Hierarchical Multi-Modal Place Categorization

Andrzej Pronobis* Patric Jensfelt*
*Centre for Autonomous Systems, Royal Institute of Technology, Stockholm, Sweden

Abstract— **In this paper we present an hierarchical approach to place categorization. Low level sensory data is processed into more abstract concept, named *properties* of space. The framework allows for fusing information from heterogeneous sensory modalities and a range of derivatives of their data. Place categories are defined based on the properties that decouples them from the low level sensory data. This gives for better scalability, both in terms of memory and computations. The probabilistic inference is performed in a chain graph which supports incremental learning of the room category models. Experimental results are presented where the shape, size and appearance of the rooms are used as properties along with the number of objects of certain classes and the topology of space.**

Index Terms— **place categoriation; graphical models; semantic mapping; machine learning**

I. Introduction

The topic of this paper is place categorization, denoting the problem of assigning a label (kitchen, office, corridor, etc) to each place in space. To motivate why this is useful, consider a domestic service robot. Such a robot should be able to "speak the language" of the operator/user to minimize training efforts and to be able to understand what the user is saying. That is, the robot should be able to make use of high level concepts such as rooms when communicating with a person, both to verbalize spatial knowledge but also to process received information from the human in an efficient way.

Besides robustness and speed, there are a number of additional desirable characteristics of a place categorization system:

C1: Categorization The system should support true categorization and not just recognition of room instances. That is, it should be able to classify an unknown room as "a kitchen" and not only recognize "the kitchen".

C2: Spatio-temporal integration The system should support integration over space and time as the information acquired at a single point rarely provides enough evidence for reliable categorization

C3: Multiple sources of information No single source of information will be enough in all situations and it is thus important to be able to make use of as much information as possible.

C4: Handles input at various levels of abstraction The system should not only be able to use low level sensor data but also higher level concepts such as objects.

C5: Automatically detect and add new categories The system should be able to augment the model with new categories identified from data.

C6: Scalability and complexity The system should be scalable both in terms of memory and computations. That is, for example, it should not be a problem to double the number of room categories.

C7: Automatic and dynamic segmentation of space The system should be able to segment space into areas (such as rooms) automatically and should be able to revise its decision if new evidence suggesting another segmentation is received.

C8: Support life-long incremental learning The robot system cannot be supplied with all the information at production time, it needs to learn along the way in an incremental fashion throughout its life.

C9: Measure of certainty There are very few cases where the categorization can be made without uncertainty due to imperfections in sensing but also model ambiguities. Ideally the system should produce a probability distribution over all categories, or at least say something about the certainty in the result.

In out previous work we have designed methods that meet C1, C3, C7 and partly C2, C4 and C9. In this paper we will improve on C4 and C9 and add C6 and C7. The main contribution of the paper relates to C4, C6 and C9.

A. Outline

In Section II presents related work and describes our contribution with respect to that. Section III describes our method and Section IV provides implementation details. Finally, Section V describes the experimental evaluation and Section VI draws some conclusions and discusses future work.

II. Related work

In this section we give an overview of the related work in the area of place recognition and categorization. Place categorization has been addressed both by the computer vision and the robotics community. In computer vision the problem is often referred to as scene categorization. Although also related, object categorization methods are not covered here. However, we believe that objects are key to understanding space and we will include them in our representation but will make use of standard methods for recognizing/categorizing them. Table II maps some of the methods presented below to the desired characteristics presented in the previous section.

In computer vision one of the first works to address the problem of place categorization is [19] based on the so called "gist" of a scene. One of the key insights in the paper is that the context is very important for recognition and categorization of both places and objects and that these processes are intimately connected. Place recognition is formulated in the context of localization and information about the connectivity of space is utilized in an HMM. Place categorization is also addressed using a HMM. In [23] the problem of grouping images into semantic categories is addressed. It is pointed out that many

	C1: Categorization	C2: Spatio/temporal	C3: Multi source	C4: Multi levels	C5: Novelty detection	C6: Scalability	C7: Segmentation	C8: Incremental	C9: Uncertainty
[19]	X	x							X
[23]	X								
[20]									x
[10]	X								
[12]	X	x	X	x			x		
[14]									
[9, 15]								X	
[13]									x
[26]	x		x	x					
[16]	X	x	X						x
[24]	X	x							
[18]									X
[17]	X	X					X	X	X
[22]	X								X
[21]		x			X			X	
This work	X	x	X	X		X	x	x	X

TABLE I
CHARACTERIZING SOME OF THE PLACE CATEGORIZATION WORK BASED ON THE DESIRABLE CHARACTERISTICS FROM SECTION I.

natural scenes are ambiguous and the performance of the system is often quite subjective. That is, if two people are asked to sort the images into different categories they are likely to come up with different partitions. [23] argue that *typicality* is a key measure to use in achieving meaningful categorizations. Each cue used in the categorization should be assigned a typicality measure to express the uncertainty when used in the categorization, i.e. the saliency of that cue. The system is evaluated in natural outdoor scenes. In [3] another method is presented for categorization of outdoors scenes based on representing the distribution of codewords in each scene category. In [25] a new image descriptor, PACT, is presented and shown to give superior results on the datasets used in [19, 3].

In robotics, one of the early systems for place recognition is [20] where color histograms is used to model the appearance of places in a topological map and place recognition performed as a part of the localization process. Later [10] uses laser data to extract a large number of features used to train classifiers using AdaBoost. This system shows impressive results based on laser data alone. The system is not able to identify and learn new categories: adding a new category required off-line re-training, no measure of certainty and it segmented space only implicitly by providing an estimate of the category for every point in space. In [12] this work is extended to also incorporate visual information in the form of object detections. Furthermore, this work also adds a HMM on top of the point-wise classifications to incorporate information about the connectivity of space and make use of information such as offices are typically connected to corridors. In [14] a vision only place recognition system is presented. Support Vector Machines (SVMs) are used as classifiers. The characteristics are similar to those of [10]; cannot identify and learn new categorizes on-line, only works with data from a single source and classification was done frame by frame. In [9, 15] a version of the system supporting incremental learning is presented. The other limitations remains the same. In [13] a measure of confidence is introduce as a means to better fuse different cues and also provide the consumer of the information with some information about the certainty in the end result. In [16] the works in [10, 14] are combined using an SVM on top of the laser and vision based classifiers. This allows the system to learn what cues to rely on in what room category. For example, in a corridor the laser based classifier is more reliable than vision whereas in rooms the laser does not distinguish between different room types. Segmentation of space is done based on detecting doors that are assumed to delimit the rooms. Evidence is accumulated within a room to provide a more robust and stable classification. It is also shown that the method support categorization and not only recognition. In [24] the work from [25] is extended with a new image descriptor, CENTRIS, and a focus on visual place categorization in indoor environment for robotics. A database, VPC, for benchmarking of vision based place categorization systems is also presented. A Bayesian filtering scheme is added on top of the frame based categorization to increase robustness and give smoother category estimates. In [17] the problem of place categorization is addressed in a drastically different and novel way. The problem is cast in a fully probabilistic framework which operates on sequences rather than individual images. The method uses change point detection to detect abrupt changes in the statistical properties of the data. A Rao-Blackwellized particle filter implementation is presented for the Bayesian change point detection to allow for real-time performance. All information deemed to belong to the same segment is used to estimate the category for that segment using a bag-of-words technique. In [27] a system for clustering panoramic images into convex regions of space indoors is presented. These regions correspond roughly with the human concept of rooms and are defined by the similarity between the images. In [21] panoramic images from indoor and outdoor scenes are clustered into topological regions using incremental spectral clustering. These clusters are defined by appearance and the aim is to support localization rather than human robot interaction. The clusters therefore have no obvious semantic meaning.

As mentioned above [12] makes use of object observations to perform the place categorization. In [5] objects play a key role in the creation of semantic maps. In [18] a 3D model centered around objects is presented as a way to model places and to support place recognition. In [22] a Bayesian framework for connecting objects to place categories is presented. In [26] the work in [12] is combined with detections of objects to deduce the specific category of a room in a first-order logic way.

A. Contributions

In this paper we contribute a method for hierarchical categorization of places. The method can make use of a very diverse set of input data, potentially also including spoken dialogue. We make use of classical classifiers (SVM in our case, building on the work [16]) and a graphical model to fuse information at a higher level. The categorical models for rooms are based on so called *properties* of space, rather than the low level sensor characteristics which is the case in most of the other work presented above. This also means that a new category could be defined without having the need to re-train from the sensor data level. The properties decouples the system. The introduction of properties also makes the system more scalable as the low level resources (memory for models and computations for classifiers) can be shared across room categorizers. The system we present still rely on the detection of doors like [16] but the graphical model allows us to add and remove these doors and thus change the segmentation of space. The system will automatically adjust the category estimates for each room taking into account the new topology of space.

III. Hierarchical multi-modal categorization

We pose the problem of place categorization as that of estimating the probability distribution of category labels, c_i, over places, p_j. That is, we want to estimate $p(c_i, p_j)$. We consider a discrete set of places rather than a continuous space. In our implementation the places are spread out over space like bread crumbs every one meter [26]. The places become nodes (representing free space) in a graph covering the environment. Edges are added when the robot has traveled directly between two nodes.

In our previous work [26] we performed place categorization by combining a room/corridor classifier (based on [10]) with an ontology that related objects to specific room types. For example, we inferred being in a living room if the classification system reported a room and a sofa and a TV set were found (objects associated with a living rooms according to the ontology). This method had some clear and severe shortcomings that made it only appropriate for illustrating ideas rather than being a real world categorization system in anything but simple and idealized test scenarios. Furthermore, because the system was unable to retract inferred information any categorization was crisp and set in stone. Conceptually the solution has several appealing traits. It allowed us to teach the system, at a symbolic level, to distinguish different room categories simply by assigning specific objects to them. It combined information from low level sensor data (to classify room/corridor) with high level concepts such as objects.

The place categorization system in this paper provides a principled way to maintain the advantages mentioned above even in natural environments. Our approach is based on the insight that what made the previous system easy to re-train was that the categorization was based on high level concepts rather than on low level sensor data. For this purpose, we introduce what we call *properties* of space where in the previous system the properties corresponded to the existence of certain types of objects. In general these properties could be related to, for example, the size, shape and appearance of a place.

The introduction of properties decomposes our approach hierarchically. The categories are defined based on the properties and the properties are defined based on sensor data, either directly or in further hierarchies. This is closely related to the work on part based object recognition and categorization [2]. The property based decomposition buys us **better scalability** in several ways. Instead of having to build a model from the level of sensor data for every new category, we can reuse the low level concepts. This **saves memory** (models for SVMs can be hundreds of megabytes in size) and **saves computations** (calculations shared across categories). The introduction of properties also **makes training easier**. Once we have the models for the properties, training the system for a new category is decoupled from low level sensor data. The properties can be seen as high level basis functions on which the categories are defined, providing a significant dimensionality reduction. The graph made up of the free space nodes can be used to impose topological constraints on the places as well and help lay the foundation for the segmentation process.

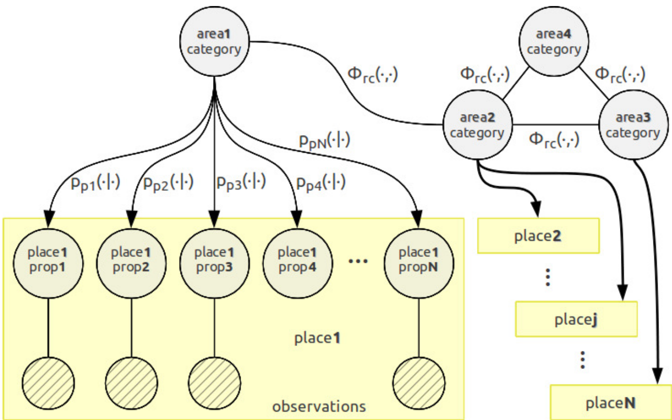

Fig. 1. Structure of the graphical model for the places showing the influence of the properties and the topology on the categorization and segmentation.

We use a graphical model to structure the problem, starting from the place graph. More precisely we will use a probabilistic chain graph model [7]. Chain graphs are a natural generalization of directed (Bayesian Networks) and undirected (Markov Random Fields) graphical models. As such, they allow for modelling both “directed” causal as well as “undirected” symmetric or associative relationships, including circular dependencies. Figure 1 shows our graphical model. The structure of model depends on the topology of the environment. Each discrete place is represented by a set of random variables connected to variables representing the semantic category of areas. Moreover, the category variables are connected by undirected links to one another according to the topology of the environment. The potential functions $\phi_{rc}(\cdot,\cdot)$ represent the knowledge about the connectivity of areas of certain semantic categories (e.g. kitchens are typically connected to corridors). The remaining variables represent properties of space. These can be connected to observations of features extracted directly from the sensory input. Finally, the functions $p_{p_1}(\cdot|\cdot)$, $p_{p_2}(\cdot|\cdot)$, $\ldots$, $p_{p_N}(\cdot|\cdot)$ model spatial properties.

The joint density f of a distribution that satisfies the Markov property associated with a chain graph can be written as [7]:

$$f(x) = \prod_{\tau \in T} f(x_\tau | x_{pa(\tau)}),$$

where $pa(\tau)$ denotes the set of parents of vertices τ. This corresponds to an outer factorization which can be viewed as a directed acyclic graph with vertices representing the multivariate random variables X_τ, for τ in T (one for each chain component). Each factor $f(x_\tau|x_{pa(\tau)})$ factorizes further into:

$$f(x_\tau|x_{pa(\tau)}) = \frac{1}{Z(x_{pa(\tau)})} \prod_{\alpha \in A(\tau)} \phi_\alpha(x_\alpha),$$

where $A(\tau)$ represents sets of vertices in the moralized undirected graph $\mathcal{G}_{\tau \cup pa(\tau)}$, such that in every set, there exist edges between every pair of vertices in the set. The factor Z normalizes $f(x_\tau|x_{pa(\tau)})$ into a proper distribution.

In order to perform inference on the chain graph, we first convert it into a factor graph representation [8]. To meet the real time constraints posed by most robotics applications we then use an approximate inference engine, namely Loopy Belief Propagation [11].

IV. Implementatation

In our implementation, each object class results in one property, encoding the expected/observed number of such objects. In addition, we use of the following properties:

- *shape* (e.g. elongated, square) –
  Extracted from laser data
- *size* (e.g. large (compared to other typical rooms)) –
  Extracted from laser data
- *appearance* (e.g. office-like appearance) –
  Extracted from visual data
- *doorway* (is this place in a doorway) –
  Extracted from laser data

In indoor environments, rooms tend to share similar functionality and semantics. In this work we cluster places into areas based on the door property of places (using door detector from [16]). The doorway property is considered to be crisp. The door places are not part of the chain graph but rather act as edges between areas. However, the graphical model allows us to easily change the topology if new information becomes available. The overall system therefore performs segmentation automatically and the dynamic nature of it is based on re-evaluating the existence of doors. Figure 2 illustrates how the places (small circles) are segmented into areas (ellipses) by the existence of doors (red small circles) and how this defines the topology of the areas.

We build on the work in [16] when defining the property categorizers for shape, size and appearance (see [16] for details). The categorizers are based on Support Vector Machines (SVMs) and the models are trained on features extracted directly from the robot's sensory input. A set of simple geometrical features [10] are extracted from laser range data in order to train the shape and size models. The appearance models are build from two types of visual cues, global, Composed Receptive Field Histograms (CRFH) and local based on the SURF features discretized into visual words [1]. The two visual features are further integrated using the Generalized Discriminative Accumulation Scheme (G-DAS [16]). The models are trained from sequences of images and laser range data recorded in multiple instances

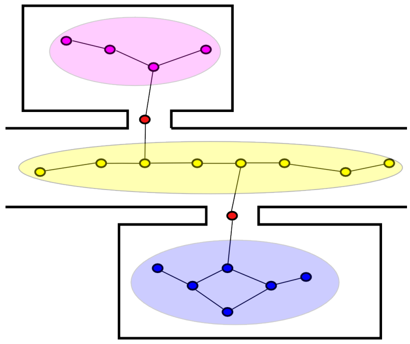

Fig. 2. The set of places, $\{p_i\}$, is segmented into areas based on the door places. The doors form the edges in the topological area graph.

of rooms belonging to different categories and under various different illumination settings (during the day and at night). By including several different room instances into training, the acquired model can generalize sufficiently to provide categorization rather than instance recognition. The estimate for the uncertainty in the categorization results is based on the distances between the classified samples and discriminative model hyperplanes (see [13] for details).

To learn the probabilities associated with the relations between rooms, objects, shapes, sizes and appearances we analyzed common-sense resources available online (for details see [6]) and the annotated data in the COLD-Stockholm database[1]. The relations between rooms and objects were bootstrapped from part of the *Open Mind Indoor Common Sense* database[2]. The object-location pairs found through this process were then used to form queries on the form '*obj* in the *loc*' that were fed to an online image search engine. The number of hits returned was used as a basis for the probability estimate. Relations that were not found this way were assigned a certain low default probability not to rule them out completely.

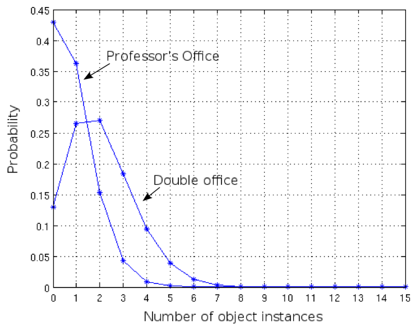

Fig. 3. The Poisson distributions modelling the existence of a certain number of objects in a room on the example of computers present in a double office and a professor's office.

The conditional probability distributions $p_{p_i}(\cdot|\cdot)$ for the object properties are represented by Poisson distributions. The parameter λ of the distribution allows to set the expected number of object occurrences. This is exemplified in Fig. 3

[1] `http://www.cas.kth.se/cold-stockholm`
[2] `http://openmind.hri-us.com/`

which shows two distributions corresponding to the relation between the number of computers in a double office and a professor's office. In the specific case of the double office, we set the expected number of computers to two. In all remaining cases the parameter λ is estimated by matching $p_\lambda(n = 0)$ with the probability of there being no objects of a certain category according to the common sense knowledge databases.

V. EXPERIMENTS

A. Experimental Setup

The COLD-Stockholm database contains data from four floors. We divide the database into two subsets. For training and validation, we used the data acquired on floors 4, 5 and 7. The data acquired on floor 6 is used for evaluation of the performance of the property classifiers and for the real-world experiment.

For the purpose of the experiments presented in this paper, we have extended the annotation of the COLD-Stockholm database to include 3 room shapes (elongated, square and rectangular), 3 room sizes (small, medium and large) as well as 7 general appearances (anteroom-, bathroom-, hallway-, kitchen-, lab-, meetingroom- and office-like). The room size and shape, were decided based on the length ratio and maximum length of edges of a rectangle fitted to the room outline. These properties together with 6 object types defined 11 room categories used in our experiments, see Figure 5.

B. Evaluation of Property Categorizers

The performance of each of the property categorizers was evaluated in separation. Training and validation datasets were formed by grouping rooms having the same values of properties. Parameters of the models were obtained by cross-validation. All training and validation data were collected together and used for training the final models which were evaluated on test data acquired in previously unseen rooms. Table II presents the results of the evaluation. The classification rates were obtained separately for each of the classes and then averaged in order to exclude the influence of unbalanced testing set. As can be seen all classifiers provided a recognition rates above 80%. Furthermore, integrating the two visual cues (CRFH and BOW-SURF) increased the classification rate of the appearance property by almost 5%. From the confusion matrices in Fig. 4 we see that the cases with confusion occurs between property values being semantically close.

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

TABLE II

CLASSIFICATION RATES FOR EACH OF THE PROPERTIES AND CUES.

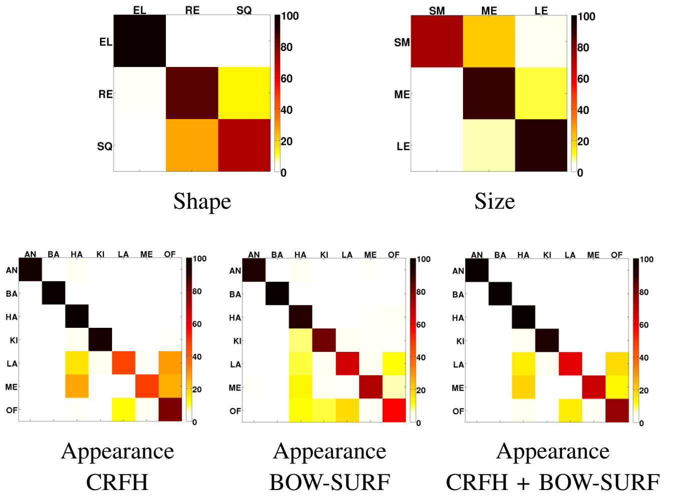

Fig. 4. Confusion matrices for the evaluation of the property categorizers.

C. Real-world experiments

In the real-world experiment the robot was manually driven through the environment using a joystick. The robot started with only the models obtained in the evaluation of the property categorizers. Laser based SLAM [4] was performed while moving and new places were added every meter traveled into unexplored space. The robot was driven through 15 different rooms while performing real-time place categorization without relying on any previous observations of this particular part of the environment. The object observations where provided by human input. The information comes into the change graph in exactly the same was as would real object detections.

Figure 5 illustrates the performance of the system during part of a run. The 11 categories can be found along the vertical axis. The ground truth for the room category is marked with a box with thick dashed lines. The Maximum a posteriori (MAP) estimate for the room category is indicated with white dots. The system correctly identified the first two rooms as a hallway and a single office using only shape, size and general appearance (no objects were found). The next room was properly classified as a double office. The MAP estimate switches to professors office for a short while when one computer is found and switches back again when a second if found. After some initial uncertainty where the MAP switches category several times the next room is classified as a double office until the robot finds a computer at which point it switches to professor's office. Later the robot enters a robot lab which according to its models is very similar to a computerlab. Initially there is a slightly higher probability for the hypothesis that it is a computerlab, but once the robot detects a robot arm the robotlab hypothesis completely dominates. The next non-hallway room is a single person office currently occupied by a bunch of Master's students. Because of its current appearance, the best match is a double office. The robot continues and the rest of the categorizations are correct. The system is able to perform the categorization in real-time as can be seen these preliminary results indicate that the accuracy is quite good.

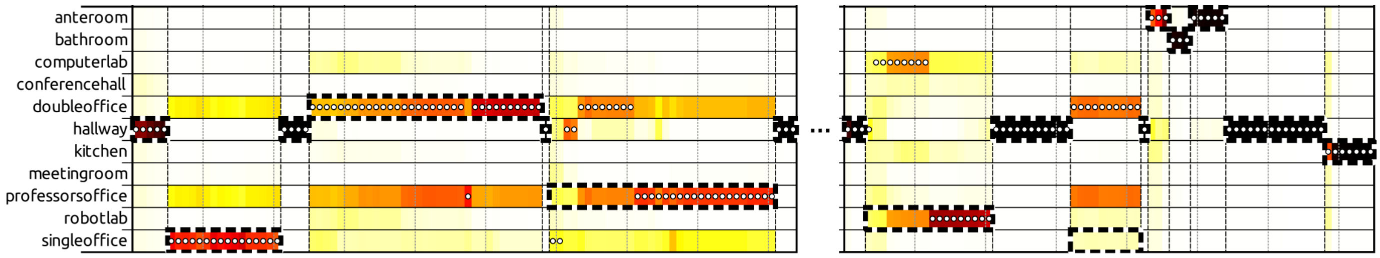

Fig. 5. Visualization of the beliefs about the categories of the rooms. The room category ground truth is marked with thick dashed lines while the MAP value is indicated with white dots.

VI. Conclusions and Future Work

In this paper we have presented a probabilistic framework combining multi-modal and uncertain information in a hierarchical fashion. So called properties were introduces as a way to model high level characteristics of the environment. These properties gave us a way to decouple the categorization into categorization of the properties based on low level sensor information and categorization of high level concepts such as rooms based on the properties. A chain graph model was used for the probabilistic inference. We provided an initial evaluation of the system which indicates that it works in well practice.

Part of the future work is to evaluate the system more thoroughly. It is important to note that we are not able to evaluate our system on other databases such as VPC [24] as it does not contain laser data. We will also investigate the use of the place categorization system in semantic mapping.

Acknowledgment

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

References

[1] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In *Proc. of ECCV'06*, 2006.

[2] G. Bouchard and B. Triggs. Hierarchical part-based visual object categorization. In *Proc. of CVPR'05*, 2005.

[3] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *Proc. of CVPR'05*, 2005.

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

[5] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J. A. Fernandez-Madrigal, and J. Gonzalez. Multi-hierarchical semantic maps for mobile robotics. In *Proc. of IROS'05*, 2005.

[6] M. Hanheide, N. Hawes, C. Gretton, A. Aydemir, H. Zender, A. Pronobis, J. Wyatt, and M. Göbelbecker. Exploiting probabilistic knowledge under uncertain sensing for efficient robot behaviour. In *Proc. of IJCAI'11*, 2011.

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

[8] H.-A. Loeliger. An introduction to factor graphs. *IEEE Signal Processing Magazine*, 21:28–41, January 2004.

[9] J. Luo, A. Pronobis, B. Caputo, and P. Jensfelt. Incremental learning for place recognition in dynamic environments. In *Proc. of IROS'07*, 2007.

[10] O. Martínez Mozos, C. Stachniss, and W. Burgard. Supervised learning of places from range data using adaboost. In *Proc. of ICRA'05*, 2005.

[11] J. M. Mooij. libDAI: A free and open source C++ library for discrete approximate inference in graphical models. *Journal of Machine Learning Research*, 11:2169–2173, August 2010.

[12] O. Martínez Mozos, R. Triebel, P. Jensfelt, A. Rottmann, and W. Burgard. Supervised semantic labeling of places using information extracted from laser and vision sensor data. *Robotics and Autonomous Systems*, 55(5):391–402, May 2007.

[13] A. Pronobis and B. Caputo. Confidence-based cue integration for visual place recognition. In *Proc. of IROS'07*, 2007.

[14] A. Pronobis, B. Caputo, P. Jensfelt, and H. Christensen. A discriminative approach to robust visual place recognition. In *Proc. of IROS'06*, 2006.

[15] A. Pronobis, Jie Luo, and B. Caputo. The more you learn, the less you store: Memory-controlled incremental SVM for visual place recognition. *Image and Vision Computing*, March 2010.

[16] A. Pronobis, O. Martinez Mozos, B. Caputo, and P. Jensfelt. Multi-modal semantic place classification. *International Journal of Robotics Research*, 29(2-3), February 2010.

[17] A. Ranganathan. Pliss: Detecting and labeling places using online change-point detection. In *Proc. of RSS'10*, 2010.

[18] A. Ranganathan and F. Dellaert. Semantic modeling of places using objects. In *Proc. of RSS'07*, 2007.

[19] A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin. Context-based vision system for place and object recognition. In *Proc. of ICCV'03*, 2003.

[20] I. Ulrich and I. Nourbakhsh. Appearance-based place recognition for topological localization. In *Proc. of ICRA'00*, 2000.

[21] C. Valgren and A. Lilienthal. Incremental spectral clustering and seasons: Appearance-based localization in outdoor environments. In *Proc. of ICRA'08*, 2008.

[22] S. Vasudevan and R. Siegwart. Bayesian space conceptualization and place classification for semantic maps in mobile robotics. *Robotics and Autonomous Systems*, 56:522–537, June 2008.

[23] J. Vogel and B. Schiele. A semantic typicality measure for natural scene categorization. *Pattern Recognition*, pages 195–203, 2004.

[24] Jianxin Wu, H. I. Christensen, and J. M. Rehg. Visual place categorization: Problem, dataset, and algorithm. In *Proc. of IROS'09*, 2009.

[25] Jianxin Wu and J. M. Rehg. Where am i: Place instance and category recognition using spatial pact. In *Proc. of CVPR'08*, 2008.

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

[27] Z. Zivkovic, O. Booij, and B. Kröse. From images to rooms. *Robotics and Autonomous Systems, special issue From Sensors to Human Spatial Concepts*, 55(5):411–418, May 2007.