Understanding Spatial Concepts by Exploiting Interaction Patterns*

Elin A. Topp

Department of Computer Science
Lund University - LTH
22100 Lund, Sweden
Elin_Anna.Topp@cs.lth.se

1 Introduction

During the last years approaches to "interactive mapping" or "Human Augmented Mapping" with and for service robots have become of increasing interest both for robotic mapping methods and for the investigation of Human Robot Interaction. The underlying idea is to take a robot on a guided tour and explain the surroundings to it. This guided tour rises questions both regarding the technological development and the cognition and interaction part of the tour. Several approaches have been reported to "interactive" or "semantic" mapping, including also works related to the representation of space that could be suitable for an interactively controlled robotic mapping process [1–3]. There is an abundance of publications to be found regarding the actual representation(s) of space used by humans in different contexts, hence the most relevant ones are those considering the description of space that humans give in an interaction and that they can remember [4, 5]. Also particular aspects of the interaction with service robots or robot companions have been investigated in user studies [6] (as example), partly conducted in "home-like" laboratory environments and "laboratory apartments". However, with the respective systems still being subject to research and development, it has become of interest to investigate *how people actually present their own home or other familiar environments to a mobile robot* to draw conclusions on the system requirements for further development. This paper describes observations that have been part of the analysis of a respective study in people's homes, aiming to find a way to improve the general understanding of a given situation for a service robot.

A previously introduced respective framework [7] proposes a cognitively inspired *generic environment model* based on hierarchically structured (spatial) concepts. This model links human and robotic environment representations, so that communication about the environment is facilitated and can then also be used to improve the robot's understanding of its surroundings. So far the model incorporates *regions* (delimited areas, typically corresponding to rooms), *locations* (workspaces, defined by large, rather stationary objects, e.g., a fridge or a coffee-maker) and *objects* (small items that can be manipulated and are not stationary, e.g., cups or books).

During a guided tour different types of ambiguities can arise, one of which occurs when the corresponding concept for a given label is not known, or, even

* The user study that served as the means for data collection leading to this particular investigation was conducted together with Helge Hüttenrauch at the School of Computer Science, Royal Institute of Technology (KTH) as "Multiple Room Study" within the European Project "COGNIRON" (concluded in 2008).

more interesting, does not match other observations. This type of ambiguity can be termed *conceptual mismatch*. A typical example of such an ambiguity is a user who presents a *region* (e.g., an office) by just pointing towards the respective door, while being – together with the robot – in the hallway *in front of the region in question*. In a pilot study [8] it could be confirmed that such situations occur in a guided tour with a robot and consequently a respective system has to be capable to detect and handle them. It can be argued that if it was possible to observe a correspondence between the spatial concept that is presented to the robot and the user's preceding or concurrent behaviour, it is also possible to generate a hypothesis for the underlying (spatial) concept category for the item from this observation.

This paper discusses results from a recent user study that investigated – among other lines of research – whether it is possible to find particular interaction and presentation patterns (or strategies) in the observable human-robot interaction in a guided tour scenario that correspond to the (spatial) concepts proposed for the Human Augmented Mapping framework.

2 User Study

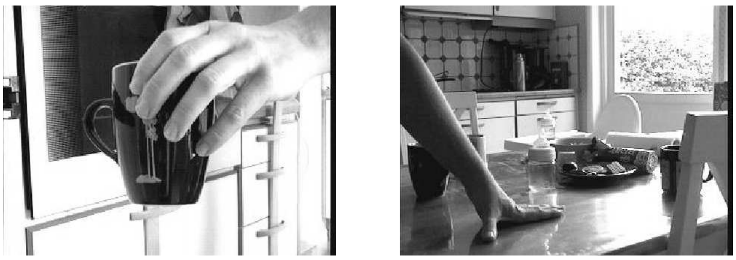

Fig. 1. Some observed presentation strategies

The investigation on potential interaction patterns was one of several lines of research for a study conducted in users' homes in the greater Stockholm area [9]. Here, only a brief summary of the study is given to explain the context for the investigation.

Eight persons were visited in their homes with a mobile robot (a Performance PeopleBot by Mobile Robots Inc.) that was equipped with a "study tool implementation" of the Human Augmented Mapping framework [7]. This particular implementation allows an experiment leader to monitor and - if necessary - to control the robot. Additionally the experiment leader can act as speech recognition wizard by interpreting the user's utterances and feeding those to the robot with the help of a graphical user interface. The study aimed in general at gathering data on the interaction between arbitrary users and the mobile service robot in a "home tour scenario" corresponding to the guided tour assumed for the Human Augmented Mapping framework. These data (both external video data, sensory data available from the robot platform, and the information given by the user in relation to the current robot's position) were stored for further investigation of the interaction. The subjects were asked to make the robot follow them and present a number of items in their home which they received a list of examples for together with the instructions on how to interact with the robot.

2.1 Environment Model and Working Hypotheses

Within the framework for Human Augmented Mapping a (partially) hierarchical generic model of space is assumed, that was implemented and used for investigations based on the three conceptual categories, *region*, *location* and *object*. To find potential correspondences between those concepts and people's presentation strategies it was necessary to guarantee a certain coverage of all the concepts in each of the trials of the study. Hence, the study subjects received quite canonical instructions, including a list of suggestions of what to present to the robot in their home, that included at least three items of each category and the additional instruction to show at least one *region* which was forbidden for the robot to enter (e.g., the bathroom). This was done to increase the probability of subjects presenting a "non-entered *region*", i.e., enforcing a situation of a *conceptual mismatch*. From observations made during previous studies the following assumptions were formulated:

- (Assumption 1, *objects*) When presenting an *object* to the robot, people tend to pick up, manipulate and move the item *to the robot* to prepare the presentation process, rather than that they navigate the robot into a particular pose. They presumably show the item by holding it directly in front of the robot's perceptive system (camera) or by placing it carefully and pointing to it with a very specific deictic gesture, a "fingertip pointing", often touching the item.
- (Assumption 2, *locations*) When presenting a *location* people tend to prepare the process by moving / near-navigating the robot into a particular ("optimal") pose and use deictic gestures, here "coarse pointing / waving with the whole hand", to direct the robot's attention.
- (Assumption 3, *regions*) When presenting a *region* while being *inside* it, people tend to omit gestures completely. A *region* that is not entered is presented by showing the door in the way a *location* would be presented.

2.2 Data Analysis, Observations, and Results

The analysis of the collected video footprint was performed as follows. The SHOW phase [10] for each item presented during the trial runs was segmented into a *preparation* and a *show event* (corresponding to a gesture and / or an introducing utterance like "This is ..."). The observations were then categorised in a number of preparations and gestures that were determined from preliminary inspection of available video material both from this and previous studies, corresponding to the three assumptions. Figure 1 illustrates two rather often observed types of gestures, a) holding an item in front of the robot's camera and b) touching, even leaning on a table to present the respective *location*.

In table 1 the observations made during the overall 128 presentation episodes (or 'SHOW phases") are summarised, showing already some of the trends that could be extracted with the help of more detailed inspection. Table 2 shows then the observations made for preparations by relating the categories of the presented items to the observed preparation. Similar tables were used to relate the categories of the presented item to specific gestures or types of gestures and

Observation category / Item category	Episode	Preparation	Gesture
all	128	122	100
Region	33	22	11
entered	24	6	6
non-entered	9	16	5
Location	61	61	62
Object	34	39	34

Table 1. Summarised numbers for observations regarding the presentation episodes with explicit preparations and gestures.

Item category / Preparation	All	Region	Location	Object	Non-entered region
all (excl. "none")	122	6	61	39	16
Move rob. to pos.	32	2	21	3	6
Turn robot	35	2	22	6	5
Position self	15	–	10	3	2
Fetch or take item	26	–	–	26	–
Arrange item	1	–	–	1	–
Remove obstacle	13	2	8	–	3
None	44	18	22	2	2

Table 2. Categories of presented items in relation to observed preparations

to relate the observed preparations and gestures to the category of the presented item [7].

From such quantifiable observations it is possible to find certain similarities in the subject's presentation strategies, hence there seem to be "interaction patterns" observable across subjects. However, the observations were made only with *one* particular robot, thus a full generalisation is not possible.

Nevertheless, the data supported assumption 1 in the sense that for the observed cases the following can be stated:
If an object is to be shown it is likely to be manipulated and it will most likely be held in front of the robot's camera and *if something is manipulated and held in front of the robot's camera, it is an object.*

Also assumption 2 seemed supported by the observations and for the observed cases it is possible to state that
If a location is to be presented, it is likely that the robot is positioned carefully and the item is presented using a coarse directed deictic gesture and *if the robot is positioned carefully and turned to a particular pose and a coarse directed gesture is used, it is likely that a location is presented, or a "location" that refers to a non-entered region.*

Assumption 3 was supported by the data in the sense that
If a region is to be presented while being inside it is unlikely to observe any

preparation or gesture apart from a "sweep" gesture and *if something is specified only verbally, neither using a preparation nor a gesture it is likely to be a region.* However, *if it is known from a linguistic categorisation that a region is presented, but a clear preparation and specific pointing gesture are observed, it is likely that the region was not entered and should only be referred to with a "link"* and *if a movement preparation and a "location gesture" (hand point) are observed but there is only an opening to be "seen" it is possible that a region is referred to.*

3 Conclusions

This paper summarised an investigation of potentially observable interaction patterns and their correspondence to a conceptual hierarchy for the representation of indoor environments, which is used as generic environment model in a framework for Human Augmented Mapping. The results of this investigation are very promising for future ideas on reasoning strategies for the learning of not only the labels of particular items but also their (spatial) concept, so that the amount of *a priori* knowledge can be reduced in a respective robotic system. However, the focus of current work is the setup of rather controlled experiments to actually investigate the applicability of the interaction pattern hypotheses.

References

1. Martínez Mozos, O., Triebel, R., Jensfelt, P., Rottmann, A., Burgard, W.: Supervised semantic labeling of places using information extracted from sensor data. Robotics and Autonomous Systems **55**(5) (2007) 391–402
2. Vasudevan, S., Gächter, S., Nguyen, V., Siegwart, R.: Cognitive maps for mobile robots – an object based approach. Robotics and Autonomous Systems **55**(5) (2007) 359–371
3. Zender, H., Jensfelt, P., Mozos, Ó., Kruijff, G.J., Burgard, W.: An Integrated Robotic System for Spatial Understanding and Situated Interaction in Indoor Environments. In: Proceedings of the National Conference on Artificial Intelligence (AAAI), Vancouver, BC, Canada. (2007)
4. McNamara, T.: Mental representations of spatial relations. Cognitive Psychology **18** (1986) 87–121
5. Holsanova, J.: Spatial language and dialogue: A multimodal perspective. In: Workshop on Spatial Language and Dialogue, Delmenhorst, Germany (2005) http://www.lucs.lu.se/people/jana.holsanova/.
6. Walters, M.L., Dautenhahn, K., Woods, S.N., Koay, K.L., te Boekhorst, R., Lee, D.: Exploratory studies on social spaces between humans and a mechanical-looking robot. Connection Science **18**(4) (2006) 429–442
7. Topp, E.: Human-Robot Interaction and Mapping with a Service Robot: Human Augmented Mapping. Doctoral Dissertation, KTH School of Computer Science and Communication (CSC), Stockholm, Sweden (2008)
8. Topp, E., Hüttenrauch, H., Christensen, H., Severinson Eklundh, K.: Bringing Together Human and Robotic Environment Representations – A Pilot Study. In: Proceedings of the IEEE/RSJ International Conference of Intelligent Robots and Systems (IROS), Beijing, China. (2006)
9. Hüttenrauch, H., Topp, E., Severinson Eklundh, K.: The Art of Gate-Crashing: Bringing HRI into Users' Homes. Interaction Studies **10**(3) (2009 (to appear))
10. Hüttenrauch, H., Severinson Eklundh, K., Green, A., Topp, E., Christensen, H.: What's in the Gap? Interaction Transitions that make HRI work. In: Proceedings of the IEEE Workshop on Robots and Human Interactive Communications (Ro-Man), Hatfield, UK. (2006)