recent work which attempts to understand the acceptability of robots among people through a user study is done in [24]. This work was done on the sidelines of [25], which was a recent large-scale demonstration of the remarkable growth of personal and service robotics. This underlines the need for compatibility between robots and humans. In the work proposed here, we propose a representation for robots that could enable them to not only perform navigation related tasks but also enable them towards becoming more spatially aware and human-compatible machines that could inhabit our homes alongside us. With the rapid increase in the importance of HRI, the need for evaluating the work through human-centred experiments was felt necessary. Further, it was felt that such experiments

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

could contribute positively to the enhancement of the work itself. Towards these objectives, an elaborate user study was conducted to understand human perception and representation of spaces. This has been detailed later in Section 5.

Thus, the main contributions of this paper include (1) the proposition of a hierarchical cognitive probabilistic representation of space for mobile robots, (2) the first definite steps towards conceptualization of the space, for mobile robots, from the learnt representation, (3) the relevant results of a user study experiment that provides a cognitive validation of some of the elements that constitute the representation and corroborates its compatibility with us humans. The contributions enable a robot to form concepts about space in a human compatible fashion. The paper demonstrates a novel effort to establish a strong link from the sensors that a robot is equipped with, to the human compatible spatial concepts that the robot forms in order to represent and understand its surroundings. First results of this work were published in [26]. This report attempts to revisit and further analyse the previously obtained results to understand their significance, how they can be improved and how they relate to existing work in mobile robotics. It also details the relevant results of a user study experiment that was conducted to support and improve the approach from a human perspective. Finally, it lays the foundation for ongoing and future work in this direction.

3. Approach

3.1. 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. A place, here, refers to a spatial abstraction; in indoor environments, these may be rooms. 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 classify its surroundings and to recognize them, when possible.

3.2. Overview

Figs. 1 and 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 to facilitate understanding of the individual processes. Section 3.3 elicits the details of the perception system — this includes the object recognition and door detection processes. The next section, specifies the details on how the representation is

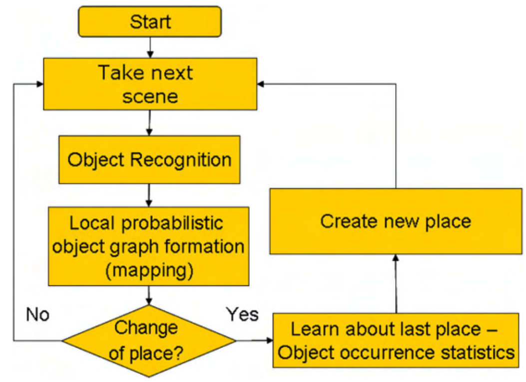


Fig. 1. The mapping process. High-level feature extraction is implemented as an object recognition system. Place formation is implemented using door detection. Beliefs are represented and appropriately treated. Together, these are encoded to form a hierarchical representation comprising places, connected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

created (process depicted in Fig. 1), both local probabilistic object graphs and individual places. The issue of learning about place categories (kitchens, offices etc.) is addressed in Section 3.5. Sections 3.6 and 3.7 explain how such a representation could be used for spatial cognition (process depicted in Fig. 2) and the manner in which the representation is updated. The remaining parts of the paper discuss the experiments conducted, the user study and the conclusions drawn thereof.

3.3. 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 is not the thrust of this work. Established or simplified algorithms have thus been used.

For this work, a SIFT-based object recognition system was developed (Fig. 3) along the lines of [12]. A stereo camera is used to recognize the object and to obtain its coordinates in 3D space. Very briefly, the SIFT approach to object recognition is a "local-features" method. It does not learn any general properties of objects — in order to categorize and classify them. It does, however, transform a set of features, obtained from the object-of-interest using a naive technique, into a robust feature set that incorporates invariance to scale and rotation changes; to a considerable extent, it deals with illumination changes and changes in viewing direction as well. In our experience, this method was found to be very effective for recognizing most textured objects. More details on SIFT-based object recognition can be obtained from [12]. The objects detected are used to represent places as explained in Section 3.4.

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. The door detection process involved the following steps:

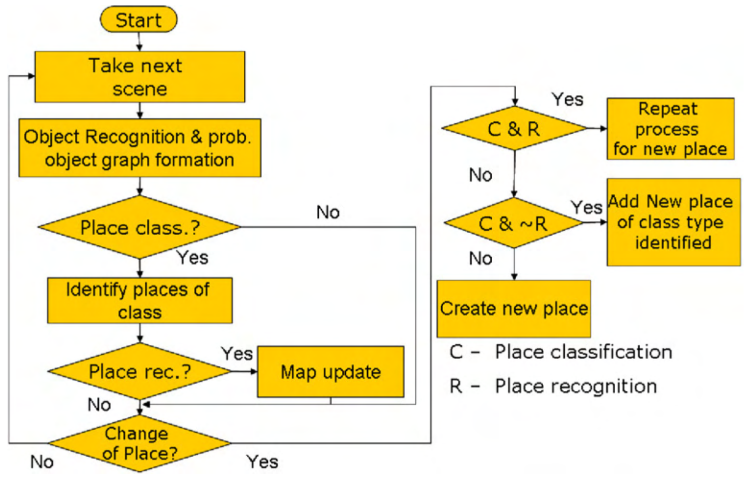


Fig. 2. The reasoning process for each place. First step is place classification — the robot uses the objects it perceives to classify the place into one of its known place categories (office, kitchen etc.). Next step is – recognizing specific instances of the place it is aware of – place recognition.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

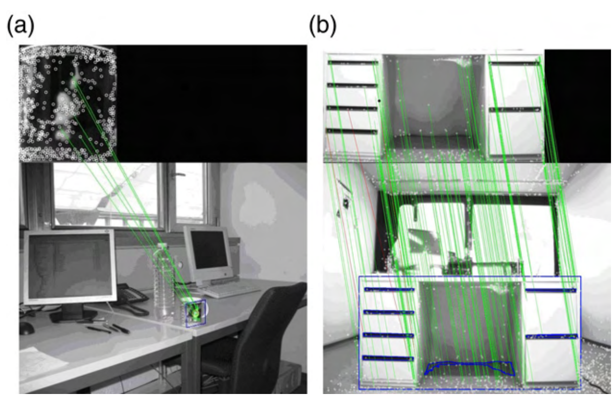


Fig. 3. Object recognition using SIFT features. Image (a) shows a mug being recognized, image (b) shows a table being recognized. Objects used in this work include cartons of different kinds, a table, chair, shelf and mug.

(i) Detect lines in 2D space (split and merge method, [27]);
(ii) Apply heuristics to identify doorways;
(iii) Track door hypotheses. Use a likelihood-based tracking process to eliminate false positives.

References for places are fixed at the end of the door that occurs first in the anticlockwise direction (generally the left) when the door is crossed. Occasionally, due to the coordinates of the door with respect to the robot position during detection, the same end of the door may be chosen as the reference both when entering the place and when leaving it. This algorithm was applied on a dataset with over 150 scans taken over several rooms of our laboratory. The robot was rigorously tested by moving it into corners, between tables, walls etc. A promising performance was observed — this is indicated in Table 1 and Fig. 4.

3.4. 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 (probabilistic graph encoding the objects and relationships between them).

Table 1
Door detection performance

Number of good detections (door detections, going into a room + leaving a room)	14 (of 15 expected)
Number of false positives (additional "noise" detections)	3

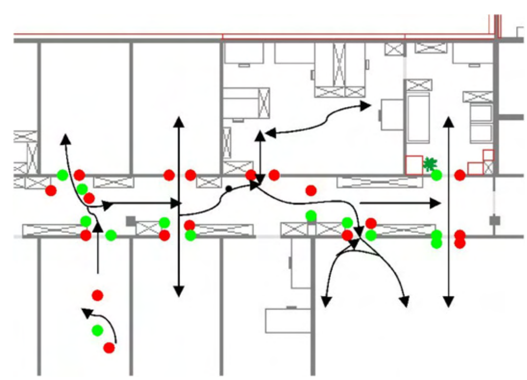


Fig. 4. Door detection algorithm tested over a part of our lab. A pair of dots represents a door. The red (darker) ones are the references of the place explored on crossing the doors. The few false positives that were obtained were observed and removed using simple techniques. The arrows describe the path through which the robot traversed while accumulating sensor data for door detection. As observed, the algorithm worked well even when the robot visited corners or cluttered areas. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Very briefly, objects and doors are detected in 3D space as shown above. Knowing the robots pose (using odometry) relative to a local reference, these objects and doors are identified in the local frame of reference. Using this information, a probabilistic graphical representation encoding the objects and the relative spatial information between them is formed as a local representation for this place. The local representations of different places are connected through the doors that connect them. In this way, the formed representation can be understood either as an extended relative metric

Let O = Local place reference frame ; C = Camera reference frame R = Robot reference frame

$$X_o = f(X_c) \quad where \quad f = M_{RO} * M_{CR}$$

is the object coordinate in O.
M_{RO} = Transformation between R & O.
M_{CR} = Transformation between C & R.

$$P_O = F_1 P_1 F_1' + F_2 P_2 F_2' \; where \; F_1 = J_{X_1}(f) \;\&\; F_2 = J_{X_2}(f)$$

is the uncertainty in object position.
$X_1 = (X_R, Y_R, \theta_R)$ is the robot pose in O.
$X_2 = (X_C, Y_C, \theta_C)$ is the object position in C.
P_1 = uncertainty in robot pose.
P_2 = uncertainty in object position.
F_1 = Jacobian of f with respect to X_1.
F_2 = Jacobian of f with respect to X_2.

Fig. 5. Belief representation for individual objects in a place. The objects in 3D camera space are transformed into local place reference coordinates. The uncertainty in its position in the place reference frame is dependent on the uncertainty in the robots pose, the uncertainty in the object pose and the dependence of the transformation on them. This is computed as shown. Position uncertainties are expressed using a covariance matrix representation.

representation (from the design perspective) or as a hierarchical metric-topological-semantic representation of space where the topological information is given by the places and the semantic content is encoded by using objects and their properties. More details are given in the paragraphs that follow.

Object graphs were used by the authors of [28]. 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 recomputable based on current viewpoint. The presented 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 (currently, inter-object relationships) 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. The odometry model employed was that of a standard differential drive model as suggested in [3]. A stereo camera is used to know the positions of various objects in 3D space. The stereo model suggested by Jung in his work on SLAM (Simultaneous Localization and Mapping) using stereo vision [29] is used here.

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, "position" beliefs are maintained in the form of covariance matrices. These beliefs are based on detailed mathematical formulations — the end result of which is shown below in Fig. 5. By representing both kinds of beliefs, such a representation will serve in the context of both, high level reasoning and 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 position beliefs — the computation of which takes the form shown in: Fig. 6.

3.5. Learning about place categories

Concepts are learnt when creating the representation of various places. Currently, these encode the occurrence statistics of different objects in different place categories (office, kitchen, etc.). These statistics serve as likelihood values for a place classification procedure discussed in the next section. The Laplace succession law is used for computing the likelihood of being in a place on observing a particular object, as shown in Fig. 7. Thus, in a future exploration task, a robot could actually understand its environment and thereby classify its surroundings based on the objects it perceives.

3.6. Spatial cognition 1 — place classification

Place classification is done in an online incremental fashion, with every perceived object contributing to one or more hypotheses of previously learnt place concepts. An informal specification of the algorithm is given in Fig. 8. Hypothesis selection uses both, distinctiveness metric and a threshold metric to classify a place. The distinctiveness metric is used as a means of distinguishing between multiple competing hypotheses. A hypothesis would only be selected if it was sufficiently different from the nearest competing one. The threshold metric simply identifies potential hypothesis based on preset thresholds (in this case the prior belief — uniform

Let $X_1(x_1, y_1, z_1)$ & $X_2(x_2, y_2, z_2)$ represent the positions of two objects.

$f(X_1, X_2) =$ relative spatial information between the objects.

$P_1 \& P_2 =$ uncertainty in object positions (covariance matrices).

Position belief -

$$Bel_1(f) = F_1 P_1 F_1' + F_2 P_2 F_2'$$

where $F_1 = J_{X_1}(f)$ & $F_2 = J_{X_2}(f)$ are the Jacobians of f with respect to X_1 & X_2 respectively.

Existential belief -

$$Bel_2(f) = min(\text{ beliefs in existence of objects }).$$

Fig. 6. Belief representation for relationships between objects. Given any two objects (in 3D space) in the place reference, the distance and angle measures are computed. The existence of a relationship is subject to the belief in the existence of the objects themselves. The precision of the relationship (position belief) is subject to the relationships dependence on the objects coordinates and the uncertainty in the object positions. These are computed as shown:

$$P(object\_type|place\_type) = (N\_object\_type + 1)/(N\_place\_type + 2)$$

$P(object\_type|place\_type) =$ likelihood of observing an object of category object_type in a place of category place_type.
N_object_type = Number of times object_type was present in places of category place_type
N_place_type = Number of occurrences of places of category place_type

Fig. 7. Likelihood of being in a place of a particular place category given the occurrence of an object. Laplace succession law is used for the computation. Each object type is treated as a separate entity and is assigned a different label/name.

Algorithm

(i) Initialize each place_type with a prior belief
For each scene that the robot observes

(ii) For each object_type 'o' in unknown_place

(a) For each place_type 'p' that could have this particular object

compute posterior belief of place_type given object_type occurrence using bayesian update rule

$$P(p|o) = P(o|p) * P(p)/P(o)$$

Note $P(o|p)$ is computed as shown in Fig. 7

$P(p)$ is the prior belief that the unknown_place is of type 'p'

$P(o)$ is the belief in the observed object.

(iii) Normalize beliefs of place categories

(iv) Hypothesis selection based on distinctiveness and threshold.

Fig. 8. Algorithm for place classification. Starting with a uniform prior for all place hypotheses, the algorithm accumulates evidence (observation of objects) incrementally as the robot explores its surroundings.

value for all hypotheses in the beginning) for deeming that a classification has indeed occurred. In the case of place classification, the distinctiveness metric is given more importance as this process is about clarification between multiple competing hypotheses.

3.7. Spatial cognition 2 — place recognition

Place recognition is done by a graph matching procedure which matches both the nodes and its relationships to identify a node match. The number of node matches serves as a measure of recognition. Again, a hypothesis is finally selected on the basis of both distinctiveness and threshold metrics as explained in the context of place classification. However, in this case, the threshold metric is given much more importance than the other metric. This is because, the aim is to find the maximal common set of identically configured objects between a place the robot knows (previously mapped) and the one it currently perceives.

From a computational perspective, performing place classification prior to actual place recognition makes the graph matching procedure much more scalable by using the semantics of a place to zero down on a select set of places that could possibly match the one in consideration. Such a methodology would be visibly useful when the representation grows in complexity, size and number of instances of places in various place categories. Also, data association at the level of objects is done by not only comparing the objects under consideration but also the relationships that the particular object obeys with its neighbors. This facilitates a distinctive representation of space.

From a scientific standpoint, place recognition identifies a specific place — for instance, it can answer the question "Is this *my* office?". This problem has been addressed by several works such as [30,6], using diverse sensory information. Place classification, however, involves the conceptualization of a place — it could answer the harder and more general question "Is this *an* office?". The representation proposed here attempts to create such a conceptualization in a manner that is compatible with humans — incorporating the semantics of space. The work demonstrated here presents the first definitive steps in this direction.

3.8. Map update

A map update operation (internal graph representation is updated) is required both for handling the revisiting of places and the reobservation of objects while mapping a place. It involves the following steps —

(i) Remove unobserved nodes and relations;
(ii) Increase belief for reobserved nodes and relations;
(iii) Add new nodes and relations;
(iv) Remove nodes and relationships with low belief.

For step (i), two options exist: (a) to remove unobserved nodes and relations, (b) to reduce the beliefs and implement a gradual "forgetting" process. Both have been tried, but for the experiments in this work, the former was used.

4. Experiments

4.1. System overview and scenario

The robot platform shown in Fig. 9 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 five rooms covering about 20 m in distance. The objects used for the representation comprised different cartons annotated in the subsequent figures as *xerox, carton, logitech, elrob* and *tea*; a chair, mug, shelf, table and a book, respectively annotated as *chair, mug, shelf, table* and *book*.

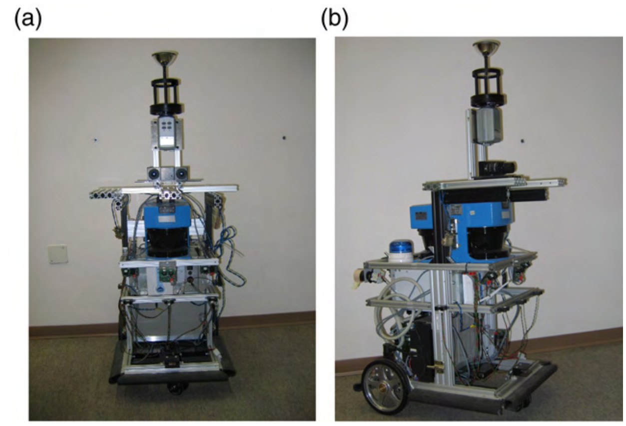


Fig. 9. The robot platform that was used for the experiments. The encoders, stereo vision system and laser scanners were used for this work.

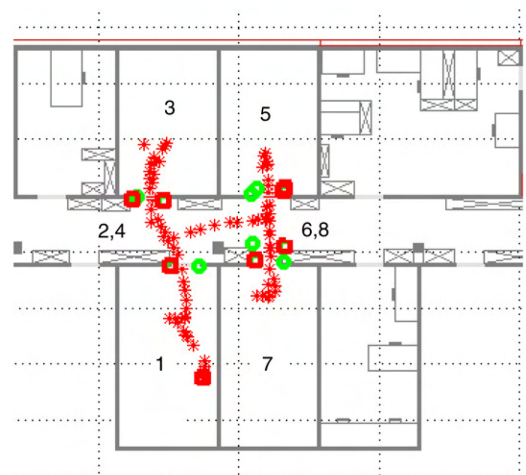


Fig. 10. Map displaying the robot path. The robot traverses four rooms crossing a corridor each time it moves from one room to another. Green (lighter)/red (darker) circles indicate the doors detected. The red (darker) circles also serve as the place references for the place explored on crossing the door. The numbers indicate the sequence in which the places were visited. The asterisks trace the actual path of the robot. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.2. Mapping

Fig. 10 shows the path of the robot. On performing the mapping process, objects and doors were recognized and the representation was formed in accordance with the methods described in the previous sections. The objects and doors recognized are shown in the object based map depicted in Fig. 11. Fig. 12 illustrates the complete probabilistic object-graph representation formed as a result of the process. Two problems were encountered in the mapping process. First, the robot occasionally observed multiple doors at the same place (due to the presence of large cupboards) on either side of the door. This is clearly seen in Fig. 10. This had no serious implications on the representation per se, but a more intuitive representation could be obtained by fusing door occurrences that are very close to each other into a single door or simply using a more sophisticated door detection

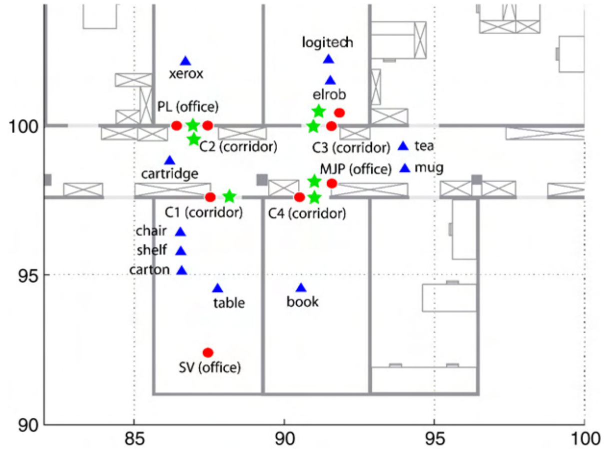


Fig. 11. Object-based map produced as a result of exploring the test environment. Red circles are the place references, blue triangles are the objects and the green stars are the doors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

methodology. Second, the robot created multiple occurrences of the corridor (clearly visible in Figs. 11 and 12) as, the topological information between places that is implicitly encoded in the representation was not used in the experiments in this work, and further, it did not see an identical set of objects through the corridor so as to be able to recognize the previously visited corridor. This topological loop closure problem (revisiting a place) will be addressed in subsequent work as discussed in Section 6.1.

4.3. *Spatial cognition — place classification/recognition*

The robot was made to traverse two previously visited places along the path shown in Fig. 13. 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–copier–carton–table–logitech–cartridge*. Fig. 14 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. Fig. 15 displays the updated internal representation of the robot.

For the second place, the robot perceived the objects in the following sequence: *door–book–cartridge–elrob–mug–tea*. The robot managed to classify the place as being a corridor but it could not recognize it as the internal representation that the robot is equipped with has multiple instances of the corridor due to the topological loop closure (revisiting the place) problem discussed earlier. Under these circumstances however, the robot would be expected to continue exploration until it crosses over

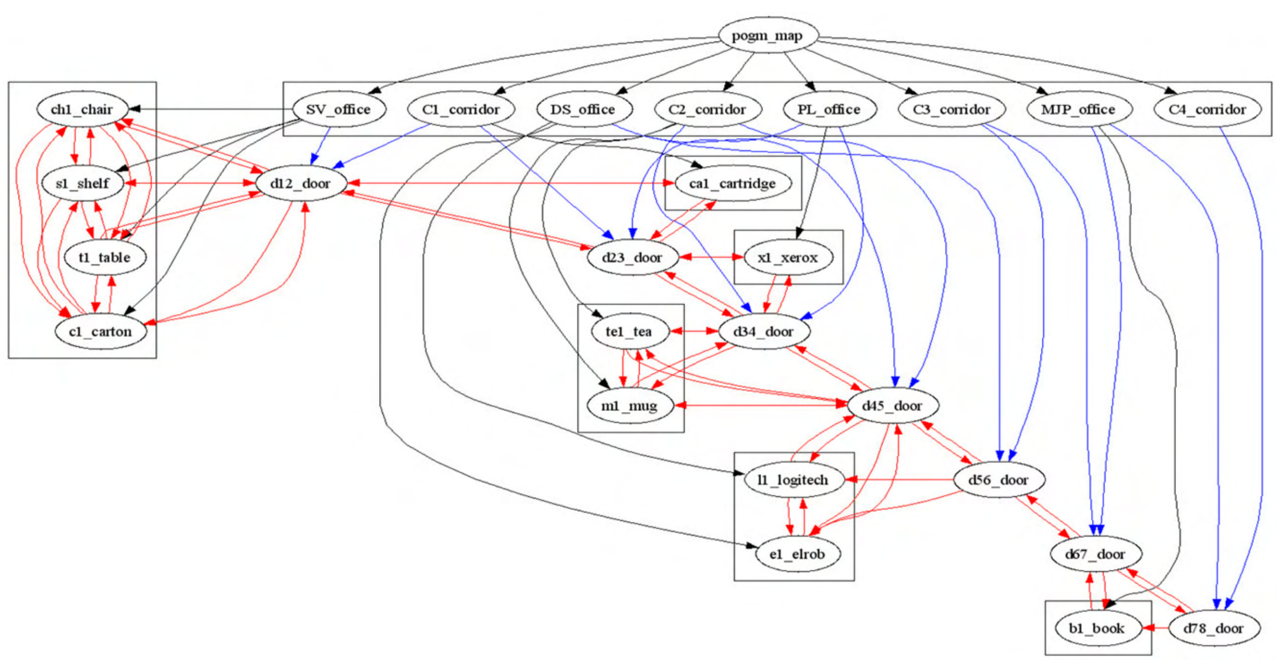


Fig. 12. Probabilistic object graph representation created as a result of exploring the path shown in Fig. 10. The map is composed of several places such as SV_office, C1_corridor etc. which correspond respectively to the places SV (office), C1 (corridor) etc. shown in Fig. 11. Each place is shown to have a set of "children" objects, these correspond directly to the objects mapped in the respective places in Fig. 11. The black lines link the place nodes to the nodes representing the objects within it. The red (lighter; between objects and door within a place) lines represent the inter-object relationships encoded. The blue (darker; between place nodes and doors) lines are meant to show the topological connection between the places through the doors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 13. The robot path for the experiment on reasoning about places. The robot traversed the path shown above. The configuration of the objects within the place was significantly changed from that in Fig. 11. The numbers indicate the sequence in which the places were visited. As before, a pair of circles represent a door detection. The red (darker) circles represent the reference for the place that is subsequently explored by the robot. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

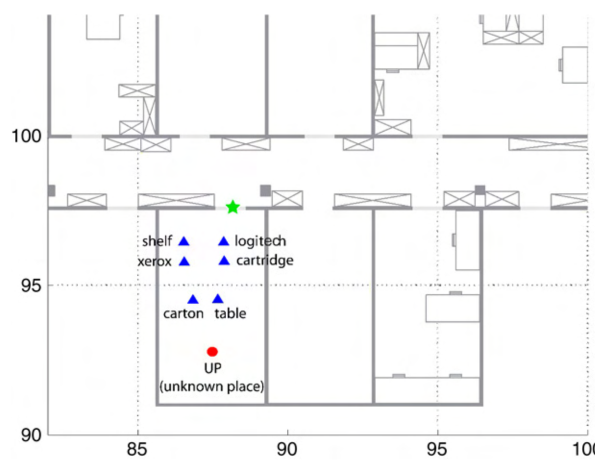


Fig. 14. First "unknown" place at the time of place recognition. As before, red circles are the place references, blue triangles are the objects and the green stars are the doors. The configuration of the objects is different from that of the same place in Fig. 11 Note — The carton is above the table and the copier is above the shelf. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to a new place, after which it would add a new node of the particular place category to its internal representation. The unknown place in this case is shown in Fig. 16.

5. User studies — A cognitive validation of the proposed representation

5.1. 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.

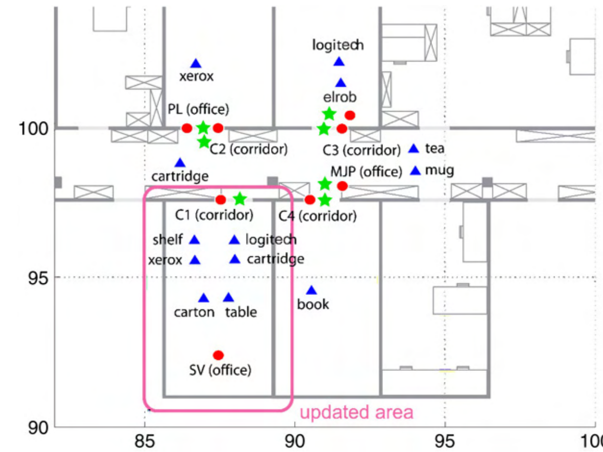


Fig. 15. Updated internal representation, as compared to the one shown in Fig. 11, of the robot after place recognition. As before, red circles are the place references, blue triangles are the objects and the green stars are the doors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

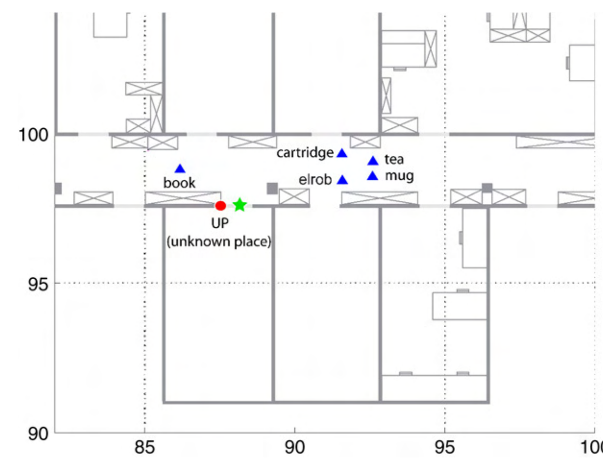


Fig. 16. Second "unknown" place (configuration of objects changed). The place is classified as a corridor. As before, red circles are the place references, blue triangles are the objects and the green stars are the doors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The study was conducted with a view of addressing issues pertaining to the larger vision of the work presented here. Thus, only results of the survey that are relevant to the aspects of the representation proposed in the current work are presented. The complete study will be reported in a more appropriate forum.

The survey comprised a questionnaire posed to fifty-two people who were taken through a course within our premises, wherein they were exposed day-to-day things and places. Their answers were categorized and tabulated/visualized. These were used to draw conclusions on various aspects of spatial cognition and representation that are relevant for this work. Questions were posed in English or French, as the user preferred. The survey was intended to be as unbiased as possible without losing the focus of the work. It was also attempted to make it as statistically representative as possible. Care was taken to ensure as far as possible that age, gender and nationality and background did not bias the survey in any way.

Table 2
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

5.2. 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 or purpose associated with a place. *Ground materials* refer to the floor material (wooden or carpeted). *Boundaries* refer to walls, doors etc. Partitions are also considered as boundary elements. 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 or represented about that place in their minds. Table 2 shows the results obtained.

As shown in the table, each of the places was best characterized by typical household objects that we find in them. 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. Some of the other factors were also found to be significant.

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 results shown in Table 3.

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. 17.

5.3. Analysis and 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 observed from Tables 2 and 3 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 and their properties) that are present in them — the underlying philosophy of this work and the direction of our future works as well. It was also found that boundaries, such as walls, doors, partitions and windows constituted an important component in describing the places and the *function* of the place (kitchen — cooking etc.) was an important descriptive element as well. The last graph (Fig. 17) seems to convey that boundary elements and the arrangement of objects are critical towards 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. The results also suggest that a representation of the kind proposed herewith will be compatible to most people. The results also go a long way in providing us with input on how our representation can be enhanced — what kinds of information can be incorporated; what states and properties of the elements could possibly be encoded; how do humans reason about their surroundings. While the results certainly are indicative of the proposed representation being human-like, the degree of similarity between the exact methodology of representing the information in our brains and that proposed here, is not measured in this study. How similar or dissimilar they are remains an open question and would need Neuroscience/Cognitive Psychology expertise to answer it. Lastly, Fig. 17 also indicates some other contributing factors towards detecting a change of place such as luminosity size, color and ground materials — these could be understood to characterize a sort of “visibility” measure of a place. Thus the results not only validate the proposed representation in a cognitive sense but provide further ideas on the future enhancements (functionality of a place etc. need to be incorporated) to this representation and how it is formed.

Table 3
Means of description of places

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

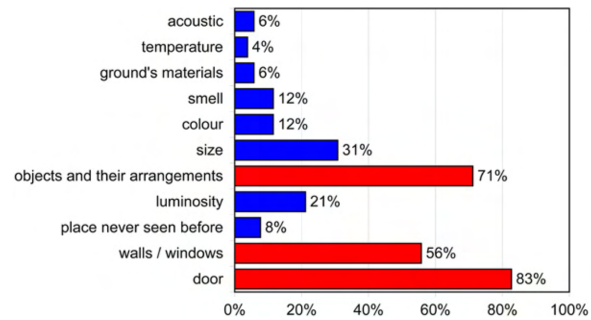


Fig. 17. Criteria to ascertain a change of place. The red (lighter) coloured bars represent the more significant factors determining a change of place. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.4. Overview of other results

The survey broadly addressed issues pertaining to spatial cognition/reasoning that are relevant to our overall approach. Only results relevant to the work presented here were mentioned earlier. Some of the issues that were also addressed in the study but not mentioned here include object representation, object recognition, the clustering of objects in space and route descriptions. A very brief overview of other results and their analysis is given here, in order to indicate the nature of the study and the general direction of the work.

It was found that the *structure* and *material* were important elements in both representing and describing objects. The *structure* of objects was the most significant recognizing element. People seemed to perceive scenes using objects and spatial relationships between them. It was found that objects were clustered into groups due to a variety of reasons including purpose and arrangement/proximity. Sometimes, a place was characterized by several smaller spatial abstractions within it, thereby indicating a hierarchy in the representation. These spatial abstractions were produced through the groups of objects and also boundary elements such as partitions in larger places. It was also observed that people specified paths through hierarchical descriptions, mostly using landmarks and spatial relationships between them in the form of directions. More details on all these results will be presented in a separate report.

6. Discussion and future work

Each subsection here discusses certain perspectives of this work and details ongoing efforts or future directions.

6.1. Localization SLAM and path planning

Localization, mapping, SLAM and path planning are fundamental problems in autonomous mobile robotics. Research in mobile robotics has put its thrust on these areas as navigation is a cardinal capability for humans and robots alike. The work presented in this paper attempts to provide a framework for performing these functions. The work also attempts to demonstrate that such a representation could be useful for going beyond these primal functions, towards conceptualization of space and actually understanding space — spatial cognition. This would be required for addressing other problems in robotics such as reasoning and interaction. It may be considered as a small step towards bridging AI and robotics.

Topological localization is done through the place classification and recognition procedures. At a local level, i.e. within any given place, the robot maintains its local pose with respect to the nearest detected reference, i.e. nearest door detection, through odometry. This implies that either the robot is equipped with well calibrated and precise encoders (the case here) or it uses poor odometry coupled with a scan matching module [31] to provide an accurate robot pose estimate within the place. Given the probabilistic object graph for the place and the observation of at least one object by the robot, metric localization of the robot with respect to the objects it has on its map is possible. If in addition, a reference origin for the place is also known, then the absolute map for the place could also be computed.

The presented work was not aimed at addressing the SLAM problem directly. It proposes a representation that could fit in the framework of the vast SLAM research that has been done and simultaneously address the deficiency in existing representations — the lack of semantics. From a technical standpoint, however, the presented work can be developed into a hierarchical (Semantic–Topological–Metric) SLAM methodology. The work presented in this report provides a preliminary insight into various aspects of the methodology. The experiments clearly showed the need for a topological loop closing (revisiting a place) mechanism. This will be immediately addressed in subsequent work to develop a full SLAM implementation that is based on objects.

Given the local probabilistic object graph of the place, path generation within it can be easily performed using basic vector algebra in accordance with the discussion on localization presented above. Path generation across places in the representation could draw on standard graph search algorithms [32]. Path planning amidst obstacles, typically household objects, could be done using a variety of approaches — a good survey of which is presented in [33].

6.2. "Cognitive" is human compatible and not necessarily human like

This work is aimed at providing an engineering solution to the general Cognitive Mapping problem. Several different works have tried to suggest formulations of the Cognitive Map [17,18]. The approach in this work follows suit. The results from the survey provide an encouraging feedback on the degree of compatibility between a Cognitive Map (the one in our brain) and the one presented in this work. The design decisions of the approach and results from the survey guarantee that the representation so formed is cognitive in that it is human compatible but are insufficient to estimate the similarity with the representation of the information in our brains. This would require expertise from neuroscience/cognitive psychology and is beyond the scope of this work. Future work will try to gain a more detailed insight on the human cognitive map and attempt to draw comparisons between it with the work presented here. Some work would also be dedicated towards conducting more detailed user studies so as to gain more insight on cognitive spatial representations — for both validation and enhancement of experimental work.

6.3. The perception system

The approach presented here depends on the objects (and any other features) observed and their inter-relationships. Thus, it would work across different places as it tries to *learn a concept or definition* of the place. In this sense it is fairly generic. However, the proposed representation is dependent on the perception system, in this case — object recognition and door detection, that provides it with the sensory data to encode. Object recognition and classification are hard and yet unsolved problems, the system has thus been demonstrated in a smaller scale with a limited number of objects. However, this does not affect the generality of the approach. Future work in this area would attempt to enhance existing perception systems so as to make them more reliable and robust.

6.4. Other future directions

The representation presented here extends the state of the art in robot mapping by building an object-based map. A robot equipped with such a human compatible representation can definitely form a first conceptualization of its surroundings. However, more work needs be done towards encoding semantic information such that the robot can be equipped with a rich, yet scalable and efficient representation. Examples of elements that are being considered for this include boundary elements such as walls, states and properties of various objects. These efforts will be carried out in the framework of addressing the fundamental scientific research directions of "representation of space", "conceptualization of space" and "spatial cognition" as applied to mobile robots.

7. Conclusion

A cognitive probabilistic representation of space based on high level features such as household objects and doors was proposed. The representation was experimentally demonstrated. First efforts towards forming a conceptualization of space using the acquired representation were elicited. Spatial cognition using the proposed representation was shown through experiments on place classification and place recognition. Relevant results of a user study that validate aspects of the approach in a cognitive sense were presented. These results also pointed towards ways of enhancing the proposed representation. In summary, a strong link was thus presented, from the sensory information acquired by a robot to the human compatible spatial concepts that the robot forms in order to represent its surroundings. The results showed that the proposed approach enables a robot to acquire a human compatible representation of space and that such a representation could be used by a robot to demonstrate spatial cognition. The proposed representation can enable a robot to acquire a greater level of spatial awareness than has been acquired by robots in the past.

Acknowledgements

The authors would like to thank Davide Scaramuzza and Laurent Winkler for their valuable help towards developing the object recognition system.

This work has been supported by the Swiss National Science Foundation (Grant No. 200021-101886) and the EU Integrated Project COGNIRON (The Cognitive Companion), 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: Exploring Artificial Intelligence in the New Millenium, Morgan Kaufmann, 2002.

[2] R. Chatila, J.P. Laumond, Position referencing and consistent world modeling for mobile robots, in: IEEE International Conference on Robotics and Automation, 1985.

[3] K.O. Arras, Feature-based robot navigation in known and unknown environments, Ph.D. Thesis, Swiss Federal Institute of Technology Lausanne (EPFL), Thesis number 2765, 2003.

[4] A. Martinelli, A. Svensson, N. Tomatis, R. Siegwart, SLAM based on quantities invariant of the robot's configuration, in: IFAC Symposium on Intelligent Autonomous Vehicles, 2004.

[5] H. Choset, K. Nagatani, Topological simultaneous localization and mapping (SLAM): Toward exact localization without explicit localization, IEEE Transactions on Robotics and Automation 17 (2001) 125–137.

[6] A. Tapus, Topological SLAM — Simultaneous localization and mapping with fingerprints of places, Ph.D. Thesis, Swiss Federal Institute of Technology Lausanne (EPFL), Thesis Number 3357, 2005.

[7] S. Thrun, Learning metric-topological maps for indoor mobile robot navigation, Artificial Intelligence 99 (1) (1998) 21–71.

[8] N. Tomatis, I. Nourbakhsh, R. Siegwart, Hybrid simultaneous localization and map building: A natural integration of topological and metric, Robotics and Autonomous Systems 44 (2003) 3–14.

[9] A. Martinelli, A. Tapus, K.O. Arras, R. Siegwart, Multi-resolution SLAM for real world navigation, in: 11th International Symposium of Robotics Research, Siena, Italy, 2003.

[10] B. Kuipers, The spatial semantic hierarchy, Artificial Intelligence 119 (2000) 191–233.

[11] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J.-A. Fernandez-Madrigal, J. Gonzalez, Multi-hierarchical semantic maps for mobile robotics, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, Edmonton, Canada, 2005, pp. 3492–3497.

[12] D.G. Lowe, Distinctive image features from scale-invariant key-points, International Journal of Computer Vision 60 (2) (2004) 91–110.

[13] D. Anguelov, D. Koller, E. Parker, S. Thrun, Detecting and modeling doors with mobile robots, in: International Conference on Robotics and Automation, ICRA, 2004.

[14] D. Kortenkamp, L.D. Baker, T. Weymouth, Using gateways to build a route map, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 1992.

[15] C. Stachniss, Ó. Martinez-Mozos, A. Rottmann, W. Burgard, Semantic labeling of places, in: International Symposium of Robotics Research, ISRR, San Francisco, CA, USA, 2005.

[16] E.C. Tolman, Cognitive maps in rats and men, Psychological Review 55 (1948) 189–208.

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

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

[19] E. Davis, Representing and Acquiring Geographic Knowledge, Pitman and Morgan Kaufmann Publishers, 1986.

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

[21] T. Fong, I. Nourbakhsh, K. Dautenhahn, A survey of socially interactive robots, Robotics and Autonomous Systems 42 (2003) 143–166.

[22] L.S. Lopes, A. Teixeira, Human–Robot Interaction through spoken language dialogue, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2000.

[23] M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, D. Brock, Spatial language for Human–Robot dialogs, IEEE Transactions On Systems, Man, and Cybernetics Part C. Applications and Reviews 34 (2) (2004) 154–167.

[24] K.O. Arras, D. Cerqui, Do we want to share our lives and bodies with robots? A 2000-people survey, Tech. Rep. 0605-001, Autonomous Systems Lab (ASL), Swiss Federal Institute of Technology Lausanne (EPFL), June 2005.

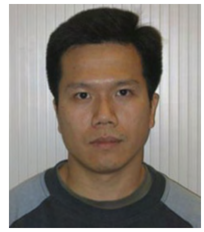
[25] R. Siegwart, K.O. Arras, S. Bouabdallah, D. Burnier, G. Froidevaux, X. Greppin, B. Jensen, A. Lorotte, L. Mayor, M. Meisser, R. Philippsen, R. Piguet, G. Ramel, G. Terrien, N. Tomatis, Robox at Expo.02: A large-scale installation of personal robots, Robotics and Autonomous Systems 42 (2003) 203–222.

[26] S. Vasudevan, V.T. Nguyen, R. Siegwart, Towards a cognitive probabilistic representation of space for mobile robots, in: IEEE International

Coference on Information Acquisition, ICIA, Weihai, Shandong, China, 2006.
[27] V.T. Nguyen, A. Martinelli, N. Tomatis, R. Siegwart, A comparison of line extraction algorithms using 2D laser rangefinder for indoor mobile robotics, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, Edmonton, Canada, 2005.
[28] A. Tapus, S. Vasudevan, R. Siegwart, Towards a multilevel cognitive probabilistic representation of space, in: International Conference on Human Vision and Electronic Imaging X, part of the IS&T/SPIE Symposium on Electronic Imaging, San Jose, CA, USA, 2005.
[29] I.K. Jung, SLAM in 3D environments using stereo vision, Ph.D. Thesis, Institut National Polytechnique, Toulouse, France, Thesis Number 04132, 2004.
[30] I. Ulrich, I. Nourbakhsh, Appearance-based place recognition for topological localization, in: IEEE International Conference on Robotics and Automation, ICRA, San Francisco, CA, USA, 2000, pp. 1023–1029.
[31] J.-S. Gutmann, C. Schlegel, AMOS: Comparison of scan matching approaches for self localization in indoor environments, in: 1st Euromicro Workshop on Advanced Mobile Robots, IEEE Computer Society Press, 1996.
[32] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, Introduction to Algorithms, MIT Press, 2001.
[33] R. Siegwart, I. Nourbakhsh, Introduction to Autonomous Mobile Robotics, MIT Press, 2004.

Shrihari Vasudevan is a Ph.D. student/Research assistant at the Autonomous Systems Lab (ASL), Swiss Federal Institute of Technology Zürich (ETHZ). He completed his baccalaureate from the University of Madras, India in May 2002. Subsequently, he studied at the University of Southern California, Los Angeles, USA, towards a Master's degree in computer science with an emphasis on intelligent robotics. On finishing his Masters degree in May 2004, he commenced work at the ASL, EPFL towards a Ph.D. degree. Since 1 July 2006, he is officially a doctoral student of the Autonomous Systems Lab, now based at ETH, Zürich. He has been a student member of the IEEE since 2000 and also a member of the Robotics and Automation Society. His primary research interests lie in the areas on cognitive spatial representations for mobile robots, spatial cognition and the conceptualization of space.

Stefan Gächter received the Master's degree in electrical engineering from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2001. After his master's studies, he spent three years as a research engineer in the field of active magnetic bearings at Koyo Seiko Co. Ltd, Japan. He joined the Autonomous Systems Lab (ASL), École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2004, where he commenced work towards a Ph.D. degree. Since July 2006, the Autonomous Systems Lab has been based at the Eidgenössische Technische Hochschule Zürich (ETHZ), Switzerland. His research interest is in computer vision for mobile robotics with emphasis on probabilistic algorithms for object classification applied to range images.



Viet Nguyen is a Ph.D. student at the École Polytechnique Fédérale de Lausanne (EPFL). He received his Bachelor's degree in software engineering from the Australian National University, Australia in 1999. He worked as a research assistant in the Laboratory for Artificial Intelligence at EPFL in 2001. In 2003, he joined the Autonomous Systems Laboratory where he started his Ph.D. study under the supervision of Prof. Roland Siegwart. Since July 2006, he has been working at the Autonomous Systems Laboratory, now at the Swiss Federal Institute of Technology Zürich (ETHZ). His research covered mobile robot navigation, SLAM, feature extraction, data association, constraint satisfaction and optimization.

Roland Siegwart is full professor for autonomous systems at ETH Zürich. He has a Diploma in Mechanical Engineering (1983) and Ph.D. in Mechatronics (1989) from ETH Zürich. In 1989–90 he spent one year as postdoctoral fellow at Stanford University. After that, he worked part time as R&D director at MECOS Traxler AG and as lecturer and deputy head at the Institute of Robotics, ETH Zürich. In 1996, he was appointed as associate and later full professor for autonomous microsystems and robots at the École Polytechnique Fédérale de Lausanne (EPFL). During his period at EPFL he was deputy head of the Nactional Competence Center for Reseach (NCCR) on Multimodal Information Management (IM2), co-initiator and founding Chairman of Space Center EPFL and Vice Dean of the School of Engineering. In 2005, he held a visiting position at NASA Ames and Stanford University. Roland Siegwart is a member of the Swiss Academy of Engineering Sciences and a board member of the European Network of Robotics (EURON). He served as Vice President for Technical Activities (2004–05) and is currently Distinguished Lecturer (2006/07) of the IEEE Robotics and Automation Society. He is member the "Bewilligungsausschuss Exzellenzinitiative" of the "Deutsche Forschungsgemeinschaft (DFG)". He is coordinator of two European Project and co-founder of several spin-off companies. His research interests are in the design and control of systems operating in complex and highly dynamic environments. His major goal is to find new ways to deal with uncertainties and enable the design of highly interactive and adaptive systems. Prominent application examples are personal and service robots, planetary exploration robots, autonomous micro-aircrafts and driver assistant systems.