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

A Conditional Random Field Model for Place and Object Classification

John G. Rogers III and Henrik I. Christensen

Abstract— **Place categorization and object recognition are competencies needed by robots to perform a variety of service tasks in the home, such as fetch-and-carry, retrieval, cleaning, meal preparation, and companionship. Context is a powerful cue for place categorization and object recognition; rooms are laid out in a specific fashion to enable comfortable and efficient living, and objects are used within rooms for tasks specific to that room. This paper will present a technique which leverages contextual cues for joint reasoning about object and room classification via a conditional random field model.**

I. INTRODUCTION

Domestic service robots will one day work in the home to perform useful tasks such as object retrieval, cleaning and organization, and security. The tireless support of these systems will enable the elderly to live independently by providing service, safety, and companionship. Despite significant automation already being present in the home such as dishwashers, washing machines, and robot vacuums, people face a steadily increasing amount of duties necessary to support their current lifestyles. In our society, people are looking to spend less time on domestic drudgery, robotic assistance in the household will be a welcomed development.

One of the most important competencies needed for domestic service robots is the ability to understand their surroundings well enough to perform their duties. Robots will be required to interact with objects in people's homes to perform tasks such as cleaning and meal preparation; an understanding of where these objects are located or belong is required to perform these tasks. People prefer to interact with robots in human terms, such as "Get the cup from the kitchen", instead of robot terms, such as "Get object 1372 from (4.2, 12.8, 1.2)". Room category representations in addition to room and object co-occurrence is needed to enable this interaction modality.

One of the most important tasks that a new domestic service robot must be capable of is the first one that it will perform when it is unpacked from its shipping crate: mapping its new home and familiarizing itself with the objects with which it will need to interact. The position of an object within the environment can be used as a cue for that object's identity. For example, the microwave oven is more likely to be found in the kitchen, and the toilet is more likely to be found in the bathroom. The knowledge of the label of the room currently inhabited by the robot can be used to narrow the potential classifications of the objects in the room. The recognition of some objects can also be used as a cue for the identity of other objects around them, such as the mouse is usually to be found to the right of the keyboard, or the light switch should be found on one side of the doorway.

The task of identifying objects in an unknown (and dynamic) environment should incorporate spatial location and object permanence. In this way, object recognition and SLAM are linked; the performance of each is improved by the other. Object recognition can provide a strong cue for data association, and spatial position can provide a strong cue for object identification. Prior identification of an object from a certain vantage point, combined with *object permanence*, the expectation that things remain where they were last seen for short periods of time, can be used to simplify the future recognition task; it limits the search space and permits less certain matches to be incorporated if they agree with previous measurements of the object.

This paper will present a technique for combining reasoning about object and room classification within a mapping framework. Object classification will consist of two components: first, a direct recognition based on SURF feature matching, and second, a bag-of-words technique for classification of objects which are not recognized by the first component. Room location and extent are measured and placed in a hybrid metric/topological map. Object recognition and classification measurements are provided along with room adjacency and object-in-room relationships to a conditional random field(CRF) model. The CRF model uses loopy belief propagation to estimate the marginals on each object and room node.

Related work will be presented in Section II. The specific algorithms and techniques developed for this paper will be presented in Section III. The experimental procedure will be outlined in Section IV and results will be discussed in Section V. Conclusions will be presented in Section VI and an outline of our future research direction will be presented in Section VII.

II. RELATED WORK

A visual place recognition technique is presented in [26]. This technique extracts SIFT features [9] and performs room recognition with a support vector machine (SVM). This technique is demonstrated to generalize across several experimental settings, and even to images taken by robots with catadioptric as well as perspective cameras. Another technique which uses SIFT features for visual localization is presented in [2] and a probabilistic model for appearance-based localization or place recognition is presented in [7]. Our approach is different in that we first segment rooms and

J. Rogers is a Ph. D. students at Georgia Institute of Technology College of Computing {jgrogers}@gatech.edu

H. Christensen is the Kuka Chair of Robotics at Georgia Institute of Technology College of Computing hic@cc.gatech.edu

then recognize objects within that room for classification in a CRF model.

A hybrid metric/topological cognitive model called the Spatial Semantic Hierarchy (SSH) is presented by Kuipers in [8]. This model incorporates representations for robot reactive behaviors and control at its lowest levels. At the higher levels of the hierarchy, the SSH makes use of a topological map representation as well as a metric map representation. Another technique by Zender *et.al.* applies the *conceptual spaces* representation of Gärdenfors [6] to mobile robots by representing topological relationships, metric maps, objects, and people in [27]. We will also be building a topological map representation to determine room adjacency and which objects appear within each room for use in the CRF model. We also build a metric map of the geometric coordinates of these objects which the robot uses in a simultaneous localization and mapping (SLAM) framework to keep track of its location, as well as to generate the topological map.

A technique for classifying places or rooms based upon range data from a laser scanner is presented in [10]. This technique demonstrates good performance in categorizing among three classes of room, corridor, and doorway; however, it is unclear how well this technique would perform at the task of classifying specific types of rooms. The authors extended this work by using this representation for exploration using semantic information in [24].

Simultaneous Localization and Mapping(SLAM) has been an active research topic since the mid 1980s. The problem is the synthesis of two simpler problems: localization in a known map, and map building with known location. Currently, many researchers believe that SLAM is a solved problem; however, few plug-in systems exist and there are few commercial applications in the marketplace today. Many details about the early approaches to the SLAM problem can be found in [3] and modern approaches in [1].

Context can be a useful cue in recognition. Global features such as bag-of-words based texture recognition can help with scene recognition and improve object recognition [13]. In [22], a technique is developed for jointly segmenting and classifying the objects on a per-pixel basis in images. This technique also uses a conditional random field (CRF) model with shape *textons*, color distributions, image locations, and edges to segment and classify objects in the image. Semantic information about object co-occurrence was added to this CRF model in [15].

III. ALGORITHM

Our group uses the Robot Operating System (ROS) developed at Willow Garage [14] for low level behavior and interprocess communication. Our mobile robot, Jeeves, is a Segway RMP200 base which has been modified to be statically stable with support wheels, which can be seen in figure 2. The robot uses an Asus Xtion Pro 3D camera for object segmentation. The robot has several laser scanners which are used for obstacle avoidance, localization, mapping, and measuring the size of rooms; however, in this application the robot is tele-operated and uses the Hokuyo UTM30

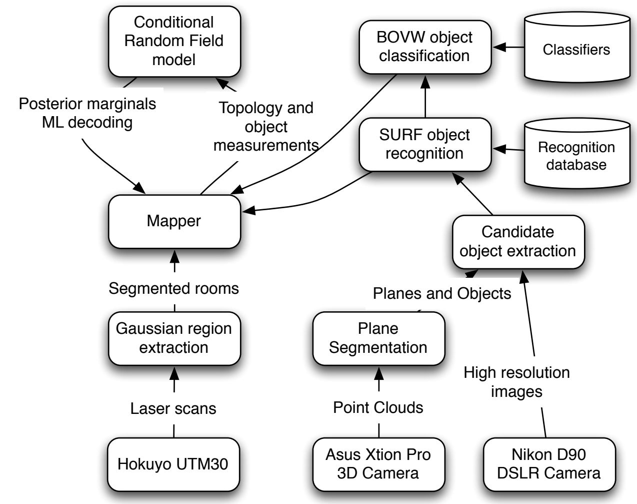

Fig. 1. A diagram of the components used in this paper.

Fig. 2. Our mobile robot ”Jeeves”, inspecting a cup on a table in our test environment.

laser scanner to measure rooms. The robot uses a Directed Perception PTU-46-70 pan-tilt unit to aim sensors on its upper extremity. Currently, the only sensor on the pan-tilt unit is a Nikon D90 DSLR camera with an 18mm lens. This high resolution camera is used to gather detailed images of target objects for classification and recognition.

An overview diagram of our system can be seen in figure 1. Laser range measurements are provided by the Hokuyo UTM30 laser scanner to the Gaussian place segmentation module (section III-E), which provides estimates of room shape and size to the mapping module(section III-F). Point cloud data is gathered by the Asus Xtion Pro depth camera and then is filtered and segmented to extract candidate object point clouds. Candidate object point clouds are then projected into high resolution images from the Nikon D90 camera to extract candidate image regions. Object recognition(section III-C) and classification(section III-D) is performed on these candidate image regions and the results are provided to the mapper. The mapper (section III-F) builds a metric map of the location of places and objects, and also builds a topological map which it sends to the conditional random field module (section III-A) along with object classification distributions and recognition results. The conditional random field module computes the posterior distribution over place and object labels, as well as the maximum likelihood configuration of the world. Detailed descriptions of these modules is provided below.

A. *Conditional random field model*

Domestic environments are organized with specific objects located in particular rooms such as toothpaste in the bathroom and calculators in the office. Rooms are also arranged in specific patterns to enable efficient and comfortable living, such as bathrooms are next to bedrooms, and dining rooms are adjacent to kitchens.

We use a conditional random field (CRF) model of room adjacency and object-room compatibility to reason about these design patterns to determine room label. CRF models express the probability of configurations of variables through a set of compatibility functions. In the case of modeling room adjacency and room-object co-occurrence, there is one type of compatibility function which favors likely pairs of rooms and combinations of rooms and objects.

$$p(y|x) = \frac{1}{Z(x)} \exp \sum_{k=1}^{K} \lambda_k f_k(y,x) \tag{1}$$

In the application of equation 1 used in this paper, feature functions represent room adjacency and object in room properties. More appropriate combinations of these values are given larger values by the feature functions. In the current implementation, we have assigned values for the feature functions which correspond to combinations that made sense to us; however, it is not important what these values are. Training feature functions from data will be the subject of future research.

The CRF model is chosen for this task instead of the more typical Markov Random Field (MRF) model because it allows us to incorporate certain pieces of evidence as absolute, and solve for the distribution of uncertain variables conditioned on this absolute evidence. The evidence which we consider absolute are objects which have been identified by SURF feature matching. The evidence which we consider uncertain are objects *classified* by a bag-of-words classifier. The true potential of the choice of CRFs over MRFs will be the subject of future work when feature functions are trained from data. Learning with CRFs can be performed discriminatively, to maximize the likelihood of the labels given the certain evidence. In contrast, learning with MRFs is generative and maximizes the joint likelihood of the labels and evidence.

We use the UGM implementation of CRFs from [21]. Room connectivity topology and object recognition/categorization measurements are sent to the CRF. Posterior marginal room label distributions and object category distributions are returned back to the mapping program.

Fig. 3. Point cloud data is projected into the camera image. Horizontal planes are extracted (yellow points) and objects are clustered points which appear above the plane (green points). The region of interest is selected based upon the projection of object points into the image (blue rectangle)

B. *Object Segmentation*

Objects are segmented from the background with 3D point clouds. We leverage the Asus Xtion Pro depth camera (which has a Primesense sensor similar to the Microsoft Kinect) to observe the 3D structure of the scene immediately in front of the robot. The camera software provides a 3D point cloud which we operate on with the Point Cloud Library (PCL)[20]. First, the point cloud is spatially filtered to contain only relevant portions which fall within a volume of interest in front of the robot. The point cloud is then downsampled to one point per cubic centimeter using a voxel filter. We then extract up to 4 planes from the remaining point cloud using a RANSAC [5] technique available from the PCL library. These planes are then analyzed to find remaining points which lie above them. These remaining points are clustered to find candidate objects. The points from each candidate object of appropriate size are projected into a high resolution camera image which was taken at the same time. The extent of these points in the camera image is used to segment the candidate object from the background. This segmented image region is then passed on to the subsequent components to perform object recognition or classification. An example selected region of interest, object points, and table surface points can be seen projected into a high resolution image in figure 3.

C. *Object recognition*

After an candidate object image is segmented using the technique described in section III-B, we attempt to recognize it using a SURF feature matching technique. First, the SURF features in the candidate object image are extracted. These SURF features are compared to the SURF features previously extracted in a model database. The set of matched features to a model element are used to compute a homography to the model image using RANSAC [5]. This homography is then used to filter out erroneous feature matches and select an inlier set. If this set is of sufficient size, then the object is said to be recognized absolutely and this information is then sent

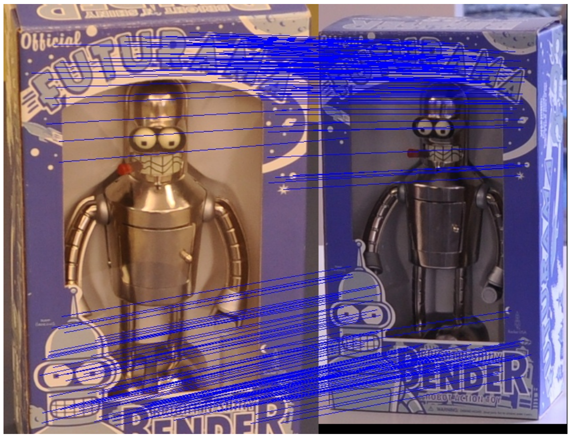

Fig. 4. A successfully recognized object from the class ”Toys”. The model image appears on the right, the candidate object image appears on the left. Blue lines indicate matches which are inliers to the homography filter. There are enough inlier matches, so the recognition algorithm accepts this as a match.

Fig. 5. An object which does not match with class ”Sundries”. The model image appears on the right, and the candidate object image appears on the left. Blue lines indicate matches which are inliers to the homography filter. There are too few inlier matches, so the recognition algorithm rejects this as a match.

to the mapping system along with geometric measurements of the object’s location. If there are not enough inlier feature matches, then the object is not recognized and must be classified by the next module described in section III-D. The use of a homography is only fully correct when the candidate object is planar; further developments include 3D feature coordinates and matches consistent with camera projections. An example of a correctly identified object can be seen in figure 4. An example where this technique identifies a non-matching object can be seen in figure 5.

D. Object classification

If a candidate object cannot be recognized by any of the models using the object recognition technique described in section III-C, then we employ a probabilistic object classification technique to give a distribution over classes. The technique for object classification is based upon the *bag of visual words* approach of [23]. This technique requires a

Fig. 6. Visual words extracted from a candidate object image. Circle color indicates which visual word a given feature is assigned to by vector quantization. The size of each circle indicates the scale parameter for the underlying SURF feature.

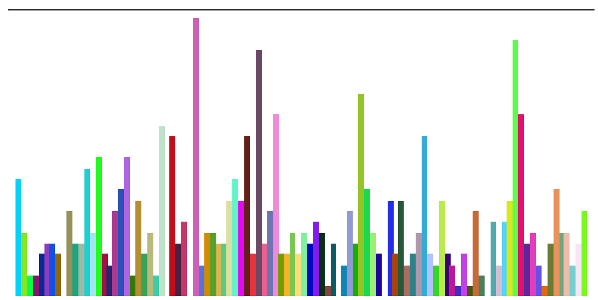

Fig. 7. The histogram of visual words appearing in figure 6

vocabulary of visual words which was learned by performing K-means clustering on SURF descriptors extracted from the *Caltech 101* [4] training set. SURF features are extracted from an image of a candidate object and are vector-quantized to the nearest visual word in the vocabulary. The visual words extracted from a candidate object image can be seen in figure 6. A histogram is built by counting the frequency of the appearance of each visual word in this image of the candidate object weighted by the term frequency inverse document frequency (TF-IDF) [16], see figure 7 for the histogram generated from the visual words seen in figure 6. A set of Relevance Vector Machines (RVMs) are trained to recognize object categories using this visual word histogram. Each of the RVMs is trained to recognize one category. The output of an RVM is a probability, unlike the output of a Support Vector Machine (SVM).

To perform object categorization on a segmented candidate object image, SURF features are first extracted and then vector quantized into their nearest visual word. The visual

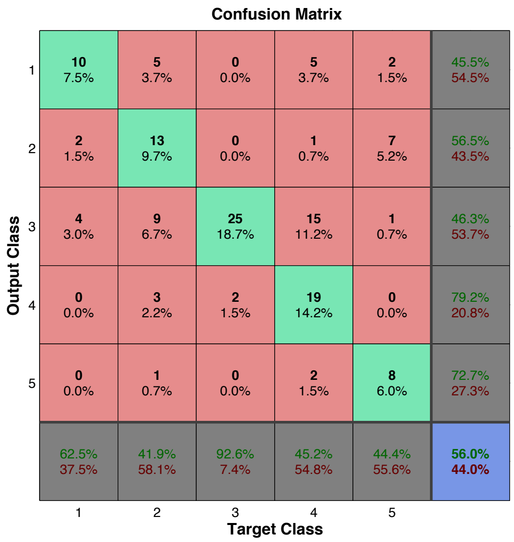

Fig. 8. Confusion matrix from a 5 class experiment on the Caltech 101 dataset

words present in this image are collected in a histogram and normalized. The set of RVMs representing the object categories each analyze the histogram of visual words and give a probability that the given candidate object is a member of that class. These probabilities are combined under the assumption that each RVM is independent by accumulating them into a multinomial distribution and re-normalizing. The resulting distribution over object classes, along with the geometric details of the object, are sent to the mapper for integration into the model. To establish the effectiveness of the RVM on bag-of-words technique, we performed an experiment on the Caltech 101 dataset [4]. Five categories were selected and objects of interest were extracted using the annotations provided with the dataset. Half of the images were used to train the RVM model using 3-fold cross-validation for RVM radius parameter selection. The resulting RVMs were used to classify images from the other half of the images reserved for the test set. The confusion matrix from this experiment can be seen in figure 8.

E. Gaussian place segmentation

We presented in [12] a straightforward technique to represent the extent of a ”room” by a Gaussian extracted from the laser range data at a given pose. The assumption made here is that the robot can mostly only see within one room or location and that the laser will sufficiently cover the area of the room. The use of this model allows us to perform a simple Mahalanobis distance test to determine if the robot is within one of our ”rooms”, or if a new room Gaussian should be created. We keep track of the rooms which are traversed by the robot to build up a topological representation of room adjacency. When object recognitions or classification distributions are observed by the robot, they are assigned to the room Gaussian where the robot currently resides. Each room Gaussian is given a distribution over room labels as a result of the CRF algorithm described in section III-A.

F. Mapping

We have developed a library for mobile robot mapping called *OmniMapper*. We have used it to develop virtual measurements [25], learned object recognition mapping [19], for multi-robot mapping [17], and to determine mapping performance with sensory degradation [18]. Interested readers should refer to these papers for more complete implementation details.

The mapper uses the GTSAM library to optimize a graph of measurements between robot poses along a trajectory, and between robot poses and various landmarks in the environment. Measurements come from various software components; in this case, measurements come from the object recognition and classification modules described in sections III-C and III-D. Measurements of simple objects like points, lines, and planes are data associated to mapped landmarks with the joint compatibility branch and bound (JCBB) technique in [11]. Measurements of richer landmarks such as objects or signs are data associated based upon interpretation of this semantic information [19].

IV. EXPERIMENT

We performed a series of experiments on log data gathered from the Aware Home test facility at Georgia Tech. The topology of the Aware Home is typical of modern American architecture. There is a kitchen with a small dining area opens into the living room. The living room is connected to a hallway which connects to an office, bathroom, two bedrooms, and a closet. The Asus Xtion Pro 3D camera is currently placed on the robot at a height that makes it difficult to observe tables and objects if they are more than about 1 meter above the ground, so we made sure that there was a counter, desk, or table at or below this height. In the future, the 3D camera will be placed on the pan-tilt unit alongside the high resolution camera to enable observations of surfaces and objects at any height.

We tele-operated the robot in a set of trajectories in the kitchen, living room, office and bathroom. In each room, the robot was made to collect high resolution camera images of the table surfaces and the objects thereupon.

When an image is captured by the high resolution camera, the 3D camera captures a point cloud. The point cloud is filtered to select a volume of interest in front of the robot. Planes are extracted from the remaining points, and objects are selected as clusters of points which lie above horizontal planes. The object points are projected into the high resolution image to find a region of interest for visual feature analysis. Each region of interest becomes a candidate object image.

Some examples of trained objects and the categories to which they belong can be seen in figure 9. The current set of object classes include *food*, *toys*, *cooking*, *sundries* (medicine, soap, etc.), *tools*, and *electronics*. This taxonomy was selected to assign objects to classes which should logically correspond to room assignment. The rooms currently available to the CRF model are *kitchen*, *bathroom*, *living room*, *office*, *hall*, and *bedroom*.

Fig. 9. Example images from recognition database. These objects should be recognized by the SURF recognition module. Additional objects were used to train the bag-of-words classifier module.

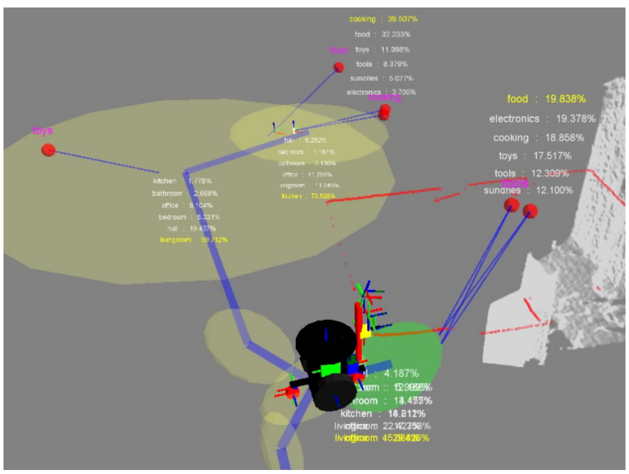

Fig. 10. Output posterior marginal distributions. Portion of topological map is rendered into metric map display. Room nodes are shown as yellow ellipsoids, except the current room is shown in green. Topological room connectivity is shown with thick blue lines between room nodes. Posterior room label distributions is shown below each room node (unlabeled nodes are Hall nodes, and are omitted for clarity). The yellow entry is the most likely marginal room label. Objects are shown as red spheres with thin blue lines linking them to the poses from which they are observed. Recognized objects have their labels displayed in pink. Objects which have not been recognized and were instead categorized have their posterior marginal distributions displayed above them.

V. RESULTS

The posterior marginal distributions over some of the objects and rooms in a test run can be seen in figure 10. The most likely decoding can be seen in figure 11.

In one test run, the first time the robot entered the office it failed to recognize the Dremel tool or the Mac mini with SURF feature matching. The Mac mini was not part of the recognition database, and it was mistakenly categorized as a Food. This resulted in the office posterior distribution favoring the label "living room". When the robot re-entered the office at a later point in the test run, the Dremel tool was recognized directly and the office was relabeled correctly.

The Gaussian regions used for place segmentation performed poorly due to the fact that the laser scanner could only observe a 180 degree arc in front of the robot. When the robot is driven in reverse, it quickly exits the current Gaussian region and creates a new one. This new region similarly favors the region in front of the robot and therefore is also quickly exited. This problem can be addressed by either making use of laser scanners covering the front and back of the robot to provide a more complete view of the room, or to come up with a new representation for place segmentation.

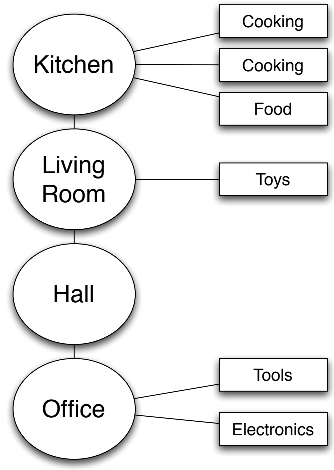

Fig. 11. The most likely configuration of rooms and places from a test run in the Aware Home (redrawn, some extraneous Hall nodes removed for clarity).

VI. CONCLUSIONS

We have demonstrated a model which is used to jointly classify objects and room labels on a mobile robot. This model incorporates information from two types of object measurements: recognition of previously seen and trained objects, as well as classification of novel objects. This model was tested on a mobile robot teleoperated in a real domestic environment with objects. The current experiments focus on determining that these recognition, mapping, and reasoning components can be made to work together to accomplish rudimentary understanding of the purpose and structure of a domestic environment. The selection of objects used in the experimental tests consisted of many of the same objects which were explicitly used for training in the recognition and classification components; however, we maintain that it is reasonable and even desirable for a mobile robot to acquire models of the objects with which it will be interacting through human interaction in addition to self training.

VII. FUTURE WORKS

The future purpose of the CRF model is to enable autonomous exploration and mapping, as well as service tasks. The current implementation establishes that this reasoning model is useful for leveraging context for classification of objects and places. The posterior marginal distribution on the CRF model can be used to direct exploration. The robot can search for additional objects in rooms where the label entropy is large, or it can examine unknown objects in greater detail. The robot can select certain unknown objects to pose a small set of questions to a human operator by selecting the objects which are most likely to provide a great deal of information to the rest of the model when they become disambiguated.

Other tasks which use the CRF model to understand human commands should use the maximum likelihood decoding (which is not in general the same as the maximum marginal likelihood) of the entire graph. In this way, the robot can use the context of the entire model to perform tasks such as "get a cup from the kitchen".

Currently, model parameters such as room adjacency and room-object compatibility are assigned to what we thought were reasonable values. In the future, we will train model parameters from observations. We can look at a corpus of architectural floor-plans to establish the room adjacency model. We would like to train the room-object compatibility model by collecting a survey of where people find classes of objects within their own homes. This training information should also be augmented by online adaptation to new observations.

The current recognition database is limited in size due to the amount of time needed to perform SURF feature matching on high resolution images. We intend to use the bag-of-words histogram comparison to filter out low probability class matches as an initial step. We will expand object classes and recognition database to cover the key components which will be useful to robots in performing their tasks as well as in understanding their environments.

VIII. ACKNOWLEDGMENTS

This work was made possible through the ARL MAST CTA project. The authors would like to thank Carlos Nieto for his assistance in data collection and Alex Trevor for helpful discussion.

REFERENCES

[1] T. Bailey and H. Durrant-Whyte. Simultaneous localisation and mapping (SLAM): Part II state of the art. *Robotics and Automation Magazine*, September 2006.

[2] O. Booij, B. Terwijn, Z. Zivkovic, and B. Krose. Navigation using an appearance based topological map. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 3927–3932. IEEE, 2007.

[3] H. Durrant-Whyte and T. Bailey. Simultaneous localisation and mapping (SLAM): Part I the essential algorithms. *Robotics and Automation Magazine*, June 2006.

[4] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. In *Computer Vision and Pattern Recognition (CVPR)*, 2004.

[5] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[6] P. Gärdenfors. *Conceptual Spaces: the Geometry of thought*. MIT Press, 2000.

[7] B. J. A. Kröse, N. Vlassis, R. Bunschoten, and Y. Motomura. A probabilistic model for appearance-based robot localization. *Image and Vision Computing*, 2001.

[8] B. Kuipers. A hierarchy of qualitative representations for space. In *Spatial Cognition*, 1998.

[9] D. G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision (ICCV)*, 1999.

[10] O.M. Mozos, C. Stachniss, and W. Burgard. Supervised learning of places from range data using adaboost. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, pages 1730–1735. IEEE, 2005.

[11] J. Neira and J.D. Tardós. Data association in stochastic mapping using the joint compatibility test. In *IEEE Transactions on Robotics and Automation*, 2001.

[12] C. Neito-Granda, J. G. Rogers III, A. J. B. Trevor, and H. I. Christiensen. Semantic map partitioning in indoor environments using regional analysis. In *IEEE International Conference on Intelligent RObots and Systems (IROS)*, 2010.

[13] A. Oliva and A. Torralba. The role of context in object recognition. In *Trends in cognitive science*, 2008.

[14] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng. ROS: an open-source robot operating system. In *ICRA Workshop on Open Source Software*, 2009.

[15] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. Objects in context. In *International Conference on Computer Vision (ICCV)*, 2007.

[16] J. Ramos. Using tf-idf to determine word relevance in document queries. In *Proceedings of the First Instructional Conference on Machine Learning*, 2003.

[17] J. Rogers, M. Paluri, A. Cunningham, H. I. Christensen, N. Michael, V. Kumar, J. Ma, and L. Matthies. Distributed autonomous mapping of indoor environments. In *SPIE Defense and Security*, 2011.

[18] J. Rogers, A.J.B. Trevor, C. Nieto, A. Cunningham, M. Paluri, N. Michael, F. Dellaert, H.I. Christensen, and V. Kumar. Effects of sensory perception on mobile robot localization and mapping. In *International Symposium on Experimental Robotics (ISER)*, 2010.

[19] J. G. Rogers III, A. J. B. Trevor, C. Nieto-Granda, and H. I. Christensen. Simultaneous localization and mapping with learned object recognition and semantic data association. In *Submitted to IEEE International Conference on Intelligent RObots and Systems (IROS)*, 2011.

[20] R. B. Rusu and S. Cousins. 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, 2011 2011.

[21] Mark Schmidt. Undirected graphical model. In *http://www.di.ens.fr/ mschmidt/Software/UGM.html*, 2011.

[22] J. Shotton, J. Winn, C. Rother, and A. Criminisi. TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *European Conference on Computer Vision (ECCV)*, 2006.

[23] J. Sivic and A. Zisserman. Efficient visual content retrieval and mining in videos. *Advances in Multimedia Information Processing-PCM 2004*, pages 471–478, 2005.

[24] C. Stachniss, O.M. Mozos, and W. Burgard. Speeding-up multi-robot exploration by considering semantic place information. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 1692–1697. IEEE, 2006.

[25] A. J. B. Trevor, J. G. Rogers III, C. Nieto-Granda, and H. I. Christensen. Applying domain knowledge to SLAM using virtual measurements. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2009.

[26] M.M. Ullah, A. Pronobis, B. Caputo, J. Luo, R. Jensfelt, and H.I. Christensen. Towards robust place recognition for robot localization. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 530–537. IEEE, 2008.

[27] H. Zender, O. Martínez 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, 2008.