RI-Small: Statistical Relational Models for Semantic Robot Mapping

1 Introduction

In the last decade, the mobile robotics community has made tremendous progress in the development of efficient yet robust techniques for dealing with noisy sensor data. A major reason for this progress is a better understanding of probabilistic algorithms, which have become the most successful and most widely applied tools for dealing with uncertainty in mobile robotics [11, 23, 144]. The key idea of these techniques is to represent uncertainty using probability distributions and to make use of structural independences to make inference tractable. Bayesian filtering techniques, for example, have been applied with great success to state estimation problems ranging from robot localization [26, 65, 40, 8, 124] to map building [31, 142, 102, 113, 56, 147] to people and object tracking [103, 37, 129, 74].

However, current robot systems are still very limited in their capabilities to reason about environments in high-level concepts such as places and objects. Such concepts are extremely important especially for applications that require robots to interact and collaborate with humans, such as health care [34, 151, 99, 150], personal assistants for the elderly [10, 101], or search and rescue [72, 91]. For instance, in search and rescue tasks, a mobile robot that can reason about objects such as doors, and places such as rooms is able to coordinate with first responders in a much more natural way, being able to accept commands such as "Search the room behind the third door on the right of this hallway", and conveying information such as "There is a wounded person behind the desk in that room". As another example, consider autonomous vehicles navigating in urban areas. While the recent success of the DARPA Urban Challenge [21, 146] demonstrates that it is possible to develop autonomous vehicles that can navigate safely in constraint settings, the successful application of such systems in more realistic, populated urban areas requires the ability to distinguish between objects such as cars, people, buildings, trees, and traffic lights.

The goal of this project is to develop a reasoning and learning framework that enables robots to build rich, semantic maps of their environments. Semantic maps describe environments in terms of places, such as rooms, hallways, streets, or parking lots, and objects, such as doors, walls, cars, people, or trees. Current mapping techniques lack the expressive power necessary to generate such representations. Our framework, called *Relational Semantic Maps (RS-Maps)*, overcomes limitations of existing mapping techniques by building on recent advances in statistical relational learning [52]. Relational models extend (propositional) probabilistic models such as Bayesian networks [114, 29, 53, 106], Markov Random Fields [49, 7, 155], or Conditional Random Fields [79, 138] to (first-order) relational domains. By using first-order logical languages to specify the structures of the underlying probabilistic models, relational techniques are extremely expressive and flexible. However, building semantic maps still requires us to address various research challenges:

- *Complex sensor data:* RS-Maps will combine camera and laser range-data to detect and model different object and place types. In addition to high-dimensional, continuous feature vectors extracted from these sensors, RS-Maps will leverage visual object detectors trained on existing object recognition datasets. So far, statistical relational learning techniques have not been applied to such complex, continuous feature sets, and we will develop new learning algorithms that can incorporate such features into relational models.
- *Complex relationships:* Individual objects are difficult to recognize when analyzed in isolation [108, 149, 123]. Therefore, a semantic mapping technique has to reason about individual

objects, their spatial relationships to other objects, and their spatial context, such as the types of places they can be found in. This can result in large, strongly connected graphical models in which efficient inference and learning is extremely challenging.

- *Uncertain model structure:* The number of objects and places in an environment is unknown initially and needs to be inferred from sensor data. Due to such uncertainties, the structure of the underlying probabilistic models can change during inference. For instance, different segmentations of space result in different numbers of places with different properties. While relational techniques are designed to reason about changing model structures, scalable inference in such models is still an open problem.
- *Learning from experience:* A robot must be able to learn from previously explored environments and apply this knowledge to unknown environments. Thus, the parameters of semantic maps must be learned in a manner that is transferable between different environments. Furthermore, the learned concepts must be compatible with the human perception of environments. While this can be achieved using human-labeled training data, the complexity of the learning task requires the incorporation of additional, unlabeled data.

Research Projects

In this project, we will develop Relational Semantic Maps (RS-Maps), a statistical relational framework for building high-level maps of indoor and outdoor spaces.

- *Building RS-Maps*: We will develop and implement the RS-Map framework in the context of building maps of indoor environments and urban spaces. Indoor data will be collected by a mobile robot equipped with an omni-directional camera and a 3D laser range-scanner; outdoor data will be collected by a car equipped with multiple cameras and laser scanners. Our RS-Maps build on and substantially extend Relational Markov Networks, a statistical relational technique developed in the context of web page classification [139]. To deal with the complexity of sensor data, we will develop learning techniques that can automatically extract useful feature combinations from thousands of continuous features describing the shape, appearance, and relationships of objects and places. In order to perform scalable inference in RS-Maps, we extend Relational Markov Networks by introducing *computational clique templates*, a novel framework that performs different types of inference (exact, sampling, Gaussian, *etc.*) in different parts of the probabilistic model. Structural uncertainty will be handled by building on recent advances in Markov Chain Monte Carlo sampling techniques. The parameters of RS-Maps will be learned from partially labeled data, leveraging training data sets for visual object recognition.
- *Task-related evaluation of RS-Maps (undergraduate research projects)*: While the graduate level research focuses on extracting RS-Maps from sensor data, we will also evaluate the usefulness of RS-Maps in the context of tasks inspired by indoor office delivery and search and rescue scenarios. One undergraduate research project will be to develop a path planner for RS-Maps. Using this planner, the task of the robot could be to find a specific object using guidance of the form: "Take a picture of the desk in the room at the end of the hallway". Another project will be to guide a person to a specific location or object using high-level information extracted automatically from an RS-Map: "Go down the hallway, go into the second hallway on the left, and enter the room next to the fire extinguisher".

Before we present our proposed framework for semantic mapping, we first discuss related work in mobile robot map building and Relational Markov Networks (RMN).

2 Research Background

2.1 Mobile Robot Mapping

Generating rich representations of environments is a fundamental problem in mobile robotics. Over the last decade, much of the research in map building has focused on the simultaneous localization and mapping (SLAM) problem, *i.e.*, the problem of estimating the joint posterior over the robot's location and the map of the environment. This research has produced various techniques that are able to efficiently build maps of large scale, cyclic environments [55, 31, 24, 102, 142, 143, 111, 57, 15, 33, 113, 69, 61, 68, 33, 57, 144, 43]. However, existing mapping techniques have only limited expressive power. For example, occupancy grids and related techniques [104, 156, 134, 54, 61] are metric maps that represent whether or not a small patch in an environment is occupied by an obstacle. Even recently introduced 3-D volumetric maps extracted from laser range-scans do not provide information beyond occupancy grid cells or planar surfaces [105, 143, 90, 58].

Most Kalman filter based SLAM techniques rely on landmarks to represent environments [27, 81, 31, 102, 111, 113]. Typically, these landmarks are lines or point features such as corners, and no meaning is associated with the features. Topological and hybrid approaches use richer landmarks to specify distinctive locations such as hallways or intersections [65, 67, 24, 70, 141, 148, 121, 95, 12]. However, these approaches only extract very coarse structural information by ignoring valuable metric information, and the parameters of these models are typically tuned manually. Vision-based mapping algorithms rely on low-level features not directly related to the physical structure of an environment [110, 63, 152, 130, 19, 125].

Recently, several research groups have applied machine learning techniques to classify objects and places in both indoor and outdoor environments. Anguelov and colleagues use hierarchical Bayesian reasoning to detect doors in a building [5, 6]. In our own work we developed an approach based on Relational Markov Networks to classify lines extracted from 2D laser scans into doors or walls in single hallways [89]. Posner and colleagues combine 3D laser range data with camera information to classify surface types such as brick, concrete, grass, or pavement in outdoor environments [116, 117]. While they label every laser scan return independently of other laser returns, we showed how to use Conditional Random Fields to jointly classify beams in 2D laser scans into seven object types (car, person, wall, tree trunk, foliage, grass, other) by combining laser shape information and camera information [32].

For indoor place labeling, Martinez and colleagues apply AdaBoost and associative Markov networks to distinguish between different types of places [95, 96]. We developed Conditional Random Fields for indoor place labeling [47]. Figure 1 shows an occupancy map of the Intel Research Lab along with the different places detected by our approach. The middle panel indicates the different types of places classified by the Conditional Random Field, the parameters of which are learned from labeled training maps. During learning, the Conditional Random Field automatically selects useful features from several hundred geometric features extracted from the occupancy map. The labeling can be used to generate a topological-metric map of the environment, describing rooms, hallways, their connections and their spatial layout (right panel).

While these existing techniques demonstrate the feasibility of laser and camera based object detection and place labeling, they only focus on limited, isolated aspects of the semantic mapping task. None of these approaches aims at building semantic maps that describe environments in terms of places and objects. Building such maps for large scale environments can involve thousands of objects of a variety of types and complex relations between them. To extract such object and place descriptions from raw sensor data, it is necessary to use probabilistic models that are flexible and compact, and at the same time support efficient inference and learning. Currently used

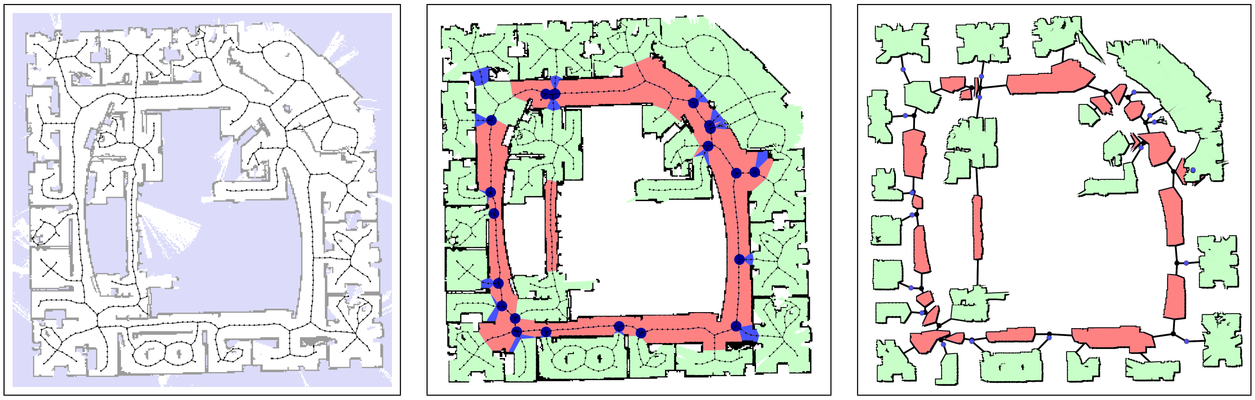

Figure 1: Map of the Intel Research Lab in Seattle: (left) Occupancy grid map built via SLAM along with automatically extracted Voronoi graph. (middle) The Voronoi graph labeled via a Conditional Random Field defines a place type for each point in the map [47]. Hallways are colored gray (red), rooms light gray (green), doorways dark grey (blue), and junctions are indicated by black circles. (right) Topological-metric map extracted from the labeled graph.

techniques are not able to solve this task. This is due to the fact that these models are *propositional* representations of a domain, that is, they only reason about a fixed, instantiated set of objects, and fixed relations between them. The goal of this project is to overcome these limitations by developing a novel inference and learning framework for semantic mapping. This framework builds on Relational Markov Networks, which we describe next.

2.2 Relational Markov Networks

Statistical relational models were introduced to overcome limitations of propositional probabilistic models [50, 4, 126, 139, 122, 52]. Relational models combine first-order logical languages with probabilistic graphical models. Intuitively, a relational probabilistic model is a *template* for propositional models such as (dynamic) Bayesian network, Markov Random Fields, or Conditional Random Fields (similar to how first-order logic formulas can be instantiated to propositional logic). Templates are defined over object classes through logical languages such as Horn clauses, frame systems, SQL, and full first-order logic. Given data, these templates are then *instantiated* to generate propositional models, on which inference and learning is performed. Relational probabilistic models use high level languages to describe systems involving complex relations and uncertainties. Since the parameters are defined at the level of classes, they are shared by the instantiated networks.

Relational Markov Networks (RMN) are *undirected* relational models. Since their introduction, RMNs have been used successfully in a number of domains, including web page classification [139], link prediction [140], information extraction [18], and activity recognition [88]. RMNs describe specific relations between objects using clique templates specified by SQL queries: each query C selects the relevant objects and their attributes, and specifies a *potential* function, or clique potential, ϕ_C, on the possible values of these attributes. Intuitively, the clique potentials measure the "compatibility" between values of the attributes. Clique potentials are defined as log-linear combinations of *feature* functions, *i.e.*, $\phi_C(\mathbf{v}_C) = \exp\{\mathbf{w}_C^T \cdot \mathbf{f}_C(\mathbf{v}_C)\}$, where $\mathbf{v}_C$ are the attributes selected in the query, $\mathbf{f}_C()$ is a feature vector for C, and $\mathbf{w}_C^T$ is the corresponding weight vector.

To perform inference, an RMN is *unrolled* into a Conditional Random Field (CRF) [139, 79], in which the nodes correspond to object attributes. The connections among the nodes are built by applying the SQL templates to the data; each template C can result in several cliques, which share the same feature weights. Given observations $\mathbf{x}$, the cliques generated by an RMN factorize the conditional distribution over the labels $\mathbf{y}$ as follows [139, 89, 138]:

$$p(\mathbf{y} \mid \mathbf{x}) \quad = \quad \frac{1}{Z(\mathbf{x})} \prod_{C \in \mathcal{C}} \prod_{\mathbf{v}_C \in C} \phi_C(\mathbf{v}_C) \quad = \quad \frac{1}{Z(\mathbf{x})} \prod_{C \in \mathcal{C}} \prod_{\mathbf{v}_C \in C} \exp\{\mathbf{w}_C^T \cdot \mathbf{f}_C(\mathbf{v}_C)\} \tag{1}$$

Here, $Z(\mathbf{x})$ is the normalizing partition function defined as $Z(\mathbf{x}) = \sum_{y'} \prod_{C \in \mathcal{C}} \prod_{\mathbf{v}'_C \in C} \phi_C(\mathbf{v}'_C)$. Th¡e first product in Eq. 1 ranges over all clique template queries, and the second product over all cliques generated by each template. As noted above, the instantiated network is a CRF. CRFs were originally developed for labeling sequence data [79], and have been shown to outperform generative approaches such as HMMs in areas such as natural language processing [79] and computer vision [71, 118]. CRFs directly model the *conditional* distribution over the hidden variables $\mathbf{y}$. Due to this structure, they can handle arbitrary dependencies between the observations $\mathbf{x}$, which gives them substantial flexibility in using very high-dimensional, overlapping feature vectors [138]

Exact inference algorithms in RMNs often become intractable except for some special graph topologies, such as sequences or trees. In general, it is necessary to use approximate techniques such as loopy belief propagation [139] or Markov Chain Monte Carlo (MCMC) [87, 89]. The goal of parameter learning in RMNs is to find the weight vector $\mathbf{w}_C$ for each SQL template so as to maximize the overall conditional log-likelihood L of the training data:

$$L(\mathbf{w}) \quad \equiv \quad \log p(\mathbf{y} \mid \mathbf{x}, \mathbf{w}) - \frac{\mathbf{w}^T \mathbf{w}}{2\nu} \tag{2}$$

The rightmost term in Eq. 2 is a regularization to avoid overfitting. It imposes a Gaussian shrinkage prior with variance ν on each component of the weight vector [139]. Since Eq. 2 can be shown to be convex, the global optimum of L can be found using modern numerical optimization algorithms, such as conjugate gradient or quasi-Newton techniques [131]. However, maximizing the conditional likelihood requires running the inference procedure at each iteration of the optimization, which can be very expensive. An alternative is to maximize the *pseudo-likelihood* of the training data [13]. This approximate technique can be evaluated extremely efficiently and has been shown to perform well in several domains [71, 122, 88].

3 Relational Semantic Maps (RS-Maps)

In this section we propose RS-Maps, a new framework for building semantic place and object-level descriptions of environments. While RS-Maps describe objects mainly by their 2D outline, they use 3D information extracted from laser range-scans and camera images to estimate the presence and types of objects and places.

RS-Maps are based upon and extend Relational Markov Networks (RMN) [139]. They describe environments by hierarchical collections of geometric primitives, objects, places, and the relationships between them. Like RMNs, RS-Maps can be instantiated into Conditional Random Fields (CRF) that describe probability distributions over objects and places. Figure 2 illustrates the concept of such a CRF instantiated for an indoor environment. Nodes in the network correspond to objects and places, and undirected edges represent probabilistic constraints between them. In this hierarchical structure, geometric primitives extracted from sensor data are modeled at the lowest level. The next levels describe basic and aggregated objects such as doors, sections of walls, or pieces of furniture. Higher levels estimate spatial areas such as hallways and rooms, and their connectivity relations. In this network, for instance, the geometric primitive with ID g_r is part of object d_s, which is of type door. This door corresponds to connector c_2, which connects places p_2 and p_4. The connector "generates" a probabilistic relationship between p_2 and p_4, indicated by a

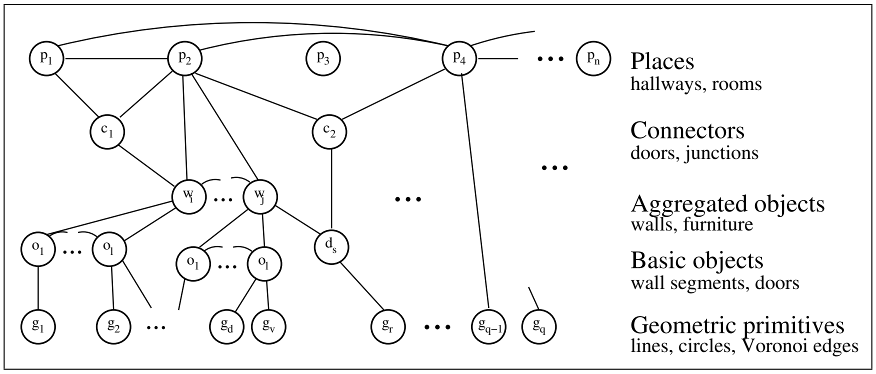

Figure 2: Part of an indoor RS-Map instantiated into a Conditional Random Field using sensor data and an RS-Map schema. Each node corresponds to an object or place in an indoor environment. Edges indicate undirected, probabilistic constraints between objects. The lowest level of the RS-Map contains geometric primitives extracted from the data. Each primitive is associated to an object such as a door or a chair. Objects can be aggregated into larger objects, and are associated to places, which are linked via connector objects.

link. The door is in spatial relationship to the wall object w_j, which itself is part of the boundary of place p_2.

3.1 Relational Schemata

Similar to RMNs, our RS-Maps will be defined by relational schemata and clique templates. In general, a relational schema defines the *entity types* (*i.e.*, classes) in a domain and their *reference relationships*. Relational database theory, especially the entity-relationship and the relational data model, provides us with powerful tools for building compact and consistent relational schemata for real domains [119]. In RS-Maps, we propose to organize entities in a multi-level structure, as illustrated in Figure 3.

Entity types

We will categorize the entities in RS-Maps into the following four categories.

Geometric primitives are the basic building blocks of RS-Maps. In our system, primitives such as lines, circles, planes, and points are extracted from 2D or 3D laser range-finder scans. As shown in the relational schema in Figure 3, each primitive has a unique ID, a description of its geometry (location, length, radius, *etc.*) and appearance (color, texture, *etc.*), and its type (person, door, wall segment, tree trunk, *etc.*). The appearance of primitives is computed by aligning camera information with laser scans. We will additionally introduce primitives that are nodes in a spatial graph representation of an environment. Such a graph could be a Voronoi graph of an indoor environment (see left panel in Figure 1), or a street and walkway map of an outdoor area. We found such graphs very useful to compactly represent the connectivity structure of environments and for reducing the complexity of probabilistic inference [47, 35].

Objects are generated hierarchically, with basic objects being built from primitives, and more complex objects being generated by physical aggregation from other objects. Each object has a specific type, such as person, door, wall segment, wall, chair, desk, car, or tree. Depending on their type, objects can have different properties and relationships. Figure 3 illustrates the relational schema of the three object types door, wall segment, and wall in an indoor

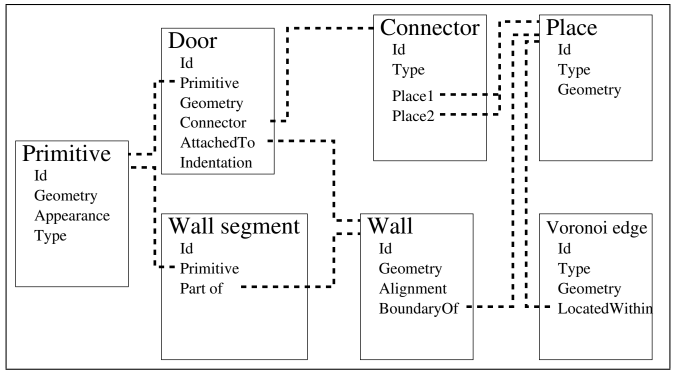

Figure 3: Part of an RS-Map entity-relationship schema. Boxes describe entity types, and dashed lines indicate reference relationships.

environment. For instance, a wall segment is a basic object that consists of a single line segment annotated with appearance information. walls consist of several wall segments. Such more complex objects are generated by a process we call *physical aggregation*, which extracts both geometric and appearance descriptions from the aggregated objects. To handle dynamic objects, each object type can have additional properties describing if and how it can change its location (*e.g.*, doors rotate around their hinge points, trees don't move, cars might drive).

Places describe spatially coherent areas such as hallways, rooms, parking lots, or streets. While such places can be defined accurately for indoor environments (*e.g.*, Figure 1), the boundaries between places in outdoor environments is less clear and we will investigate different types of places suitable for outdoor mapping. Each object in an environment is associated with at least one place. It can either be located in a place (*e.g.*, chair in a room, car on the street), or be part of the boundary of a place (*e.g.*, wall). The link between objects and places is important, since the interpretation of geometric primitives strongly depends on the place they are in. For instance, it is extremely unlikely to find a tree in the middle of a street.

Connectors: In order to reason about connectedness and neighborhood on a place level, we propose to introduce the notion of connectors, which are special entities that describe transitions between places. They can be either associated to physical objects such as doors, or place concepts such as hallway or street intersections. Structures such as Voronoi graphs or street maps provide a convenient way to define such connectors.

Reference relationships

Relational schemata specify the reference relationships among the entities using *reference attributes* (*i.e.*, foreign keys). For example, in Figure 3, the attributes Place1 and Place2 in class connector refer to a pair of places that are linked by the connector. Different reference attributes can have different semantic meanings. They can imply the relation of *physical aggregation*; for example, the "Part of" attribute in wall segment indicates a wall object is generated by a set of wall segments. They can also indicate "Is A" relation so as to represent *class hierarchies*; for example, the connector attribute in class door indicates that door is a kind of connector, just like junctions and other kinds of connectors.

Structural uncertainties

In RS-Maps, the schema can involve structural uncertainties, such as *existence uncertainties* over both objects and links [51]. For example, the numbers of objects and places in an environment are unknown initially and have to be inferred by probabilistic reasoning. Link uncertainty is due to the fact that the unknown type of an object specifies which other objects it is connected to.

3.2 Defining Features via Clique Templates

We now discuss how to define the feature functions that are used by RS-Maps to estimate semantic representations of an environment. A key advantage of the Conditional Random Fields underlying RS-Maps is their ability to incorporate large numbers (many thousands) of features, even when there are strong dependencies among the individual features [79, 138]. This ability adds substantial flexibility to RS-Maps.

In order to perform probabilistic inference in RS-Maps, the relational schemata and the data are used to generate a Conditional Random Field that models the probability distribution over the objects in an environment. This process of unrolling includes generating the nodes and the link structure of a CRF, along with the functions that describe the potentials of the cliques in the network. We propose to define RS-Map cliques and their potentials via clique templates specified by SQL queries on the object database. Clique templates provide an extremely flexible and concise language to define network structures and features. We will develop the following types of features.

Local features describe the *geometry* and *appearance* of objects and places. In RS-Maps, such features can be modeled by generating a clique for each object and the corresponding attribute. As done in current techniques, we will extract various shape features from both 2D and 3D laser range scans [117, 32, 120, 96]. The right panel in Figure 4 illustrates how a 2D laser scan can be enhanced with image information by projecting the scan points into a camera image (such a projection can also be performed for 3D scans [117]). In [32] we showed that such a combination of laser shape and visual appearance information can achieve superior classification results on individual laser returns. Due to the flexibility of Conditional Random Fields, RS-Maps will be able to incorporate thousands of low level visual features such as color histograms, SIFT descriptors [92], steerable pyramids [132], Haar features [154], or surface geometry [59, 60].

Local features can also be defined for aggregated objects and places. For example, a wall object can have a feature that measures the alignment of its wall segments. Such a feature allows an RS-Map to ensure that only well aligned line segments are associated to a wall. A local feature of a place could be a description of its shape and size, for instance.

Visual object detectors will be incorporated to leverage existing large, annotated vision data sets (*e.g.*, [3, 1, 2]), thereby reducing the need for labeling data collected with our robots. Vision data sets can be used to learn specialized detectors for objects such as people, faces, cars, or parts thereof. The output of such detectors can then be incorporated as additional information into the RS-Maps. The combination of low level features and detectors will result in strong overlap and dependencies between individual feature components. Fortunately, this will not pose a significant challenge to our approach, since the ability to deal with such dependencies is a key property of the Conditional Random Fields underlying RS-Maps, in contrast to generative approaches such as Bayesian networks or Markov Random Fields.

Spatial features describe spatial dependencies between objects and places. For example, trees are extremely unlikely to be found on streets, desks are more likely to be found in rooms, and doors are very likely to be near doorways. Such dependencies will be modeled in RS-Maps by incorporating links between nearby objects and places into the underlying graphical

Figure 4: (left) One of the vehicles we will use to collect data for semantic mapping in urban environment. The vehicle, built by the Australian Centre for Field Robotics, is equipped with multiple cameras and laser range-finders. (right) Camera image along with a projected laser scan. Colors of mapped laser returns indicate the type of object each return points at (see also [32]).

model. The potentials of the resulting cliques then depend on the specific spatial relationships between specific objects and places (*e.g.*, distance, relative angle, is-inside).

Global features will be used to describe properties of objects that are potentially far apart. For instance, an important class of global features are those that measure regularities within an environment. Such a feature might describe how similar the widths, indentations, or colors of the doors in a certain hallway are. An example of how a global door width feature can be described using the SQL language we will develop for RS-Maps is

```
SELECT Variance(d.Width)
FROM Door d, Wall w
WHERE d.AttachedTo=w.Id
```

Global features are very powerful for expressing complex relationships between objects, but they also pose challenging problems for efficient inference, since they generate large cliques in the unrolled Conditional Random Fields.

4 Inference and Learning in Relational Semantic Maps

In addition to developing the general framework and features used in RS-Maps, an important contribution will be the development of suitable learning and inference mechanisms.

4.1 Inference

Hierarchical inference under structural uncertainty

The key to scaling our approach to large environments with many objects and places is to develop techniques that make use of the hierarchical structure underlying RS-Maps. Hierarchical representations [36, 107, 17] have been shown to enable very efficient inference and learning techniques in various domains, including activity recognition [16, 85, 88], robot mapping [141], and speech recognition [14]. However, these existing approaches assume that the structure of the hierarchical model is fully specified, which is not the case for RS-Maps. For instance, different segmentations of an environment might generate a different number of hallways and rooms in an RS-Map, which would be instantiated into CRFs with different place nodes. A robust inference system for RS-Maps must thus be able to reason about multiple structures representing an environment.

While MCMC sampling techniques are well suited to perform inference under structural uncertainty [89, 97, 98, 115], they do not scale to very complex models. We will therefore investigate an alternative approach for efficiently reasoning about multiple model structures. Initially, this approach instantiates an RS-Map only partially. It then performs inference in the resulting CRF model and generates a set of likely interpretations of the data. These interpretations can be generated efficiently using either k-best belief propagation [157] or MC-SAT, a recently introduced MCMC technique that combines probabilistic sampling with SAT solving [115]. MC-SAT is particularly well suited to handle deterministic dependencies such as transitive closure: "if primitives i and j are part of the same object, and primitives j and k are part of the same object, then primitives i and k must also be part of the same object". Each sampled interpretation can then be used to instantiate additional parts of the RS-Map, thereby generating larger CRF models. For instance, if we start by segmenting an environment into places, then k-best inference will generate different segmentations, which produce different CRF models. Once place nodes are known, it is possible to connect all geometric primitives to the places (nodes) they are located in, thereby generating more complex CRFs. We can repeat this process of adding nodes based on inference until a set of complete CRFs is generated, inference in which will produce multiple possible interpretations of an environment.

The key advantage of this approach is that at every iteration, the structure of the individual CRF models is completely known, which enables the use of efficient inference techniques such as belief propagation. In essence, our proposed approach performs efficient search in the set of possible models by iteratively constructing more and more complex structures based on likely interpretations of the data.

Computational clique templates

The Conditional Random Fields instantiated during RS-Map inference will contain a variety of cliques with different complexity and containing both continuous and discrete states. However, current inference techniques are defined over a complete network, applying the same approach to all parts of the network. They are thus not flexible enough to adjust to local substructures typically present in the large networks instantiated by RS-Maps. To overcome this limitation, we will develop a hybrid inference framework that performs different types of inference in different parts of a network. The RS-Map inference system will be based on (loopy) belief propagation, which is an inference technique that sends local messages between neighbors in a network in order to infer posterior probabilities or maximum *a posteriori* (MAP) assignments [109, 157, 88]. Each node computes a message to a neighbor based on messages it receives from other neighbors, the clique potential that specifies the connection to this neighbor, and the clique potentials modeling the impact of local features.

We propose to develop a hybrid inference system in which the messages sent between cliques are computed depending on the structure of the clique. For instance, while exact inference might be adequate to compute messages for rather small cliques of an RS-Map, such inference does not scale to large cliques; as those needed to compute global features. In such cliques, MCMC sampling such as MC-SAT [115] might be more adequate, or aggregation via Fast Fourier Transform techniques [86]. Messages involving continuous hidden states can be handled via Gaussian approximations or sampling, as done in non-parametric belief propagation [137].

To specify such locally optimized inference techniques in a single framework, we will extend the relational language underlying RS-Maps and RMNs by adding *computational clique templates*. Such templates not only specify the structure of cliques in an RS-Map, but also which type of inference should be performed within this clique in order to compute the probabilistic messages sent in belief propagation. The resulting relational model will enable highly efficient inference, since each clique computes its messages using the most adequate algorithm.

Reasoning about dynamic objects

Certain types of objects, including people, cars, doors, and furniture, are dynamic and might move after or even during the mapping process. While handling such dynamics will not be the focus of this project, we will investigate techniques for object-level tracking once the basic RS-Map framework is developed. A key component of tracking is temporal data association, which is the process of determining which measurements are caused by the same object. An advantage of the object level representation of RS-Maps is that we can perform data association at the object level, taking shape and appearance information into account. We recently showed how Conditional Random Fields can match individual beams of laser range-scans using shape and appearance information [120]. We will extend this technique to object-level matching and incorporate it into RS-Maps. Modeling dynamics within the RS-Map framework will enable it to take into account which types of objects can move (*e.g.*, people, cars, chairs), and which ones are fixed (*e.g.*, walls, buildings, and trees). As a result, the fact that an object is moving can help to determine its type.

4.2 Learning

Since inference is an inherent part of learning in Relational Markov Networks, the approaches used for scaling up inference can be readily applied to speed up learning. Furthermore, as shown in different contexts, pseudo-likelihood is an extremely scalable and robust method for learning parameters of complex models [13, 47, 122, 71]. However, RS-Maps pose additional challenges which we propose to address as follows.

Dealing with continuous features

The majority of relational techniques have only been applied to discrete, non-physical domains, in which features are mainly boolean indicator functions. However, RS-Maps need to consider many continuous features, such as visual appearance and geometry of objects, relative locations of objects, and shapes of places. While the log-linear models underlying RS-Maps could incorporate thousands of continuous features, the resulting probability distributions would correspond to uni-modal Gaussian likelihoods in generative models [112]. Since the relationships between hidden states and features are typically far more complex, such a straightforward incorporation of continuous features is thus not flexible enough to achieve good classification results.

Recently, several research groups found that boosting [46] can be used to infer features in the context of Conditional Random Field training [149, 30, 47, 83]. These approaches learn discretizations of features that result in good CRF classification performance. While these techniques achieve very promising initial results, they are not able to model or learn the complex cliques found in our RS-Maps. For instance, they cannot learn neighborhood cliques that depend on features observed at different nodes in the model. Such types of cliques are crucial to model spatial relationships between objects, which is an important component of RS-Maps.

In this project, we will build on boosting-based CRF training to develop complex feature induction techniques for Relational Markov Networks. We will additionally investigate the use of l_1 regularization for feature learning [153]. The key idea of this approach is to define a very large set of features and replace the regularization term in Eq. 2 by l_1 regularization on the model weights so as to learn only a small subset of features with non-zero weights. While Vail and colleagues demonstrated that this approach works very well in the context of activity recognition using $\approx$1,000 features, we will investigate how it can be scaled to thousands of continuous features used in RS-Maps.

Learning from partially labeled data

The probabilistic models underlying RS-Maps are trained discriminatively [139, 138]. While such an approach has advantages over generative training techniques [112], it typically requires fully labeled training data, which can be extremely tedious to generate. In the context of Conditional Random Fields, several researchers showed that learning techniques such as expectation maximization (EM) or entropy regularization can be applied successfully to discriminative models [64, 133, 22, 93, 80], and we will investigate their application to RS-Maps. To further reduce the burden of manual data labeling, we will use existing vision datasets to train a variety of object detectors. The output of these detectors can be readily incorporated as additional features into RS-Maps.

Knowledge transfer through hierarchical Bayesian features

A very important question with respect to learning is the ability to transfer information between different environments. While Relational Markov Networks are well suited to learn models that can be applied to different environments, current techniques do not model regularities between different environments in a satisfying way. For instance, while the widths of doors might differ substantially between different environments, they typically are extremely similar within one particular hallway. While such information can be captured using variance features, it might be more adequate to use feature functions that take these variabilities into account. For instance, instead of using the raw value of the door width as a feature, one can use the log-likelihood of the door width under a Gaussian distribution estimated from test environments. In order to capture variability between environments, we intend to extend this Gaussian feature function to a *hierarchical Gaussian model* [48, 20, 128, 82]. Such a model can capture both the variability between environments and the variability within a hallway, for example. In addition to Gaussian models, we intend to use log-likelihoods under discrete distributions, which can be handled within a hierarchical model using Dirichlet distributions [100, 44]. The parameters of these feature functions can be adapted to unknown environments, thereby increasing the flexibility of RS-Maps.

5 Experimental Evaluation

The key goal of this project is to develop a relational learning and inference framework that enables the generation of semantic maps of both indoor and outdoor environments. To build such maps, we will equip our existing indoor robots with an omni-directional camera and a manipulator that allows us to collect 3D laser scans. For building semantic outdoor maps, we will focus on urban settings and data collected by cars equipped with various types of laser range-finders and cameras. For instance, the left panel in Figure 4 shows a vehicle developed by the Australian Centre for Field Robotics. We have permission to collect data with this vehicle, which carries several 2D laser scanners and cameras. Furthermore, the PI of this project is in close contact with several of the teams that participated in the DARPA Urban Challenge, and we will be able to use data collected by their vehicles, including data collected with Velodyne scanners (`http://www.velodyne.com/lidar/`). These sensors provide extremely rich 3D laser scans, and played an important role in the successful outcome of the Urban Challenge [146].

We propose to assess the results of our research using several performance criteria. Different aspects of RS-Maps will be evaluated using standard approaches such as classification and detection accuracy, confusion matrices, and string edit distance, which provides a measure of segmentation error [47, 94]. Using such measures, we will evaluate how many different types of objects the approach can distinguish when using different feature functions and relations, how robust the approach is with respect to changing lighting conditions, or how the quality of the learned models decreases with the sparseness of the labeled training data.

While the accuracy in labeling individual objects and places is an important measure of performance, we will perform additional evaluations that are focused on the potential applications of RS-Maps. For instance, we will evaluate the accuracy and usefulness of the learned models with respect to guiding people to locations and objects, or following human path descriptions to a specific object. These evaluations will be performed as undergraduate research projects, the goal of which will be to implement and test a high-level planner for RS-Maps. This planner will be used to perform robot tasks inspired by indoor office delivery and search and rescue scenarios. Key questions will be how well the place and object information contained in RS-Maps enables a robot to perform tasks specified semantically, such as "Take a picture of the desk in the room at the end of the hallway", and how well it enables robots to guide people to specific locations, such as "Follow this hallway, go into the second hallway on the left, and enter the room next to the fire extinguisher". To perform tests using real robots, the high-level planner will be connected to our existing robot exploration and navigation system [43].

6 Results from Prior NSF Support

Title: *CAREER: Probabilistic Methods for Multi-Robot Collaboration*
Award number: IIS-0093406; **amount**: $440,000; **period of support**: March 2001–2006

The goal of this CAREER project was to bridge the gap between the success of probabilistic methods for single-robot systems and their successful application to collaborative multi-robot systems.

Technical contributions: We developed novel adaptive, real-time particle filter approaches for state estimation [38, 39, 77, 76, 78, 84, 127], reinforcement learning for active sensing [75], and Rao-Blackwellised particle filters for target tracking [74] and map building [57]. We furthermore developed an efficient, decision-theoretic technique that allows multiple robots to explore environments from different, unknown start locations. The approach avoids the exponential complexity of this most difficult instance of the exploration problem by actively verifying hypotheses for the relative locations of robots [66, 69, 135, 44, 43]. Our proposed work will use the SLAM mapping techniques developed in this project to combine sensor information in a spatially consistent way. Furthermore, the high-level planner we will develop for RS-Maps will use the navigation routines of this project to support low level navigation tasks.

RoboCup as undergraduate research opportunity: The PI mentored 14 undergraduate students during their participation in the RoboCup robot soccer challenge (AIBO legged robot league). This undergraduate teaching and research effort included participation in the RoboCup competitions 2001–2004 [9, 28, 73, 74].

Teaching and curriculum development: The PI introduced a graduate course on "Probabilistic mobile robotics", and developed an undergraduate project course entitled "Mobile robotics capstone". This lecture focuses on a hands-on experience in robotics using robots such as legged AIBO robots or an autonomous blimp as teaching platforms. With S. Thrun and W. Burgard, the PI wrote a textbook entitled "Probabilistic Robotics" [144].

7 Education and Outreach Plan

Curriculum Improvement and Outreach

In this project we will build on the PI's robotics course to develop an advanced graduate level course on statistical relational techniques for real-world data analysis. This course will provide an in-depth treatment of relational learning techniques in the context of areas such as robotics and activity recognition. The proposed research will play an important role in assessing the key lessons learned from applying relational reasoning to these novel domains.

Robotics is an ideal tool to educate the wide public about how exciting engineering research can be. Furthermore, we believe that robotics researchers have the obligation to keep the public informed about the potential benefits robots can provide to society, but also about which progress is realistically achievable within the next decade. Thus, we will demonstrate our research at public events such as the annual SRS Robothon (`http://www.seattlerobotics.org/robothon`), a national event that showcases robotics.

In addition to these local activities, we will deepen our collaboration with Dr. Andrew Williams at Spelman College, with whom the PI of this project recently started a collaboration as part of "ARTSI: Advancing Robotics Technology for Societal Impact", an NSF Broadening Participation in Computing Alliance Grant. The goal of the ARTSI project is to expose female African-American students to research in robotics, artificial intelligence, and computer science. In addition to the activities funded under the existing grant, the PI will give guest lectures in Dr. Wiliams' course and guide the integration of a particle filter software package for ARTSI related activities. We believe that methods such as particle filters are ideally suited to provide an intuitive idea of fundamental problems in robotics and AI, thereby motivating young students to get stronger involved in these research areas.

Dissemination of Data Sets

This project will generate data sets that go well beyond those typically available for research in mobile robotics. While the Robotics Data Set Repository [62] contains data sets of various environments, it typically provides laser range-data only. Our sets will contain 2D and 3D laser range-finder data along with aligned camera information. Many of these data sets will be labeled manually, which will make them extremely useful for other research groups. All data sets will be made available to the research community via our own web site and via our contribution to the Robotics Data Set Repository [62].

8 Research Project Work Plan

We will focus on building RS-Maps from a combination of laser range-data and camera images. To collect such data, we will acquire an omni-directional camera and a manipulator that allows us to collect 3D laser-scans in indoor environments. Camera data will be used to enhance both 2D and 3D laser shape information with rich appearance information, similar to [116] and our preliminary work on classifying 2D laser beams into seven different outdoor object classes (car, person, wall, tree, foliage, grass, other) [32]. We will collect outdoor data with the vehicle shown in the left panel of Figure 4. We will have access to additional 3D laser and camera data collected by the CMU, Stanford, and Cornell DARPA Urban Challenge vehicles. While such laser data could be used to build complex, 3D volumetric models of an environment, we will restrict RS-Maps to generating 2D layouts models of objects. The 3D data will still be extremely useful to provide rich shape information for object detection and classification.

To generate spatially consistent data sets, we will use existing SLAM and scan alignment techniques (*e.g.*, [33, 57, 58, 25, 43, 145]). The reliance on pre-aligned data is not due to a conceptual limitation of RS-Maps or the algorithms we will develop, but rather due to the time constraints of the project.

The two graduate students funded by this project will collaborate to develop the general RS-Map framework. One student will work on indoor environments, with a focus on reasoning about spatial relationships between objects and places. An important aspect of this work will be the development of inference techniques suitable for the complex, hierarchical structure underlying RS-Maps. These techniques will be integrated into the RS-Map framework by introducing *computational clique templates*, which enhance Relational Markov Networks to reason about continuous hidden states and to perform different inference algorithms in different parts of an RS-Map.

The second student will work on outdoor environments with an initial focus on extracting features from 2D and 3D laser data and from camera images. In addition to generating a variety of low-level features, this student will leverage labeled vision data sets to learn specific object detectors (*e.g.*, [3, 1, 2]). The student will then develop boosting related learning techniques that can incorporate these features and detectors into RS-Maps. To further reduce the burden of manual data labeling, the student will also investigate semi-supervised learning techniques for RS-Maps.

Once the RS-Map framework is developed, the relational structure underlying our approach will enable us to rapidly explore different features, spatial relations, and learning and inference techniques. At this stage, we will also start the undergraduate research project that aims at investigating the usefulness of RS-Maps for high-level planning and human robot communication. Toward the end of the project, the graduate students will enhance the basic RS-Maps to detect and track moving objects and to allow more complex, 3D shape models for objects.

9 Broader Impact

If successful, the techniques developed in this project will greatly increase the reasoning and interaction capabilities of robotic systems. Such capabilities will have significant impact on a variety of robotic applications. For instance, in search and rescue tasks, a mobile robot equipped with RS-Map reasoning will be able to communicate with first responders in a much more natural and therefore much more effective way [72, 91]. As another example, the ability to distinguish a variety of objects is crucial for safe navigation in populated urban areas, with the potential to avoid a large number of fatalities due to unsafe human driving [21, 146].

Beyond enabling more capable and robust robotic systems, the research performed in this project will have impact on several other research communities. It will further enhance the capabilities of statistical relational models by developing novel learning and inference techniques that make them applicable to domains characterized by complex, continuous relationships between objects, and by high-dimensional, continuous sensor data. Specifically, the computational clique template framework and the boosting-related learning approach developed in this project will be completely independent of the specific application, and will thus be readily applicable to other areas involving probabilistic reasoning. We conjecture that the techniques developed in this project will be particularly useful in areas such as sensor-based human activity recognition [136] and computer vision [108, 149].

In addition to the general benefit to robotics and related areas, our project promotes teaching and training both at the graduate and undergraduate level. Besides the graduate research projects, we will support undergraduate projects in the area of high-level planning and human robot communication. To improve the understanding of the concepts underlying this research, we will offer research seminars and develop the curriculum of a new graduate level course on relational probabilistic models for real-world reasoning. We will also collaborate with Dr. Andrew Williams at Spelman College to enhance lectures and existing software infrastructure in the context of the NSF-funded project "ARTSI: Advancing Robotics Technology for Societal Impact". Furthermore, the outcome of this research will be presented to the broad public at events such as the annual SRS Robothon, a national event that showcases developments in robotics.

References

[1] Caltech data set. `http://www.vision.caltech.edu/html-files/archive.html`.

[2] Labelme data set. `http://labelme.csail.mit.edu/`.

[3] UIUC image database for car detection. `http://l2r.cs.uiuc.edu/~cogcomp/Data/Car/`.

[4] C. R. Anderson, P. Domingos, and D. S. Weld. Relational Markov models and their application to adaptive web navigation. In *Proc. of the 2002 Conference on Knowledge Discovery and Data Mining*, August 2002.

[5] D. Anguelov, R. Biswas, D. Koller, B. Limketkai, S. Sanner, and S. Thrun. Learning hierarchical object maps of non-stationary environments with mobile robots. In *Proc. of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 2002.

[6] D. Anguelov, D. Koller, E. Parker, and S. Thrun. Detecting and modeling doors with mobile robots. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2004.

[7] D. Anguelov, B. Taskar, V. Chatalbashev, D. Koller, D Gupta, G. Heitz, and A. Ng. Discriminative learning of Markov random fields for segmentation of 3D scan data. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[8] D. Austin and P. Jensfelt. Using multiple Gaussian hypotheses to represent probability distributions for mobile robot localization. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2000.

[9] D. Azari, J.K. Burns, K. Deshmukh, D. Fox, D. Grimes, C.T. Kwok, R. Pitkanen, P. Shon, and P. Tressel. Team description: UW Huskies-01. In A. Birk, S. Coradeschi, and S. Tadokoro, editors, *RoboCup-2001: Robot Soccer World Cup V*. Springer Verlag, 2002.

[10] G. Baltus, D. Fox, F. Gemperle, J. Goetz, T. Hirsh, D. Magaritis, M. Montemerlo, J. Pineau, N. Roy, J. Schulte, and S. Thrun. Towards personal service robots for the elderly. In *Proc. of the Workshop on Interactive Robotics and Entertainment (WIRE)*, 2000.

[11] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. John Wiley, 2001.

[12] P. Beeson, N.K. Jong, and B. Kuipers. Towards autonomous topological place detection using the extended Voronoi graph. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2005.

[13] J. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24, 1975.

[14] J. Bilmes, G. Zweig, T. Richardson, K. Filali, K. Livescu, P. Xu, K. Jackson, Y. Brandman, E. Sandness, E. Holtz, J. Torres, and B. Byrne. Discriminatively structured graphical models for speech recognition: JHU-WS-2001 final workshop report. Technical report, CLSP, Johns Hopkins University, Baltimore MD, 2001. `http://www.clsp.jhu.edu/ws2001/groups/gmsr/GMRO-final-rpt.pdf`.

[15] M. Bosse, P. Newman, J.J. Leonard, M. Soika, W. Feiten, and S. Teller. An ATLAS framework for scalable mapping. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2003.

[16] H. Bui. A general model for online probabilistic plan recognition. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2003.

[17] H. Bui, D. Q. Phung, and S. Venkatesh. Hierarchical hidden markov models with general state hierarchy. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 2004.

[18] R. Bunescu and R. J. Mooney. Collective information extraction with relational markov networks. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2004.

[19] R. Bunschoten and B. Kröse. Robust scene reconstruction from an omnidirectional vision system. *IEEE Transactions on Robotics and Automation*, 19(2), 2003.

[20] B.P. Carlin and T.A. Louis. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall / CRC, second edition, 2000.

[21] DARPA Urban Challenge. `http://www.darpa.mil/grandchallenge/index.asp`.

[22] H. Chieu, W. S. Lee, and L. P. Kaelbling. Activity recognition from physiological data using conditional random fields. In *Proc. of the Singapore-MIT Alliance Symposium*, 2006.

[23] H. Choset, K.M. Lynch, S. Hutchinson, G. Kantor, W. Burgard, L.E. Kavraki, and S. Thrun. *Principles of Robot Motion: Theory, Algorithms, and Implementations*. MIT Press, 2005.

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

[25] D. Cole and P. Newman. Using laser range data for 3D SLAM in outdoor environments. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2006.

[26] I. J. Cox. Blanche—an experiment in guidance and navigation of an autonomous robot vehicle. *IEEE Transactions on Robotics and Automation*, 7(2):193–204, 1991.

[27] I. J. Cox and J. J. Leonard. Modeling a dynamic environment using a Bayesian multiple hypothesis approach. *Artificial Intelligence*, 66:311–344, 1994.

[28] Z. Crisman, E. Curre, C.T. Kwok, L. Meyers, N. Ratliff, L. Tsybert, and D. Fox. Team description: UW Huskies-02. In G. Kaminka, P.U. Lima, and R. Rojas, editors, *RoboCup-2002: Robot Soccer World Cup VI*. Springer Verlag, 2003.

[29] T. Dean and K. Kanazawa. Probabilistic temporal reasoning. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 1988.

[30] T.G. Dietterich, A. Ashenfelter, and Y. Bulatov. Training conditional random fields via gradient tree boosting. In *Proc. of the International Conference on Machine Learning (ICML)*, 2004.

[31] M.W.M. Dissanayake, P. Newman, S. Clark, H.F. Durrant-Whyte, and M. Csorba. A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on Robotics and Automation*, 17(3), 2001.

[32] B. Douillard, D. Fox, and F. Ramos. A spatio-temporal probabilistic model for multi-sensor multi-class object recognition. In *Proc. of the International Symposium of Robotics Research (ISRR)*, 2007.

[33] A. Eliazar and R. Parr. DP-SLAM: Fast, robust simultaneous localization and mapping without predetermined landmarks. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2003.

[34] J. F. Engelberger. Health-care robotics goes commercial: The 'helpmate' experience. *Robotica*, 11:517–523, 1993.

[35] B. Ferris, D. Hähnel, and D. Fox. Gaussian processes for signal strength-based location estimation. In *Proc. of Robotics: Science and Systems (RSS)*, 2006.

[36] S. Fine, Y. Singer, and N. Tishby. The hierarchical hidden Markov model: analysis and applications. *Machine Learning*, 32, 1998.

[37] A. Fod, A. Howard, and M.J. Mataric. Laser-based people tracking. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2002.

[38] D. Fox. KLD-sampling: Adaptive particle filters. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14 (NIPS)*, Cambridge, MA, 2001. MIT Press.

[39] D. Fox. Adapting the sample size in particle filters through KLD-sampling. *International Journal of Robotics Research (IJRR)*, 22(12), 2003.

[40] D. Fox, W. Burgard, F. Dellaert, and S. Thrun. Monte Carlo Localization: Efficient position estimation for mobile robots. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 1999.

[41] D. Fox, W. Burgard, H. Kruppa, and S. Thrun. A probabilistic approach to collaborative multi-robot localization. *Autonomous Robots*, 8(3), 2000.

[42] D. Fox, W. Burgard, and S. Thrun. Active Markov localization for mobile robots. *Robotics and Autonomous Systems*, 25:195–207, 1998.

[43] D. Fox, J. Ko, K. Konolige, B. Limketkai, and B. Stewart. Distributed multi-robot exploration and mapping. *Proc. of the IEEE*, 94(7):1325–1339, 2006. Special Issue on Multirobot Systems.

[44] D. Fox, J. Ko, K. Konolige, and B. Stewart. A hierarchical Bayesian approach to mobile robot map structure learning. In P. Dario and R. Chatila, editors, *Robotics Research: The Eleventh International Symposium*, Springer Tracts in Advanced Robotics (STAR). Springer Verlag, 2005.

[45] D. Fox, S. Thrun, F. Dellaert, and W. Burgard. Particle filters for mobile robot localization. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo in Practice*. Springer-Verlag, New York, 2001.

[46] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, 28(2), 2000.

[47] S. Friedman, D. Fox, and H. Pasula. Voronoi random fields: Extracting the topological structure of indoor environments via place labeling. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2007.

[48] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd edition, 2003.

[49] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 6, 1984.

[50] L. Getoor, N. Friedman, D. Koller, and A. Pfeffer. Learning probabilistic relational models. In S. Dzeroski and N. Lavrac, editors, *Relational Data Mining*. Springer-Verlag, 2001.

[51] L. Getoor, N. Friedman, D. Koller, and B. Taskar. Learning probabilistic models of link structure. In *Journal of Machine Learning Research (JMLR)*, volume 3, 2002.

[52] L. Getoor and B. Taskar, editors. *Introduction to Statistical Relational Learning*. MIT Press, 2007.

[53] Z. Ghahramani. *Lecture Notes in Artificial Intelligence*, chapter Learning Dynamic Bayesian Networks. Springer-Verlag, 1998.

[54] R. Grabowski, P. Khosla, and H. Choset. An enhanced occupancy map for exploration via pose separation. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2003.

[55] J.S. Gutmann and K. Konolige. Incremental mapping of large cyclic environments. In *Proc. of the IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA)*, 1999.

[56] D. Hähnel, W. Burgard, D. Fox, K. Fishkin, and M. Philipose. Mapping and localization with RFID tags. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2004.

[57] D. Hähnel, W. Burgard, D. Fox, and S. Thrun. An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2003.

[58] D. Hähnel, W. Burgard, and S. Thrun. Learning compact 3D models of indoor and outdoor environments with a mobile robot. *Robotics and Autonomous Systems*, 44(1), 2003.

[59] D. Hoiem, A. Efros, and M. Hebert. Geometric context from a single image. In *Proc. of the International Conference on Computer Vision (ICCV)*, 2005.

[60] D. Hoiem, A. Efros, and M. Hebert. Putting objects in perspective. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006.

[61] A. Howard, L.E. Parker, and G.S. Sukhatme. The SDR experience: Experiments with a large-scale heterogenous mobile robot team. In *Proc. of the International Symposium on Experimental Robotics (ISER)*, 2004.

[62] A. Howard and N. Roy. The robotics data set repository (radish), 2003. `radish.sourceforge.net`.

[63] C. Jennings, D. Murray, and J.J. Little. Cooperative robot localization with vision-based mapping. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 1999.

[64] F. Jiao, S. Wang, C. Lee, R. Greiner, and D. Schuurmans. Semi-supervised conditional random fields for improved sequence segmentation and labeling. In *Proc. of Joint Conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics*, 2006.

[65] L. P. Kaelbling, A. R. Cassandra, and J. A. Kurien. Acting under uncertainty: Discrete Bayesian models for mobile-robot navigation. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1996.

[66] J. Ko, B. Stewart, D. Fox, K. Konolige, and B. Limketkai. A practical, decision-theoretic approach to multi-robot mapping and exploration. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2003.

[67] S. Koenig and R. Simmons. Passive distance learning for robot navigation. In L. Saitta, editor, *Proc. of the International Conference on Machine Learning (ICML)*, 1996.

[68] K. Konolige. Large-scale map making. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 2004.

[69] K. Konolige, D Fox, C. Ortiz, A. Agno, M. Eriksen, B. Limketkai, J. Ko, B. Morisset, D. Schulz, B. Stewart, and R. Vincent. Centibots: Very large scale distributed robotic teams. In M. Ang and O. Khatib, editors, *Experimental Robotics: The 9th International Symposium*, volume 21 of *Springer Tracts in Advanced Robotics (STAR)*. Springer Verlag, 2006.

[70] B. Kuipers and P. Beeson. Bootstrap learning for place recognition. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 2002.

[71] S. Kumar and M. Hebert. Discriminative random fields: A discriminative framework for contextual interaction in classification. In *Proc. of the International Conference on Computer Vision (ICCV)*, 2003.

[72] V. Kumar, D. Rus, and S. Singh. Robot and sensor networks for first responders. *IEEE Pervasive Computing*, 3(4), 2004. Special Issue on Pervasive Computing for First Response.

[73] C.T. Kwok, M. Burkhart, G. Hazen, N. Mohebbi, M. Stipanovich, and D. Fox. Team description: UW Huskies-03. In D. Polani, A. Bonarini, B. Browning, and K. Yoshida, editors, *RoboCup-2003: Robot Soccer World Cup VII*. Springer Verlag, 2004.

[74] C.T. Kwok and D. Fox. Map-based multiple model tracking of a moving object. In *RoboCup 2004: Robot Soccer World Cup VIII*, volume 3276, 2004.

[75] C.T. Kwok and D. Fox. Reinforcement learning for sensing strategies. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2004.

[76] C.T. Kwok, D. Fox, and M. Meilă. Adaptive real-time particle filters for robot localization. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2003.

[77] C.T. Kwok, D. Fox, and M. Meilă. Real-time particle filters. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15 (NIPS)*, Cambridge, MA, 2003. MIT Press.

[78] C.T. Kwok, D. Fox, and M. Meilă. Real-time particle filters. *Proceedings of the IEEE*, 92(2), 2004. Special Issue on Sequential State Estimation.

[79] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. of the International Conference on Machine Learning (ICML)*, 2001.

[80] C. Lee, S. Wang, F. Jiao, D. Schuurmans, and R. Greiner. Learning to model spatial dependency: Semi-supervised discriminative random fields. In *Advances in Neural Information Processing Systems (NIPS)*, 2006.

[81] J. J. Leonard and H. J. S. Feder. A computationally efficient method for large-scale concurrent mapping and localization. In *Proc. of the International Symposium of Robotics Research (ISRR)*, 1999.

[82] J. Letchner, D. Fox, and A. LaMarca. Large-scale localization from wireless signal strength. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 2005.

[83] L. Liao, T. Choudhury, D. Fox, and H. Kautz. Training conditional random fields using virtual evidence boosting. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2007.

[84] L. Liao, D. Fox, J. Hightower, H. Kautz, and D. Schulz. Voronoi tracking: Location estimation using sparse and noisy sensor data. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2003.

[85] L. Liao, D. Fox, and H. Kautz. Learning and inferring transportation routines. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 2004.

[86] L. Liao, D. Fox, and H. Kautz. Location-based activity recognition. In *Advances in Neural Information Processing Systems (NIPS)*, 2005.

[87] L. Liao, D. Fox, and H. Kautz. Location-based activity recognition using relational Markov networks. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2005.

[88] L. Liao, D. Fox, and H. Kautz. Extracting places and activities from GPS traces using hierarchical conditional random fields. *International Journal of Robotics Research (IJRR)*, 26(1), 2007.

[89] B. Limketkai, L. Liao, and D. Fox. Relational object maps for mobile robots. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2005.

[90] Y. Liu, R. Emery, D. Chakrabarti, W. Burgard, and S. Thrun. Using EM to learn 3D models with mobile robots. In *Proc. of the International Conference on Machine Learning (ICML)*, 2001.

[91] K. Lorincz, D. Malan, T. Fulford-Jones, A. Nawoj, A. Clavel, V. Shnayder, G. Mainland, S. Moulton, and M. Welsh. Sensor networks for emergency response: Challenges and opportunities. *IEEE Pervasive Computing*, 3(4), 2004. Special Issue on Pervasive Computing for First Response.

[92] D. Lowe. Discriminative image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 2004.

[93] M. Mahdaviani and T. Choudhury. Fast and scalable training of semi-supervised CRFs with application to activity recognition. In *Advances in Neural Information Processing Systems (NIPS)*, 2007.

[94] C. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing.* MIT Press, 1999.

[95] O. Martinez-Mozos, C. Stachniss, and W. Burgard. Supervised learning of places from range data using Adaboost. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2005.

[96] O. Martinez-Mozos, R. Triebel, P. Jensfelt, A. Rottmann, and W. Burgard. Supervised semantic labeling of places using information extracted from sensor data. *Robotics and Autonomous Systems*, 55(5), 2007.

[97] B. Milch, B. Marthi, D. Sontag, S. Russell, D. Ong, and A. Kolobov. General-purpose mcmc inference over relational structures. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2005.

[98] B. Milch and S. Russell. General-purpose mcmc inference over relational structures. In *Proc. of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 2006.

[99] D. P. Miller. Assistive robotics, an overview. In V. Mittal, H. Yanco, Aronis. J., and R. Simpson, editors, *Assistive Technology and Artificial Intelligence*, Lecture Notes in Artificial Intelligence, pages 126–136. Springer Verlag, 1998.

[100] T. Minka. Estimating a Dirichlet distribution. Technical report, MIT, 2000.

[101] M. Montemerlo, J. Pineau, N. Roy, S. Thrun, and V. Verma. Experiences with a mobile robotic guide for the elderly. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 2002.

[102] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 2002.

[103] M. Montemerlo, S. Thrun, and W. Whittaker. Conditional particle filters for simultaneous mobile robot localization and people-tracking. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2002.

[104] H. P. Moravec. Sensor fusion in certainty grids for mobile robots. *AI Magazine*, Summer 1988.

[105] H. P. Moravec and C. Martin. Robot navigation by 3D spatial evidence grids. Mobile Robot Laboratory, Robotics Institute, Carnegie Mellon University, 1994.

[106] K. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning.* PhD thesis, UC Berkeley, Computer Science Division, 2002.

[107] K. Murphy and M. Paskin. Linear time inference in hierarchical HMMs. In *Advances in Neural Information Processing Systems (NIPS)*, 2001.

[108] K. Murphy, A. Torralba, and W. Freeman. Using the forest to see the trees: A graphical model relating features, objects and scenes. In *Advances in Neural Information Processing Systems (NIPS)*, 2003.

[109] K. Murphy, Y. Weiss, and M. Jordan. Loopy belief propagation for approximate inference: An empirical study. In *Proc. of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 1999.

[110] J. Neira, M.I. Ribeiro, and J.D. Tardós. Mobile robot localization and map building using monocular vision. In *Proc. of the International Symposium on Intelligent Robotic Systems (SIRS)*, 1997.

[111] P. Newman and J.J. Leonard. Consistent convergent constant time SLAM. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2003.

[112] A. Ng and M. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Advances in Neural Information Processing Systems (NIPS)*, 2002.

[113] M.A. Paskin. Thin junction tree filters for simultaneous localization and mapping. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2003.

[114] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, Inc., 1988.

[115] H. Poon and P. Domingos. Sound and efficient inference with probabilistic and deterministic dependencies. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 2006.

[116] I. Posner, D. Schroeter, and P. M. Newman. Describing composite urban workspaces. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2007.

[117] I. Posner, D. Schroeter, and P. M. Newman. Using scene similarity for place labeling. In *Proc. of the International Symposium on Experimental Robotics (ISER)*, 2007.

[118] A. Quattoni, M. Collins, and T. Darrell. Conditional random fields for object recognition. In *Advances in Neural Information Processing Systems (NIPS)*, 2004.

[119] Raghu Ramakrishnan and Johannes Gehrke. *Database Management Systems.* McGraw-Hill Higher Education, 2nd edition, 2000. ISBN 0072465352.

[120] F. Ramos, D. Fox, and H. Durrant-Whyte. CRF-matching: Conditional random fields for feature-based scan matching. In *Proc. of Robotics: Science and Systems (RSS)*, 2007.

[121] A. Ranganathan and F. Dellaert. Inference in the space of topological maps: An MCMC-based approach. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2004.

[122] M. Richardson and P. Domingos. Markov logic networks. *Machine Learning*, 62(1-2), 2006.

[123] A. Rottmann, O. Martinez-Mozos, C. Stachniss, and W. Burgard. Semantic place classification of indoor environments with mobile robots using boosting. In *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 2005.

[124] S.I. Roumeliotis and G.A. Bekey. Bayesian estimation and Kalman filtering: A unified framework for mobile robot localization. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2000.

[125] P. Sala, R. Sim, A. Shokoufandeh, and S. Dickinson. Landmark selection for vision-based navigation. *IEEE Transactions on Robotics*, 22(2), 2006.

[126] S. Sanghai, P. Domingos, and D. Weld. Dynamic probabilistic relational models. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2003.

[127] D. Schulz, W. Burgard, and D. Fox. People tracking with mobile robots using sample-based joint probabilistic data association filters. *International Journal of Robotics Research (IJRR)*, 22(2), 2003.

[128] D. Schulz and D. Fox. Bayesian color estimation for adaptive vision-based robot localization. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2004.

[129] D. Schulz, D. Fox, and J. Hightower. People tracking with anonymous and id-sensors using Rao-Blackwellised particle filters. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2003.

[130] S. Se, D. Lowe, and J. Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *International Journal of Robotics Research (IJRR)*, 21(8), 2002.

[131] F. Sha and F. Pereira. Shallow parsing with conditional random fields. In *Proc. of Human Language Technology-NAACL*, 2003.

[132] E. Simoncelli and W. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *Proc. of the International Conference on Image Processing*, 1995.

[133] N. Smith and J. Eisner. Contrastive estimation: Training log-linear models on unlabeled data. In *Proc. of the Joint Conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics*, 2005.

[134] C. Stachniss and W. Burgard. Exploring unknown environments with mobile robots using coverage maps. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2003.

[135] B. Stewart, J. Ko, D. Fox, and K. Konolige. The revisiting problem in mobile robot map building: A hierarchical Bayesian approach. In *Proc. of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 2003.

[136] A. Subramanya, A. Raj, J. Bilmes, and D. Fox. Hierarchical models for activity recognition. In *Proc. of the International Workshop on Multimedia Signal Processing (MMSP)*, 2006.

[137] E. Sudderth, A. Ihler, W. Freeman, and A. Willsky. Nonparametric belief propagation. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[138] C. Sutton and A. McCallum. An introduction to conditional random fields for relational learning. In L. Getoor and B. Taskar, editors, *Introduction to Statistical Relational Learning*, chapter hi. MIT Press, 2006.

[139] B. Taskar, P. Abbeel, and D. Koller. Discriminative probabilistic models for relational data. In *Proc. of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 2002.

[140] B. Taskar, M. Wong, P. Abbeel, and D. Koller. Link prediction in relational data. In *Advances in Neural Information Processing Systems (NIPS)*, 2003.

[141] G. Theocharous and S. Mahadevan. Learning the hierarchical structure of spatial environments using multiresolution statistical models. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2002.

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

[143] S. Thrun, W. Burgard, and D. Fox. A real-time algorithm for mobile robot mapping with applications to multi-robot and 3D mapping. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2000.

[144] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, Cambridge, MA, September 2005. ISBN 0-262-20162-3.

[145] S. Thrun and M. Montemerlo. The GraphSLAM algorithm with applications to large-scale mapping of urban structures. *International Journal of Robotics Research (IJRR)*, 25(5/6), 2005.

[146] S. Thrun, C. Urmson, R. Rojas, and W. Uther. NIPS workshop *The Urban Challenge – Perspectives of Autonomous Driving*. 2007.

[147] Y. Thrun, S. Liu, D. Koller, A.Y. Ng, Z. Ghahramani, and H. Durrant-Whyte. Simultaneous localization and mapping with sparse extended information filters. *International Journal of Robotics Research (IJRR)*, 23(7–8), 2004.

[148] N. Tomatis, I. Nourbakhsh, and R. Siegwart. Hybrid simultaneous localization and map building: a natural integration of topological and metric. *Robotics and Autonomous Systems*, 44(1), 2003.

[149] A. Torralba, K. Murphy, and W. Freeman. Contextual models for object detection using boosted random fields. In *Advances in Neural Information Processing Systems (NIPS)*, 2004.

[150] J.K. Tsotsos, G. Verghese, S. Dickinson, M. Jenkin, A. Jepson, E. Milios, F. Nuflo, S. Stevenson, M. Black, D. Metaxas, S. Culhane, Y. Ye, and R. Mann. PLAYBOT: A visually-guided robot for physically disabled children. *Image & Vision Computing, Special Issue on Vision for the Disabled*, 16(4), 1998.

[151] S. Tzafestas. Guest editorial, special issue on autonomous mobile robots in health care services. *Journal of Intelligent and Robotic Systems*, 22(3-4):177–179, 1998.

[152] I. Ulrich and I. Nourbakhsh. Appearance-based place recognition for topological localization. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2000.

[153] D. Vail, J. Lafferty, and M. Veloso. Feature selection in conditional random fields for activity recognition. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2007.

[154] P. Viola and M. Jones. Robust real-time ob ject detection. In *International Journal of Computer Vision*, volume 57, page 2, 2004.

[155] C. Wellington, A. Courville, and T. Stentz. Interacting Markov random fields for simultaneous terrain modeling and obstacle detection. In *Proc. of Robotics: Science and Systems (RSS)*, 2005.

[156] B. Yamauchi and R. Beer. Spatial learning for navigation in dynamic environments. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, Special Issue on Learning Autonomous Robots, 1996.

[157] C. Yanover and Y. Weiss. Most probable configurations using loopy belief propagation. In *Advances in Neural Information Processing Systems (NIPS)*, 2003.