Clarification dialogues in human-augmented mapping

Geert-Jan M. Kruijff & Hendrik Zender
Language Technology Lab
DFKI GmbH
Saarbrücken, Germany

{gj,hendrik.zender}@dfki.de

Patric Jensfelt & Henrik I. Christensen
Center for Autonomous Systems
Royal Institute of Technology
Stockholm, Sweden

{patric,hic}@nada.kth.se

ABSTRACT

An approach to dialogue based interaction for resolution of ambiguities encountered as part of Human-Augmented Mapping (HAM) is presented. The paper focuses on issues related to spatial organisation and localisation. The dialogue pattern naturally arises as robots are introduced to novel environments. The paper discusses an approach based on the notion of Questions under Discussion (QUD). The presented approach has been implemented on a mobile platform that has dialogue capabilities and methods for metric SLAM. Experimental results from a pilot study clearly demonstrate that the system can resolve problematic situations.

Keywords

Human-augmented mapping; natural language dialogue; mixed initiative; clarification

1. INTRODUCTION

In human-augmented mapping (HAM), a human and a robot interact to establish a correspondence between how the human perceives the spatial organization of the environment, and what the robot autonomously learns as a map [22]. This helps to bridge the difference in perspective: robots usually construct metric maps, whereas humans adopt a more topological perceive of spatial organization. Being able to overcome this difference is crucial for interactive, mobile service robots.

Existing dialogue-based approaches to human-assisted mapping usually implement a *master/slave* model of dialogue: the human speaks, the robot listens. Such dialogues are sufficient when the only goal is for the human to tell the robot the names of different locations. However, situations naturally arise in which interaction should be more flexible, allowing also the robot to take the initiative in the dialogue.

We present an approach that enables the robot to initiate a subdialogue to clarify an issue. This is one important form

of *mixed-initiative* interaction, to enable a robot to recognize when help is needed from a human user, and learn from this interaction [5, 6, 17]. In human-assisted mapping, several situations may arise that require clarification, for example:

Uncertainty in automatic classification: Doorways provide important knowledge about spatial organization, but are difficult to recognize robustly and reliably.

Inconsistency between classification and description: The semantic classification of a location which a robot can automatically establish [19] may be inconsistent with the description provided by the human user.

Diagnosed faults in perception: One source of faults lies in the localisation system, which may not always generate classifications for new locations, or the robot may enter into areas that only appears to be novel. Another source of faults are odometric errors, which may cause slippage in mapping. Finally, the robot may be unable to perceive certain obstacles with its given suite of sensors, e.g. doorsteps.

Clarification dialogues can help to improve the quality of the representation that the robot constructs of the spatial organization of the environment, and to increase the robot's robustness to dealing with uncertain or incomplete information. The contribution of this paper is that we present an approach that enables a robot to carry out clarification dialogues, triggered by the types of situations described above.

For our method for clarification we start from an approach that is concerned with questions and clarifying communication in multi-modal dialogue systems for information-seeking tasks [15]. Applying it in human-robot interaction yields the novel situation that modalities other than communication actually can trigger a need for clarification. Although the need for robots to be able to take initiative when requiring help has been acknowledged [6, 17], existing systems that allow for mixed-initiative either focus on issues in control ("adjustable autonomy") using mostly graphical interfaces [5], or do not have the rich perceptual input our system deals with [20]. Our approach is similar to [20] in that we also adopt a planning-based model of dialogue, which is more flexible than the finite-state methods employed in e.g. [12, 11].

An overview of the paper is as follows. In §2 we present

our approach, discussing how clarification dialogues are triggered, and how they are processed at the levels of communication and mapping. §3 presents the implementation iin an integrated architecture, including the communication subsystem, the mapping & localization subsystem, and the mediation between these subsystems. In §4 we present the results of a pilot study, in which we investigated the viability and effectiveness of cross-modal integration we achieved between communication, and mapping & localization. The results clearly demonstrate the potential for the approach to increase the quality of the maps the robot creates, but also raise several -still outstanding- issues. The paper closes with a discussion of possible extensions of the approach in §5, and conclusions.

2. GROUNDING AND CLARIFICATION

To solve the issues that arise in the situations sketched above, we propose to conceive of these issues as *grounding problems* which *require clarification*. Contemporary dialogue systems use clarification mechanisms primarily to address *communicative grounding problems*. A grounding problem arises when the hearer has a problem figuring out what the speaker's utterance meant. This may be due to e.g. a *lack of perception or understanding* regarding what the utterance seems to refers to, *ambiguity* in the understood meaning of the utterance, or *conflicts* between the hearer's beliefs and the understood meaning. The situations we deal with in this paper can also be seen as grounding problems, which require clarification – except that in our case grounding problems arise outside (and sometimes irrespective) of the current dialogue. They represent a problem with grounding the robot's own unmediated perceptions, relative to an intended representation that relates to the perspective of a human user.

We adopt an approach to dealing with questions and their function in grounding inspired by Ginzburg, and Larsson, who define for information states-based dialogue management a datastructure QUD (Questions Under Discussion) to keep track of open questions or issues that still need to be handled, cf. [15]. Similarly, we define a structure XMQUD to store issues that have been raised by a modality other than communication, and which need to be addressed cross-modally. XMQUD is an ordered set, with ordering based on recency, in which each question has a unique identifier provided by the issuing modality.

If processing in a modality (e.g. mapping) runs into a grounding problem, it can submit a *question* to the XMQUD. Abstractly, we can think of a question as having the following form: $?x.\textbf{content}(u, x)$, meaning there is some aspect x of the content u which is under discussion. We call x the *scope* of the question, and u the *restrictor*: x is what the question is about, i.e. we are looking for an aspect x that is circumscribed by the content of u.

For the purposes of this paper, we are interested in two kinds of scopes, namely over *things* (e.g. locations "where is the desk?" or objects "what is near me?") and over the *truth* of a proposition (e.g. "is there a door here?"). This is represented as in Example 1: (a) specifies for a question q_1 a ?-scope over x being of sort *location*, and (b) gives a ?-scope over the statement of a *state* s being **true** for a question q_2.

(1) a. $q_1 = ?x : location$
     b. $q_2 = ?\textbf{true}(s : state)$

The restrictor we specify as a relational structure, consisting of a conjunction of *elementary predications* [13]. The most basic elementary predication is an identifier n_j, which may be sorted as illustrated in Example 1, with a proposition **p** that holds for that identifier:

$$@_{n_j}(\textbf{prop}), \text{ "at } n_j, \textbf{prop} \text{ holds".}$$

We specify a feature f with value **v** for n_j as

$$@_{n_j}(\langle\text{f}\rangle\textbf{v}).$$

Finally, a relation $\langle\text{R}\rangle$ holds between two identifiers n_i and n_j. For example,

$@d1 : doorway(\textbf{door}$
$\quad \wedge \langle\text{Location}\rangle(n1 : region \wedge \textbf{near}$
$\qquad \wedge \langle\text{Proximity}\rangle\textbf{proximal}$
$\qquad \wedge \langle\text{Dir:Anchor}\rangle(i1 : person \wedge \textbf{speaker})))$

represents a doorway $d1$ being a door, which is in a location $n1$ in that is proximal to an anchor $i1$, being the speaker – i.e "door near me/here".

Example 2 illustrates then the two types of questions we are interested in. The question in (a) scopes over things near the speaker ("what is near me?"), whereas (b) is after the truth of the statement that there is a door near the speaker ("is there a door near me/here?").

(2) a. $q_1 = ?x : thing.$
         $@x : doorway(\langle\text{Location}\rangle(n1 : region \wedge \textbf{near}$
         $\quad \wedge \langle\text{Proximity}\rangle\textbf{proximal}$
         $\quad \wedge \langle\text{Dir:Anchor}\rangle(i1 : person \wedge \textbf{speaker})))$
     b. $q_2 = ?\textbf{true}(d1).$
         $@d1 : doorway(\textbf{door}$
         $\quad \wedge \langle\text{Location}\rangle(n1 : region \wedge \textbf{near}$
         $\quad \wedge \langle\text{Proximity}\rangle\textbf{proximal}$
         $\quad \wedge \langle\text{Dir:Anchor}\rangle(i1 : person \wedge \textbf{speaker})))$

When a modality submits a question to the XMQUD, the question is stored with its identifier. For the purposes of this paper, there is no planning involved in how to deal with a question: we always request the communication subsystem to try and resolve the question, through a dialogue with the human. (See also §5 for how this can be extended.) The communication subsystem plans a communicative goal and the content to express the question, and returns the XMQUD an identifier for the utterance that is being planned. The XMQUD establishes a connection between the question to this identifier (fusion).
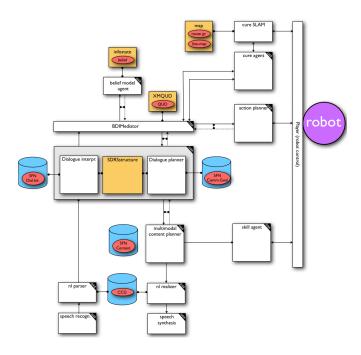
Figure 1: Architecture (partial) with communication, mapping & localization, mediation subsystems

The robot subsequently plans a communicative goal (i.e. ask the human), and the content to realize this goal. The robot then generates a string expressing this content, utters it, and at the same time adds the planned content to a model of the dialogue context, to record the fact that the robot has asked a question. After the robot has posed the question to the human, we can assume that the human answers the question. Once the robot has parsed the answer, it tries to interpret the answer in the current dialogue context: That is, we try to establish a rhetorical relation between the answer, and a question that has been raised in the dialogue.

The result of this analysis is a relational structure, connecting the content of the answer to that of the question the robot beliefs it is an answer to. Now this analysis is passed back to process that maintains the XMQUD. The identifier of the question the answer was bound to is resolved against any outstanding questions on the XMQUD. Provided a matching question on the XMQUD is found, the modality which raised the question is informed of the answer, so that it can (intra-modally) resolve the outstanding issue.

3. IMPLEMENTATION
We have implemented the approach of §2 in a distributed architecture for integrating different perceptual and deliberative skills that deal with a variety of modalities. The architecture is inspired by multi-level distributed cognitive architectures like [21].

Figure 1 illustrates the relevant aspects of the architecture: the communication subsystem, the mapping subsystem, and the BDI-based mediation between the different subsystems. By saying that beliefs *mediate* we mean that they provide a common ground between different modalities, rather than
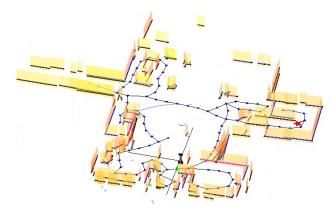

Figure 2: Simultaneous Localization And Mapping

being a layer on top of the different modalities. Beliefs provide a means to perform forms of global information fusion [23], in its minimal form by co-indexing references to information in individual modalities. Furthermore, it is at this mediating level that we deal with requests to clarify issues that have arisen in a particular modality, through communication or through another modality.

3.1 Communication
The communication subsystem consists of several components for the analysis and production of natural language. It has been implemented as a distributed architecture using the Open Agent Architecture [7], following the idea of agent-based models for multi-modal dialogue systems [1].

On the analysis side, we use the Sphinx4 speech recognition engine[1] with a domain-specific JSAPI speech grammar. The string-based output of Sphinx4 is then parsed with OpenCCG[2]. OpenCCG employs a combinatory categorial grammar [3] to yield a representation of the linguistic meaning that the string (i.e. the utterance) represents. We represent linguistic meaning as an ontologically rich, relational structure (see also §2) in a description logic-like formalism. Finally, in structural dialogue analysis we relate the linguistic meaning of an utterance to the current dialogue context, in terms of how it rhetorically and referentially relates to preceding utterances. This yields an updated model of the (situated) dialogue context [2, 4].

On the production side, we use dialogue planning to enable flexible, contextually appropriate interaction. Based on a need to communicate, established either by the current dialogue flow or by another modality, the dialogue planner establishes a communicative goal. In turn, we plan the content to express this communicative goal, possibly in a multimodal way using non-verbal means (pose, head moves) next to verbal communication. In these planning steps, we can inquire the models of the situated context (e.g. dialogue context, visually situated context) to ensure that the content we plan is contextually appropriate [14]. We realize verbal content using the OpenCCG realizer, which generates a string for the utterance, and then synthesize this string using a

---

[1]http://cmusphinx.sourceforge.net/sphinx4/
[2]http://openccg.sf.net

text-to-speech engine[3].

For example, let us consider the case that the mapping subsystem informs the BDI Mediator that it system is uncertain about the presence of a door in a particular location. The mapping system sends a *content specification* (e.g. $@d1 : doorway(\mathbf{door})$) and the status of this content in the model that the mapping system maintains of the environment. In our architecture we classify this status as *known, ambiguous* to indicate the uncertainty of an already made interpretation.

The BDI mediator in turn informs the dialogue planner of this content, its status, and the modality it originates in. The dialogue planner generates a communicative goal for resolving the issue, by determining the question type that needs to be communicated: We have built an ontology of abstract assertion, command, and question types that mediates between linguistic content and content in other high-level cognitive processes. The utterance planner then constructs the logical form which expresses this type of question, using the content and information about the modality.

An important aspect of the resulting subsystem is that it interacts in a distributed fashion with other modalities (cf. also [1, 20]), instead of playing a centralized controling role in the overall architecture [4]. Backed by the flexibility that dialogue planning offers over e.g. scripted responses in a finite-state approach, we can exploit this distributed interaction to enable other modalities trigger a need for communication – e.g. for clarification.

3.2 Mapping and Localization

The subsystem for SLAM uses a feature based representation where the main features are lines, typically corresponding to walls in the environment. The underlying feature representation is flexible and other types of features can easily be incorporated [9]. The basis for integrating the feature observations is the extended Kalman filter (EKF).

A feature based map is rather sparse and only captures structures that fit the predefined feature description (e.g. lines). One cannot distinguish free space from areas where the structures do not fit the feature model. There is thus no explicit information about where there is free space such as in an occupancy grid. Here we use a technique as in [18] and build a navigation graph (also called routegraph) while the robot moves around. When the robot has moved a certain distance, a node is placed in the graph at the current position of the robot. Whenever the robot moves between two nodes, these are connected in the graph. The nodes represent the free space and the edges between them encode paths that the robot can use to move from one place to another.

On top of the navigation graph there is a graph where each area corresponds to one node and the edges tell which areas are connected. The graph is automatically generated from the navigation graph by labeling the nodes into different areas and thus partitioning it. Our strategy rests on the simple observation that the robot passes a door to move between rooms. Therefore, whenever the robot passes a door a

---

[3]http://mary.dfki.de

node marked as a door is added to the navigation graph and consequetive nodes are given a new area label. Currently, door detection is simply based on detecting when the robot passes through a narrow opening. The fact that the robot has to pass through an opening removes a lot of false doors that would result from simply looking for narrow openings. However, this alone will still lead to some false doors in cluttered rooms. Assuming that there are few false negatives in the detection of doors we get great improvements by enforcing what was stated above, i.e. it is not possible to change room without passing a door. For example, while moving around in a room the robot may detect a narrow passage and falsely assume that a door was passed. It will initially put a door label on that particular node. The robot continues to move around in the room and eventually reaches the nodes from before adding the false door. These nodes will then have different room labels, that is, the room has changed without passing a door. If this happens, an inconsisteny is found and a dialogue with the user is triggered to clarify the situation.

Note that our method for constructing the topological graph is independent of the representation used for the rest of the map. The estimate of the robot position can come from any source. Furthermore, the sensors needed to detect when the robot is standing in a narrow opening can be very simple. In [8] a feature based map is also used for localization. However, for generating the topological map an occupancy grid has to be constructed. In [10] the topological graph is also extracted from an occupancy grid. Constructing the occupancy grid is computationally intensive. An alternative to the grid based method is presented in [19] where Boosting is used to train classifiers to recognize different types of environments, such as doorways, corridor and rooms from laser data. The boosting method would make a suitable compliment to the method used in this paper. The integration of the two methods is therefore underway.
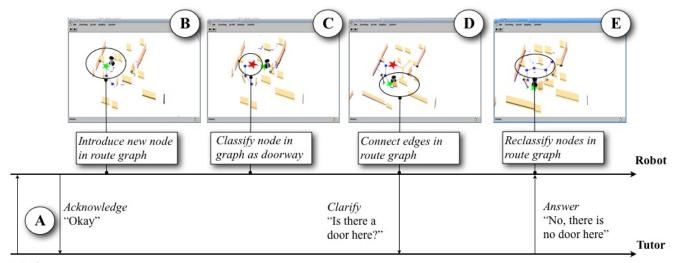
4. PILOT STUDY

Using the implemented system, we have performed a pilot study to investigate the viability and effectiveness of cross-modal integration between communication, and mapping & localization. The results, discussed below, clearly demonstrate the potential for the approach to increase the quality of the maps the robot creates. We discuss results on the basis of sample interactions, and we outline still outstanding issues.

4.1 Investigations

The pilot study consisted in repeated runs through two types of indoor environment: *connected cluttered spaces*, in the form of an office and a laboratory connected through a door-inbetween and a corridor; and a *large open space*, in the form of a lobby. In the experiments, we focused on using clarification to ensure the proper classification of nodes in the route-graph as *door* or *room*. We used an ActivMedia PeopleBot as our mobile platform, equipped with a SICK laser range finder.

In the first experimentation setting, connected cluttered spaces, the robot travelled approximately 20 meters through an office, a laboratory, and a corridor. The route graph for the map has on the average 22 nodes, 3 of which correspond to

Figure 3: Sample interaction timeline to clarify mapping issue

doors; see Figure 4. Both the office and the laboratory were cluttered with several larger objects, e.g. a P3-AT robot as illustrated in Figure 5.
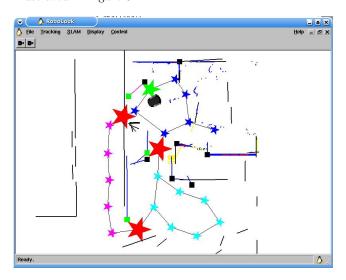

Figure 4: Sample route graph for cluttered spaces

The timeline in Figure 3 shows a prototypical interaction sequence between the tutor and the robot. When entering a new room, the tutor describes the room to the robot, e.g. "This is the office." (A). Adopting a collaborative model of dialogue, the robot duly acknowledges that it has understood.

While travelling through the office, the robot introduces new nodes into the route graph. The green star indicates the most recently added node. (B) corresponds roughly to the scene in Figure 5, with the position of the green star being near the white square to the right behind the P3-AT.

When the PeopleBot passes through the narrow opening be-

Figure 5: Real situation for Figure 3

tween the P3-AT and the bags, it wrongly classifies this as a doorway and consequently adds a door node (the red star). (C) Because passing through a door means entering a new room, the robot now assumes it is no longer in the office. It accordingly labels new nodes in the routegraph with a new area identifier, indicated by a different color.

Once the robot has arrived at its position indicated in (D), the robot is able to reconnect the subgraph of the "new room" with a node it previously visited – which was in a different room. This, however, means that the robot would have re-entered the 'previous' room without having passed through a door. It is at this point that the mapping & localization subsystem indicates that there might be an issue with a node it classified as a door. The robot asks a clarification question, which -in this case- is answered in the negative. Based on this information, the robot corrects its map. (E)

The internal processing that happens at stages (D) and (E) is given in more detail in Figure 6. Once the mapping subsystem realizes it is uncertain about the presence of a door, it informs the BDI Mediator that it is unsure about the
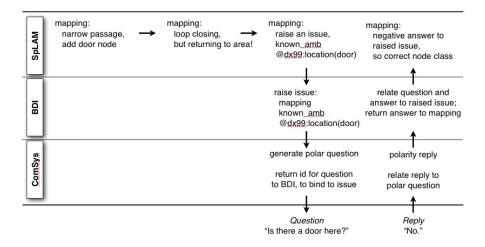
Figure 6: Internal processing for Figure 3(D)-(E)

status of location $dx99$ being indeed a door (*known_amb* indicates *known*, *ambiguous* in the multi-valued 'truth' system we employ). The communication subsystem translates this complex request into a question of the abstract type *informative.polar.endurant.perspective.spatial*, asking after the truth of locating an object (*door*) in the given spatial perspective (*here*). The utterance planner in turn plans a logical structure to realize this communicative goal, by constructing the logical form for a polar question of the form "is there ⟨object⟩ ⟨spatial perspective⟩?" Once we have construed the logical form has been construed, its identifier is returned to the BDI Mediator to indicate that this is the question posed to resolve the issue, we store it in the dialogue model, and we realize the logical form as a string using the OpenCCG realizer.

Then, when the human tutor provides an answer, we try to resolve that answer against outstanding questions in the dialogue model (i.e. performing a basic form of rhetorical resolution). As answer to a polar questions we need an assertion that expresses a valuation, possibly with a correction. Our example just gives the simplest case, namely "No". In the dialogue model, we then state that content of this assertion provides an answer to the previously asked polar question. This information (i.e. answer, plus antecedent question) is then passed on to the BDI Mediator for further processing. Since in the BDI Mediator we kept track of the identifier of the question that was communicated to resolve the raised issue, we can now return the assertion provided as answer to the modality that raised the QUD.

We repeated the experiment 5 times. In total, 105 route-graph nodes were generated. 22 nodes were classified as doorways, 14 of them being correct door nodes, and 8 being false positives. We only had one false negative, i.e. a node wrongly classified as a room whereas it should have been a door. Clarification detected, and helped resolve, 7 out of the 8 false positives. Hence, after clarification, 15 nodes remained classified as doors, 14 of them being correct. In percentages, 77.78% of the misclassified nodes over all runs were reclassified, resulting in an increase from 92.38% to 98.10% nodes correctly classified.

We performed a similar experiment in the lobby of our building. The lobby is a large space with several relatively static landmarks, e.g. an information board, and reception desks. Like the loose objects in the previous experiment, these landmarks often yielded false positives which could be accurately recognized, and corrected through clarification questions.

4.2 Issues

Human-augmented mapping can clearly improve the quality of a map that the robot automatically acquires – provided though the robot itself can recognize false positives, and that there is a human tutor to answer the questions. As we already described for the sample interaction in Figure 3, in our system the robot only properly recognizes false positives if it is able to "close the loop" in a graph. It needs to explicitly realize it has returned to a -presumably different- area it was in before, without having passed through a doorway.

If there is no loop, the robot does not recognize it has returned to the same room without having passed through a door. Instead, it will consider the falsely recognized doors as positive evidence for passing through doorways to get to different rooms. Figure 7 ilustrates such a situation, where we have an extremely narrow, cluttered room with a turning space of less than one meter.
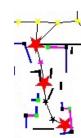

Figure 7: Non-looping subgraph with misclassified doornodes

Moreover, even if the route graph does loop, it is not guaranteed that false positives can be recognized. Figure 8 shows the situation of a conference room where the small space between rows of chairs and cupboards was often recognized as a doorway. Although the robot did close the loop, we had
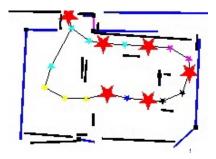
Figure 8: Looping subgraph with misclassified doornodes

the same self-delusion as in the previous example: there were too many false positives, making the robot wrongly believe it indeed had passed through doors.

It is interesting though to consider, what the interaction between the robot and the human would look like should the robot be able to recognize the false positives in Figure 8: Every other meter, the robot would ask the human "Is there a door here?" We would like to avoid such tedious interactions, e.g. asking seemingly endless sequences of questions about wrongly recognized doors in rapid succession. One way to deal with this is to provide the robot with more linguistic capability ("There are no doors here!" – i.e. do not ask until you are explicitly told you are in a new room). Another way is to mix in other modalities, e.g. the robot can try to verify visually whether there is a door, or whether it is still in the same room, before asking a clarification question. (At the same time, the quality of visual recognition would obviously affect false positives and false negatives due to differences in reliability of vision across situations.)

A final issue raised in the experiments was the possible inappropriateness or ambiguity of "here" in clarification questions like "Is there a door here?". The robot will only raise an issue once it has closed a loop. If that loop covered a large area, then "here" is clearly inappropriate, because the robot is not likely to be in the vincinity of the questionable node. But even if the robot is, there may be a problem for the speaker to resolve "here". For example, in the situation illustrated in Figure 5 and Figure 3, the robot closed the loop at a point where it was close to the door leading out of the office. In this case, the tutor could interpret the question "Is there a door here?" to refer to the office door, and thus wrongly answer "Yes, there is a door here." In other words, the robot needs to use a deictic reference that can properly be resolved by the human.

5. EXTENSIONS

We are currently investigating several means for improving the automatic classification and verification of locations, so that the robot needs to rely less on a human tutor to be present. As we noted earlier, Boosting may be used to train classifiers to recognize different types of environments from laser data (based on the polygonal shape of the frontier), such as doorways, corridors and rooms. We would like to use this approach to complement the methods we are currently using to classify areas and doorways – complement, because although it may improve the recognition of door-

ways, it may also introduce novel errors such as semantic misclassifications of locations.

Another approach we are investigating is the use of visual feedback. Using a SIFT-based approach to visual recognition [16], the robot can dynamically construct a short-term memory of scalable visual models that characterize a local situation. Each time the tutor tells the robot it enters a new room, the robot looks around and constructs three models representing visual perspectives on the room. Then, as long as the robot travels on a similar heading as it enters the room, it can try to use the models it constructed to recognize whether it is still in the room. Provided visual recognition yields a high enough confidence score, this is an easy and fast way to verify whether it has just passed a door or not. When a significant change in heading occurs, the robot needs to construct additional models. Preliminary experiments show that visual models of perspectives tend to be useful for about four meters of (straight) travel, and that learning new models takes about two seconds including moving the pan/tilt unit.

All these extensions help us continue along the main ideas we discussed in this paper. Namely, how do we go beyond integrating communication and mapping & localization, generalizing to n modalities, so that we can equip the robot with efficient cross-modal clarification strategies?

6. CONCLUSIONS

In this paper we presented an approach to clarification, which enables a robot to initiate a dialogue with a human to clarify an issue that has arisen in one of its modalities. We focused on issues that can arise in mapping & localization in novel environments, and discussed a pilot study which showed that clarification can clearly contribute to increasing the quality of the maps that a robot constructs. The pilot study also identified several issues for future research, such as the need to complement communication with other modalities that can assist in clarifying an issue, and to have flexible and intelligent dialogue strategies to avoid repetitive interactions. At the end of the paper, we pointed out how recently developed techniques in automatic location classification (Boosting) and visual recognition (SIFT) could be used to extend our approach to clarification with additional modalities for verification.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] James Allen, Donna Byron, Myroslava Dzikovska, George Ferguson, Lucian Galescu, and Amanda Stent. An architecture for a generic dialogue shell. *Journal of Natural Language Engineering*, 6(3):1–16, 2000.

[2] Nicholas Asher and Alex Lascarides. *Logics of Conversation*. Cambridge University Press, 2003.

[3] Jason Baldridge and Geert-Jan M. Kruijff. Multi-modal combinatory categorial grammmar. In *Proceedings of EACL'03*, Budapest, Hungary, 2003.

[4] J. Bos, E. Klein, and T. Oka. Meaningful conversation with a mobile robot. In *Proceedings of the Research Note Sessions of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL'03)*, Budapest, Hungary, 2003.

[5] D. J. Bruemmer, J. L. Marble, D. D. Dudenhoeffer, M. O. Anderson, and M. D. McKay. Mixed-initiative control for remote characterization of hazardous environments. In *Proc. of Hawaiian International Conference on Computer Science 2003*, Waikoloa Village, Hawaii, 2003.

[6] D.J. Bruemmer and M. Walton. Collaborative tools for mixed teams of humans and robots. In *Proc. of the Workshop on Multi-Robot Systems*, Washington, D.C., 2003.

[7] Adam Cheyer and David Martin. The open agent architecture. *Journal of Autonomous Agents and Multi-Agent Systems*, 4(1):143–148, March 2001.

[8] A. Diosi, G. Taylor, and L. Kleeman. Interactive SLAM using laser and advanced sonar. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA'05)*, Barcelona,Spain, April 2005.

[9] John Folkesson, Patric Jensfelt, and Henrik Christensen. Vision SLAM in the measurement subspace. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA'05)*, 2005.

[10] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J.A. Fernández-Madrigal, and J. Gonález. Multi-hierarchical semantic maps for mobile robotics. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'05)*, pages 3492–3497, 2005.

[11] A. Haasch, S. Hohenner, S. Huewel, M. Kleinehagenbrock, S. Lang, I. Toptsis, G. A. Fink, J. Fritsch, B. Wrede, and G. Sagerer. Biron - the Bielefeld robot companion. In E. Prassler, G. Lawitzky, P. Fiorini, and M. Haegele, editors, *Proc. Int. Workshop on Advances in Service Robotics*, pages 27–32, Stuttgart, Germany, 2004. Fraunhofer IRB Verlag.

[12] H. Ishiguro, T. Ono, M. Imai, T. Maeda, T. Kanda, and R. Nakatsu. Robovie: an interactive humanoid robot. *Int. J. Industrial Robotics*, 28(6):498–503, 2001.

[13] Geert-Jan M. Kruijff. *A Categorial-Modal Logical Architecture of Informativity: Dependency Grammar Logic & Information Structure.* PhD thesis, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic, 2001.

[14] Geert-Jan M. Kruijff. Contextually appropriate utterance planning for CCG. In *Proc. of the 9th European Workshop on Natural Language Generation*, Abderdeen, Scotland, 2005.

[15] Staffan Larsson. *Issue-Based Dialogue Management.* Phd thesis, Department of Linguistics, Göteborg University, Göteborg, Sweden, 2002.

[16] D. G. Lowe. Distinctive image features from scale-invariant keypoints. In *Int. Jnl. Computer Vision*, pages 91–110, 2004.

[17] J.L. Marble, D.J. Bruemmer, and D.D. Dudenhoeffer D.A. Few. Evaluation of supervisory vs. peer-peer interaction with human-robot teams. In *Proc. of the Hawaiin International Conference on Computer Science*, 2004.

[18] P. Newman, J. Leonard, J.D. Tardós, and J. Neira. Explore and return: Experimental validation of real-time concurrent mapping and localization. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA'02)*, pages 1802–1809, Washington D.C., USA, 2002.

[19] Axel Rottmann, Oscar Martinez Mozos, Cyrill Stachniss, and Wolfram Burgard. Semantic place classification of indoor environments with mobile robots using boosting. In *Proc. of the National Conference on Artificial Intelligence*, Pittsburgh, PA, July 2005. AAAI.

[20] C.L. Sidner, C.D. Kidd, C.H. Lee, and N.B. Lesh. Where to look: A study of human-robot engagement. In *Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI)*, pages 78–84, 2004.

[21] Aaron Sloman. Beyond shallow models of emotion. *Cognitive Processing*, 2(1):177–198, 2001.

[22] E.A. Topp and H.I. Christensen. Tracking for following and passing persons. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'05)*, pages 70–76, 2005.

[23] W. Wahlster. Smartkom: Fusion and fission of speech, gestures, and facial expressions. In *Proceedings of the 1st International Workshop on Man-Machine Symbiotic Systems*, pages 213–225, Kyoto, Japan, 2002.