Information Theoretic Question Asking to Improve Spatial Semantic Representations

Sachithra Hemachandra, Matthew R. Walter, and Seth Teller
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139 USA
{sachih, mwalter, teller}@csail.mit.edu

Abstract

In this paper, we propose an algorithm that enables robots to improve their spatial-semantic representation of the environment by engaging users in dialog. The algorithm aims to reduce the entropy in maps formulated based upon user-provided natural language descriptions (e.g., "The kitchen is down the hallway"). The robot's available information-gathering actions take the form of targeted questions intended to reduce the entropy over the grounding of the user's descriptions. These questions include those that query the robot's local surround (e.g., "Are we in the kitchen?") as well as areas distant from the robot (e.g., "Is the lab near the kitchen?"). Our algorithm treats dialog as an optimization problem that seeks to balance the information-theoretic value of candidate questions with a measure of cost associated with dialog. In this manner, the method determines the best questions to ask based upon expected entropy reduction while accounting for the burden on the user. We evaluate the entropy reduction based upon a joint distribution over a hybrid metric, topological, and semantic representation of the environment learned from user-provided descriptions and the robot's sensor data. We demonstrate that, by asking deliberate questions of the user, the method results in significant improvements in the accuracy of the resulting map.

Introduction

Robots are increasingly being deployed in human-occupied environments. In order to be effective partners, robots need to reason over representations of these environments that model the spatial, topological, and semantic properties (e.g., room types and names) that people associate with their environment. An efficient means of formulating these representations is through a guided tour in which a human provides natural language descriptions of the environment (Zender et al. 2008; Hemachandra et al. 2011; Walter et al. 2013; Hemachandra et al. 2014). With these approaches, the robot takes a passive role whereby it infers information from the descriptions that it fuses with its onboard sensor stream.

The challenge to learning is largely one of resolving the high-level knowledge that language conveys with the low-level observations from the robot's sensors. The user's descriptions tend to be ambiguous, with several possible interpretations (*groundings*) for a particular environment. For example, the user may describe the location of the kitchen as being "down the hall," yet there may be several hallways nearby, each leading to a number of different rooms. Furthermore, language grounding typically requires a complete map, however the robot may not yet have visited the regions that the user is referring to. It may be that the user is describing a location known to the robot or a new location outside the field-of-view of its sensors.

Figure 1: A user giving a tour to a robotic wheelchair designed to assist residents in a long-term care facility.

Rather than try to passively resolve the ambiguity in the inferred map, the robot can take active information-gathering actions. These may take the form of physical exploration of the environment or, as we consider in this paper, targeted questions asked of the user. There are several challenges to using dialog in order to improve the accuracy of the inferred map in an effective manner. The first involves context. It would be beneficial if the algorithm was not restricted to questions that query the robot's current location. However, asking the user about temporally and spatially distant locations necessitates that the questions provide the user with sufficient context. Second, the questions should be structured in such a way that the answers are as informative as possible. Third, it is important that the method accounts for the social cost incurred by engaging the user in dialog, for example, by not asking too many questions.

In this paper, we consider the scenario in which a robot acquires an environment model through a guided tour (Wal-

ter et al. 2013; Hemachandra et al. 2014), where the human shows the robot around the environment while providing natural language descriptions. During the tour, the robot maintains a distribution over the *semantic graph*, which is a metric, topological and semantic representation of the environment, using a Rao-Blackwellized particle filter. At each timestep, the robot decides between actions that either follow the guide or that ask a question to improve its representation. We formulate the decision process as a QMDP (Littman, Cassandra, and Kaelbling 1995), where we evaluate actions as a Markov Decision Process (MDP) for each possible configuration of the world (particle), and select the best action using the QMDP heuristic. This allows us to balance the information gained by asking questions of the user with their associated cost. The algorithm reasons over the natural language descriptions and the current learned map to identify the (possibly null) question that best reduces ambiguity in the map. The algorithm considers egocentric and allocentric binary (yes/no) questions that consist of spatial relations between pairs of regions. These regions may be local to the robot in the case of situated dialog (e.g., "Are we in the kitchen?", "Is the lab on my right?") or distant in the case of non-situated dialog (e.g., "Is the lounge next to the conference room?"). We associate with each question a cost that reflects the burden on the user and a reward based on the information gain for each possible answer. The algorithm selects the best action based upon the expected Q value using the QMDP formulation.

We demonstrate that this question asking policy reduces the ambiguity in natural language descriptions and, in turn, results in semantic maps of the environment that are more accurate than the current state-of-the-art.

Related Work

Several approaches exist that construct semantic environment models using traditional robot sensors (Kuipers 2000; Zender et al. 2008; Hemachandra et al. 2011; Pronobis and Jensfelt 2012), while others have looked at additionally integrating natural language descriptions to improve the semantic representations (Walter et al. 2013; Williams et al. 2013; Hemachandra et al. 2014). With most of these techniques, however, the robot only passively receives observations, whether they are from traditional sensors or user-provided descriptions.

Related work exists that endows robots with the ability to ask questions of the user in the context of following guided tours (Kruijff et al. 2006) and understanding a user's commands (Deits et al. 2013). Kruijff et al. (2006) outline a question asking procedure mainly to determine robust room segmentation by asking about the presence of doorways. However, they do not tackle allocentric descriptions, reason about uncertainty over groundings, maintain multiple hypothesis, nor reason about entropy. More recently Deits et al. (2013) have looked at question asking from an information-theoretic perspective in the scenario of following natural language manipulation commands. They use an information gain-based evaluation method to evaluate the best questions to ask to reduce the entropy over the grounding for a given natural language command. However, the questions they ask are more straightforward, and do not explicitly provide context to the human. While we use a similar information gain metric to drive our approach, we formulate the problem as a decision problem, where the robot has to decide between continuing the tour or interrupting the tour to ask a question. Furthermore, Deits et al. (2013) do not reason over when to ask the questions, since they immediately follow the corresponding command. In our case, a question can simultaneously refer to areas that the user described at distant points in time. This necessitates that we consider when it is most meaningful to ask the question and that it be phrased in a manner that provides sufficient context. Our expected information gain metric is similar to the work of Stachniss, Grisetti, and Burgard (2005), who decides the best exploration-based motion actions to improve the entropy over the map.

Semantic Graph Representation

Spatial-Semantic Representation

We define the semantic graph (Hemachandra et al. 2014) as a tuple containing topological, metric and semantic representations of the environment. The topology G_t is composed of nodes n_i that denote the robot's trajectory through the environment (with a fixed 1 m spacing) and edges that denote connectivity. We associate with each node a set of observations that include laser scans z_i, semantic appearance observations a_i based on laser l_i and camera i_i models, and available language observations $\lambda_i \in \Lambda$. We assign nodes to regions $R_\alpha = \{n_1, .., n_m\}$ that represent spatially coherent areas in the environment intended to be compatible with human concepts (e.g., rooms and hallways).

The vector X_t consisting of the pose x_i of each node n_i constitutes the metric map, which takes the form of a pose graph (Kaess, Ranganathan, and Dellaert 2008) according to the structure of the topology. The semantic map L_t is modeled as a factor graph with variables that represent the type (e.g., office, lounge) and colloquial name (e.g., "Carrie's office") of each region in the environment. The method infers this information from observations made from scene classifiers (image and laser) as well as grounding the user's natural language descriptions (Hemachandra et al. 2014). In this paper, we consistently segment the regions using spectral clustering (compared to sampling segments in Hemachandra et al. (2014)). We also use a template-based door detector to segment regions.

Grounding Natural Language Descriptions

We consider two broad types of natural language descriptions provided by the guide. Egocentric descriptions that involve the robot's immediate surround are directly grounded to the region in which the description was provided. Allocentric descriptions that provide information about distant regions require more careful handling.

We parse each natural language command into its corresponding Spatial Description Clauses (SDCs), a structured language representation that includes a figure, a spatial relation and possibly a landmark (Tellex et al. 2011). For example, the allocentric description "the lounge is down the hall-

way," results in an SDC in which the figure is the "lounge," the spatial relation is "down from," and the landmark is the "hallway". With egocentric descriptions, the landmark or figure are implicitly the robot's current position.[1]

Algorithm 1: Semantic Mapping Algorithm

Input: $P_{t-1} = \left\{ P_{t-1}^{(i)} \right\}$, and $(u_t, z_t, a_t, \lambda_t)$, where $P_{t-1}^{(i)} = \left\{ G_{t-1}^{(i)}, X_{t-1}^{(i)}, L_{t-1}^{(i)}, w_{t-1}^{(i)} \right\}$

Output: $P_t = \left\{ P_t^{(i)} \right\}$

1) Update Particles with odometry and sensor data.

for $i = 1$ *to* n **do**

1. Employ proposal distribution to propagate the graph sample based on u_t, λ_t and a_t.
   (a) Segment regions
   (b) Sample region edges
   (c) Merge newly connected regions
2. Update the Gaussian distribution over the node poses $X_t^{(i)}$ conditioned on topology.
3. Reevaluate language groundings and answered question and update the semantic layer L_t.
4. Update particle weights.

end

2.) Normalize weights and resample if needed.
3.) Evaluate action costs and carry out minimum cost action.

In order to ground each expression, the algorithm first identifies regions in the map that may correspond to the grounding based upon their semantic label likelihood. We normalize these likelihoods to compute the landmark grounding probability for each of these regions

$$p(\gamma_l = R_j) = \frac{p(\phi^l_{R_j} = \mathtt{T})}{\sum_{R_j} p(\phi^l_{R_j} = \mathtt{T})}, \tag{1}$$

where γ_l is the landmark region grounding and $\phi^l_{R_j}$ denotes the binary correspondence variable that specifies whether region R_j is the landmark. For each potential landmark region, the algorithm then calculates the likelihood that each region in the map corresponds to the figure based on a model for the spatial relation SR. We arrive at the overall figure grounding likelihood by marginalizing over the landmarks

$$p(\phi^f_{R_i} = \mathtt{T}) = \sum_{R_j} p(\phi^f_{R_i} = \mathtt{T} | \gamma_l = R_j, SR)\, p(\gamma_l = R_j), \tag{2}$$

[1]We make the assumption that the descriptions are provided with respect to the robot's reference frame and not the user's.

where $\phi^f_{R_i}$ is the correspondence variable for the figure. We normalize these likelihoods for each potential figure region

$$p(\gamma_f = R_i) = \frac{p(\phi^f_{R_i} = \mathtt{T})}{\sum_{R_i} p(\phi^f_{R_i} = \mathtt{T})}. \tag{3}$$

This expresses the likelihood of the correspondence variable being true for each figure region R_j in the factor graph in the semantic layer. However, when there is uncertainty over the landmark or figure grounding, the likelihood of the label associated with the figure region can become diluted.

In our previous approaches (Walter et al. 2013; Hemachandra et al. 2014), we commit to a description once the likelihood of its grounding exceeds a threshold. We improve upon this in this paper by continuously re-grounding the language when relevant regions of the map change. These changes can be in the form of updates to the metric position of the figure or landmark regions (e.g., due to a loop closure), or new potential landmark or figure regions being visited and added to the map.

Algorithm

Algorithm 1 outlines the process by which robot updates its representation and decides on the optimal action. At each time step, the system integrates the odometry and sensor information to update the distribution over the semantic graph. This includes reevaluating the language descriptions and answers to questions from the guide. Then, the algorithm evaluates the cost of each valid (possibly null) dialog action, and executes the one with the highest expected Q value. The following section elaborates on our action selection procedure.

Action Selection

In this section, we outline the action selection procedure employed by the algorithm. We treat the guided tour as an MDP, with associated costs for taking each action. These actions include following the person, staying in place, and asking a particular question. We define an additional set of question asking actions dependent on the current number of allocentric descriptions provided by the guide. We introduce a cost function for these question asking actions based upon the expected information gain for each question as well as a measure of social burden.

We define the state S_{t+1} as a tuple of $\{P_t^{(i)}, a_t, z_t^a\}$, where $P_t^{(i)}$ is particle i at time t, a_t is the action taken, and z_t^a is the resulting observation. For a single particle, we define the Q value as

$$\begin{aligned} Q(S_t, a_t) &= \sum_{S_{t+1}} \gamma V(S_{t+1}) \times p(S_{t+1} | S_t, a_t) - \mathcal{C}(a_t) \\ &= \sum_{S_{t+1}} \gamma \mathbb{E}(V(S_{t+1})) - \mathcal{C}(a_t), \end{aligned} \tag{4}$$

where the value of S_{t+1}

$$V(S_{t+1}) = \mathcal{F}(I(a_t)) \tag{5}$$

is a function of the information gain, and the cost of question asking action a_t

$$\mathcal{C}(a_t) = \mathcal{F}(f(a_t)) \tag{6}$$

is a function of the feature set of each action. We use a discounting factor $\gamma = 1$.

At each time step, the robot takes the best action a_t^B from the available set of actions using the QMDP heuristic.

$$a_t^B = \arg\max_{a_t} \sum_{S_t} p(S_t)Q(S_t, a_t), \tag{7}$$

where $p(S_t)$ is the particle weight $w_t^{(i)}$.

Action Set

The action set consists of the "Follow Person" action $A_{\mathcal{F}}$, "Stay-In-Place" action $A_{\mathcal{S}}$, and the valid set of question asking actions. The "Follow Person" action $A_{\mathcal{F}}$ is available at all times except when the robot is waiting for an answer to a question, when only $A_{\mathcal{S}}$ is available for selection. We derive our questions from a templated set for each grounding entity in a natural language description. These templates can be categorized into two basic types.

I The simple template takes a spatial relation from the set of spatial relations (near, away, in front, behind, left of, right of) and a grounding variable to create a question of the type "Is the kitchen in front of me?". For such questions, the possible answers are "yes," "no," and "invalid" (for questions that do not make sense given a spatial entity).

II The more complex template defines questions in terms of spatial relations between non-local locations in the environment. If the robot is highly confident of the semantic label of a particular location, it could generate a question about regions close to that entity to resolve uncertainty. For example, when the robot is uncertain about the location of the "lounge," but thinks one possibility is the space in front of the "conference room," while several are not, it could ask "Is the lounge in front of the conference room?".

The robot can only use questions of the first type to ask about spatial regions in its immediate vicinity. As such, the ability to receive useful information is limited to instances when the robot is near a potential hypothesized location. Questions of the second type allow the robot to reduce its uncertainty even when a hypothesized location is not within its immediate vicinity. However, this may place a higher mental burden on the user who must then reason about spatial entities outside their immediate perception range.

Value Function

We define the value of the next state as a linear function of the information gain for each action. We define the next state S_{t+1} as the question and answer pair. Each next state is assigned a value based on the information gain for the related language grounding. Since there is a distribution over the set of answers that could be received for a given question, we evaluate the expected likelihood of transitioning to a particular state given a question. The likelihood of transitioning to each state is the likelihood of receiving a particular answer given the question.

Information Gain The information gain $I(a, z^a)$ for action a, as shown in Equation 8 is defined as the reduction in entropy by taking action a and receiving observation z^a. In our framework, the entropy is over a grounding variable γ_f created for a natural language description provided by the guide. Calculating the exact entropy is infeasible since the map might not yet be complete, and also because it is inefficient to calculate the likelihood of some spatial regions that are too far outside the local area. Therefore, we approximate the distribution based on the spatial regions considered during the language grounding step for the language description.

$$I(a, z^a) = H(\gamma_f|\Lambda) - H(\gamma_f|\Lambda, a, z^a) \tag{8}$$

In this paper, we concentrate on questions that can result in a discrete set of answers. This allows us to better model the expected change in entropy given the answer to the question (unlike an open ended answer which could be drawn from a large space of possible answers). However, in general, we can use the same approach for open ended questions as long as we can evaluate the expected information gain from these questions.

Given the answer, we evaluate the change it has on the distribution over the particular grounding variable. For most spatial relations, we define a range over which a particular question can be applied in a meaningful manner. For example, we only consider regions within a 20 m distance when evaluating a question. As such, we limit the entropy calculation to the regions for which the question is expected to be meaningful.

$$p(\gamma_f = R_i|\Lambda, a, z^a) = \frac{p(z^a|a, R_i) \times p(\gamma_f = R_i|\Lambda)}{\sum_{R_i} p(z^a|a) \times p(\gamma_f = R_i|\Lambda)} \tag{9}$$

The expected value of the next state is based on the transition function from the current state to the next state.

$$\mathbb{E}(V(S_{t+1})) = \sum_{z_j^a} \mathcal{F}(I(a|z_j^a)) \times p(z_j^a|S_t, a) \tag{10}$$

For the action $A_{\mathcal{F}}$, we assume that there is no change in the entropy as we are not modeling the expected change in the language groundings based on spatial exploration. Thus, the Q value for $A_{\mathcal{F}}$ is only the cost of the action.

Transition Likelihood

The transition function is the likelihood of receiving each answer given the state and the question asking action. We arrive at this value by marginalizing out the grounding variable. This results in a higher expected likelihood of receiving a particular answer if there were spatial regions that had a high likelihood of being the grounding and also fit the spatial relation in the question.

$$p(z_j^a|S_t, a) = \sum_{R_i} p(z_j^a|S_t, R_i, a) \times p(R_i|\Lambda) \tag{11}$$

Cost Function Definition

We define a hand-crafted cost function that encodes the desirability of asking a given question at each timestep. The

cost of each question asking action is a function of several relevant features. For this implementation, we have used the following:

i Time since last question asked

ii Time since last question asked about grounding

iii Number of questions asked about entity

In our current implementation, we use a linear combination of these features to arrive at a reasonable cost function. The weights have been set such that they result in negligible burden on the user and do not impeded the conducting of the tour. Ideally, these weights would be learned from user preferences based upon human trials.

For the person following action $A_{\mathcal{F}}$, we assign a fixed cost such that only a reasonably high expected information gain will result in a question being asked. The value was set empirically to achieve a reasonable level of questions.

Integrating Answers to the Representation

We couple each of the user's answers with the original question to arrive at an equivalent natural language description of the environment. However, since the question is tied to a particular spatial entity, we treat the question and answer pair together with the original description, according to Equation 9. As such, each new answer modifies the distribution over that grounding variable, and any informative answer improves the robot's representation.

When new valid grounding regions are added, we reevaluate both the original description as well as the likelihood of generating the received answer for each new region, and update the language grounding. Figure 2 shows the grounding likelihoods before and after asking three questions.

Results

We evaluate our algorithm on an indoor dataset in which a human gives a robotic wheelchair (Fig. 1) (Hemachandra et al. 2011) a narrated tour of MIT's Stata Center building. For this experiment, we inject three natural language descriptions at locations where the descriptions are ambiguous. We ran the algorithm on the dataset and a human provided answers to the questions. We outline the resulting semantic map and compare it with a semantic map that does not integrate language, and one that integrates language but does not ask questions of the guide.

Overall, the dataset contains six descriptions of the robot's location that the algorithm grounds to the current region, and three allocentric expressions that describe regions with relation to either landmarks in the environment (e.g., "the elevator lobby is down the hall") or to the robot (e.g., "the lounge is behind you"). The robot asked a total of five questions of the guide, four of which were in relation to itself, and one in relation to a landmark in the environment. In this experiment we ran the algorithm with one particle.

As can be seen in Table 1, the semantic map that results from integrating the answers received from the guide has much less uncertainty (and lower entropy) over the figure groundings. For all three descriptions, the robot was able to significantly reduce the entropy over the figure groundings by asking one to three questions each.

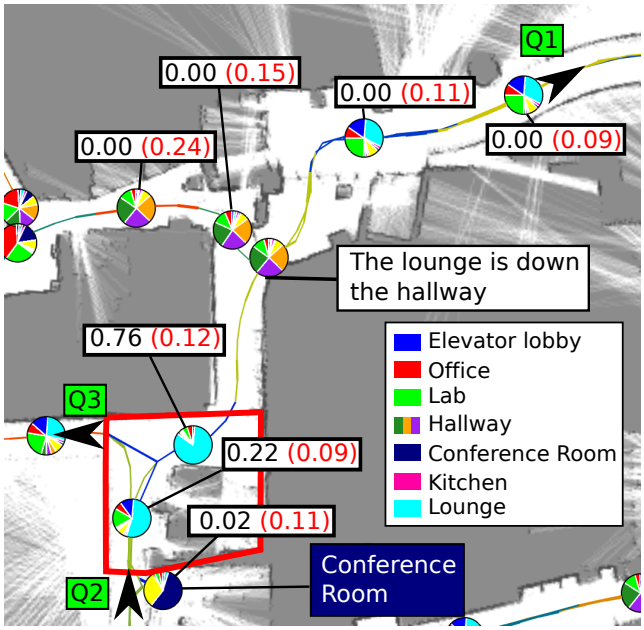

Figure 2: Language groundings for the expression "The lounge is down the hall". Grounding likelihood with questions is in black and without questions in red. Questions asked (and answers), Q1: "Is the lounge near the conference room?" ("Yes"); Q2: "Is the lounge on my right?" ("No"); Q3: "Is the lounge behind me?" ("Yes"). The ground truth region boundary is in red. Pie charts centered in each region denote its type while path color denotes different regions.

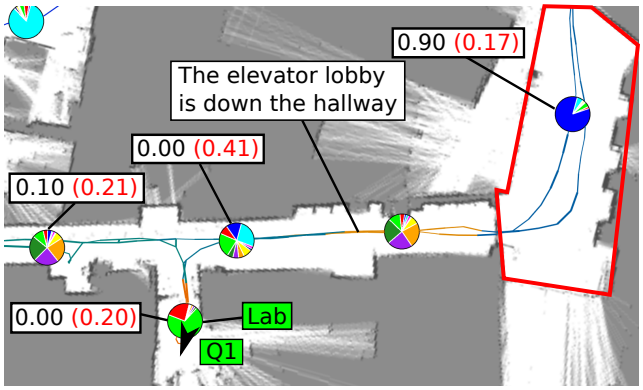

Figure 3: Language groundings for the expression "The elevator lobby is down the hall". Grounding likelihood with questions is shown in black and without questions in red. Question asked (and answer), Q1: "Is the elevator lobby near me?" ("No"). The ground truth region is outlined in red.

Conclusion

We outlined a framework that enables robots to engage a human in dialog in order to improve its learned semantic map during a guided tour. We provided an initial demonstration of its ability to successfully reduce uncertainty over the groundings for natural language descriptions.

Table 1: Entropy over figure groundings with and without questions

Language Event	Entropy Without Questions	With Questions	No. of Questions
"The lounge is down the hallway" (Fig. 2)	2.015	0.620	3
"The elevator lobby is down the hallway" (Fig. 3)	1.320	0.325	1
"The lounge is behind you" (Fig. 4)	0.705	0.056	1

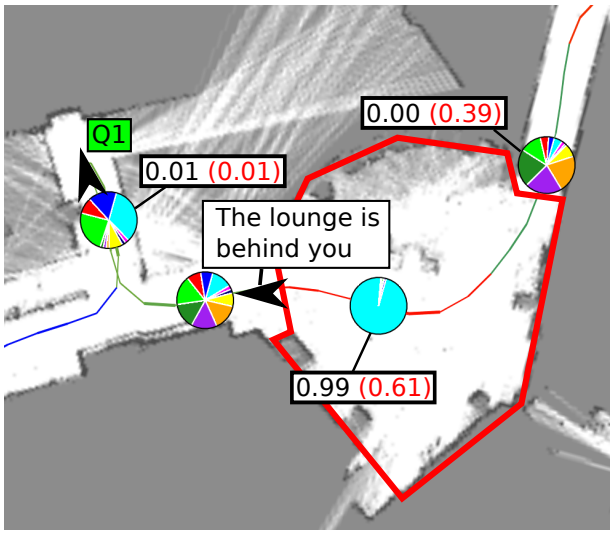

Figure 4: Language groundings for the expression "The lounge is behind you". Grounding likelihood with questions is shown in black and without questions in red. Question asked (and answer), Q1: "Is the lounge near me?" ("Yes"). The ground truth region is outlined in red.

Going forward, we plan to conduct extensive experiments both on collected datasets as well as through live trials to assess both the effectiveness of the approach over diverse settings as well as the human factors aspect of this interactive tour model. We also plan to extend the current approach to ask additional types of questions that can provide more information than simple yes/no type questions.

A drawback of the current approach is that the system does not account for the likelihood that the figure to which a question refers may correspond to a yet unvisited, and thus unknown, part of the environment. A more comprehensive approach would be to model the likelihood that figure references ground to unvisited regions in the environment, and evaluate the affect of the questions on these regions as well.

The current framework only considers questions that reduce the entropy over language groundings. However, as the robot integrates semantic information from other sources, such as room appearance classifiers and object detectors, it would be beneficial to ask questions about spatial regions even in the absence of language. For example, upon observing a computer monitor, it could ask whether it is in an office.

There are a number of extensions that could be carried out to enhance this framework such that it expands the scope of actions available to the robot to improve its model of the world. Currently, the framework only allows the robot to take exploration actions by asking questions. We could expand the scope of actions available to the robot by including navigation, such that the robot can actively explore the environment (possibly after the tour) to reduce its uncertainty over the space of entities described during the tour.

References

Deits, R.; Tellex, S.; Thaker, P.; Simeonov, D.; Kollar, T.; and Roy, N. 2013. Clarifying commands with information-theoretic human-robot dialog. *J. Human-Robot Interaction* 2(2):58–79.

Hemachandra, S.; Kollar, T.; Roy, N.; and Teller, S. 2011. Following and interpreting narrated guided tours. In *Proc. IEEE Int'l Conf. on Robotics and Automation (ICRA)*, 2574–2579.

Hemachandra, S.; Walter, M. R.; Tellex, S.; and Teller, S. 2014. Learning spatial-semantic representations from natural language descriptions and scene classifications. In *Proc. IEEE Int'l Conf. on Robotics and Automation (ICRA)*.

Kaess, M.; Ranganathan, A.; and Dellaert, F. 2008. iSAM: Incremental smoothing and mapping. *Trans. on Robotics* 24(6):1365–1378.

Kruijff, G.-J. M.; Zender, H.; Jensfelt, P.; and Christensen, H. I. 2006. Clarification dialogues in human-augmented mapping. In *Proc. ACM/IEEE Int'l. Conf. on Human-Robot Interaction (HRI)*.

Kuipers, B. 2000. The spatial semantic hierarchy. *Artificial Intelligence* 119(1):191–233.

Littman, M. L.; Cassandra, A. R.; and Kaelbling, L. P. 1995. Learning policies for partially observable environments: Scaling up. In *Proc. Int'l Conf. on Machine Learning (ICML)*.

Pronobis, A., and Jensfelt, P. 2012. Large-scale semantic mapping and reasoning with heterogeneous modalities. In *Proc. IEEE Int'l Conf. on Robotics and Automation (ICRA)*, 3515–3522.

Stachniss, C.; Grisetti, G.; and Burgard, W. 2005. Information gain-based exploration using rao-blackwellized particle filters. In *Proc. Robotics: Science and Systems (RSS)*.

Tellex, S.; Kollar, T.; Dickerson, S.; Walter, M. R.; Banerjee, A. G.; Teller, S.; and Roy, N. 2011. Understanding natural language commands for robotic navigation and mobile manipulation. In *Proc. Nat'l Conf. on Artificial Intelligence (AAAI)*, 1507–1514.

Walter, M. R.; Hemachandra, S.; Homberg, B.; Tellex, S.; and Teller, S. 2013. Learning semantic maps from natural language descriptions. In *Proc. Robotics: Science and Systems (RSS)*.

Williams, T.; Cantrell, R.; Briggs, G.; Schermerhorn, P.; and Scheutz, M. 2013. Grounding natural language references to unvisited and hypothetical locations. In *Proc. Nat'l Conf. on Artificial Intelligence (AAAI)*.

Zender, H.; Martínez Mozos, O.; Jensfelt, P.; Kruijff, G.; and Burgard, W. 2008. Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems* 56(6):493–502.