Human Augmented Mapping – a pilot study

Elin Anna Topp* and Helge Hüttenrauch†
School of Computer Science and Communication (CSC)
Royal Institute of Technology (KTH)
10044 Stockholm, Sweden
{topp,hehu}@csc.kth.se

March 14, 2006

Abstract

Human interaction with a service robot requires a shared representation of the environment for spoken dialogue and task specification where names used for particular locations are depending on personal preferences. A question is how such human oriented models can be tied to the geometric robotic models needed for precise localisation and navigation. We assume that this integration can be based on the information potential users give to a service robot about its working environment. We further believe that this information is best given in an interactive setting (a "guided tour") in this particular environment. This report presents a pilot study that investigates how humans present a familiar environment to a mobile robot. The study is set up within our concept of Human Augmented Mapping, for which we assume an initial "guided tour" scenario to teach a robot its environment. Results from this pilot study are used to evaluate a proposed generic environment model for a service robot.

Figure 1: Illustration of a user showing the kitchen to her robot

*Centre for Autonomous Systems/Computational Vision and Active Perception Laboratory
†Interaction and Presentation Laboratory

Contents

1 Introduction

Service robots are often mobile platforms that provide assistance to humans. Thus, a basic competence for such a mobile robotic system is the ability to move from one place to another to provide its services. This requires navigation and localisation functionalities. Also, the robot has to share the environment with its potential users, which means that it has to move around and reason about its whereabouts in a way that is comprehensible. Mobile robots can navigate on the basis of metric, often feature based maps, and they can build those maps autonomously while exploring an environment for the first time. Methods in robotics research are dealing with this issue of Simultaneous Localization and Mapping (SLAM) [3, 4, 6]. Humans have a topological, (partially) hierarchical, view on their environment [13]. To enable a service robot to perform tasks for users in arbitrary environments well known to the user, but initially unknown to the robot, a spatial representation that is understandable for both, the robot and the user, is needed. Assuming an indoor environment such as a home or office building we mean by a commonly understandable map representation, that the robot's notion of the environment appears to be the same as the one the user might refer to. In other words, we need to build a "shared mental model"[1]. Such a model is likely to depend on a very personal view a potential user has on the environment. We propose that this individual user perspective and understanding of the physical environment can be used to "personalise" the robot's understanding of it. Provided that the robot has some general world knowledge, the robot is adapting the specific details based upon the user's preferences. We assume a scenario of a "guided tour" to be an appropriate way to "teach" the robot its environment and to test for the possible effects of such a user controlled personalization of the robot's knowledge representation. The user can guide the robot around and name important places (e.g., rooms or large, mostly static, objects). Concurrently the robot can build a (metric) map of the environment. This map is augmented by the user's information which allows to integrate the robot's metric, feature-based map with the topological map representation of the user. Figure 1 illustrates a scene of such a guided tour.

Central issues for the comprehensible representation of environments are the questions

a) what strategies to present an environment would be used by different users, and

b) how the given information can be incorporated into an environment model to actually satisfy the requirements for a shared representation.

In earlier work we introduced the concept of *Human Augmented Mapping* (HAM,[17]) which allows us to subsume different aspects of Human-Robot Interaction (HRI) and robotic mapping. Different aspects of interaction as well as posture and positioning of subjects in relation to a robot were studied [7] in an earlier experiment with 22 participants. In this case the scenario was also a "guiding the robot around" scenario, but the environment was limited to one room and the robot used in the study was controlled remotely.

The present report describes a user study in which subjects guide around an autonomous mobile robot in a floor of a research department that they are familiar with. The study investigates how different users present a well known environment to a robot. We suggest

[1] We are aware that the term "shared mental model" is used in other contexts already and might not be entirely appropriate here. Nevertheless it allows to express pragmatically, what we want to achieve: A model that represents a common ground for task oriented communication.

a generic robotic environment model and demonstrate with results from an initial pilot study, how this model can be personalised to different individual representations of a given environment.

1.1 Outline of the report

The rest of this report is organised as follows. We give an overview of related work and refer to hierarchical environment representations motivated from results in Cognitive Science and Psychology to propose a general robotic environment representation in sections 2 and 3. Sections 4 and 5 explain the design and results of our study, and in section 6 we draw conclusions on this pilot study and its results.

2 Representation of space

In “The intelligent use of space” [9] Kirsh stated that in order to understand complex (human) models of an environment, we have to observe the interaction of the (human) agent with and within this environment. Based on those observations, corresponding *robotic* models can be obtained. Transferred to the interaction of two agents in and about a certain environment, observations from human-human interaction could be the base for a general robotic environment model. The strength of such a derived model for robots lies in in its incorporation of information given by users that in turn understad and experience this HRI as a method of personalising the robotic system to their preferencesas and expectations. Personalisation along the taxonomy of Blom [2] means in this context to *accommodate work goals* (to “customise” the robotic system for certain tasks) and to *accommodate individual differences* (of different users in the explicitly stated representation). We propose the following method to achieve this adaptation of the robotic environment model: A user and a robot are observed during their interaction when the user is showing an environment to the robot. The analysis of the observations will allow to better understand what robotic model can be used to build a “shared mental model” that both the user and the robot can refer to later.

Kyriakou *et al.* investigate in a study that uses a miniature robot on a table top street “map” how computer vision can be used to follow verbal guiding directions [12], by having subjects guide the robot with commands like “follow this road to the station, then turn left”. This is another, spoken dialogue based, form of “guiding a robot” without actually being part of a collaboratively operating duo in a co-present, embodied interaction. Furthermore an important condition for such a setup is the a priori availability of a map representing the environment in question that contains all items a potential user considers important at the respective position. Since we wanted to learn about what users present in a given environment and how they do that, we consider such a setup too limiting for our purpose. Additionally such a simulation setup would not allow to take into consideration the jointly experience complexity and ad-hoc opportunities of the user and the robot being in a shared work space.

Kuipers *et al.* presented a mapping approach that represents the environment as a combination of global topological and local metric maps [11]. The main aspect of this work however is the handling of large scale maps, that can be achieved by representing the environment as local metric maps that are linked in a global, topological (and as such hierarchical) representation. Also in other approaches the segmentation of metric maps and/or organisation of them into hierarchies has been studied as part of research in SLAM, but primarily as a way to limit computational complexity [3, 15, 16, among others].

Interesting techniques to interactive robotic mapping have been reported by Diosi *et al.* [4] as well as by Althaus and Christensen [1]. Diosi *et al.* obtain a purely metric spatial representation of an office environment by guiding a robot around and defining labelled rooms. Althaus and Christensen model the environment rather as a topological graph, but do not consider different levels of granularity in their representation. We believe that not only rooms are needed, but also a lower level of complexity has to be integrated in a topological model. This allows, for example, to integrate specific places (objects) into the specified areas.

2.1 Motivation for an environment model

A number of different theories on how spatial relations are acquired and represented have been proposed throughout the years. According to McNamara [13] those theories can be grouped different the dimensions of

a) format (analog vs. propositional),

b) functionality (spatial configuration vs. semantic or logical knowledge),

c) structure (flat vs. strongly hierarchical), and

d) contents (encoded information vs. procedural knowledge to compute information).

McNamara used this categorisation to design a psychological study on spatial representations that concentrated only on the two latter characteristics (structure and contents). Subjects were given recall and distance estimation tasks on items that were spread out in physically separated regions on a “map”. The results indicated, that distance between two items matters as well as co-coexistence in one region. In other words, if two items were close to each other, but in different regions, it was still possible for the subjects to recall and estimate their spatial relation. If the distance was large, this recall and estimation worked better within the same region. Thus McNamara came to the conclusion, that a *partially hierarchical model* supported his findings most appropriately.

Following these findings, we assume that users would not necessarily follow a hierarchical order when explaining the environment to the robot, e.g., explain a certain room first and then give information about specific places or items in this room. Transferring the idea of a partially hierarchical human mental model to our guided tour implies that the assumed robotic environment model has to be able to handle spatial information given in arbitrary order. Thus we propose a hierarchical structure that incorporates the required flexibility with generic entries on each level in which places can be represented. We express this assumption as well in a number of working hypotheses for the pilot study in section 4.3.

To incorporate other dimensions, particularly the *functionality*, the hierarchy needs to be extended. Galindo *et al.* [5] propose *Multi-Hierarchical Representations* to incorporate semantic information into their environment model used for mobile robotics. In their work two hierarchies, one conceptual, the other spatial, are linked with anchoring to enable reasoning. Their spatial hierarchy is build from local map representations obtained from sensory data, that are interpreted as “open spaces” (rooms, corridors) connected in a topological structure. The conceptual hierarchy incorporates concepts such as *workspace*, *room*, *object* and instances of those categories. A semantic model that links objects conceptually to rooms is given a priori. For example it is assumed that an object “bathtub” is to be found in a room called “bathroom”. By observing objects, the conceptual hierarchy is then used to assign a specific concept (“bathroom”) to a local map representation in the spatial hierarchy. As stated above we assume a hierarchical representation of the environment, but do not incorporate any semantics so far.

Along with the functionality one issue is the personalisation [2] of a particular environment representation. From intuition one would expect, that individuals have different preferences and ideas how to interpret and use their surroundings, however we equally believe that these individual perspectives are based upon a (more or less vague) common understanding and agreement, e.g., on the semantics of the concept (room and term) “kitchen”. We consider the fact that different users might give different information to the very same robot as an issue of future work. Our environment model though is flexible enough to model those individual differences within the same framework.

3 Human Augmented Mapping

With our concept of *Human Augmented Mapping* [17] we can establish the link between a robotic map that enables the robot to navigate and the environment representation of a user (also referred to as "cognitive map" [10]). We use a graph based model of the environment, described in section 3.1 to incorporate the information that is given interactively. Our assumption is that a "guided tour" is an appropriate way to give the user the possibility to personalise the robot's generic environment representation on-line. An "off-line" personalisation that could be achieved by pointing out items and places on a metric map representation that was *autonomously* created by the robot does not seem as useful, since the user would have to remember *exact* spatial relationships between items. When the robot and its user instead are allowed to share the same workspace in an interactive setting, it is presumably easier to determine for the user, if the robot "understood" a piece of information correctly. From the robotic point of view we see the advantage of an interactively controlled mapping process in the disambiguation possibilities that arise from the interaction. We further do not assume a full, initial environment model that allows the robotic system to instantiate content entities by autonomous exploration, but consider a more general, structural model that can be filled with personalised information and that can be revised if necessary.

3.1 A hierarchical graph structure

We model the environment by using a hierarchy of graphs. The main concepts we incorporate so far are *locations* (or places) and *regions*, as depicted in figure 2.

We define *locations* as
Specific positions/areas that can represent the position of large objects that are considered static.
Such *locations* can for example be a closet, a refrigerator or a sofa.
A *region* is then
Any portion of space that is large enough to allow for different *locations* in it, or at least large enough to navigate in it.
Typically this would be rooms or corridors or parts of those.

A natural extension to a higher level of the representation would be *floor* or *building*, but this was considered an extension for later work. Accordingly smaller objects, such as milk bottles in the fridge or brooms in a closet, that can (hypothetically) be manipulated and change their position frequently could be integrated on a lower level of the hierarchy. *Regions* are represented by local (metric) maps that can be used for navigation. The local maps are linked metrically by pairs of internal *connector nodes* that have an absolute position with respect to the local map they are in. Since those maps are built at the same time as the graph structure is filled with the information from the user, an initial internal hypothesis needs to be introduced. To maintain the hierarchical structure but allow for partially hierarchical representations as well, we assume a "generic *region*" in which we start the mapping process. With the "generic *region*" we can guarantee that all mapped areas are represented as a *region* on the respective level of the hierarchy. As soon as a *region* is assigned a name (label) by the user, it is stored together with the corresponding local representation that might already contain information on specific *locations*. When a specified *region* is left, and the adjacent area was not explored before, this "new" and yet unexplored *region* becomes a "generic *region*". Note that the "generic *region*" can consist of several, topologically delimited *regions* as an autonomous mapping approach would possibly determine. Only specified *regions* are entities in the hierarchy that form

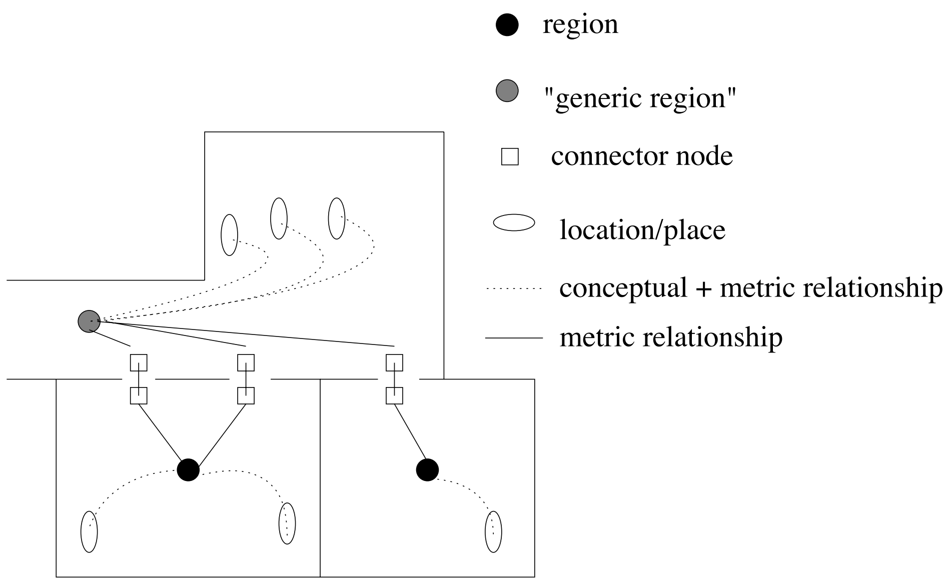


Figure 2: Our graph structure visualised in 2D

a new branch from the respective level downward. This makes it possible to define a specific *location* in a *region*, that is not (yet) specified, for example, to point out the "entrance" in the "generic *region*" (e.g., the corridor in this situation).

4 The pilot study

We conducted a pilot study to test our proposed robotic environment model against the information on a specific office area given explicitly by a human user to a mobile robot. Additionally the pilot study serves as a proof-of-concept for a more comprehensive user study currently prepared. The pilot study comprised trial runs with five subjects of about 45 minutes duration each. Within this time period the subjects spent about 20 minutes interacting with the robot, the rest of the time was used for instructions before and short interviews after the sessions. All participants received a cinema ticket voucher as compensation for their participation.

4.1 Scenario

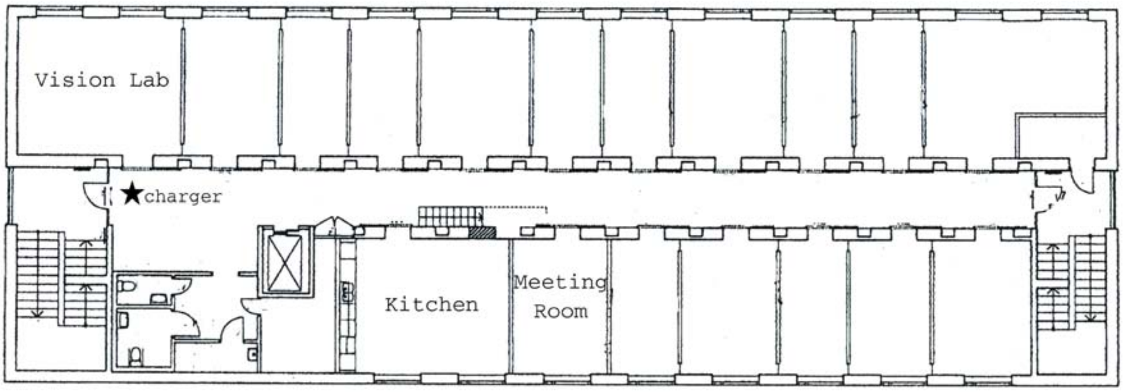


Figure 3: The floor plan of our office environment on which the experiments took place. The star marks the starting point, where subjects encountered the robot

The scenario of the pilot study was a "guided tour" through a portion of an office building. Figure 3 shows the floor plan with offices (not marked), the kitchen, the meeting room and the computer vision laboratory of our office building where the trials were conducted. Subjects were instructed to show the robot around in the environment so that it later could perform non-specified service tasks and in order to do this needed to have "seen" the respective *locations* (a more detailed description of the instructions and the technical realisation is given in sections 4.2.2 and 4.2.3).

4.2 Method

In the following section we explain our selection of subjects, the instructions given to them, and the methods used for data collection.

4.2.1 Subjects

As important precondition to our pilot study we assumed subjects to know the environment they would guide the robot around in. This assumption on user qualification and experience is important and based on the belief that potential users will "add" service robots to their (to them already well known) homes and offices. Subjects were therefore recruited from the laboratory environment the experiments took place in. We make this a requirement also for our future study as it will ensure that subjects are not primed or biased by information that would have to be given to participants unfamiliar with the given robotic trial environment. To require familiarity with the robot's operation area is thus a design choice that differs from other human-robot interaction studies, where subjects often are invited into an unfamiliar or even "simulated" environment. The deliberate choice comes at a price however: in our own office environment some subjects of the pilot study were expected to be familiar with the internals of robotic systems. As a consequence we plan on (also) using a different environment for our future user study to make sure that the familiarity with our robotic system is counterbalanced by subjects without experience in robotics research.

To assure at least some variety in familiarity with robotic systems we selected our five subjects actively among the members of the Computer Vision and Active Perception Laboratory[2] that hosts a part of the Centre for Autonomous Systems[3] on our campus. The group of pilot subjects included one secretary (familiar with robots from films, presentations and frequent encounters in the office environment, but not familiar with their internals), three computer vision researchers, one of them somewhat familiar with the internals of robotic systems, and one robotics researcher from the field of robotic mapping. Thus, the participants represented the full range of robot expertise available at the laboratory. All subjects had been working in this particular office environment for about two years.

4.2.2 Instructions

Our subjects were given an instruction sheet (included in the report as appendix A) that explained the task and the functionalities and abilities of the robot. The task was to use a number of commands (`follow me, go to <target>, stop, turn left, turn right`) and explanations (`this is <item>`) to make the robot follow and to point out everything that the subject considered important for the robot to know on the floor the experiments took place on. The time frame given to the subjects for the completion of their task was about 20 minutes (15 minutes for the guided tour and five minutes to test the robots "memory"). In the instruction none of the words *region, location, position* or *place* was named. We referred to *"everything, that you think the robot needs to know", "whatever you pointed out before", etc.*, so that subjects were completely free to decide, what they would present to the system and how they would name it. Neither did we give any example (e.g., "You can name for example the coffee maker"), to avoid priming the participants on items that a particular subject would not have considered important in the first place. Nevertheless all subjects were informed that we were not interested in small objects, since the robot had no object recognition abilities, it just would need to know "where" to go. The instruction sheet included a drawing that showed, how the field of view of the robot looked like, and explained that the robot used a laser range finder to detect the subject for following and "looking around". This information was important to the users' understanding of the robot's capabilities and behaviour, since the laser range finder only offers a forward field of view of 180° with a range of 8 meters (for the detection of users we reduced it further to 3 meters). The subjects could thus understand, how the robot perceived its environment.

A particular instruction given to the subjects regarded the approach to the robot and initiation of the robot's following functionality. We explained that in order to be detected and classified as user the subject had to move a few steps in front of the robot. Further, in order to make the robot start following them, the user had to gain a distance to the

[2]http://www.nada.kth.se/cvap
[3]http://www.nada.kth.se/cas

robot of at least one meter, to give it the space to actually move.

Subjects were also informed that the robot was moving autonomously throughout the trial but all commands were interpreted by an experiment leader and fed manually into the system. Since we did not incorporate any object recognition we stated that a service task (`go to <target>`) would be successfully completed when the robot could find its way to the *location* where the task would have been performed. Also for the actual presentation of an item, the robot was assumed to "see", when it was "facing" the item. The instruction sheet was very honest about the robot's abilities: we clearly stated which of the functionalities of the robot were in fact simulated or remotely controlled (see 4.2.3 for details) by an experiment leader that followed the (subject and robot) pair. We also explained what the subjects should not try to do, as for example to send the robot around to explore the environment on its own, or to use the elevator. Subjects were offered to ask for help before and during the actual experiment, and knew that they could abort the experiment at any time.

4.2.3 Technical realisation

The study was performed with a commercially available Performance PeopleBot by ActivMedia[4]. In a previous study this robot was used in a Wizard-of-Oz-setting [7], where the robot's functionalities were remotely controlled or simulated by two experiment leaders. For the technical realisation of our pilot study scenario we used a laser range data based tracking and following system [17], which has been extended to incorporate a metric laser range data feature based SLAM method [6] and an input option to label *regions* or *locations* with name tags. Basic platform control and access to the sensors and to a text-to-speech system (Festival[5]) are provided by the Player/Stage[6] software library.

The system represents labelled *regions* and *locations* in a simple graph structure that distinguishes between specifically labelled positions ("defined place") and internal navigation nodes. The internal nodes are used to build a navigation graph on which the system can perform a graph search to plan a path[7] to a previously named position. Note that this system does not yet implement the hierarchical model we proposed above, but enables a user to act and interact with the robot according to our scenario to test the validity of our proposed model.

The verbal interaction of the user with the robot was still controlled by the experiment leader, i.e., utterances from the subject were interpreted by the experiment leader and labels of *locations* or *regions* as well as user commands were fed into the system via a graphical user interface (GUI) running on a laptop. This allowed us to avoid problems due to miscommunication (as studied by Green *et al.* [8]), which otherwise could have interfered with the actual task. For verbal feedback though we used the text-to-speech system with precoded utterances, so that the robot could give spoken feedback about its own state and the task given to it (e.g., "I will follow you", "Stopped following", "I think I have lost you", "Stored <item>").

As the experiment took place on an entire floor of the building one experiment leader (the robot's supervisor) had to follow the subjects to observe the experiment including

[4]http://www.activmedia.com

[5]http://www.cstr.ed.ac.uk/projects/festival/

[6]http://playerstage.sourceforge.net

[7]implementation part of the CURE library (© 2005 Patric Jensfelt and John Folkesson, Centre for Autonomous Systems, Royal Institute of Technology, Stockholm, Sweden)

all utterances and in general, to assure the subject's safety at any time. The implemented system allowed switching from autonomous navigation to full remote control immediately by invoking a soft joystick implementation. Thus potentially inconvenient situations could be solved by the experiment leader as in a Wizard-of-Oz setting without having consequences for the mapping process and the labelling.

We provided the robot with two different behavioural strategies for the labelling of either a *location* or a *region*. If a *location* (including a "link" to a *region*, e.g., a doorway) was presented, the robot did not move and stated immediately, that it stored the given information. If on the other hand a *region* was presented, the robot stated, that it needed to have a look around and performed a 360° turn before confirming the information. The decision, which behaviour to choose, was made by the experiment leader according to our generic environment model and the respective definitions of *regions* and *locations*.

4.2.4 Observation methods and data collection

By storing the data provided by the robot's sensory systems we could get a full "real time" (graphical) representation of each of the experiments. Additionally we recorded the experiments with two digital video cameras each. One video was recorded from the robot's point of view by mounting the video camera on its upper platform. The other camera recorded from an external perspective by being moved, accompanying the user and the robot. After their experiments our pilot subjects were asked to answer a number of questions (see appendix B) on the experiment in a short interview. This interview was scripted with a list of prepared questions on the motivation for naming or not naming certain *locations* or *regions* and for the handling of the tour scenario. We were particularly interested in whether subjects had perceived the behaviour of the robot differing depending on what was pointed out (a *location* or a *region*) and what they thought about this difference.

4.3 Hypotheses

We wanted to study how different individuals present a known environment to a mobile robot and relate the resulting information to an environment model we consider appropriate in the context of Human Augmented Mapping. We assumed that humans do not necessarily follow a hierarchical structure when they present a known environment to a robot (see section 2). Thus, we started out with a number of working hypotheses (WH) about the way subjects would present the *regions* and *locations* they considered relevant, as well as about the entities that would be named:

WH1: "users do not name all *regions* in the environment",

WH2: "users point out *locations* in *regions* they did not name before", and

WH3: "users point out *regions* without entering them".

We use these hypotheses to test whether the observations from the pilot study can be related to our environment model. We did not formulate a specific hypothesis for the dependency "familiarity with robotic systems vs. way of explaining the environment to a robot" to explore this issue. Nevertheless we expected a robotic researcher particularly familiar with map representations to be more explicit than subjects not familiar with robotic environment representations. Further we speculated, along the argumentation of Sidner *et al.* [14], that the difference in the robot's behaviour would allow the subjects to "understand" the robot's internal processes, when storing either a *region* or a *location*.

5 Results from the experiments

In this section we present the results from our pilot study. We are aware that the data set is small and consequently not entirely representative. However, it is possible to analyse the outcome of the experiments in terms of *occurrence* of different phenomena. Additionally, our observations and the subjective answers we obtained in the short interviews allow us to investigate how subjects reasoned about their strategy to show *regions* and *locations* and to improve the system for further studies. As one outcome we gained an increased confidence that the methodology for conducting the pilot study actually can be applied to show the validity of our approach in getting information on individually different ways of building map representations in an interactive, joint process. Furthermore we believe that the soundness of our environment model seems to be supported by its demonstrated ability to handle the diverse situations observed. In table 1 we summarise the quantifiable results to give an overview over our observations and statements from the interview.

Observation Subject	VR	VR	VR	SE	RR
Interaction time	22min	19min	11min	25min	24min
# *regions*	4	2	–	2	2
# *locations*[I]	4	4	5	4[II]	8[III]
# *regions* w o loc.	3	2	–	1	1
# loc. w o *region*	3	4	5	2	3[IV]
# *regions* w o entering	1	2	1	1	–
Behaviour noticed	Yes	Yes	–	No	Yes
– appropriate	Yes	Yes	–	–	Yes
– appears smart	Yes	No	–	–	Yes

VR: Vision researcher, SE: Secretary, RR: Robotics Researcher

I: including *regions* that were only pointed to
II: including one small object (a salt shaker)
III: including one person and two doorways to respective rooms
IV: excluding doorways

Table 1: Quantifiable results from the pilot study

5.1 Observations

All subjects but one used the full time frame to present the environment to the robot. The "tour" started for each experiment at one end of the corridor (see Figure 3), where the robot awaited its user. An initial *location* (the "charger") was generated automatically directly after the system was initialised to enable the robot to go back to this starting point. As a consequence we refrain from taking this automatically generated *location* into consideration for the analysis – despite the fact that subjects were informed about its existence and used it, e.g., to send the robot back when finishing their trial.

All subjects took the robot into the kitchen, probably because this is a central room in our office environment, both from a topological, a functional, and a social point of view. However, the observed diversity in strategies to introduce the kitchen to the robot was quite large, ranging from the pure introduction of *the kitchen* over some combination of *specific locations in the kitchen and the kitchen itself* to *specific locations only*. Already from our small sample of data we can thus conclude that the variety of explicitly stated information that a robotic system in an interactive mapping process would have to cope with is large and needs to be handled by the robot's environment representation. More specifically, these differences in naming observed for the kitchen and its *locations* correspond to our expectations expressed in hypotheses `WH1` and `WH2`.

We also noted that none of the subjects named the corridor or hallway – leading towards and being traversed on the way to the kitchen – itself as a *region*, but all of them pointed out specific *locations* in it, which gives us further evidence for our hypotheses `WH1` and `WH2`. One frequently presented *location* in the corridor was the "elevator" (or "lift") (named by four of the five subjects), which was however only shown by positioning the robot in front of it and pointing to the *doors*. This pattern was equally observed for rooms that were indicated only by pointing to the respective door, confirming our expectation expressed in hypothesis `WH3`.

When asked about their strategy in the post-trial interview, most subjects stated that they had pointed out those *locations* or rooms they personally considered important. Other rooms or *locations* were therefore left out on purpose and not presented to the robot. In some cases the subjects stated that the time constraints given by the experiment leaders kept them from presenting more to the robot. We see this as a sign for a strategy to personalise the robot's environment representation to personal needs and preferences while trying to adhere to the trial settings. A possible consequence to this observation is to increase the time limit for the interaction with the robot in the respective scenario or to run multiple trial sessions with the same subject.

We asked all subjects that had presented a mixture of *regions* and *locations* (four out of five), if they had perceived the difference in reaction of the robot (turning by 360° for a *region* vs. not turning for a *location*). Three out of those four answered that they had observed the difference in behaviour. All three stated that this behaviour seemed *appropriate* and/or made the robot *look smart*, since it obviously wanted "to understand its surroundings". One subject did not notice the difference in behaviour, possibly because only two rooms were presented, and the subject stated to have been busy figuring out, "why the robot sometimes needed a long time to understand me, and sometimes not". Note that this was stated despite the fact that written information had been given to all subjects, stating clearly that all dialogue features were to be simulated by the experiment leader.

Despite some technical problems (see section 5.4 for details) and the above mentioned timing problem that made it difficult for one subject to understand the robot's reactions, all subjects expressed their satisfaction with the flow of interaction and communication as well as the robot's performance.

5.2 Particular situations

Even with the limited number of subjects we were able to observe some interesting strategies for the presentation of the environment. We relate the observations to statements from the short interviews where possible and try to establish the relationship and importance to our environment model.

5.2.1 Pointing out persons

In two cases subjects tried to point out a person. In one case the person was sitting at her desk and the robot was made to store the respective *location* (from where the robot had been shown the person) by the experiment leader. In the other case the subject reacted spontaneously to someone suddenly walking out of the elevator right in front of the robot. Here the robot was not made to store the information from the introduction. Nevertheless these situations show that the system would have had to handle introductions of persons as well, since the introductory phrase "this is < *name* >" was exactly the same as for the kitchen[8]. Another interpretation of the observed behaviour is that it is necessary to consider the influence of spontaneous encounters and opportunities for interaction which might prompt users to incorporate them into the interaction with the robot.

5.2.2 Possessive pronouns and relations

One of the subjects presented an office (a *region*) as "my office" to the robot. In such a case a dialogue system would actually have to analyse "my" and relate it to the subject's name. Since this was beyond the scope of our pilot study, the robot was later referring to "my office". Nevertheless the dereferencing of possessive relations is an issue for both the dialogue and the representation of possessiva in the generated map representation.

5.2.3 Extreme personal point of view

In one experiment session we observed that two rooms were pointed out, but no *locations* in them. In the corridor, none of the service points (pigeon holes, printer, etc.) was named, but the two entrances/exits (named as "exit") to either side of the building and the elevator as well as the experiment leader's office (only the door was pointed to) were shown to the robot. When asked why no other *locations* in e.g., the kitchen were named the subject stated that the exits were considerably important, as well as the kitchen as a room in case that guests had to be met and/or served. However, the coffee machine and the refrigerator were not equally important to the subject who stated not to drink coffee at all nor to have used the refrigerator. The observation holds both evidence for our hypothesis `WH2` and an extreme personal point of view on the environment.

5.2.4 Explaining no rooms at all

One of our subjects concentrated only on *locations* (e.g., pigeon holes, coffee machine, refrigerator) and did not name any room (or other *region*). On the question, why not, for example, the kitchen as a whole was named, the answer was that the robot should rather know about the whereabouts of the places (*locations*), where it should *do* something. Just sending it to *the kitchen* would by no means help to get a coffee, the subject stated. We see this as a strong evidence for hypothesis `WH2`.

5.2.5 Explaining doorways

We expected our subject with robotic research and mapping experience to be more precise and explicit than other subjects. This expectation could be confirmed by the fact that the doors to shown rooms were pointed out explicitly *when the robot was standing exactly in the door opening*. We could also observe that both named rooms were actually entered. Since only two rooms were presented during this experiment, we can of course

[8] In this particular experiment the subject left out all articles when presenting items, e.g., uttering: "this is kitchen"

not generalise, but we consider at least our expectations for the robotics researcher's strategy confirmed. On a more abstract level we might need to look for effects caused by different levels of understanding the robot's internals and provide plausible metaphors to be guiding the interaction strategy design accordingly.

5.3 Relation to our environment model

Our observations show that even with a small group of subjects different ways to show and explain the environment are to be expected and dealt with, depending on the individual view and use of particular items and rooms. We see these differences as a proof of concept for our proposed environment model (as introduced in section 3.1) which we consider usable for a robotic map representation.

A general assumption is that a given robotic system has the ability to perceive *regions* that are delimited from other *regions* autonomously. This could for example be achieved by door detection or a method like the watershed algorithm [4, as an example of use]. We also assume that we have a general knowledge model distinguishing between *regions* and *locations* and a dialogue model that uses this knowledge base. From the experiments we collected evidence on the strategy of users to point out a *region* by only showing the respective door leading into the *region* to be named. In these observed cases subjects positioned the robot with the help of "turn commands" so that it was facing the particular "link" (doorway or elevator doors), before naming the *region*. If these subjects on the other hand presented the *region* they were currently in they just stated that this was "the $<$ *name* $>$" without positioning the robot with "turn commands". The detection of such differences in the user's behaviour and spoken utterances could give a signal on the actual intention. We hope to find further evidence for such a differing behaviour in our future study.

Departing from our detailed observations we can postulate some key situations that need to be handled by our robotic environment model (see figure 3.1) and suggest some possible solutions to cope with them. Those suggestions can be taken into account when improving the system for further studies.

5.3.1 Presenting persons

Given an appropriate dialogue model, it would be possible to ask, if actually the *region*/room the person is in should be named accordingly (e.g., "Elin's office", in case "Elin" was introduced to the robot).

5.3.2 *Locations* in an unnamed *region*

If a *location* is named before the *region* it is in, or the *region* is not named at all, this *location* would end up in the branch of the "generic *region*" in our hierarchy. If later the information about the *region* is given, the *region* needs to be delimited and separated from the generic *region*. All *locations* within the observed delimiters (e.g., walls, doorways) are now associated to this new branch in the hierarchy.

5.3.3 Links to *regions*/rooms

With the "connector nodes" of our representation links to rooms (pointed out doorways) can be handled. In the current *region* (which might be the "generic *region*") a connector node with a virtual directed edge to the named *region* is created. Thus the system knows, that it can find the way *to* a certain *region*, without knowing anything about its appearance. Such a process requires obviously the knowledge that a) a *region* is presented, and b) that it is not the one the robot is currently located in. We observed in the cases where a *region* was shown to the robot by pointing out the door leading to this *region* that the respective subjects made the robot face the doorway (by using turn commands) as if they were presenting a *location*. This behaviour was not observed when the current *region* (the one user and robot were in) was presented. In further studies we plan to look for more evidence for such behaviour that would allow to infer about the intention of the user. In all cases the system would need knowledge about the classification of labels for either *regions* or *locations* anyway.

5.3.4 Pointing out doorways explicitly

The environment model could cope with explicitly pointed out doorways by generating a *location* with the respective name. There are several possibilities to represent it in the hierarchy though. One option is to decide which *region* it belongs to, based on the name of the respective *region* (e.g., as observed "this is the door to the kitchen"). The second option is to keep the *location* in both *regions*, with a relative position to the respective local map that relates to the same absolute position (if possible). A third option would be to generate an entry of the generic *region*, that would allow to state that the robot is "in between two *regions*". However, since we could observe the respective strategy only with the robotics researcher, we assume this to be rarely observable with a differently structured sample of users to test with.

Summarising we believe that our model holds at least for the variety of strategies to present a known environment to a robot observed in our pilot study.

5.4 Interaction issues

During the pilot experiments we observed several issues of the technical realisation that had consequences for the actual interaction between subjects and the robot.

Despite the instruction to give the robot space when it was about to follow, subjects waited standing still for the robot to move. The robot's verbal indication to follow ("I will follow you") was obviously not enough to indicate that it would actually follow. From carefully studying the recorded interaction on video we concluded that the robot actually needs to indicate with a body (movement) gesture like turning toward the user that it has seen the user and is ready to follow. A similar problem occurred, when subjects made the robot face something to "look at it" and wanted to continue the tour afterwards. We plan to make the robot turn back toward the user to indicate, that it is ready to continue after storing a presented item.

6 Conclusion and future work

In this report we presented two important aspects of our concept of Human Augmented Mapping, namely the environment representation of the robot and the interactive context that allows to build a shared mental model of an environment. We explained our approach to a robotic map representation and showed, to what extend this representation holds in different situations within an interactive mapping process. A pilot study was conducted to investigate strategies of users to present a for them well known environment to a robot.

Despite the small number of subjects in the study we were able to observe a rather large variety of strategies to present a known environment to the robot in a "guided tour". Partially diversity might be due to differing knowledge in robotics or the individual interest in the robot that our subjects had. However, we can state that all the different situations or strategies characterised in a number of hypotheses we formulated actually occurred at least once. The variety in presentation strategies we observed and the self reflecting comments on them showed us, that there is a need for quite flexible representations, when one robotic system should be used and guided around by different users.

We got mostly positive feedback on the behaviour of the robot, especially on a "*region* observation strategy" we implemented to enable subjects to understand the internal processes of the robot to some degree. This assures us to keep such a behavioural strategy for further studies to allow subjects to understand more of the internal procedures of the robot.

The results from the pilot study encourage us to use the proposed setup in a more comprehensive user study and to investigate the applicability of the proposed environment model in a robotic framework in more detail.

7 Acknowledgments

The work described in this report was conducted within the EU Integrated Project COGNIRON ('The Cognitive Robot Companion' - www.cogniron.org) and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.

We also thank Patric Jensfelt and John Folkesson for a their technical help with the implementation of the system, especially the SLAM method used.

A The instruction sheet

The following text was given to our subjects as instructions for their trials.

HRI Pilot study: Explicit environmental representations in the context of Human Augmented Mapping

Elin A. Topp and Helge Hüttenrauch,
School of Computer Science and Communication (CSC),
Royal Institute of Technology (KTH), Stockholm, Sweden

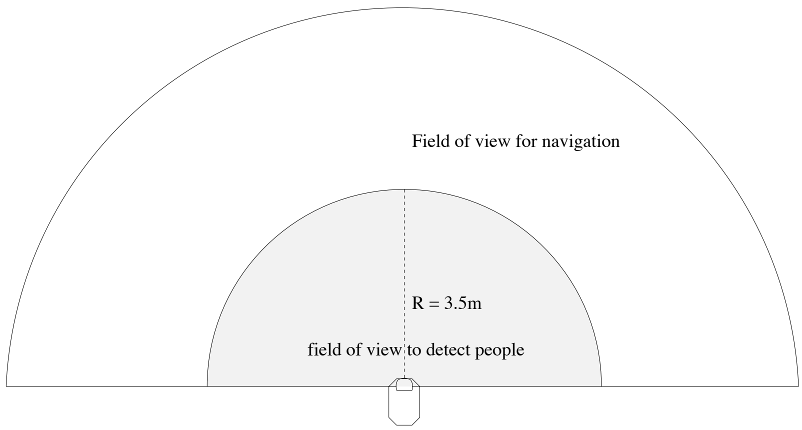
Hello and welcome to this pilot study. This document should give you the information you need to participate in our study on human robot interaction and interactive (robotic) mapping.

The most important information here is that
whenever you are not sure about the task or feel uneasy during the experiment, please TELL us.
It is as well okay to interrupt or even abort the experiment if you do not feel comfortable at any time.

Further we have to make another thing clear: We are not testing you, but our robotic system and some assumptions on the interaction you will have with it. Please, do not feel stupid if something goes wrong, it is probably the robot that does not work properly.

What is this all about?

a)



b)

You will be introduced to your new service robot "Minnie". See Figure a) to get an idea of how the robot looks like. Once Minnie is activated, it will be able to provide services for you, like fetching things, or checking the state of a window for example. Minnie has

a pretty good idea of what certain things can look like and it also has some general ideas what to expect in an office environment. Still, it needs to know, how exactly this, i.e., your office building looks like, and where to find the respective places to perform its tasks. Minnie will detect you and can follow you around, so that you can present the building by showing it around. To detect you, Minnie uses its laser range finder (the blue “coffee maker” thing on the lower platform). Figure b) gives you an idea on the field of view of this device. In fact, Minnie cannot detect you, when you are standing “behind” the baseline of the laser range finder. The laser range finder is actually also the device (“eye”, so to speak) that Minnie uses to find its way.
To make sure that Minnie knows what you presented you can ask it to go there. One place that it will know about immediately is where it will find its charger - which is, where it is activated. Sometimes Minnie does not seem too smart and it might lose you, when it is supposed to follow. Be forgiving, it is a “young robot”! And do not wonder about the gripper that looks not very useful for fetching anything, that is something to deal with later, first Minnie needs to find its way!

Your task

Please take Minnie on a tour around the sixth floor in your office building. Point out everything that you think is important for Minnie to know. Check its memory by sending it around to whatever you already pointed out whenever you like. Please try to show everything important within fifteen minutes (if you need less, do not worry...!). Afterwards we will check if Minnie is able to solve three tasks. One will be a fetch-and-carry type task (that means to go to a certain place, pick up something and bring it to you), the second one a conditioned “fetch-and-carry” (if there is something to fetch, bring it, otherwise report) and the third is a “check”-task (check the state of something and report). The task is successfully performed, when Minnie reaches where it needs to go to perform the required action.
Now, you will be asked to answer some questions in a short interview about the experiment. And after that, you are done. Thank you very much for your cooperation!

Commands and options

You can give the following commands to Minnie:

- “Follow me” will make Minnie come after you.
- “Stop” or “Stop following” will make Minnie stop immediately.
- “Turn left” and “Turn right” will make the robot turn on the spot in the respective direction (robot’s left and robot’s right).
- “This is < whatever you want to present >” will make Minnie store the information.
- “Go to < whatever you already presented >” makes it go to what you pointed out.

Things that do not work

Some things, that you should not try are to:

- send the robot to anything that you know was not presented,
- direct Minnie around “remotely”, or
- use the elevator.

Some notes on technical issues

Nobody is perfect. We are not perfect and therefore the robot is not either. But we will do whatever we can to help you.

Control of the robot The experiment leaders - we - will follow you and the robot for several reasons. One is, that we want to observe and videotape the experiment not only from the robot's point of view, but also from a more general point of view.
The other reason is, that we want to assure your safety and comfort. We can interrupt the robot's automated control at any time and switch to manual control. Or abort the experiment completely. That is one reason for a laptop being carried around. The second is, that we do not want to rely on speech recognition. The contents of your utterances are translated into the respective commands and informations "by hand" and given to the system by typing. That is the second reason for the laptop.

Tracking and following You need to move around a bit (walk some steps) in front of the robot, before it initially detects you. It will start to move only when you are about one meter away already, but it will come a bit closer when you stop, before it stops itself. Do not walk too fast, it might lose track then. At the moment, the maximum distance you can have is a little more than three meters, as shown in Figure b). It might happen, that the tracking system gets confused by objects in your vicinity. In that case we will tell you how to solve the situation and help you if necessary.

Passing doors, other narrow passages and cluttered areas The robot should not collide with anything, neither you, nor a door frame. Therefore the maximum speed in the vicinity of "things" is reduced quite a bit. This means, that Minnie needs a while to go through a door or other narrow passages.

Turning and "seeing" When you ask Minnie to turn left or right, it will turn about 45°, but sometimes (due to technical reasons) it will in fact turn quite a bit more, just repeat the command or ask for the opposite direction. If you want to point out something, Minnie should face this item roughly. It is not necessary that the robot is placed immediately "in front of" the item. A distance of one to one and a half meter is fine.

Your privacy

We are going to videotape the experiment. Additionally the system will log the data from the experiment. These data will be used for an evaluation and as a basis for further research. We will be referring to the data in an anonymous fashion, and we will only do that in a research context.

B Interview questions

The following questions formed a loose guideline for the interviews.

HRI Pilot study: Interview questions

Elin A. Topp and Helge Hüttenrauch, Testperson:

1. Did you notice the difference in the reactions of the robot to regions/rooms and places/locations?

2. Do you think this difference was appropriate?

3. Why do you think the robot had these differences in its behaviour?

4. Did this give you the impression that the robot was "thinking" of the same thing as you were?

5. Why did you not show the ...?

6. Why did you show the ..., but not the ...?

7. When you headed for the (room) and presented the (something), were you planning to present the (something) or were you just planning to go to the (room) and look for things to present there?

References

[1] P. Althaus and H.I. Christensen. Automatic map acquisition for navigation in domestic environments. *in Proc. of the International Conference on Robotics and Automation,* September 2003, Taipei, Taiwan.

[2] J. Blom. Personalization: a taxonomy. *CHI'00 extended abstracts on Human factors in computing systems,* 313-314, ACM Press.

[3] M. Bosse, P. Newman, J. Leonard, S. Teller. SLAM in Large-scale Cyclic Environments using the *Atlas* Framework. *International Journal of Robotics Research,* 23(12):1113-1140, December 2004.

[4] 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,* April 2005, Barcelona, Spain.

[5] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J.A. Fernández-Madrigal, and J. González. Multi-Hierarchical Semantic Maps for Mobile Robotics. *in Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* August 2005, Edmonton, AB, Canada.

[6] J. Folkesson, P. Jensfelt, and H.I. Christensen. Vision SLAM in the Measurement Subspace *in Proc. of the IEEE International Conference on Robotics and Automation,* April 2005, Barcelona, Spain.

[7] A. Green, H. Hüttenrauch, K. Severinson-Eklundh. Applying the Wizard-of-Oz Framework to Cooperative Service Discovery and Configuration. *in Proc. of the 13th IEEE International Workshop on Robot and HumanInteractive Communication,* September 2004, Kurashiki, Okayama, Japan.

[8] A. Green, H. Hüttenrauch, K. Severinson-Eklundh, B. Wrede, S. Li. Integrating Miscommunication Analysis in Natural Language Interface Design for a Service Robot. *submitted*

[9] D. Kirsh. The intelligent use of space *Artificial Intelligence,* 73:31–68, 1995.

[10] R.M. Kitchin. Cognitive Maps: What Are They and Why Study Them? *Journal of Environmental Psychology,* 14(1):1–19, 1994.

[11] B. Kuipers, J. Modayil, P. Beeson, M. MacMahon, and F. Savelli. Local Metrical and Global Topological Maps in the Hybrid Spatial Semantic Hierarchy. *in Proc. of the IEEE International Conference on Robotics and Automation,* April 2004, New Orleans, LA, USA.

[12] T. Kyriakou, G. Bugmann, and S. Lauria. Vision-based urban navigation procedures for verbally instructed robots. *Robotics and Autonomous Systems,* 51:69–80, January 2005.

[13] T.P. McNamara. Mental Representations of Spatial Relations. *Cognitive Psychology,* 18:87–121, 1986.

[14] C.L. Sidner, C. Lee, C.D. Kidd, N. Leash, C. Rich, Explorations in engagement for humans and robots. *Artificial Intelligence,* 166(1):140-164, 2005

[15] N. Tomatis, I. Nourbakhsh, and R. Siegwart. Hybrid Simultaneous Localization and Map Building: a Natural Integration of Topological and Metric. *Robotics and Autonomous Systems,* 44:3–14, 2003

[16] A. Tapus, G. Ramel, L. Dobler, and R. Siegwart. Topology Learning and Place Recognition using Bayesian Programming for Mobile Robot Navigation. *in Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* September 2004, Sendai, Japan.

[17] E.A. Topp and H.I. Christensen. Tracking for following and passing persons. *in Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* August 2005, Edmonton, AB, Canada.