Human-Robot Interaction and Mapping with a Service Robot: Human Augmented Mapping

ELIN ANNA TOPP

Doctoral Thesis
Stockholm, Sweden 2008

TRITA-CSC-A 2008:12
ISSN-1653-5723
ISRN-KTH/CSC/A--08/12--SE
ISBN 978–91–7415–118–3

KTH
School of Computer Science and Communication
SE-100 44 Stockholm
SWEDEN

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framläggs till offentlig granskning för avläggande av teknologie doktorsexamen i datalogi måndagen den 13 oktober 2008 klockan 10.15 i sal F3, Kungl Tekniska högskolan, Lindstedtsvägen 26, Stockholm.

Tryck: Universitetsservice US AB

Abstract

An issue widely discussed in robotics research is the ageing society with its consequences for care-giving institutions and opportunities for developments in the area of service robots and robot companions. The general idea of using robotic systems in a personal or private context to support an independent way of living not only for the elderly but also for the physically impaired is pursued in different ways, ranging from socially oriented robotic pets to mobile assistants. Thus, the idea of the personalised general service robot is not too far fetched. Crucial for such a service robot is the ability to navigate in its working environment, which has to be assumed an arbitrary domestic or office-like environment that is shared with human users and bystanders. With methods developed and investigated in the field of simultaneous localisation and mapping it has become possible for mobile robots to explore and map an unknown environment, while they can stay localised with respect to their starting point and the surroundings. These approaches though do not consider the representation of the environment that is used by humans to refer to particular places. Robotic maps are often metric representations of features that can be obtained from sensory data. Humans have a more topological, in fact partially hierarchical way of representing environments. Especially for the communication between a user and her personal robot it is thus necessary to provide a link between the robotic map and the human understanding of the robot's workspace.

The term Human Augmented Mapping is used for a framework that allows to integrate a robotic map with human concepts. Communication about the environment can thus be facilitated. By assuming an interactive setting for the map acquisition process it is possible for the user to influence the process significantly. Personal preferences can be made part of the environment representation that is acquired by the robot. Advantages become also obvious for the mapping process itself, since in an interactive setting the robot can ask for information and resolve ambiguities with the help of the user. Thus, a scenario of a "guided tour" in which a user can ask a robot to follow and present the surroundings is assumed as the starting point for a system for the integration of robotic mapping, interaction and human environment representations.

A central point is the development of a generic, partially hierarchical environment model, that is applied in a topological graph structure as part of an overall experimental Human Augmented Mapping system implementation. Different aspects regarding the representation of entities of the spatial concepts used in this hierarchical model, particularly considering *regions*, are investigated. The proposed representation is evaluated both as description of delimited *regions* and for the detection of transitions between them. In three user studies different aspects of the human-robot interaction issues of Human Augmented Mapping are investigated and discussed. Results from the studies support the proposed model and representation approaches and can serve as basis for further studies in this area.

Preface

The present doctoral thesis consolidates results of four years of work with a conceptual design for an approach to hierarchical, interactively controlled robotic mapping and localisation, conducted at the Centre for Autonomous Systems (CAS) hosted by the Computational Vision and Active Perception (CVAP) group of the School of Computer Science and Communication (CSC) at the Royal Institute of Technology (KTH) in Stockholm. "Human Augmented Mapping" (HAM) is a central term which is introduced and used to subsume the discussed aspects of robotic mapping and human robot interaction. The general concept of HAM is described and discussed together with results that have been achieved with a respective experimental system implementation. Those results also include observations made in three different user studies, that were conducted to test the assumptions underlying the models used for the work.

The work contributed to a large extent to the integrated EU-project "COGNIRON - The Cognitive robot companion"[1], in particular to the Key Experiment 1 – "The Home Tour", that served as a demonstrator experiment to document the results achieved in the project's research activities on "Models of space" and "Multimodal dialogue". Additionally the work was to a large extent funded through the project by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020. The funding is gratefully acknowledged.

Within and related to the project work a number of collaborations inside and outside the Royal Institute of Technology could be established, that contributed to the results presented in this thesis. My own contributions within these collaborations are pointed out specifically in the respective chapters and sections.

Anders Green and Helge Hüttenrauch, at the time of writing both affiliated with the Human-Computer Interaction group of CSC at KTH, conducted a user study early in the project that inspired my thoughts about the environment model and the first user study setup described in this thesis. It was also possible for me to use data collected during this first user study to evaluate the tracking method that is part of my approach to a complete system to Human Augmented Mapping. At this point I want to thank Anders and Helge for these opportunities and the inspiring discussions, particularly with Anders on "spatial prompting", a concept he developed and is about to publish in his doctoral thesis.

Further, I designed and conducted all three user studies described in this thesis in cooperation with Helge Hüttenrauch. Thanks again to Helge, in this case particularly for having the somewhat "different" idea of taking the robot out of the lab and into people's homes to conduct one of the studies, this was a great experience.

One of the studies conducted in the laboratory was actually assigned as a master's project to Farah Hassan Ibrahim, jointly supervised by Helge and myself, and at the time of writing a registered student in the computer science programme of CSC/KTH. Other master's projects and one short-term undergraduate project

[1] www.cogniron.org (URL verified August 19, 2008)

(a German "Studienarbeit") related to my work and supervised by myself were assigned to and conducted by Alvaro Canivell García de Paredes, Maryamossadat Nematollahi Mahani, and Stephan Platzek. Thanks for working with me and coping with the often vague and exploratory ideas for your tasks!

The COGNIRON project's aim to demonstrate results from different research activities as integrated demonstrators in different Key Experiments built the basis for a very fruitful collaboration with the Applied Computer Science group at the University of Bielefeld, Germany. The efforts put into the transfer of a significant part of my experimental implementation to an integrated interactive framework made by Marc Hanheide, Julia Peltason, Frederic Siepmann, Thorsten Spexard, and Sebastian Wrede (in alphabetical order), and myself led to a prototypical fully integrated interactive system for Human Augmented Mapping, that could be demonstrated successfully in the context of the project. Thanks, for helping to get all that code to work!

A joint effort within the research activity on "Models of space" made me travel to the Intelligent Autonomous Systems group at the University of Amsterdam with a SICK laser range finder in my carry-on luggage – also this being an interesting experience in itself – to collect a data set that was made public to give research groups dealing with some form of semantic or interactive mapping a basis to compare their approaches. This cooperation with Olaf Booij, Bas Terwijn, and Zoran Zivkovic led to one of the publications listed as related to this thesis. Thanks!

The cooperations already indicate that this thesis project involved a lot of travelling: A robot travelled through the greater Stockholm area, students travelled back and forth to pursue their international study programmes, a laser range finder got to fly to Amsterdam, and I myself travelled, well ...

I was warned right before I travelled back to Germany after my master's project (which I already travelled to Stockholm for), that I would be travelling *a lot* during my time as a PhD student, due to the European project. I thought "great, I love travelling". After more than 30 take-offs and landings – both for professional and for private reasons – during my first year in the graduate program and actually recognising members of the SAS in-flight staff on particular routes after the second I started to revise that statement slightly. I was also warned that my adviser was a person difficult to meet, due to *him* travelling a lot more than I did anyway. "If you want to talk to him, book the same flight", I figured – I tried, it did not work, he would end up being seated somewhere completely different on the plane.

Nevertheless, I managed to achieve during 48 months of doctoral studies spread over five years what I planned to, actually a bit more, this "bit" now being roughly $1\frac{1}{2}$ years old. I want to thank my advisers Henrik I. Christensen and Kerstin Severinson Eklundh who facilitated the work with their ideas and questions, and particularly for their understanding when I notified them of my pregnancy and plans for maternity leave.

I also want to thank all the other people I have met during my time at the CVAP-group, particularly Patric Jensfelt, John Folkesson and Mattias Bratt for their help and support with hard- and software. I do not know how often I found myself

unscrewing loads of tiny screws on the robot, looking puzzled into its internals and then walking up to Patric's office to ask for advise. He would usually claim that he had no idea what to do either and show up about 30 minutes later with a freshly soldered customised cable that did the trick or a tar-ball of code that he "happened to have sitting somewhere on the computer". Thanks!

Thanks also to Danica Kragic, who acted as an additional adviser in my adviser group, even if she was not really encouraged to look for collaboration opportunities for me with her own PhD students due to the COGNIRON project's intellectual property right agreements. Thanks also for making a last minute attempt to read my thesis draft within a couple of hours as secondary internal reviewer, before we eventually found out that it was allowed to have the secondary adviser sign it off.

A major thanks goes to Jeanna Ayoubi, Friné Portal and Mariann Berggren, who had to deal with all my travel orders and reimbursement forms, often involving some special work when I combined private with work-related trips. Another big thank you for general advise, interesting chats during lunchtime or at the coffee-maker, opportunities to socialise outside the office, sharing an office at CVAP or a hotel room at conferences, or even participating in one or the other experiment or study goes to current or former members of the CVAP/CAS group: Daniel Aarno, Niklas Bergström, Mårten Björkman, Lars Bretzner, Martin Erikson, Jeannette Bohg, Barbara Caputo, Stefan Carlsson, Hugo Cornelius, Oskar Danielsson, Johan Edén, Jan-Olof Eklundh, Staffan Ekvall, Daniel Fagerström, Simone Frintrop, Fredrik Furesjö, Javier Romero Gonzales, Monica Gretzer, Eric Hayman, Kai Hübner, Ronnie Johansson, Hedvig Kjellström, Oskar Linde, Gareth Loy, Peter Nillius, Elena Pacchierotti, Andrzej Pronobis, Maria Ralph, Ola Ramström, Babak Rasolzadeh, Per Rosengren, Kristoffer Sjöö, Christian Smith, Josephine Sullivan, Alireza Tavakoli Targhi and all of you I probably have missed now, there are just too many people I met during the years, including all the master students ...

I already mentioned the extra "bit" that entered my life during my time in the CVAP-group, which made my everyday life a little more resembling a jigsaw puzzle. I want to thank the people involved in the "Future Faculty" initiative at KTH for their brilliant idea of starting "Quottis", the short–term day-care facility that is open for children younger than 12 months – usually the age for them to start in regular day-care, which made the jigsaw somewhat easier to solve. A particular thank you at this point goes to Elisabeth Mosqueda, who took care of our son so often at "Quottis".

There are actually a couple names that I definitely want to be in these acknowledgements. I want to thank my family, first of all my parents Annemarie Topp-Hinterthür and Günter Topp for supporting me through these years, particularly for offering me to start a "writing camp" in their house during four weeks to assemble quite some part of this thesis, acting themselves as full-time babysitters – sometimes the big jigsaw puzzle cannot be solved otherwise. Thanks to my grandmother Anni Hinterthür for still being there – I told you you would see me get my doctoral degree! Last but most importantly I want to thank my own little family, my son Maximilian "Mäxchen" Topp for coping with me dragging

him around to work – actually taking him with me to some of the participants of the second user study, where he could watch me literally "at work", for being my sunshine, almost always able to cheer me up, and showing me once in a while what really matters; and my husband Ludwig Seitz, who had to live with all my uncertainties and many ups and downs during the last couple of years, and who managed that certainly struggling but very successful. Thank you!

Stockholm, September 2008, Elin A. Topp

Contents

List of Figures

List of Tables

Chapter 1

Introduction

An issue widely discussed in politics and economics is the ageing society of the industrialised world. Resources have to be and are assigned to (research) institutions that support various kinds of care giving developments and innovations. Consequently, there is a growing interest for investigations in the field of service robotics and the idea of the personal general purpose service robot or "robot companion" does not seem too far-fetched, given the already existing applications of, e.g., robotic vacuum cleaners and lawnmowers. Opportunities for developments in the area of service robots and robot companions are continuously generated.

There is a wide range of definitions for a "robot" in general and a "service robot" in particular to be found in today's literature, and adding the term "personal" or "companion" makes definitions subsume even more aspects that have to be taken into account. The general idea of using robotic systems in a personal or private context to support an independent way of living not only for the elderly is pursued in different ways, ranging from socially oriented robotic pets (Wada *et al.*, 2005) to mobile assistants (Montemerlo *et al.*, 2002) or "robot companions" (e.g., Haasch *et al.*, 2004) in various, often cooperative projects.

However, in the following a "personal service robot" is assumed to be a mobile robotic platform equipped with some form of manipulator and a suitable interface for interaction with its user that is able to provide general services in a domestic (or office) context, e.g., perform fetch-and-carry or basic cleaning tasks.

The respective area of research has in fact many aspects to consider, both regarding the technological advances necessary to actually build appropriate hard- and software systems, but also regarding the more complex aspects arising when the results of technological developments are encountered by their potential users. Since the "personal service robot" is assumed to be sharing the environment it is supposed to work in with its users, it is inevitable that such a robotic system becomes much more present and central in the live of its users than for example the average entertainment or cleaning equipment would do. Similar to a butler or housekeeper the assumed service robot would need to "understand" the overall

situation in its working environment, also regarding the personal preferences of its users. Creating this type of "understanding" raises a number of issues that have to be investigated in the context of "personal service robots" entering people's homes or offices. One particular aspect is the conceptual and spatial understanding of the surroundings that is presumably needed by the personal service robot, which will be elaborated on in the following, building the motivational background for the work presented in this thesis.

1.1 Motivation

Given the fact that the assumed service robot has to work in an environment that is inhabited by its user(s), how should it move around, communicate and adapt to both the environment and the inhabitant? What is to be expected when a service robot enters an environment, that has not been particularly designed for robots, but for humans? This thesis focuses on particular aspects of these general questions, regarding the representation of the environment that is necessary for the robot, as discussed in the following. To provide services, thus to move around and "work" with and in the environment, it is necessary for the robot to have a certain knowledge about the surroundings, thus, some kind of map or respective representation of the surroundings. Preferably it should also be able to communicate about its whereabouts in a way that is comprehensible for the user, hence the question, what kind of underlying model for the surroundings is needed to allow for both, use and communication, in appropriate ways? How would the robot gather the information necessary to build a usable representation of the particular environment and how can the user assist the robot in doing this?

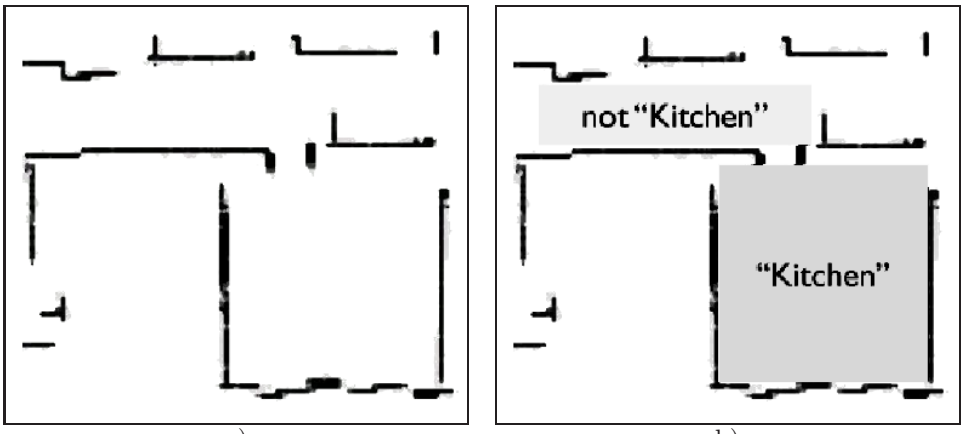
The following scenario illustrates a hypothetical situation in which a "new" service robot is taken into service and needs to be instructed.

Alice and Bob are an elderly couple living in a rather large bungalow. They both are still mobile and capable of handling their daily life in the house quite well, but it gets more and more difficult for them. They decide to get one of these new "ButlerBots" that have been on the market since a couple of months now – and obviously the producing company has fixed the initial flaws. The descriptions they get from the shop state that this new robot does not require any equipment in the house, like sending or receiving units that could be abused. Bob would not want to fiddle around with such things anyway, the house is fine as it is and he does not want to rip out the floor to put in those tiny sensor thingees that they had to put into his sister's house to make her (older) robot work properly.

Two weeks later Alice and Bob receive the package with the shiny new home service robot "ButlerBot". The robot is supposed to help with fetch-and-carry tasks in the house, occasionally it should open the door for visitors, help those around the house if necessary and check the status of windows, for example. Bob has a doctor's

appointment and will not be around for a couple of hours, but Alice decides to get started with the new toy right away – why shouldn't she be capable of getting this thing to work?

After ripping off all the plastic stuff around the robot, she reads the instructions and presses the friendly blue self-test button. Yes – the thing seems happy, according to the description. Then she reads the instructions for "The ButlerBot needs to know its working area". "Pretty obvious", Alice thinks, "any housekeeper would need such an instruction. Well, for a housekeeper it would be sufficient to guide him or her around. Wonder what I have to do to help the robot? I hope I do not have to learn "Robot language" now... Uh-oh, there we go, here is a section on ... how the robot perceives the environment ...? What does this funny drawing mean?" The manual continues to explain that together with her help the robot can fill this



a) b)

Figure 1.1: a) How your "ButlerBot" sees the environment with the help of its laser range finder. b) How your "ButlerBot" can see the environment if you help it to find its way.

funny and totally useless drawing with information that she and Bob will be able to understand. Even names for rooms that they are using only between themselves – like "the blue room" for the little sitting corner behind the kitchen will be okay to use. There is another drawing corresponding to what she would expect the robot to know. Alice thinks that it is really nice of this company that they illustrate this process. But how can she "help" the robot? She carries on reading and finds out – positively surprised – that the robot will ask her to do pretty much what she would do with any other (human) housekeeper. It will ask to be shown around in the house. She follows the instructions to switch on the robot and – "Hello, I am a ButlerBot – show me around please"...

The scenario with Alice and Bob could continue and raise more and more questions. What if Bob wants to name something differently? What happens if they have their house redecorated? What if one of the granddaughters moves in for two years and the house has to be reorganised? One central issue for the design of such a service robot is thus, how to build an environment representation that can cope with all those different particular scenarios, while being useful for the robot and understandable for its users. This thesis subsumes those questions as the guiding research questions for the development of the conceptual framework of "Human Augmented Mapping" (HAM) that is a central term for the considerations and investigations contributing to the presented work. In the following the background ideas behind "Human Augmented Mapping" are explained to give an answer to the questions mentioned above.

Robotic mapping and human concepts

Many approaches in current robotics research focus on the problem of simultaneous localisation and mapping (SLAM) (e.g., Folkesson, 2005; Thrun, 2002; Thrun *et al.*, 2005), enabling a mobile robot to move in an initially unknown environment while creating and updating a map of the area concurrently. Those maps are often based on the extraction of geometric features, e.g., lines or the alignment of raw sensory data, e.g., scan matching (Lu and Milios, 1994). The "funny drawing"(figure 1.1) Alice discovers in the robot manual corresponds to a visualisation of a geometric feature based map.

In both cases the resulting map provides the robot with the ability to localise itself in terms of geometrically defined positions, or coordinates. Such coordinates though do not necessarily correspond to the conceptual understanding a human user might have in mind when referring to a given environment. These references are needed to instruct and control the robot, when it is given a service request.

A topological description seems to be more appropriate for the representation of an environment that is understandable for humans (e.g., McNamara, 1986; Kuipers, 1982). Since on the other hand the robot needs to navigate in a sufficiently precise manner to provide its services, a pure topological map seems not entirely adequate either. Existing approaches to hybrid mapping systems (Thrun and Bücken, 1996) suggest the use of map hierarchies with geometric representations of spatial entities, e.g., rooms, on the lower level, and a topological representation on top of the geometry. Yet, such a representation does not necessarily establish a link to human concepts, terms and semantics that are needed for the communication presumably taking place between a service robot and its user(s).

Furthermore individual users are likely to have personal preferences and terms for entities in the environment that they would want the robot to know about. In the previously described scenario Bob could discover that Alice completely forgot to tell the robot about "his" room – that one where he prefers to sit and read.

Thus, the robot would need to update its knowledge and learn about the "reading room", even if Alice calls it the "yellow room" because of the wallpaper.

Summarising a joint representation for the environment is needed that is understandable and usable for both, the robot and the user. The representation must be designed to incorporate individual information depending on the preferences of particular users.

Acquiring information

The information to be incorporated in a map for the robot is highly depending on the user and the particular purpose the robot has. The environment for the scenario has to be assumed unknown, not only to the robot, but also for a hypothetical developer of the service robot as "product".

A very effective way to introduce humans to a previously unknown indoor environment is to show them around and explain certain items to them *while those can be perceived visually*. For the service robot scenario this seems to be an appropriate way of acquiring the necessary information as well.

In a "guided tour" or "home tour" the user can take the robot from room to room, while labelling names to rooms, specific places, and objects. While the user explains the environment, the robot can build a map representation that incorporates the given information together with the spatial information that can be extracted from sensor readings *in situ*. This corresponds more or less to what humans do to instruct somebody working in their house – just as Alice in the scenario refers to a housekeeper.

Another issue to consider is the fact that a service robot might not be assigned to one single person, but to several individuals that might have different opinions on which services the robot should provide for them and consequently what it needs to know about the environment. This can lead to ambiguities for the robotic mapping and the labelling process.

Ambiguities

The working environment of the assumed service robot is dynamic, initially unknown, and possibly customised as well as ambiguities might be evolving from the layout or the simple fact that spatial entities of the same type are named differently by different users. In those cases it is helpful if not necessary to offer more information to the robotic system than this can be expected to obtain itself from sensory data and previously given descriptions. Confirmation or rejection of generated hypotheses can also be part of such disambiguating information.

Alice wonders what she can do to *help the robot*, and in fact interaction can be used to disambiguate a given situation and to help the robotic system to overcome certain limitations of mapping methods used for the representation of the environment. The information obtained from the user *augments* the data that are perceivable for the robot with the help of its sensors.

Interaction and mapping: Human Augmented Mapping

The term "Human Augmented Mapping" (HAM) describes the conceptual framework suggested in this thesis to deal with the mentioned difficulties in a mapping process for a service robot. A robotic mapping approach is augmented interactively with the information a user can provide about the environment. This augmentation facilitates disambiguation and allows for a map representation that includes both human centric and robot centric information in the sense that it is understandable and useful for both.

HAM as it is used for this thesis is not a concept for a new autonomous robotic mapping approach, but it is a concept for the integration of robotic mapping with human robot interaction. The idea of this thesis is to provide a design that can integrate different environment representations in a hierarchical framework, facilitating task completion as well as meaningful communication about the obtained representation. The power of such an augmented mapping process lies in the possibility to integrate information that can be given interactively in an on-line fashion by arbitrary users, including those who never have interacted with a robot before.

The ideas, concepts and implementation are discussed before the background of the "home tour" scenario, an assumed interactive guided tour with a service robot through an arbitrary indoor environment, which served also as a demonstrator experiment of the COGNIRON project.

Considering all mentioned aspects of the integration of robotic mapping approaches and human-robot interaction it becomes obvious that it is possible to tackle evolving issues in at least two ways. One is to ignore the overall context and concentrate only on one single area, e.g., the interaction and dialogue components for a HAM-system. The other way is to consider the complete framework and pick several issues to investigate, accepting some drawbacks in the possible depth of investigation for each of them, but keeping in mind the idea of an overall integrating framework.

The present thesis discusses HAM in the latter way, by giving a suggestion for an integrating framework and presenting results in different fields of robotics and human-robot interaction research. Yet, all those areas have in common that they contribute to the overall, integrating concept.

1.2 Steps toward Human Augmented Mapping

The previous sections describe a motivating scenario and the background for the doctoral thesis project on "Human Augmented Mapping". A previously presented licentiate thesis (Topp, 2006) dealt with the initial steps toward a system for Human Augmented Mapping. These steps were the design of a conceptual framework for an overall system, the implementation of a subsystem with a focus on person tracking and following that facilitates joint exploration of an environment for robot and user

and the design and carrying out of a pilot study to find out about how users would actually present an environment to a robot. The present doctoral thesis extends this previous work by proposing an environment model and a mapping subsystem that implements this model, focusing on the representation of *regions* (rooms) and the detection of transitions between them. Further, two additional user studies were conducted using the initially implemented system to learn more about the way users present an indoor environment to a robot, also in comparison to guiding around humans. This thesis has thus to be considered a fundamentally revised extension of the previously published licentiate thesis, as is outlined in the following.

1.3 Contributions

The main contributions of this thesis can be summarised as

- A conceptual design for Human Augmented Mapping, an interactive approach to robotic mapping that considers aspects from autonomous robotic mapping and human-robot interaction;
- A model for a flexible high level environment representation suitable to incorporate individual information and applicable in an interactive context;
- Results from an experimental system implementation concentrating on a) the subsystem for tracking and following a user, and b) on the mapping subsystem with a focus on representing *regions* and detecting ambiguities in the spatial layout; and
- Results from three user studies that were conducted to support assumptions made in the design of the environment model and to investigate the applicability of the implemented system in an actually interactive context.

The concept and results have been presented at international conferences and have been published or are submitted for review as follows.

- Elin A. Topp and Henrik I. Christensen, "Tracking for Following and Passing Persons", *in Proc. of IROS 2005* (Topp and Christensen, 2005)
- Elin.A. Topp, Helge Hüttenrauch, Henrik I. Christensen, and Kerstin Severinson Eklundh, "Acquiring a Shared Environment Representation", *extended abstract and poster. In proc. of HRI 2006* (Topp *et al.*, 2006a)
- Elin A. Topp, "Initial Steps Toward Human Augmented Mapping", Licentiate Thesis, 2006 (Topp, 2006)
- Elin A. Topp, Helge Hüttenrauch, Henrik I. Christensen, and Kerstin Severinson Eklundh, "Bringing Together Human and Robotic Environment Representations - A Pilot Study", *in Proc. of IROS 2006* (Topp *et al.*, 2006b)

- Elin A. Topp and Henrik I. Christensen, "Topological Modelling for Human Augmented Mapping", *in Proc. of IROS 2006* (Topp and Christensen, 2006)
- Elin A. Topp and Henrik I. Christensen, "Detecting Structural Ambiguities and Transitions during a Guided Tour", *in Proc. of ICRA 2008* (Topp and Christensen, 2008)
- Elin A. Topp and Henrik I. Christensen, article on general HAM framework with focus on detection of region transitions (chapters 4 and 5, section 5.3), submitted for review to "IEEE Transactions on Robotics"
- Helge Hüttenrauch, Elin A. Topp and Kerstin Severinson Eklundh, article on the Multiple Room Study in people's homes (chapter 6, section 6.3), in preparation for submission for review to Special Issue of "Interaction Studies" on "Robots in the Wild: Exploring Human-Robot Interaction in Naturalistic Environments"
- Elin A. Topp and members of the Applied Computer Science group, University of Bielefeld; article on the integration of the mapping subsystem in a different framework (chapter 5, section 5.4), in preparation for submission for review to the International Conference on Robotics and Automation (ICRA) 2009

Other related publications:

- Anders Green, Helge Hüttenrauch and Elin A. Topp, "Measuring Up as an Intelligent Robot: On the Use of High-Fidelity Simulations for Human-Robot Interaction Research" *in Proc. of PerMIS 2006* (Green *et al.*, 2006a)
- Helge Hüttenrauch, Kerstin Severinson Eklundh, Anders Green, Elin A. Topp, and Henrik I. Christensen, "What's in the gap? Interaction transitions that make HRI work", *in Proc. of RoMan 2006* (Hüttenrauch *et al.*, 2006b)
- Helge Hüttenrauch, Kerstin Severinson Eklundh, Anders Green, Elin A. Topp, "Investigating Spatial Relationships in Human-Robot Interaction", *in Proc. of IROS 2006* (Hüttenrauch *et al.*, 2006a)
- Anders Green, Helge Hüttenrauch, Elin A. Topp, and Kerstin Severinson Eklundh, "Developing a Contextualized Multimodal Corpus for Human-Robot Interaction", LREC 2006 (Green *et al.*, 2006b)
- Thorsten Spexard, Shuyin Li, Britta Wrede, Marc Hanheide, Elin A. Topp, Helge Hüttenrauch, "Interaction Awareness for Joint Environment Exploration", *in Proc. of RoMan 2007* (Spexard *et al.*, 2007)
- Zoran Zivkovic, Olaf Booij, Ben Kröse, Elin A. Topp, and Henrik I. Christensen, "From Sensors to Human Spatial Concepts: An Annotated Data Set", *IEEE Transactions on Robotics, 2008* (Zivkovic *et al.*, 2008)

Other (not peer-reviewed) reports that relate to the work for the thesis:

- Elin A. Topp, "Evaluation of a multiple target tracking approach for following and passing persons", Technical Report (Topp, 2005)
- Elin A. Topp and Helge Hüttenrauch, "Human Augmented Mapping - a pilot study", Technical Report (Topp and Hüttenrauch, 2006)
- Elin A. Topp, "Design study: A control system for a mobile service robot", Course report for course (4F5109): Design of Embedded Real Time Control Systems, KTH, Department of Machine design, Winter term 2004/2005, unpublished

1.4 Organisation of the thesis

The organisation of the thesis follows rather closely the steps in developing and investigating the concept of HAM from the conceptual design of the framework and the proposal of a suitable generic environment model that forms the link from mapping to interaction over the implementation to carrying out user studies with the system. However, it has to be noted that the order of the thesis chapters does not reflect the chronological order in which the work was conducted, as different investigations and developments were continuously informing and contributing to each other, switching between model development, technical implementation and evaluation, and investigations regarding applicability and usefulness of these developments in user studies.

Chapter 2 - Maps and Interaction

Chapter 2 gives an overview of existing work in the research areas contributing to the concept of HAM and approaches that can be related immediately to the work presented in this thesis. A large part of the technological background has its roots in the field of robotic mapping, which involves both geometric, topological, and hybrid approaches, while the design of the proposed environment representation has its background in psychology and cognitive science. Furthermore an overview of relevant work in human-robot interaction is given. Since the goal of the approach to the environment modelling can be compared to the problem of establishing common ground in communication, also a short overview to research in the respective area is presented.

Chapter 3 - Human Augmented Mapping

In chapter 3 the author's concept for Human Augmented Mapping (HAM) is presented, including the requirements and situations that can be faced in the assumed "guided tour" scenario. The main aspects to be discussed in the following chapters are characterised together with some central terms used in the respective context.

The ideas are presented in a high level, conceptual architecture design with interacting modules for mapping, navigation and interaction.

Chapter 4 - Hierarchical environment representation

Chapter 4 explains the hierarchical approach to the representation of domestic or office-like indoor environments. The hierarchy establishes the link between the robotic mapping system and the cognitively inspired representations understandable for the user. Two of three user studies focused on establishing and confirming assumptions about the proposed model and its use in an interactive context. Another central question for the representation of an arbitrary indoor environment was to find a suitable way of segmenting the given space into the spatial entities used in the hierarchical model. One possible approach to region segmentation and transition detection is described also in this chapter.

Chapter 5 - Empirical studies

A design and an implementation for an initial HAM-system are discussed. The system implements the topological graph structure proposed in the previous chapter and integrates this into the suggested architectural framework. One central component of the system is the tracking and following mechanism, and even more important is the mapping subsystem. Both are in the focus of the evaluation described in chapter 5, which concentrates on the components implemented for tracking, the segmentation of *regions* and the detection of transitions between *regions*. Furthermore the integration of the mapping subsystem into a fully integrated interactive framework is described.

Chapter 6 - User studies

Three user studies were conducted to validate the assumptions about the environment representation and the scenario of the "guided tour" used for the Human Augmented Mapping system. They are reported in chapter 6. The first (pilot) study explored how humans guide a robot around in a familiar environment. A more comprehensive follow up study with slightly changed conditions investigated, in how far the hierarchical environment model proposed in chapter 4 is represented in the way humans show particular (spatial) entities to the robot. The third study investigated to which extent people tend to compare a mobile robot to a human or an individual in the context of the "home tour" or "guided tour" scenario.

Chapter 7 - Summary and concluding discussion

Chapter 7 summarises the main aspects of the thesis and presents the conclusions that can be drawn from the results achieved with the presented work. Furthermore, some future ideas and open issues are outlined.

Chapter 2

Maps and Interaction

The idea of Human Augmented Mapping (HAM) assumes the integration of two main aspects, one dealing with robotic mapping and the other one handling the interaction with the user, that are combined by a shared conceptual environment model. This model is assumed to be applied in each of the subsystems to establish an appropriate link between them. Consequently, Human Augmented Mapping relies on results and ideas from different research areas including cognitive science, cognitively inspired robotic systems, robotic mapping, human-robot interaction, and (spatial) language and communication. In this chapter those different research areas and their findings will be discussed with respect to their relation to Human Augmented Mapping.

Various approaches to robotic mapping are based on findings from psychological studies and cognitive science. Nevertheless those approaches mostly aspire to model space and the process of map acquisition and exploration of the surroundings according to what can be observed and assumed in human behaviour for autonomous robotic mapping. Since the cognitively inspired modelling of the link (the environment model) between robotic mapping and interaction is a central point for this thesis, the respective area of *cognitive models and space representation* will be the first one to be discussed in section 2.1.

One of the two main aspects in the concept of HAM is the robotic mapping, which in itself subsumes an abundance of approaches. Given the findings from psychology and cognitive science, approaches to *hybrid mapping*, thus those, that combine high-level topological representations with low-level, often geometric, feature based local maps, relate rather closely to these findings and are consequently quite central for the concept of HAM. Thus, a number of relevant contributions to hybrid solutions of the robotic mapping problem, i.e., simultaneous localisation and mapping (SLAM), are presented in section 2.2. A fundamental question is then how to build a topological layer that can establish the link between geometry and the higher level conceptual ("human") model. A number of approaches to pure *topological mapping and space segmentation* are explained to motivate the ideas for

the framework and system presented in this thesis in section 2.3.

The second main aspect of the concept of HAM is the communication and interaction with the user. HAM assumes an interaction subsystem that applies the higher-level linking environment model in a *language and communication* framework, to establish a basis for communication about the environment with the user. This can be related to the idea of establishing common ground in communication. Additionally a lot of work has been done in the resolution of deictic (spatial) references and the use of spatial language in the field of human-robot interaction. Those language related aspects will be discussed more thoroughly in section 2.4.

The interaction aspect of HAM does of course not only rely on (verbal) communication, but also more general issues in *human-robot interaction* have influenced the development of the idea and concept. Particularly the user studies conducted within the work this thesis is based on investigate a number of aspects in this broad field of research. Section 2.5 discusses a number of findings in the field of human-robot interaction, focusing also on its technical aspects relevant to HAM and the idea of the "home tour" scenario, i.e., the tracking of humans.

Despite the fact that the term "Human Augmented Mapping" (to the best knowledge of the author) was mentioned in the field of robotics or human-robot interaction initially by the author of this thesis (Topp and Christensen, 2005), there are other approaches to interactive mapping or "semantic mapping", partly referencing the author's work. Earlier approaches to interactive mapping are discussed in the context of the segmentation of space they provide in section 2.3 or in relation to the aspects in language and communication that are investigated in section 2.4 but one particular, *strongly related approach* is discussed separately in section 2.6.

2.1 Cognitive models and space representations

The representation of space and the development of spatial memory have been areas of interest in psychology and cognitive science for a long time and are still investigated in the context of neuropsychology (McNamara and Shelton, 2003, as an example). Parallels to the animal world are drawn and investigated to find out about exploration and path finding strategies in mammals (Wang and Spelke, 2002). General approaches to cognitive models for learning processes such as ACT-R[1] (Anderson and Lebiere, 1998) include also models for spatial reasoning. Direct use of ACT-R for the modelling of learning processes in robots has been reported recently, e.g., by Trafton *et al.* (2006). Given the abundance of literature in the area it is close to impossible to give more than an overview of the work and publications relevant to the ideas and concepts underlying this thesis. An overview of cognitive models for space representations and robotic mapping is given by Bakker *et al.* (2005).

[1] ACT-R stands for "Adaptive Control of Thought–Rational", "a cognitive architecture: a theory for simulating and understanding human cognition" (http://act-r.psy.cmu.edu/, URL verified August 19, 2008)

The cognitive map

Tolman (1948) defined the *cognitive map* after an experiment with rats that allowed him to conclude that there obviously exists a cognitive representation of an explored area, allowing the animals to find their way to a specific position from arbitrary points in the (observed) environment. Still, this finding does not allow to draw conclusions on how and to what extent spatial information is represented in humans and other mammals. Various, somewhat controversial, theories on spatial knowledge acquisition and representation have been proposed and discussed.

Specifically the *Map in the Head* Metaphor – indicating that spatial knowledge is exclusively represented isomorphic to a (graphical) map – has been declared obsolete by Kuipers (1982). He suggests a multi-modal and multidimensional representation, encoding different types of knowledge depending on the task to fulfil. Those dimensions for the representation of space have been described by McNamara (1986) as

- format,
- functionality,
- structure, and
- content.

The *format* of the cognitive map is according to discussed theories either *analogue* or *propositional*, or a hybrid of both. An analogue representation would capture spatial knowledge in a form of image isomorphic to the "real world". Propositional models assume representations that refer to the associations going along with spatial entities or objects. McNamara argues that analogue representation is helpful to encode configurations of objects, but does not allow for representation of semantic or logical knowledge. The latter is better encoded in a propositional, more abstract format.

This distinction forms already the second dimension: The *functionality*. Depending on the format of the information to be encoded, the format of the encoding representation adapts.

In the third dimension the *structure* of the cognitive map is described. The structure can be either *hierarchical* or *flat*, where McNamara even suggests a distinction between *strongly hierarchical* and *partially hierarchical*. A strongly hierarchical model would suggest that an environment is represented in a graph structure of *regions* where spatial relations can only be resolved by traversing the hierarchy in which information is encoded. For example, the (spatial) relation between the TV in the living room and the bathtub in the bathroom can only be estimated by knowing the positions of both objects within the respective rooms and the spatial relation between the rooms. With partially hierarchical encoding it is possible to establish direct links between entities on one level in such a hierarchy. A completely flat encoding would correspond to the exclusively analogue representation of space.

The fourth dimension McNamara refers to is the *content*. He gives the example of spatial distances. In a mental representation with a content of *encoded* information the distances between objects would be stored explicitly. The other possibility is *procedural knowledge*, describing *how to estimate the distances between objects*.

Anderson and Lebiere (1998) suggest that knowledge is encoded in two ways, as *production rules* and *chunks*, which is modelled in Anderson's cognitive architecture ACT-R accordingly. The two types of representation differentiate in fact between declarative (or explicitly encoded) and procedural knowledge. A chunk thus describes an information "entry" as a set of slots with associated values. A production rule describes with a set of conditions and tests the cognitive process of obtaining new information from given chunks. In a learning process more and more information is then encoded explicitly in chunks. Thus, a hybrid form of content in a mental representation of space can be assumed as well. In general it turned out that the seemingly controversial theories on space representations rather complement than exclude each other.

Starting from those theories, McNamara conducted a study on the representation of spatial relations (McNamara, 1986), which allowed him to confirm a *partially hierarchical model* for the representation of *spatial* information. He found that overall subjects could remember and estimate the relations between objects in delimited regions better, when those objects were in the same region than in different ones, confirming the assumptions for a hierarchy. Still, his subjects managed to handle close spatial relations *across region borders* better than distant ones *within one region*, which confirms the assumption of a partially hierarchical representation. These findings are the basis for the environment model proposed in this thesis.

Exploration and the use of maps

In his "Spatial Semantic Hierarchy" (SSH) Kuipers and Byun (1988, 1991); Kuipers (2000) captures various aspects of the different theories for space representations by proposing a hierarchical structure for the modelling of environments for robots which encodes both explicit spatial information and procedural knowledge (e.g.,"how to travel from A to B" and "how to explore the environment"). He uses five layers – metrical, topological, causal, control, and sensory – to describe the distinct representations he considers necessary for a complete model. Further he distinguishes between qualitative and quantitative representations across the hierarchy. With the *sensory* level the perception of the surroundings by robotic sensors is modelled and the *control* layer describes the control laws for exploration and navigation. The *causal* layer allows to incorporate the abstraction of the continuous world in terms of (sensory) views and actions and their causal relationships. In the topological representation Kuipers incorporates *paths*, *places* and *regions*, which in itself forms a hierarchical structure for the environment modelling. Such a hierarchical structure formed by the introduction of *regions* is used as well in the models presented in this thesis. Kuipers refers to the *metrical* level as a global metric map which "may be helpful, but seldom essential" (Kuipers, 2000, p195). A number of

publications refer to these suggestions when a hybrid, hierarchical (topological and metrical) model for the environment is proposed for robotic mapping.

Similar in terms of using hierarchies for the modelling of strategies, but different in terms of the representations of navigation behaviours is the approach of "Route Graphs" suggested by Werner *et al.* (2000) after previous investigations by Krieg-Brückner *et al.* (1998). The idea of route graphs is based on observations of the navigational behaviours of different animals and insects, but does also include human strategies. The authors refer to the fact that different behaviours for navigation are triggered depending on the perception of previously learnt route marks. Building on this, a robotic system (as the observed animals) learns directed routes for the navigation from A to B which are recalled depending on the observations made in a given situation. Since the work this thesis is based on deals mostly with the representation of (global) overview knowledge and not so much with strategies for way-finding, a more content oriented approach to the representation of environments is investigated.

2.2 Hybrid mapping

One issue with most existing approaches to simultaneous localisation and mapping (SLAM) in robotics research is the computational complexity to be handled. A significant number of methods are based on probabilistic or statistic filter implementations which can require the computation of large covariance matrices to describe the uncertainty of the current robot position related to observed features (Castellanos *et al.*, 1999, as an example). These are growing with the number of map features ("landmarks") used for the continuous localisation. Thus, often the number of features is limiting the process in terms of the size of the environment that can be handled. The same problem occurs for other types of metric maps, e.g., grid maps that model the raw sensory data in occupancy grids (Thrun and Bücken, 1996), or maps obtained by scan matching (Lu and Milios, 1994). Hence, besides improvements within the methods or choice of alternative approaches, a possible solution to the problem is to split the built map into several local sub-maps and to link them in a global, topological framework. Some general reflections on the use of hybrid mapping approaches are given by Buschka and Saffiotti (2004).

Hybrid maps for autonomous systems

One implementation that built topological maps on top of in this case grid based metric maps was presented by Thrun and Bücken (1996). They mention the orthogonality of the advantages and disadvantages of metric and topological maps and suggest the integration of both as an approach that overrides the respective disadvantages. Their idea for the separation of the regions in the grid map that form the nodes of the topological graph is to use what they call *critical lines*. A critical line is specified by a narrow passage, e.g., a doorway and computed with

the help of the Voronoi graph[2] of the free space defined by the grid maps. In the respective article the work is done quasi off-line as far as the actual integration of the maps is concerned. The topological graph is superimposed on an already existing metric grid map. Still, the article suggested criteria for the consistency and correctness of the integration that have influenced later attempts to the generation of hybrid maps.

The more recently proposed *Atlas* framework is such a hybrid framework (Bosse *et al.*, 2004). Local maps obtained with existing, arbitrary geometric mapping methods are linked in a topological graph. As delimitation criterion the complexity of a local map is used. This means that neither information on the spatial extent of the local maps nor their relation to human spatial concepts and understanding can be extracted. The framework itself emphasises the uncertainty propagation necessary for localisation with respect to adjacent or even distant maps.

Both articles refer to work related to the SSH by Kuipers (Kuipers and Byun, 1988, 1991, 1990; Kuipers, 2000), who also proposed an extension to the SSH that integrates local metric and global topological maps into one map representation (Kuipers *et al.*, 2004). The topological representation in this case consists of places, paths and regions, where a *place* is a point location on one or more paths. It describes how travel actions and turns link distinctive states that are assigned to places. The SSH itself provides control laws for the exploration of an environment that are based on trajectory following and hill-climbing[3]. With the hybrid approach these control laws for localisation are replaced by a metric localisation in *Local Perceptual Maps*. In this case thus the local maps do not represent an absolutely specified and delimited region with a spatial extent, they are limited by the perceptual capabilities of the used robotic (sensory) system. The local perceptual map for a place describes the associated paths and thus states of the place. The SSH and this hybrid mapping approach relate closely to the cognitive representation of space as such in humans, but do not reflect the human concepts that are used to describe space in terms of functionality and semantics of specific locations (see also section 2.1), and are thus not entirely reflecting the idea of establishing a link that facilitates communication with a human user.

Another, explicitly hybrid approach was proposed by Chong and Kleeman (1997). Their method uses a metric, sonar data based, SLAM method and triggers the switch to a new local map whenever the positioning error of the system exceeds a certain threshold. Again, in this case the focus was on the reduction of complexity for the mapping of large scale spaces and not on spatial modelling related to the conceptual models of space in humans.

Tomatis *et al.* (2003) propose a framework, where features describing a certain area are explicitly grouped and linked in a topological graph. They focus on a

[2]The Voronoi graph of the free space is the set of all points that are equidistant to two or more obstacles (Russel and Norvig, 2003, p922)

[3]Hill-climbing is a local search algorithm that optimises the current system's state by continually moving in the direction of increasing value, thus, "uphill" (Russel and Norvig, 2003, pp110–112)

strategy on when to switch between topological and metric localisation. They assume an environment model consisting of places (described by metric maps) and locations (topological nodes) that are connected by travel paths.

The focus of the previously mentioned hybrid approaches is autonomous mapping, in some cases also autonomous exploration. This does not require underlying models with direct links to human semantics and concepts. Other approaches, not so much focusing on the aspects of complexity reduction, concentrate on these links.

Semantics and concepts in hybrid maps

Recent work, more closely related to the concepts and semantics used by humans is a multi-hierarchical approach presented by Galindo *et al.* (2005). They use two orthogonal hierarchies to describe an arbitrary indoor environment. One hierarchy models the space and the other one the concepts. The spatial hierarchy is build on local, metric grid maps that are linked in a (metric) graph. Those maps are assumed to be acquired previously. In a conceptual hierarchy that classifies rooms and objects (with entities like "bedroom", "TV", etc.) a semantic localisation ability is achieved. The link from concepts to spatial representation is established with anchoring.

The overall framework discussed in this thesis builds on a hierarchical, hybrid approach to represent the environment. In contrast to a number of the mentioned approaches though the focus is on representing an arbitrary indoor environment with respect to the concepts used by humans for spatial references and communication. The cognitively motivated concepts of, for example, the SSH (Kuipers, 2000, as the main article describing this idea), are in fact exceeding the assumed requirements for the purpose of communication, as the SSH aspires to model a robotic system according to human models to investigate methods for autonomous exploration. Conceptual/semantic hierarchies as used by Galindo *et al.* on the other hand come much closer to the ideas for the environment modelling in HAM. Nevertheless, their system relies on previously acquired information for the local maps and the semantic links from *room* entities to *object* entities. Those links precode the assumption that it is likely to find, e.g, a "TV" in a room called "living room".

The categorisation of regions (rooms) with the help of objects that are typical for a given environment is suggested by Vasudevan *et al.* (2007), who do not only consider the categorisation but also use the constellation of objects to specify particular places, and use this information for localisation. An approach for object based localisation, thus considering the high-level information obtainable from objects, was also recently presented by Gálvez López *et al.* (2008). HAM allows for a more open setting as a starting point, being independent from precoded knowledge or the *a priori* categorisation of regions in this respect. Hence, a method for the autonomous generation of hypotheses for delimited regions has to be provided, which can also be used for interactive specification. Similar approaches are provided by methods for topological mapping and space segmentation.

2.3 Topological mapping and space segmentation

The interesting issues of topological mapping can be distinguished in the control laws needed to travel in an obtained map and the segmentation of the space to be represented. The general idea is to represent the environment in a graph structure with nodes and paths. The nodes might represent concepts as rooms or other delimited regions or they might also be significant, distinguished positions. Existing approaches often concentrate on one of those issues, but the delimitation is floating, since the map always needs both, procedural knowledge for traversing the graph and a segmentation procedure defining the nodes. Topological maps can be generated with different levels of autonomy in the process. Obviously, the more autonomously the system works, the less it is likely to represent the environment in a way that corresponds to human concepts and semantics for communication, if not the strategy used for the space segmentation corresponds closely to the one a human uses.

Topological maps

Nourbakhsh *et al.* (1995) proposed with DERVISH an approach to the exclusive use of a topological map for navigation in an indoor environment. In their case, though, the map was precoded down to the level of measurements for width and depth of door frames or other openings in the corridor to be traversed. They suggested to place the nodes of the topological graph at corridor junctions and door openings, and had implemented procedural knowledge for how to move in certain situations to get out of one room, travel along the corridor, re-plan the path in case of a blocked way and enter a goal room.

Choset and Nagatani (2001) proposed to use the Voronoi graph of a traversed environment to define similar nodes, building upon the suggestions of Kuipers and Byun (1988). Later also this approach was extended with an explicit hierarchical model (Lisien *et al.*, 2005). All these approaches assume a fully autonomous process, in which no semantics or concepts are involved in terms of communication with humans. Beeson *et al.* (2005) extend the idea of using the Voronoi graph for the autonomous learning of places. Still no relation to human verbalised concepts is given.

Althaus and Christensen (2003) suggested an approach to interactive mapping in which a user could take a robot on a tour and present an indoor environment. This was done with the help of a rather simple but effective tracking and following behaviour for the robot, making it approach the closest object in front of it. Their graph assumed places (rooms) and gateways, where the gateways were doors. The places were associated with activities which made the system switch to respective control for, e.g., navigation in interactive contexts. A clear disadvantage for a naïve user though was that the gateways had to be specified explicitly as a position in a geometric map, when the robot was placed on this position. In the user studies referred to in chapter 6 such a strategy for teaching a mobile robot the gateway from

one area to another was in fact observed once, but in this single case the observed "user" was definitely not naïve, since it was a robotics researcher working with robotic mapping and SLAM. Thus, it can safely be assumed, that potential users who are not particularly experienced with the internals of robots would not apply this as an intuitive strategy. Additionally, the system of Althaus and Christensen did not try to model the observable area according to human concepts other than the gateways.

The segmentation of space

In general the segmentation of space into nodes of a topological graph structure can be obtained by either using appearance based approaches (mostly relying on image data) or by taking the spatial (geometrical) properties or the layout into account for the description of a node. Some methods are actually more focused on the categorisation of those nodes (e.g., into categories like office, kitchen, living room, or corridor) or even the recognition of particular places, but since those categorisations can also be used for the segmentation of the environment they are mentioned in this context as well.

Appearance based approaches

With a biologically inspired approach to topological mapping Tapus *et al.* (2004a,b) introduce the use of *fingerprints* in this context. A fingerprint is a concise string description of the perception of the surroundings. They use images obtained with an omni-directional camera and encode observed colours in the environment according to their angular order in the image. Additional features are vertical lines, corners derived from laser range data and a code for "nothing". Such a string thus describes the appearance of a certain area and changes whenever the respective robot moves into an area with significantly different appearance. Whenever that happens a new node for a topological map is generated. This segmentation of the space is very efficient for describing places in terms of topological localisation, but it does not describe the spatial extent of the surroundings at all. A topological layer for a system approach to Human Augmented Mapping would need to do this in order to describe certain regions as "containers" for specific objects or locations.

More recently two purely image analysis based approaches to the representation and recognition of particular regions or rooms have been presented. One is a SVM[4]-based approach for place recognition, using large data bases of images acquired at the relevant places over certain periods of time and under different conditions (Pronobis *et al.*, 2008). The other approach relies on image matching (based on SIFT[5]) and clustering, where the clusters represent regions (often corresponding to rooms) as the nodes of a graph structure. This graph is kept sparse to reduce the complexity of path planning operations and localisation (Booij *et al.*, 2006, 2007).

[4] **S**upport **V**ector **M**achine

[5] **S**cale-**i**nvariant **f**eature **t**ransform, (Lowe, 1999)

Approaches based on spatial properties

An interactive approach to obtain a map representation that reflects delimited regions and thus a topology of an environment was suggested by Diosi *et al.* (2005) with a system for interactive SLAM. Also they use a tracking system to allow the respective robot to follow a user through an office environment, where specific locations are interactively defined by the user. Initially the geometric position at which the user gave the information is used as a labelled landmark in the map. In an off-line step the regions, in which those landmarks are to be found, are delimited from the rest of the map with the help of a watershed algorithm[6]. Adjacent regions without landmarks in them are integrated in the existing areas. Compared to the ideas used for HAM this offline approach is limiting, since one of the assumptions is that the presentation of the given environment can happen incrementally, which seems not possible with such a rigid approach. Additionally the initiative for a specification can only be on the user's side and it is not obvious, how ambiguous spatial configurations in the environment can be handled.

Two suggestions to actually capture the *spatial* properties of a given environment were made by Kröse (2000) and later by Martínez Mozos *et al.* (2005), more recently (in Martínez Mozos *et al.*, 2007) also referred to by Zender *et al.* (2007) and Friedman *et al.* (2007), who make use of the categorisation approach. Kröse (2000) proposes to describe convex areas (e.g., rooms) with features derived from a principal component analysis on laser range data obtained in the area. He draws the conclusion that such a method only holds for convex areas. Martínez Mozos *et al.* (2005) use a list of features including geometric descriptions and indices for, e.g., clutter computed from a 360° laser range data set for the classification of specific areas (room, doorway, corridor). In a supervised process the data sets collected by the robot are continuously classified with the respective type-label. The method then uses AdaBoost[7] for learning and classification of new examples.

Friedman *et al.* (2007) describe their Voronoi Random Field approach to segmenting an environment into a topological graph using the categorisation into the concepts room, corridor, junction, and doorway based on the work by Martínez Mozos *et al.* (2007). Each point (according to an occupancy grid map) of the environment is labelled with its category and neighbouring points of the same category form a node in the topological structure in an off-line process. Again, the idea of HAM is to provide a segmentation of an arbitrary indoor environment, not relying on previously learnt concepts, that should be applicable in an on-line manner. However, the previously mentioned approaches integrate the geometric layout of a certain area into a very concise description. Such a description seems useful in the context of the system proposed in this thesis, since it allows to reduce complexity

[6] A watershed approach delimits convex areas from each other with borders that are similar to the critical lines (Thrun and Bücken, 1996). In an indoor environment most of the borders are in fact doorways, but also parts of "L-shaped" rooms can be separated

[7] AdaBoost is a classification technique that uses a number of rather simple classifiers in a cascading structure to deliver better results than one sophisticated classifier could do alone (see also Russel and Norvig (2003, p666ff.)).

in the map representation and thus facilitates an on-line, interactive setting for communication about the surroundings.

2.4 Language and communication

The idea for Human Augmented Mapping is based on robotic mapping approaches as they were discussed previously, and interaction and communication aspects. User and robot need to build a shared understanding of the environment to generate a basis for communication, a shared terminology. This issue is related to the idea of establishing *common ground* (Clark and Brennan, 1991), which is a well investigated area in communication. Common ground for communication of spatial aspects can be achieved, for instance, with the use of graphical maps or drawings (Holsanova, 2005). Nevertheless, such communicative aides can only be helpful when mutual understanding and common sense or at least a conceptual pact (Brennan and Clark, 1996) can be assumed or established. This could be, for example, the use of an "X" in a drawing that marks the location of a (for humans) salient landmark – as a gas station at the junction where to turn left to reach a certain destination. Such abstractions are not possible to use when a robot is involved as one partner in the interaction. The use of landmarks as such to describe a way has been investigated by Kyriakou *et al.* (2005) in a study in which subjects were told to send a toy robot around on a table top "map" that represented an urban scene. They observed how subjects used landmarks – salient buildings and structures – to give directions to the robot and investigated in how far these directions could be applied to a navigation system for a mobile platform.

Tellex and Roy (2006) presented a recent system that interprets directional instructions ("move left", "go across the room") according to the context the respective robot – in this case a mobile platform, but aimed for, e.g., robotic wheelchairs – is in. "Move left" results in their case in a trajectory leading into the first opening available to the left of the current pose of the system. Given a position next to a wall with a wall to the left, the system would move along the wall until sensor readings allowed to conclude that there is an option to turn left. In a situation facing the wall the system would turn left and follow the wall, now on the right. In this case the presented system draws conclusions on the actual intention of the user, considering the options it has at the time it receives a command in the context of its spatial situatedness.

The work of Blisard *et al.* (2006) concentrates on how the use of spatial language can be modelled so that spatial references like "close by", "next to", "under", etc. are possible to interpret with a robotic system.

While most of these approaches to the achievement of "common ground" or mutual understanding do this in a more or less different context, Kruijff *et al.* (2006) do in fact deal with communication for Human Augmented Mapping. They use a model for their system that relates strongly to what is presented in this thesis. The focus, however, is set for this particular work on the dialogue management and

the communication with the user. Their article presents clarification dialogues in the context of ambiguous situations arising during a "guided tour". The system uses a gateway detection for space segmentation which generates a hypothesis for being in a "new room" whenever a door-like passage has been travelled. If the system finds itself back on a previously travelled path after going through such a hypothesised gateway, but without having travelled explicitly "back" through it, a dialogue is invoked to resolve the ambiguity. The scenario assumed by Kruijff *et al.* delivered the motivation for the author to investigate means for the detection of transitions between two relevant areas other than just gateway (door) detection.

The framework for Human Augmented Mapping presented in this thesis assumes a generic environment model that is applied and expressed in both the robotic mapping process and in the conceptual knowledge used in the interpretation of the user's utterances to categorise the information given about the environment in the "guided tour" scenario. Thus, a conceptual framework is provided that aspires to enable user and robot to establish common ground in their discourse about the surroundings, finding a mutual understanding of, e.g., the delimiters of a particular area and reasoning about being "inside" or "outside" this area.

Language and communication might not be in the direct focus of this thesis, but they are aspects of any system that considers the interaction of a human with a service robot. Apart from them also other aspects of human-robot interaction are relevant to HAM.

2.5 Human-Robot Interaction

The field of human-robot interaction as dedicated research area is rather young and still in the process of being established. Nevertheless an abundance of literature deals with this area evolving from Human-Computer Interaction, Cognitive Science, Robotics, Psychology and Artificial Intelligence, to name the most important influences. One already well established area is social interaction with robots. Robotic systems are developed and used in studies to learn about the effects of emotions, personality traits, or behaviour changes displayed by robots in the interaction with humans (e.g., Gockley *et al.*, 2005). Fundamental studies on social aspects of human-robot interaction have been presented throughout the years by, Breazeal *et al.* (Breazeal, 2000, 2002; Brooks and Breazeal, 2006; Thomaz *et al.*, 2006). Other work has not only a focus on the social components but investigates also the cooperation and interaction patterns in embodied human-robot interaction. Since the interaction assumed in the HAM framework has to be seen as embodied, these aspects will be discussed more thoroughly in the following.

As soon as a form of physically expressed interaction can be established, a crucial ability of the respective robotic system is to keep track of the user, or at least certain body parts, e.g., the hands in order to track a pointing gesture or the head in order to detect a nod. Some approaches to tracking systems for the purpose

of motion coordination (following) are presented (see page 24), to give an idea of the technical aspects the interaction in the home tour scenario offers.

Embodied interaction

As soon as an interactive situation occurs in a realistic or real environment, the situation influences and is influenced by this environment. Hüttenrauch and Severinson Eklundh (2002) presented observations from a long term user study with a service robot in an office environment. The robot was used for fetch–and–carry tasks and had one particular user throughout the complete time period the trial was running. Besides issues of the interaction with the user, the authors noticed interesting effects in the interaction with so called bystanders, persons in the office environment that happened to be in the same room or corridor the robot was in and interacted with it. One particular observation revealed the need for a self reflection and appropriate feedback abilities for the robot, when a cleaning cart blocked the robot and vice versa. The cleaner did not know what to do to get the robot out of the way and the robot could only state that the way was blocked, but had no plan for such a situation. Appropriate feedback and navigation functionalities could have resolved the situation.

Althaus *et al.* (2004) presented an approach to navigation in an interactive context, where a robot joins a group of people and adapts its dynamic behaviour to the configuration changes of this group. Starting from such observations and systems, the coordination of user and robot in an interactive scenario is an issue recently investigated in a number of user studies (Green *et al.*, 2006a; Hüttenrauch *et al.*, 2006a,b).

Sidner *et al.* (2005) investigated the role of physical feedback for the engagement of human user in an interaction with a toy size penguin robot. They had the robot display different response behaviours in a short discussion with a user, either acknowledging information with engagement gestures like a nod or just verbally. They concluded that appropriate feedback that is displayed not only verbally but also physically makes it easier for human users to understand the robot's behaviour and thus improves their attitude toward the robot and their willingness to interact. These findings had direct influence on the choice of a behavioural strategy displayed by the robot used for the work presented in this thesis.

Recent work by Wang *et al.* (2006) deals with similar aspects. They investigate the effects of head movements of a robot on the perception of the system. Since their robot has a strongly technical appearance, the movements helped their study participants to interpret the appearance differently and more easily. Also the work of Powers and Kiesler (2006) deals with the mental model (and thus understanding) of a robot that users develop depending on the appearance of the system. An interesting aspect is that the authors used an animated graphical robot face instead of a real robot for their studies. They claim that the use of such a virtual robot does not affect the reactions of the subjects so that it is a valid replacement for a real robot. This might be possible for certain purposes, but this thesis relies on the

assumption that nothing can replace the embodied, *in situ* interaction between a human user and a robot when the communication in and about the environment is to be investigated.

Kirsh (1995) 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 basis for a general robotic environment model which is needed for the concept of HAM. Furthermore, Reeves and Nass (1996, 2003) fount that humans tend to compare computers to social agents and address them accordingly. Again, the strategy proposed in this thesis is to observe a human user interacting with a robot rather than another human, since it does not seem obvious to adopt all findings from human-computer interaction or human-environment interaction for the interaction with robots. One of the studies described in this thesis actually compares observations from human-human and human-robot interaction experiments based on a "guided tour" scenario, giving in fact evidence for differences between human-human and human-robot interaction in this context, still supporting the findings of Reeves and Nass regarding some general interaction strategies.

As mentioned above embodied interaction that includes coordinated movement, particularly when the robot is supposed to follow its user, in an indoor environment requires a robotic system to be able to keep track of a user. A number of approaches to tracking of humans or human body parts are referred to in the following.

Tracking and motion coordination

Tracking the user is crucial for the achievement of natural interaction with a service robot. Various approaches to tracking have been presented, often with different purpose. Face and gaze trackers based on computer vision algorithms are used to monitor the physical or emotional state of, e.g., a car driver to detect signs of fatigue or for the detection of pedestrians in a street scenery (Fletcher *et al.*, 2003). The tracking of limbs (often hands) is used for the recognition of gestures, which can, e.g., accompany verbal deictic references or command a robot in terms of a sign language (Brooks and Breazeal, 2006, as an example). Further, the tracking of a user is necessary for, e.g., activity or action recognition, in this case not only body parts but also the complete body configuration has to be tracked and analysed over a time period (Knoop *et al.*, 2006). For the coordination of movement in an arbitrary environment it is not that crucial to know about the configuration of particular body parts. In this case, as also for the work presented in this thesis, it is important to know the position of the user relative to the robot. Unlike the (most often) computer vision based approaches mentioned above, a pure position tracker can be applied based on rather sparse data, e.g., range data from a laser range finder. The position data can be used to coordinate the motion of the robot with the tracked user (e.g., Prassler *et al.*, 2002).

The work presented in this thesis assumes in fact such an ability as a component to facilitate the interaction with the robot in the environment. Since for the experimental implementation presented in this thesis only the rather coarse coordination of the robot's movement with the user has to be handled, a laser range data based approach seems adequate. Hence, some more detailed information on techniques in this area is given in the following.

2D position tracking

Tracking in general is a sequence of four steps, where steps 2–4 are iterated over a sequence of time steps t starting with t_0.

1. Initialisation/Measurement (in $t = t_0$): The target (in this case a user) is detected and the tracker is initialised with its state, represented by the 2D position relative to the robot.

2. Prediction (in t): According to a motion model assumed for the user a prediction is made about the target's expected state (i.e., position) in time step $t + 1$.

3. Measurement (in $t + 1$): A new round of data is generated and analysed for respective target features, to determine possible positions of the target.

4. Update/Correction (in $t + 1$): The prediction is updated/corrected with the recently gathered information about possible states of the target.

The updated values can then be used as current output for further analysis and at the same time be the basis for a new prediction.

Choosing the filtering algorithm that produces the prediction is a central issue when setting up a tracking system. One method is the Kalman filter, a popular data filtering method used in many areas (see, e.g., Gustafsson (2000) for details). Since the Kalman filter is limited to linear process functions and Gaussian noise models, other Bayesian filters can be more appropriate. Arulampalam *et al.* (2002) give an overview of different approaches. They propose particle filtering as an appropriate technique for tracking. The advantage of particle filters is their flexibility in terms of the process to model. A particle filter models the possible (predicted) states of a target in a set of weighted samples. The weights are updated in the correction step and the set of samples is redistributed based on the updated weights to make a new prediction. Thrun *et al.* (2005) discuss different filter techniques in the context of probabilistic robotics in detail.

Both the initialisation and the measurement step rely on the detection of the target. If the target is marked with special equipment this is rather easy. If the target though does not reveal itself, the available data, in this case laser range data, has to be analysed and target hypotheses have to be generated.

Target detection

In order to detect humans in laser range data very often a pattern matching based approach is used. Depending on the height the respective scanner is mounted in either leg or body patterns are relevant. Both are convex patterns that can be segmented in the data, as suggested by Kluge (2002, for example) and used in previous work (Topp, 2003; Topp *et al.*, 2004) as well as for the work presented in this thesis. Laser range data are sparse and – as used in most cases – one dimensional in the sense that each data sample point only represents a single value. Hence, when searching for particular features or patterns in a laser range data set a rather large number of hypotheses can be generated of which only a small number in fact are correct in terms of representing the type of target looked for. The combination of different methods for the target detection, e.g., laser range data based and computer vision based, has been shown very effective to improve the robustness of proposed tracking methods (Kleinehagenbrock *et al.*, 2002; Topp, 2003). The tracking functionalities for the work described here are important for the complete system approach to work, however, the improvement of these functionalities is not in the focus of the work.

Motion models

Especially in populated indoor environments it has to be assumed that not only one target ("the user") is in the proximity of the robot at a given time. Thus, given a number of hypotheses generated by a detection method, the motion model chosen to represent the movements of the target can be crucial to recover from ambiguous situations. Bennewitz *et al.* (2002) suggest to use Expectation Maximisation to predict directions that possible targets might choose given the spatial context they are in. Similar to that Bruce and Gordon (2004) propose a method to improve tracking with the help of context dependent motion prediction. As for the improvement of the target detection the choice of the motion model is not one of the central issues of the work described here. Still, handling multiple targets is one aspect of the scenario underlying this work.

Multiple target tracking

When multiple targets have to be considered, especially when the number is not known in advance, the tracking of one particular target among other correct "person" hypotheses is an issue. Schulz *et al.* (2001) implemented a multiple target tracker with the help of sample based probabilistic data association, i.e., a particle filter method particularly helpful for multiple hypotheses. This tracking approach was modified in terms of the target detection and the environment model to match the requirements of the concepts for HAM presented in this thesis. The tracking approach will thus be described in chapters 3 and 5 together with the system design for an experimental implementation of a Human Augmented Mapping system.

2.6 A strongly related approach

Recently a growing interest can be noted in considering not only cognitive aspects for robotic (mapping) systems, but also the integration of human-robot interaction into the process. Work in interactive mapping, strongly related to this thesis has been reported by Zender *et al.* (2007)[8] where in a similar context (a "guided tour") a robot and a user explore an environment together. In an earlier publication (Kruijff *et al.*, 2006) the authors adopted the term "Human Augmented Mapping" with a reference to one of the publications relevant for this thesis (Topp and Christensen, 2005). The work reported by Zender *et al.* investigates spatial representations and reasoning about concepts and semantics. In parts, the work presented in this thesis and the work reported by Zender *et al.* are based on directly related ideas (e.g., the region segmentation presented in chapter 4 is based on an idea presented by Martínez Mozos *et al.* (2005) while the work presented by Zender *et al.* integrates a version of that work directly) and share the use of parts of a basic software package ("CURE", see chapter 5). Since the focus of the work presented in this thesis is more on the link between sensory data, their interpretation and human concepts than higher level semantics and reasoning, the two approaches to Human Augmented Mapping complement each other rather than compete.

2.7 Summary

The concepts and ideas presented in this thesis cover a wide range from psychological reflections to control issues for a mobile platform interacting with humans. Hence, this chapter presented a similarly wide range of background information and related publications from cognitive science, psychology, robotic mapping, communication theory, and human robot interaction. Some of them are directly related to the work presented in this thesis in terms of chosen methods, others refer to rather remotely related literature that stimulated the process of creating the framework for the work presented here. All in all, this chapter reflects the broad variety of interesting issues to consider when trying to join a human's and a robot's spatial representations in an interactive robotic system.

[8] within the EU project FP6-004250-IP "CoSy" – "Cognitive systems for cognitive assistants", www.cognitivesystems.org (URL verified July 10, 2008), which was closely related to the COGNIRON project that the work reported in this thesis contributed to

Chapter 3

Human Augmented Mapping

Through the years a lot of effort has been put into autonomous robotic mapping. With the development and improvement of SLAM methods the need for user-provided information seems reduced to a minimum. Exploration strategies have been developed that propose complete autonomy also for initial mapping processes. Models derived from findings in cognitive psychology allow to build robotic systems with human-like strategies for path finding and exploration. Hence, one could wonder why a framework for the integration of human concept information and robotic maps is of any use to robotics (and users, for that matter). This chapter suggests an answer to that question by explaining what Human Augmented Mapping (HAM) aspires to achieve and what the advantages compared to autonomous mapping approaches are – given an appropriate context. Taken out of such a context an interactive mapping approach might not be useful at all, and most likely would cause more problems than it could solve. Thus, the limitations of the framework will be considered as well.

Since the context is needed to understand the idea of Human Augmented Mapping, this will be the first thing to be explained together with advantages and limitations. Following this grounding section the used spatial concepts, possible situations and evolving requirements for an HAM-system are described. Starting from those requirements a schematic architecture that integrated all the necessary parts is sketched and the main aspects described in this thesis are pointed out.

3.1 Context, advantages, and limitations

The framework presented in this thesis is designed for a personal or domestic service robot. It is thus assumed that such a robot is working in close proximity to humans in a populated, but usually not crowded, environment. Additionally it has to be assumed that the environment is dynamic to a certain extent. Furniture might be moved and small "every-day" objects tend to change the appearance frequently by being moved around. As the working environment most likely exists already when a

hypothetical robot is brought into it, it seems not appropriate to require special instrumentation of the environment. Thus, the use of artificial landmarks, e.g., RFID tags[1], that would assist the robot in navigation and conceptual "understanding" is not considered in the work presented here.

Given this situation and an arbitrary indoor environment the idea of Human Augmented Mapping is to enhance the robot's mapping abilities with the information that can be obtained from the user interacting with the robot in a way as natural as possible. As a natural way of communicating information about the environment an interactive guided tour, the "home tour", is assumed as an initial scenario for Human Augmented Mapping. By integrating the user into the mapping process, the resulting map can integrate personal preferences and general "human concepts" that facilitate the communication – and reasoning – about the environment in a way comprehensible for humans. This together with the fact that the information is obtained in a common setting for the human user is a clear advantage for the user.

The acquired representation is assumed to be used and updated also in other scenarios (e.g., a fetch-and-carry scenario). Here again interaction is facilitated with the help of the representation, while updates are facilitated by interaction. Nevertheless most of the possible challenging situations (i.e., ambiguities to resolve) can be integrated in the initial "guided tour" scenario. Thus, this scenario will be the basis for the reflections presented here.

Not surprisingly, also for the robotic mapping a number of advantages compared to traditional, autonomous mapping approaches become obvious. Still assuming a situation where the result of the mapping process is not to be measured in terms of absolute accuracy, but in terms of usability for the service context, the system can take advantage of external, explicit information to resolve ambiguities. Such ambiguities could be a loop closing[2] situation, in which a hypothesis can be confirmed or rejected interactively. False positive or false negative loop closing hypotheses can be reduced in a such a setting. Also in a "kidnapped robot" or "waking up" scenario[3] the uncertainty in the system can be significantly reduced with the help of a clarification dialogue with a human user. Hypotheses generated on the current location and position can be confirmed or rejected interactively. Thus, the

[1] Radio Frequency Identification (RFID) is an automatic identification method, relying on storing and remotely retrieving data using devices called RFID tags or transponders (from Wikipedia, http://en.wikipedia.org/wiki/Rfid, URL verified August 27, 2008).

[2] Loop closing is considered a quality measure for autonomous SLAM approaches. If the system can handle the situation of coming to a previously encountered location on a loop in the environment, i.e., it recognises the location and corrects accumulated errors in the map by aligning map features accordingly, it is considered successful.

[3] With a "kidnapped robot" scenario a situation is described in which a localised robot is lifted from the ground and taken to a different location. This causes erroneous perceptions from wheel encoders – the wheels continue to rotate, but not according to the relocation process. The system has to recognise the fact, that it is no longer localised and has to correct its positioning hypothesis. The latter is also relevant to the "woken up" situation, when the robot is switched on in an arbitrary position and needs to localise with respect to a previously acquired and stored map.

interaction with the user can help to produce a sufficiently consistent environment representation that is usable in a service context.

Human Augmented Mapping (HAM) thus does not aspire to provide a sophisticated approach to robotic mapping, but aims to approach a robotic mapping process from a user's view point, integrating the user into the process of mapping.

The concept requires a number of functionalities and building blocks linked in a conceptual architecture which will be explained in the following.

3.2 Spatial concepts, situations, and requirements

The requirements for a system that aims to work in a framework for interactive mapping, or HAM, mostly arise from the interaction abilities and mapping, i.e., space representation, capacities to be provided. To illustrate the requirements, a number of specified scenarios or situations are proposed that have to be resolved. These situations relate to the information flow in the system which can be user (concept) driven or data (perception) driven, i.e., top down or bottom up. A central link between a user's concepts and the robot's map is an environment model representing the human concepts in robotic terms. Thus, one very central question is, which spatial concepts need to be represented and how this can be done.

Spatial concepts in HAM

The following descriptions refer briefly to the two most central spatial concepts that will be introduced in more detail in chapter 4, as they are reflected in the following discussion of *situations* assumed as central in Human Augmented Mapping.

Location The area from where a large, not as a whole manipulated object is reachable/visible (sofa, fridge, pigeon-holes). Also "the place where the robot is supposed to do something or look for objects".

Region A container for one or several locations. Offers enough space to navigate (rooms, corridors, delimited areas in hallways).

With those two main concepts of *regions* and *locations* a (hierarchical) space representation for indoor environments can be established, as will be explained further in chapter 4. This present chapter concentrates in the following on the situations and requirements, leading to the overall architecture.

Situations, tasks and requirements of a guided tour

As a basic scenario for Human Augmented Mapping a "guided tour" is assumed, in which a user can take his or her robot on a tour around a particular environment and

present this to the robot. Such a scenario is not necessarily limited to permanent initiative from the user's side to present information to the robot. Autonomous interpretation of the surroundings and hypothesis generation can make it possible for the robotic system to ask explicitly for information or help. In the following the presentation of the environment is assumed as limited to *regions* and *locations* with an emphasis on *regions*. The interactive approach to mapping of the HAM-concept allows thus for a number of smaller scenes or *situations* that can occur and *tasks* that the robot has to handle during such a tour.

The following is the attempt to organise possibly occurring situations, limited to the context of the "guided tour" scenario and questions regarding the environment that can arise from ambiguous spatial layouts detected by the respective system. The grouping does not by any means cover the full complexity of possible interactions and situations for a service robot. Figure 3.1 shows the resulting hierarchical grouping of situations considered most important for a HAM-system, that will be explained in detail in the following paragraphs. In general it is assumed, that "explicit information" is given by the user, and "implicit information" is generated by the robotic system, while "tasks" are also given to the robot by the user.

In each of the particular situations the system is gathering new information which can be seen as filling slots. The general description of the entries (*regions* or *locations*) can be summarised as

```
Current region:
Label:            name                   (string)
Description:      region descriptor      (geometric features)
Localisation                             (double measurement,
  confidence:     region confidence      (double measurement,
                                          classification
                                          confidence)
Closest location:
Label:            name                   (string)
Description:      location descriptor    (position relative
                                          to region, pose)
Localisation      metric confidence      (double measurement,
  confidence                              measurement,
                                          metric localisation
                                          confidence)

Overall           f(reg conf, met conf)  (summarised
  confidence                              confidence)
```

Map acquisition/update

The acquisition of an initial map is crucial for the system for navigation purposes. It is assumed to happen at least to a certain extent as a first step of the "home

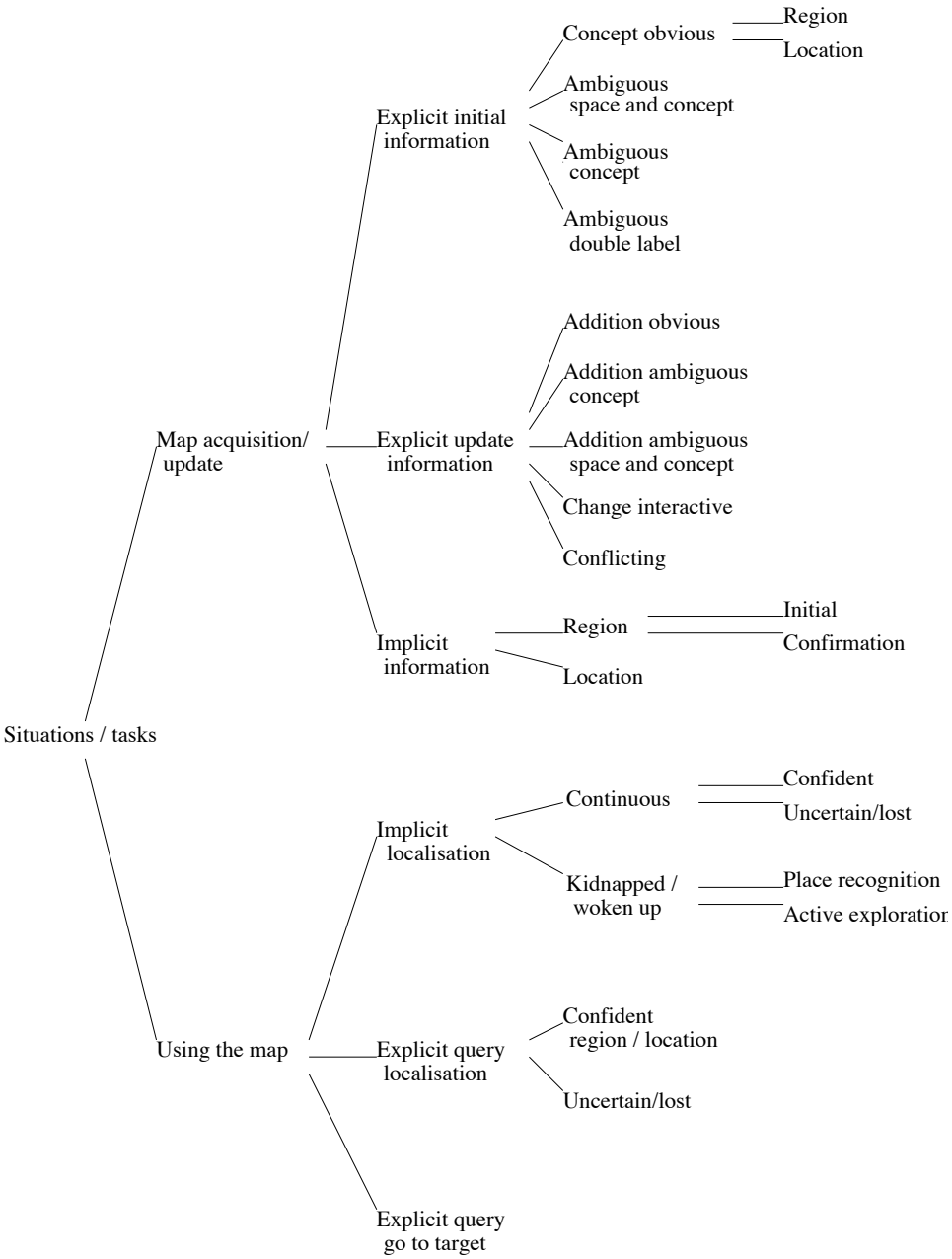


Figure 3.1: *The relevant interaction situations and tasks of a guided tour scenario*

tour". The information acquisition can be triggered in two ways throughout the complete process – explicitly/externally vs. internally/implicitly, or "user driven" vs. "data driven". In both cases the information is supposed to be given by the user, but in the first case this is done with the initiative on the user's side, while in the second case the robot would have to ask for information after an internal triggering event.

The slots of the description summarised above that need to be filled during the complete process are assumed to be

```
     Current region:
     Label:            name                  (string)
     Description:      region descriptor     (geometric features)
                                              confidence)
OR
     Current location:
     Label:            name                  (string)
     Description:      location descriptor   (position relative to
                                              region, pose)
```

In the cases of implicit information gathering (system driven) the slot for the region description is filled with hypothetical information, but the label is missing, and the description might have to be corrected. In the update cases, information can already exist and might have to be overwritten.

Explicit initial information – user driven The "standard case" in an initial "home tour" scenario. The user guides the robot and gives information to the system on *regions* and *locations*. General requirements: User tracking / following ability, dialogue / communication means, perception of the environment / space representation, incorporation of labels for spatial entities (*regions* and *locations*).

Concept obvious For the spatial concepts it is assumed that information given by the user in most cases can be classified with "common sense" knowledge into *region* or *location* information. There might be cases though where this is not obvious. Requirement: Dialogue model / knowledge base with respective categorisations.

- **Region:** The essential information "This is the <region_name>" is given to the robot. <region_name>is known to the dialogue and is thus already known as *region*. Requirement: Space segmentation to represent the labelled spatial entity correctly (delimitation of the presented area from the "rest of the world"), appropriate feedback (show/confirm that area was perceived and stored, not only the spot).
- **Location:** Information "This is the <location_name>" is given. <location_name>is known to the dialogue and is thus already known as *location*.

Requirement: Perception abilities (geometrical situation, images) for the *location* presented, commands to align sensors with information to be given (e.g., near navigation, turn), appropriate feedback (show/confirm that particular entity was perceived and stored).

Ambiguous – space and concept It is known that a *region* is presented, but actually only the "link" (e.g., the door) to it is shown. Requirements: Complete information on the interaction status, information on the previously given information, action interpretation for the presentation, ability to store "links" (gateways) in the space representation, interpretation of the spatial situatedness for hypothesis generation.

This situation is a very special case that is mentioned as a challenging situation for which it would have to be investigated, whether it can be identified with the help of the available sensory input and interpretation tools. One of the studies contributing to this thesis actually aimed to find patterns in people's ways of presenting different spatial entities to generate a basis for understanding the concept to be conveyed.

Ambiguous – concept "This is the <name>" – <name>neither known as *region* nor as *location*. Requirement: Situational knowledge ("Show"-situation Hüttenrauch *et al.* (2006a)), dialogue abilities to disambiguate.

Ambiguous – double label The information given to the system at a certain point is already stored, but for a seemingly (or truly) other *region* or *location*. Requirement: Check for existence of label in graph, dialogue functionality to determine, if the continuous localisation failed or two entries with the same label need to be stored (e.g., two "bathrooms" of the same type).

Figure 3.2 summarises the interaction flow for the previously mentioned situations. One is omitted though: since it is not clear at the current state if and how the very particular situation "Ambiguous – space and concept" can be detected, it remains as a challenging situation to be resolved (see also chapter 6 on this).

Explicit update information – user driven It has to be assumed that the "home tour" might get interrupted, either because the user thinks she has presented everything relevant, or because of an external event, a disturbance in the process. Thus, the information acquisition has to be resumed, which could also happen with a different user or after a change in the environment. There are several situations that can occur with respect to updates or resumed presentation. The simpler ones concern additional information that can thus be considered initial.

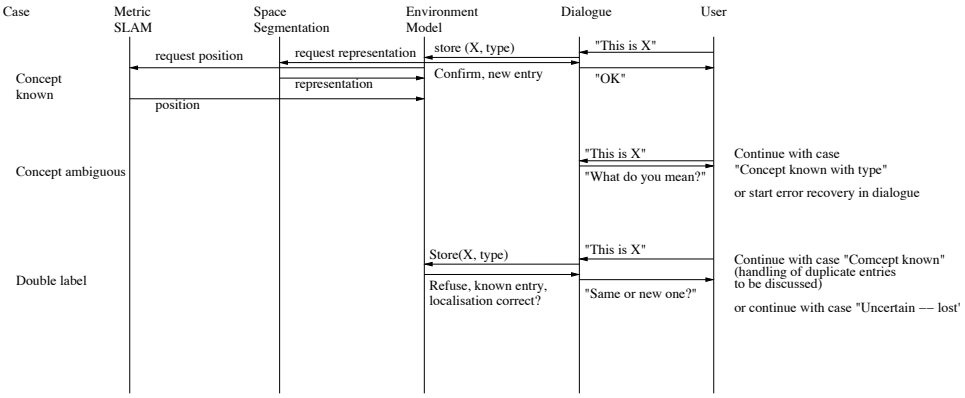


Figure 3.2: *The interaction flow for the initial user utterance for user driven annotation. Depending on user responses (in case those are necessary), the interaction flow corresponds to other, initial situations*

Addition obvious Relates to the standard situation for *region/location* presentation given that the robot is localised in the previously acquired map. "This is the <name>" – <name>known as *region/location*. Requirements: Localisation ability in a previously stored map/environment description.

Addition ambiguous – concept This situation relates to the ambiguous concept information situation for the initial tour, given that a map exists and the system is localised. "This is the <name>" – <name>neither known as *region* nor as *location*.

Addition ambiguous – space and concept Relates to the "ambiguous – space and concept" situation of the initial tour, given the localisation in the existing map. Given that it is known that a *region* is presented, but the spatial configuration suggests that actually only the "link" to the *region* is shown, the system needs to store a connection (connector node) to a yet undefined *region*. Requirement: Situation interpretation ("Show", Hüttenrauch *et al.* (2006a)), spatial situatedness interpretation, interaction interpretation, dialogue.

Change interactive The user informs the robot of a change in the environment. In an office a person might have left or moved to another room or a new coworker has arrived. The coffee machine might have changed its position, or it might be exchanged and look totally different. Requirement: Update functionality in space representation and dialogue.

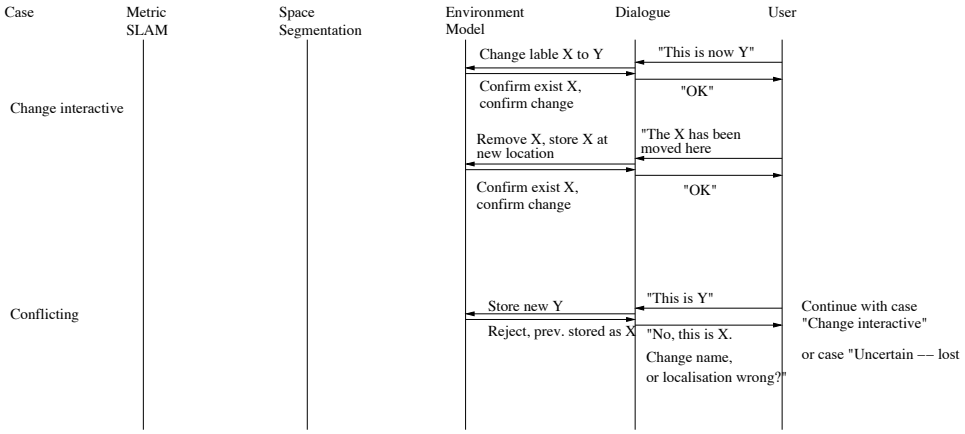


Figure 3.3: *The interaction flow for updates – those cases that correspond to the initial cases are omitted here*

Conflicting Information initially given is overridden by new "initial" information. No indication is given that the user (it might be a second user involved) knows about the previously given information. Requirement: Localisation in previously acquired map, dialogue to clarify if label needs to be changed or localisation is wrong.

Figure 3.3 shows the interaction flow for particular update situations that might occur when something in the environment has changed and when information is conflicting with the previously stored representation.

Implicit information – data driven During the user driven process the robot builds the map of the environment. Hence, the detection of a change in the environment might need to be commented by the user to avoid ambiguous annotations.

Region The delimiter of a *region* can be defined internally in different ways. One option is to assume a gateway detection (Kruijff *et al.* (2006)), or, as done in the work presented here, by using a set of features that represent a *region*'s spatial properties. Thus, the system assumes to have left a certain *region* and entered a new one when the continuously observed feature representation changes significantly, and generates hypotheses on "being still in the same *region*" vs. "having left the *region*". This latter approach has been investigated in detail and is discussed in chapters 4 and 5 respectively.

- **Initial:** The system assumes to have left a recently specified *region*, the entered *region* was not specified before, and confidence is low. Requirements:

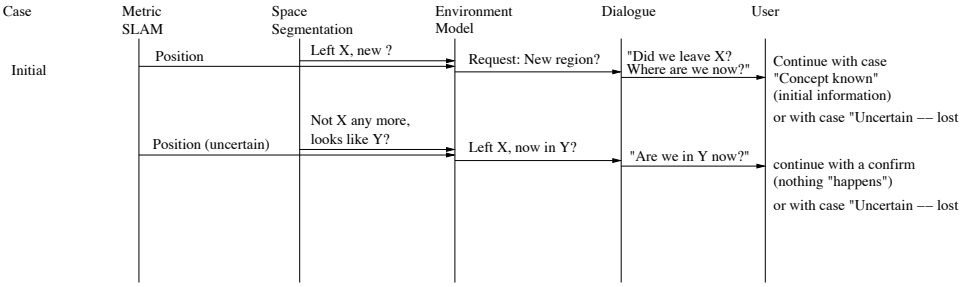


Figure 3.4: *The interaction flow for cases in which the system detects something and needs to get more information about it*

Confidence measure, clarification dialogue, space representation update, re-localisation.

- **Confirmation:** The system observes a change into a different *region* (both *regions*, the one left and the one entered were specified), but the confidence is low. Requirements: Confidence measure, dialogue, re-localisation.

In the situations for which figure 3.4 shows the interaction flow the system detects itself that the environment differs from the previous surroundings, but it is not confident enough to just rely on this information. Clarification dialogues have to be invoked.

Location The moving robot maintains a continuously updated reference to the "closest *location*" in the current *region* (if available), so that it can report its whereabouts also in the context of this *location*.

Using the map

The use of a previously acquired map is obviously also involved in the data driven information acquisition, but in the following cases it can be assumed as crucial to the situations.

All slots are assumed to be filled in the normal case for known parts of the environment. Missing information expresses thus a need for repair mechanisms, in most cases a re-localisation.

Implicit localisation "Implicit" localisation is needed continuously during any tour process or service run. Still, some different situations can occur.

Continuous The continuous localisation is assumed to be running, incrementally updating a low-level geometric representation of each of the topological graph entries (in this framework corresponding to the *regions*) to gain more and more confidence.

- **Confident:** In case of a confident localisation process nothing particular happens. It is assumed that the current position of the robot is continuously communicated to the rest of the system. Requirement: Confidence measure on continuous localisation process, decision ability to determine uncertain situation, dialogue, re-localisation.
- **Uncertain/lost:** Due to external influences the mapping process might be disturbed so that confidence gets low. Requirement: Confidence measure on continuous localisation process, decision ability to determine uncertain situation, dialogue, re-localisation. Missing information / low confidence on

```
     Current region:
     Description:      region descriptor     (geometric
                                              features)
     Localisation      LOW                   (double
       confidence                             measurement,
                                              classification
                                              confidence)
OR

     Description:      location descriptor   (position relative
                                              to region, pose)
     Localisation      LOW                   (double
       confidence                             measurement,
                                              metric loc.
                                              confidence)
AND

     Overall           LOW                   (summarised
       confidence                             confidence)
```

Figure 3.5 refers to situations that can occur in relation to continuously running localisation.

Kidnapped/woken up In this case the system needs to recognise a severe change (kidnapped robot) or has to localise in a previously acquired map after a restart (waking up). Different options to resolve the situation can be suggested. Missing information:

```
Current region:
```

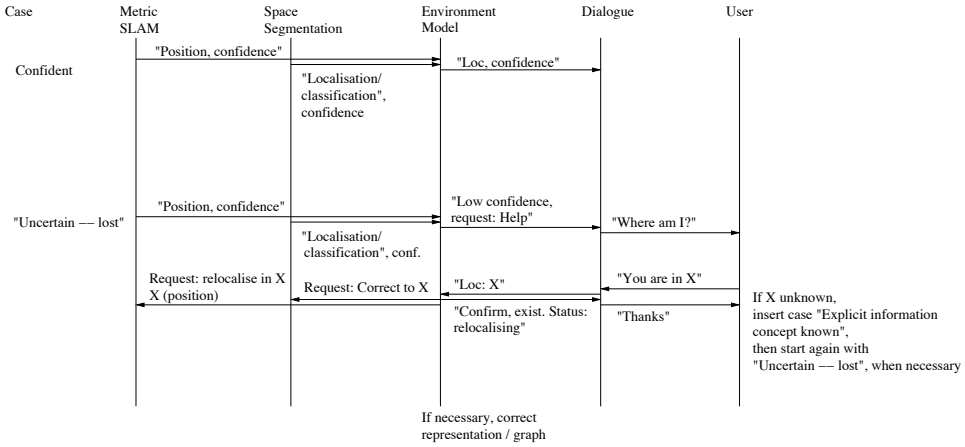


Figure 3.5: *The interaction flow for cases in which the system handles localisation continuously*

```
     Description:      region descriptor     (geometric features)
                                              confidence)
AND
     Current location:
     Description:      location descriptor   (position relative to
                                              region, pose)
```

- **Place recognition:** The system tries to recognise its surroundings and localise itself in a certain *region* – if specified. Requirements: Place recognition (*region* recognition), dialogue for confirmation/help, space representation update, re-localisation.

- **Active exploration:** The system cannot localise exactly but can give a list of possible current *regions*. Requirements: Place categorisation (*region* recognition), active exploration abilities, dialogue for confirmation/help, space representation update, re-localisation.

An illustration for the cases in which the system has to start a localisation "from scratch" is given in figure 3.6.

Explicit query – localisation An explicit query situation evolves when the user explicitly asks the system about its whereabouts. The answers depend on the confidence level of localisation.

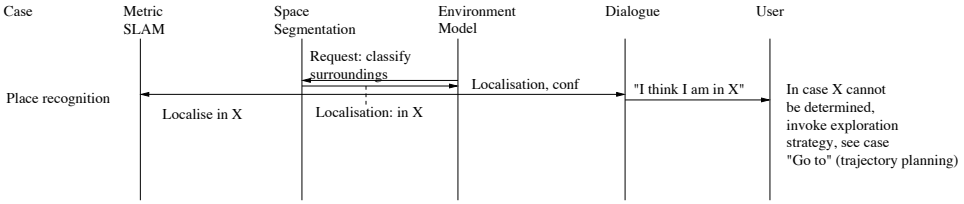


Figure 3.6: *The interaction flow for explicitly invoked localisation, after waking up the system or kidnapping it*

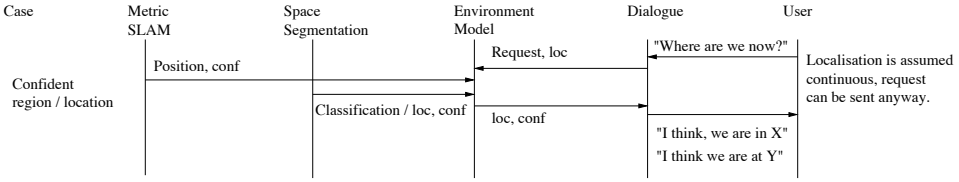


Figure 3.7: *The interaction flow for explicit user queries regarding the current whereabouts*

Confident – *region* and/or *location* "I am in <region_name>" or "I am in <region_name>, close to <location_name>". The system is in a previously specified *region* and can determine the closest object (*location*). Requirements: Dialogue to convey query and response, localisation.

Uncertain – lost Related to uncertain implicit localisation. In this case confirmation on the current position might be needed. Requirements: Dialogue to convey query and response, localisation, re-localisation.

Figure 3.7 illustrates the dialogues for explicit queries about the current whereabouts. In general (since a continuous localisation is assumed) the status of the localisation process is available at dialogue level already. An explicit request to the environment model level only needs to be sent, when the available information is outdated.

Explicit query – going to target The second big class of query situations is that the robot is sent to a (known) target. A target can be a *location* or a *region*. Unknown targets need to be handled at higher levels in dialogue. A list of way points relative to the current positions is generated / updated. The trajectory planning from point to point needs to be handled accordingly. Requirements: Localisation, way-point generation (path generation), navigation abilities, dialogue for confirmation, confidence measure for localisation, update abilities. In the normal case, none of the needed information is missing. In case a crucial entry is missing,

the respective situation occurs and the information base has to be filled accordingly, before a "go to" can be realised. The interaction flow in this refers mostly to planning components, otherwise it resembles the flow for an explicit localisation query. Since it is not much different in terms of the information patterns, a respective diagram is omitted here.

The requirements for an initial approach to HAM (particularly the guided tour scenario) informally mentioned above can be specified as follows:

- Metric positioning system. For precise navigation and localisation a metric method (e.g., a metric SLAM method) is considered useful, if not necessary[4].
- Space segmentation and classification. For the specification of new *regions*, user driven or data driven, the ability to segment the specified area from the rest of the environment is crucial. An approach to such a *region* segmentation is described in chapters 4 and 5.
- Topological representation. In order to relate spatial entities to each other a topological representation (a graph) is needed. The generation of a respective graph is explained together with the method for the segmentation of *regions*.
- Conceptual environment model. For the communication with the user the topological model needs to be associated with a conceptual framework.
- Conceptual category knowledge. The category, i.e., *region* or *location*, of a specified spatial entity must be provided either from the dialogue model or from a clarification dialogue. This involves prior knowledge that has to be fed into the system. See chapter 6, section 6.3, and chapter 7 for a reflection on *a priori* knowledge in the context of HAM.
- General dialogue abilities. A spoken dialogue system is assumed to convey verbal information from user to system and vice versa. Clarification dialogues must be considered in this system to resolve ambiguities. Chapter 5 suggests an approach to the detection of structural ambiguities that triggers the respective dialogue.
- Interaction monitoring/supervision. Depending on the flow of the "guided tour" process, interaction schemes change (Hüttenrauch *et al.* (2006b) consider different episodes that categorise the interaction). A respective monitoring system needs to keep track on the state of interaction and the current user of the system.

[4] Approaches to topological mapping often claim that metric methods can be ignored completely. The author on the other hand supports the opinion that for exact navigation (e.g., for mobile manipulation) a metric localisation method is of much greater use than a pure topological one.

- User position knowledge. The user has to be singled out from other humans being in the vicinity. This has to be kept valid for the complete period during which a user wants to interact with the system with the help of a tracking method. Some particular challenges for such a tracking system are described in section 3.4 of this chapter.

Given that those basic requirements are fulfilled functionalities like localisation, navigation to a target, or following the user can be achieved with respective functional components. The following section describes an overview of a general architecture for the space representation and interaction components that construct the framework for a HAM-system.

3.3 Architectural framework

The components for HAM can be grouped functionally into two bundles, one representing the representation of space involved and the other as being responsible for the interaction. Both parts are then linked by the conceptual environment model component that integrates space representation and dialogue. For the space representation part of the overall architecture a hybrid, hierarchical approach is assumed best, since it allows the handling of the underlying mapping process in as efficient a way as possible.

Figure 3.8 shows the design of an approach to HAM. Chapter 5 refers to the system design and implementation that were used to achieve a system applicable to particular investigations considering different modules of the general framework.

As metric positioning system any feature based SLAM approach is considered useful to establish metric links between the nodes of the topological graph. This topological graph is generated on top of a metric system with the help of a space segmentation method. For space segmentation different approaches can be considered, one of which is a *region* segmentation based on features extractable from percepts of the environment. This way of segmenting the environment is a central part of the work done for this thesis and is discussed in chapters 4 and 5.

On top of the topological layer a conceptual model links labels to the nodes of the graph and forms the connection to the dialogue, which is considered the entry point to the interaction related parts of the general architecture. The environment model uses the concepts described above in section 3.2 and is described in more detail in chapter 4.

The interaction part of the system is mainly controlled by an assumed dialogue model or interface for comprehensible communication. Functionalities as the tracking and following of humans are connected to this central interaction control. Particular issues and challenges for such a tracking system are described in the following section.

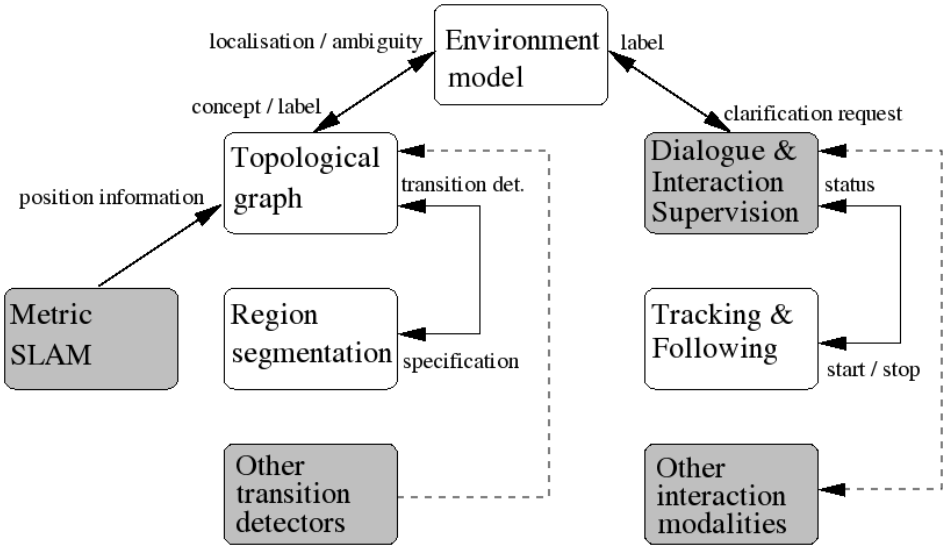


Figure 3.8: *A concept for integrating all the necessary parts for Human Augmented Mapping. The white modules are discussed in detail in the following chapters, grey ones represent parts of the concept that contribute to the discussion but are not of main interest*

3.4 Tracking for following

One of the main contributing conference articles for this thesis investigated the use and usefulness of a particle filter based multiple target tracking method in the context of human-robot interaction, i.e., for a following scenario and the adequate navigation in presence of bystanders (Topp and Christensen, 2005). Given the number of existing and published tracking approaches, the interesting issues are in fact caused by the use of a tracking method in a certain context. Thus, a number of specific challenges for the tracking system could be defined.

Multiple targets and tracking issues

Given the fact that only one user is assumed to interact with the robot at a given time, the use of a multiple target tracker might seem superfluous. Nevertheless, the robot is moving in presumably populated but not crowded environments, and might have to distinguish between the current user and *other persons* in the vicinity, that should be treated appropriately by an adequate obstacle avoidance, or a social navigation method. Such approaches have been investigated by, e.g, Pacchierotti *et al.* (2005). However, those approaches all need to keep track of the persons to pass, which makes in the present case a combination of both abilities (following

and passing based on the same multiple target tracker) an obvious solution. The different purposes though suggest slightly different criteria for quality measures of the tracking approach.

Two general types of tracker failure can occur with respective effects on the reliability of the system. These two types are the loss of a target and a confusion of different targets due to ambiguous data association.

In case of a following scenario the output of the tracker needs to be analysed to find the one and only person to follow. In this case the purpose of the multiple target tracking is to find all user hypotheses and later distinguish the user from other persons. One possible tracker failure would be to lose the target associated with the user due to detection failures over a certain period of time. If in such a case the target is removed and after a while replaced by a new one, the tracker output can still be used to define this new target as the user, as the target is probably still in the area where a user is expected. Otherwise a complete target loss could lead to an error state and thus be handled appropriately.

More critical for the following scenario is the confusion of the user target with another one, possibly another person moving in a different direction. This is a situation that is to be avoided by any means, as the system would not detect any error and would start following the wrong person.

For the purpose of passing persons the criteria are slightly different. Here it is not important to know which of the persons is associated with which target, but targets must not be lost when in fact the respective person is still around. In such a situation a person would appear as an arbitrary obstacle to a general obstacle avoidance routine. Considering those aspects, a robust tracker that allows to distinguish between targets is a solution to both problems, as they could occur at the same time. This would be the case when the system is following one person around while reacting appropriately to the presence of other persons.

In order to provide an appropriate approach, a multiple target tracker (Schulz *et al.*, 2001) was implemented and evaluated. The original approach is based on leg detection and occupancy grids to detect people and distinguish them from other objects by movement. Detected features are associated to tracked targets with a sample based joint probabilistic data association filter (SJPDAF). Using this, multiple targets can be tracked from a mobile robot, while motion compensation is done by scan matching.

For the work underlying this thesis the idea of using the SJPDAF approach for tracking and associating was adopted, but in contrast to Schulz *et al.* the detection and tracking methods allow handling of people standing still, which is needed for interaction.

A central aspect of a tracking system is the actual detection of targets, which was also reflected quite thoroughly. These aspects and the implementation and evaluation of the multiple target tracking system are described in chapter 5.

3.5 Main aspects for the thesis

The design and implementation of a complete system for Human Augmented Mapping would exceed the scope of this thesis, particularly considering the complexity of a suitable dialogue management system. However, large parts of the suggested general architecture have been covered in an experimental implementation, focusing on the mapping subsystem and the tracking and following module that dealt particularly with the (assumed) central situations discussed previously, i.e., the ones related to *map acquisition*. Within a project cooperation the mapping subsystem was transferred to a more sophisticated integrated interaction framework that provided also the higher level dialogue and interaction supervision capabilities.

Thus, the focus of this thesis is the partially hierarchical environment model (see the following chapter) used to achieve the integration of human spatial understanding and robotic mapping as well as the instantiation and evaluation of this model in the implementation of the mapping subsystem. A central part of the latter is the region segmentation and the detection of (structural) ambiguities. To investigate the applicability of both model and implementation three user studies were carried out with different guiding questions. These studies are also considered a central part of the work described in this thesis.

3.6 Summary

This chapter presented the framework of Human Augmented Mapping (HAM) with its requirements and situations to be expected in an interactive context for a mapping approach. An idea for a schematic architecture has been sketched and explained.

HAM is to be seen as a concept for the integration of human-robot interaction and SLAM. This means that it does neither aspire to be a sophisticated approach to robotic mapping or SLAM, nor does it represent a new interface for human-robot interaction. The advantages for robotic mapping arise thus from the interaction in terms of opportunities for disambiguation in uncertain situations. Furthermore interaction (communication) about the robot's workspace can be facilitated with the integrating concept, as information can be retrieved and used in a human comprehensible way. Two central aspects of the general concept have been worked on as the main issues for the doctoral project and are discussed in the remainder of this thesis.

Chapter 4

Hierarchical environment representation

Human Augmented Mapping aspires to integrate human and robotic environment representations so that both the robotic mapping process and the communication between robot and user can be facilitated. The idea is not to enable a robot to explore an environment using the strategies a human would have, which is often the case in cognitively inspired approaches for robotic mapping (Beeson *et al.*, 2005; Choset and Nagatani, 2001; Kuipers, 2000). One central question is how such a joint environment model can be built, given that it needs to be a base for communication in terms of common ground (Clark and Brennan, 1991), and at the same time has to represent a map model useful for robotic mapping. Additionally it is assumed, that the robot should acquire such a representation in an interactive setting.

A number of different theories on how spatial relations are acquired and represented in humans have been proposed throughout the years. According to McNamara (1986) those theories can be grouped along 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-existence in one region. In other words, if two items were close to each other, but in different regions, it was

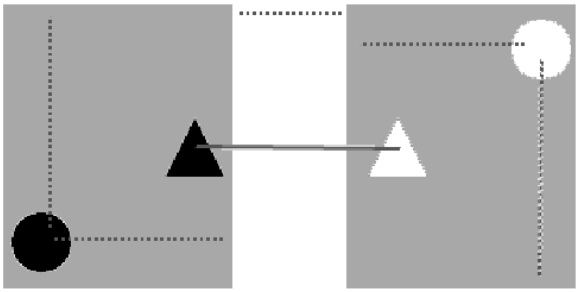


Figure 4.1: *The spatial relationship between two objects far from each other (circles) are rather estimated with the help of the relation to the region and the relation between the regions (dotted lines). For close objects (triangles) the relation is estimated directly, ignoring the "border" between the regions (solid line).*

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.

4.1 A partially hierarchical environment model

Partially hierarchical (in contrast to strongly hierarchical) implies that a relationship between two entities assigned to one level of a hierarchy can be described directly. This means to ignore the much more complicated way of first relating one entity to its "parent"-node in the hierarchy, and then relating this to the other entity. Figure 4.1 shows the two different ways of relating spatial entities with each other.

Given that a partially hierarchical model can be assumed, first of all a hierarchy has to be established to describe an environment. A natural hierarchy is a conceptual one, as already pointed out in chapter 3 as

- **Object** Small objects that can be manipulated (cup, plate, remote control). (*Objects* are not incorporated in the graph structure described in the following, but they are relevant to the user study setups discussed later in this chapter and particularly in chapter 6.)
- **Place** A distinct state with a certain set of paths and a specified view (according to Kuipers (2000)). (not incorporated in the concepts used for this thesis to avoid confusion with this definition)
- **Location** The area from where a large, not manipulated object is reachable/visible (sofa, fridge, coffee-machine, pigeon-holes). Also "the place

where the robot is supposed to do something or look for objects", thus a "workspace"[1] (The term "location" has also been used in the context of spatial cognition to describe "a view of the surroundings from this position" or a "snapshot" (Krieg-Brückner *et al.*, 1998)).

- **Region** A container for one or several locations. Offers enough space to navigate (rooms, corridors, delimited areas in hallways). (The concept of "region" is also used by Kuipers (2000), but not conflicting, since it allows to group places in a hierarchy and deliver a local description of an area, which reflects the idea used for the framework discussed in this thesis.)

- **Floor** A collection of regions, distinct by the level (in height). Requires start of a new map to deal with similar structures if no altitude information is available[2].

The common concept of "room" is explicitly not used in this context since indoor environment architectures do not always strictly follow the idea of a clear separation into rooms. Often one "architectural" room is used for different purposes and is thus separated functionally into different areas (*regions*). Additionally the architectural understanding of "room" might not always correspond to the common use of that term, technically speaking a "corridor" or "hallway" is a room as well. For the use in HAM it is important to allow all those particular "rooms" to be modelled on the same conceptual level as those commonly understood by the term. Thus, the term *region* seemed more general and easier to comprehend in the way it is defined and used for HAM.

As it is assumed, that the service robot is supposed to "work" in indoor environments, thus more or less in one building, further concepts are omitted. It could be assumed though that the concept of *buildings* or, in less architectural terms *blocks* exceeds the hierarchy after the concept of *floors* (or *levels*). Once a representation of one *floor* is generated, a respective topological structure of several *floors* can be linked by a representation of "using the elevator" to connect them. In the following though the concepts of *region*, *location* and *object* are most central to the discussion.

Once the hierarchy is established the second central idea is to use the concept of partiality to come as close as possible to the assumed human understanding of space to establish a common basis for communication.

Partiality can presumably be expressed in the use of a graphical model that describes the hierarchy. An assumption is that if the mental representation of a known environment is partially hierarchical, a human would not follow a strongly

[1] Over the time the work was conducted the author realised that the term "workspace" would actually have been better than the term "location", but it was decided to keep the original terminology that was used also in the previous licentiate thesis (Topp, 2006).

[2] Here it was also decided to keep the term "floor" instead of switching to the much more appropriate term "level", to keep the terminology consistent with the previously published licentiate thesis.

hierarchical strategy when presenting and explaining this environment to a robot. In some cases entities on one level might even be left out, but entities from a lower level might be considered important for the tour. Thus, the model has to cope with this particular type of partiality, which can be established by "generic" entities on all levels of the hierarchy but the last (lowest) one (in the previously mentioned list this corresponds to "object"). In a simple two-level hierarchy as used so far for the implementation of a Human Augmented Mapping system this would mean to have a "generic *region*" that can incorporate *locations* just as any other *region* but does not need to be specified explicitly before the *locations* are named. Figure 4.2 illustrates this simple two-level hierarchy with the "generic" entity.

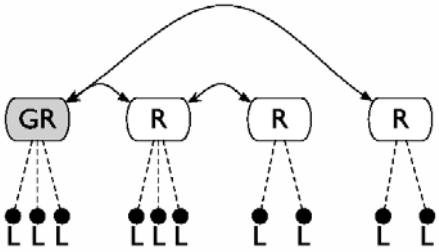


Figure 4.2: *The two-level model with "regions" (white in general (R), grey for the "generic region" GR) and "locations" (large black dots, L). The solid curved double arrows represent the links between the entries of one level (between "regions"), the dashed lines represent the relation of the "locations" to their surrounding "region".*

Topologically speaking the *regions* are connected with each other by links (edges) that can be represented with endpoints in the respective *region*, "connector nodes". Those nodes allow to express links to unknown environments as well as to known ones.

The applicability of the proposed environment model to human-robot communication and its contribution to the mutual understanding of used spatial concepts was tested in a user study setup to inform the further design of the topological graph that had to be built for the mapping subsystem of a respective complete implementation of a system for HAM, according to the general architecture proposed in chapter 3. A pilot study with five subjects was conducted (Topp *et al.*, 2006a,b, and chapter 6, section 6.2) with an initial prototypical implementation that allowed the study subjects to present a known environment to a mobile robot in a "guided tour" scenario and was used to collect sensory data (laser range data and odometer readings) to investigate the approach to the representation of *regions* that is discussed in section 4.4 of this chapter.

The study showed this initial implementation to be sufficiently robust and the environment model adequate in terms of the correspondence to the intentions of the subjects, regarding what they wanted the robot to "understand" during the tour. Particularly the concept of the "generic *region*" turned out to be successful to incorporate significantly different presentation strategies, that expressed at least to some extent the type of partial hierarchy described above. Figure 4.3 (in a

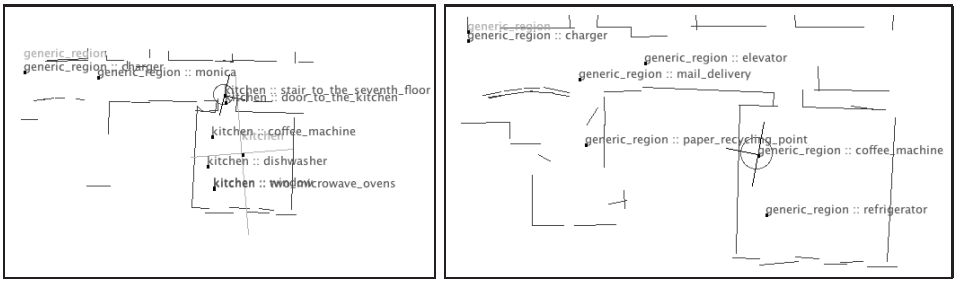


Figure 4.3: *Two different representations of the same environment generated after two runs with subjects of the pilot study. Regions are marked with the label, also showing the representation of the spatial properties with the two axes of the ellipse describing a respective data set and locations are marked together with the label of the region they are assigned to.*

coloured version also on page 120) shows two different representations generated for the same environment by two subjects of the study. One of them was very specific and presented both regions and locations, while the other subject did not present any regions at all – because these seemed not relevant for the robot to know in the subject's understanding.

In the following the topological graph structure is described, that resulted from the idea of the partially hierarchical environment model and incorporates the previously suggested concepts.

4.2 Building a topological graph structure with *regions* and *locations*

As mentioned before only two of the four proposed spatial concepts are considered central for the work presented here. The main concept for the segmentation of indoor environments is assumed to be the *region*. This means that the representation of an indoor environment that can be built by the robotic system so far corresponds to the representation of one *floor* of a building. The second relevant concept is the *location* as particular part of a *region*, which can be assumed sensible to a robotic mapping process. *Objects* according to the used model and conceptual hierarchy are mobile, thus there is no immediate reason to include them in the topological graph structure. It makes more sense to represent them separately, incorporating all needed information on, e.g., purpose or preferable grasping approaches, and add a tag LAST_SEEN_AT for the most probable *location* at which this object is located. For easy retrieval it might be helpful to maintain a list of "current *objects*" for each *location*, but this is independent of the general structure for the graph representation described in the following.

The graph that implements the proposed hierarchical environment model consists of four different types of nodes and two types of edges, building a hierarchical "graph-in-graph" structure that represents topological links on a higher level and metric, viable links on a lower level:

- High-level nodes and edges correspond to a topological graph representation of the environment:
  - High-level nodes correspond to *regions*. The initial graph contains the "generic *region*", which has an empty representation of its spatial properties. Besides the description of spatial properties each *region* contains a list with *locations* and a subgraph with *navigation nodes*.
  - High-level edges (termed *abstract edges*) express the links between regions. If there are more than one viable links (e.g, doors) between two *regions* available / observed, still only one *abstract edge* is used. Each *abstract edge* has a list of viable ("implemented") links that refer to pairs of connecting nodes on a lower level of the graph structure.
- Low-level nodes and edges form the subgraphs that belong to each *region*. Metric links and positions are expressed in the *region's* local coordinate system and are thus decoupled from the global, metric mapping / positioning process assumed to be part of the framework.
  - Low-level nodes are termed *navigation nodes*. Their concept is mainly to facilitate navigation and their representation is a purely metric position (x, y) in 2D. Each navigation node has a unique identifier and an indicator which *region* it belongs to. The *navigation nodes* are set when a certain distance from the previous node has been covered or when a connection between *regions* has to be established that cannot be expressed with existing nodes (see also "connector nodes" below).
  - Low-level *edges* connect the *navigation nodes* by relative metric, directed links (bi-directional, i.e., one link in each direction is used), thus forming a viable subgraph (*navigation graph*) inside the respective *region*. In the hypothetical case of a link that for some reason only can be passed in one direction, this can be expressed by omitting one of the directed links of the edge.
  - *Locations* are represented by a pose in 2D (x, y, θ), expressed in "their" *region's* coordinate system. *Locations* are technically speaking not part of the subgraph, but can be reached by computing a path to the closest *navigation node*. This decision was made to keep the main concepts of the environment model independent from the navigation aides on their respective level of the graph hierarchy.
  - A particular type of *navigation nodes* are *connector nodes*. These do not only have an indicator which *region* they belong to but also one indicator for the *region* they connect to. Typically those nodes are located

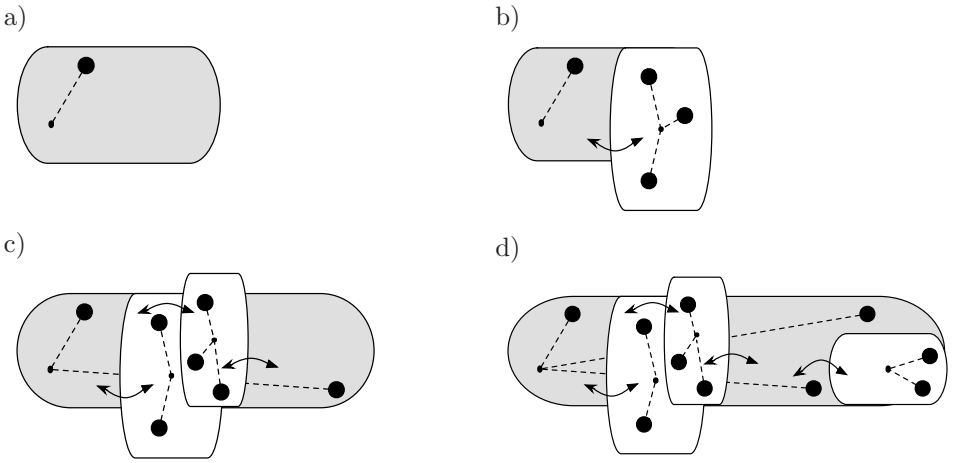


Figure 4.4: *The graph is built while the robot is travelling and gathers the information given by the user (illustration). a) The tour starts in the "generic region", where also one "location" is specified. b) The user specifies a particular "region" which is linked to the generic region. There are also three "locations" inside this "region". c) The user specifies a second "region" that overlaps slightly with the previous "region", consequently one of the "locations" is moved. The link between the two "regions" is direct, not leading into the "generic region". d) After two locations in the "generic region" the user specifies the third "region", which has no direct connection to the previous one(s). Thus, the connection is established via the "generic region".*

close to the border of the respective *region* and have a corresponding *connector node* in the neighbouring *region* (if this has been specified or corresponds to the "generic *region*"). Low-level edges between two *connector nodes* are omitted, due to the different coordinate systems the nodes are expressed in. They are computed when needed in the global coordinate system. To assure viability only observed (actually travelled) links are added as viable connection to the respective high-level *abstract edge*.

Initially, only the "generic *region*" exists, represented with its origin at the starting point of the "home tour". Only one "generic *region*" is needed, assuming that ambiguities are resolved by the user by specifying *regions* that are of importance to her. Thus, the "generic *region*" is underlying the topological structure as a kind of "gap filler", also guaranteeing consistency with the actual environment. One could

wonder about clearly identifiable *regions* that are not specified by the user, but indicated by the transition detection. The author would here rather consider to have generically named but specified (in terms of their spatial properties) *regions* instead of having several "generic *regions*". This allows also to incorporate the presentation and specification strategies of different people in one graph representation. In case an overlap between two neighbouring *region* representations occurs, it is assumed that the newer entry is dominant, thus the (geometrically speaking) overlapping part is assigned to the *region* specified last.

Figure 4.4 illustrates the graph building process, including the reassignment of one *location*, that is first specified in one *region* but has to be changed to a new *region* due to an overlap. Connections between *regions* are depicted as bi-directed arrows (solid lines), and the *locations* (large black dots) are assigned with dashed lines to the respective *region*, i.e., to the origin (small black dots). The obtained graph corresponds to the overview example given in figure 4.2 on page 50.

A question arising from observations made in the pilot study and a previous exploratory study (Green *et al.*, 2006a; Hüttenrauch *et al.*, 2006a,b) was whether there was any correspondence observable between human presentation strategies for particular items (corresponding to their concept along the lines of the previously described hierarchy) and the spatial concepts used in the previously described model. This idea is explained in the following.

4.3 A conceptual hierarchy in presentation strategies

The setup of a previous user study (Green *et al.*, 2006a; Hüttenrauch *et al.*, 2006a,b) was exploratory in the sense that no clear hypotheses were set to start with – the idea was to observe people interacting with a robot in a certain scenario and learn more about possible situations to deal with. Subjects were asked to interact with a Performance PeopleBot in a "home tour" setting in one room. The robot was remotely controlled in a Wizard-of-Oz framework[3] and the subjects received a number of suggestions on what to present to the robot. Particularly for small objects[4] they were instructed to place them on a clear and flat surface so that the robot was able to segment them from the background.

Despite these rather specific instructions the subjects acted freely enough for the experiment leaders to observe individual "presentation strategies" for different

[3]The term "Wizard-of-Oz" describes an experimental setup in which the subjects interact with a computer program (in this case a mobile robot) that acts seemingly autonomously but is controlled by an experiment leader (the "Wizard of Oz", who pulls the lines) according to a clearly specified prototypical system setup.

[4]the experiment leaders tried to instruct the subject to present "places" and "objects", but the distinction was not always clear to them – thus it became obvious that an environment model for HAM needed to postulate a clear distinction with respect to the robot's ability to make use of the model

items to be presented to the robot. In general it seemed that there was a tendency to manipulate small objects, or point to them very explicitly, while larger objects (also "places" like a corner of the room) were presented with rather coarse, almost "waving" gestures.

Similar observations were made in the pilot study (see chapter 6 for details), where subjects were asked to present an office environment to the same robot, which was in this case running partly autonomously. They were completely free to decide what to present and how to do that, but were informed that the robot would not be able to recognise small objects.

In this case it seemed that the subjects tended to present *locations* (according to the concepts used in the HAM framework) with some – often coarse – pointing gesture, while they did not perform any specific gesture when presenting a *region*. *Objects* were virtually omitted in the experiments due to the information about the lacking abilities for object recognition of the robot.

An obvious idea to further confirm the applicability of the proposed model was to investigate, whether there was an observable correspondence between human presentation strategies for different (spatial) items and the conceptual category of the particular items. The closer analysis of the observations from the mentioned studies, however, turned out to be impossible, since both studies had been set up under completely different conditions and could thus not be compared appropriately. Hence, one aspect of the second user study described in chapter 6, section 6.3, was to provide data for the analysis of possible presentation strategies that correspond to the hierarchy of (spatial) concepts used for the HAM framework and proposed previously. The study results suggested, that there is some correspondence to be observed, which confirmed again the applicability of the proposed model for the investigated "home tour" scenario.

A central question remaining for the implementation of a Human Augmented Mapping approach is how to actually segment the given space and represent entities of the proposed concepts. The following section focuses on this issue.

4.4 Segmenting and representing an indoor environment

According to the previously described environment model a *region* would most often correspond to an area that is somehow delimited from other areas. This delimitation can be achieved by walls, furniture placement, plants or any other delimiter that creates the feeling of being "inside" or "outside" a particular area. Such a delimiter could in fact be a change in the type of flooring material, particularly in cases where there is more of a functional than structural or architectural delimitation generated. Such a case is used as an example for a structural ambiguity and is discussed later in this section. As long as the concept of a *region* corresponds to what humans typically refer to as "room", one intuitive way of describing a *region* for a robotic

system could thus be to use a gateway or door detector. As anecdotal evidence for this an observation from the second user study (see chapter 6, section 6.3) can be cited. One of the study subjects presented the limits of the "living room" of an apartment by showing all the entrances/exits leading to and from the room, which were quite many in this case, since the room needed to be crossed to reach two other rooms and the balcony.

A gateway detector was used, for instance, by Kruijff *et al.* (2006), where the proposed system generated a clarification question whenever a door-like passage was traversed to build a graph with nodes belonging to particular clusters, representing a "room". One of the drawbacks of such a gateway detection is obviously, that the gateways have to be passed to be able to find the delimitation of the given area. Furthermore only the travelled paths between the gateways can be declared as part or not part of a particular cluster of known positions, but it is close to impossible to describe the area of the room (or *region*) as such.

In the following the approach to segmentation of *regions* and *locations* used for this thesis is discussed before the background of other approaches.

Range data based segmentation of indoor environments

One central issue of the work conducted for this thesis was to find a method to segment a given indoor environment into *regions* based on a representation for the *regions* that captured their spatial properties instead of the gateways to the neighbouring ones. This implies a rather strong decision for the type of representation. Related approaches to the integration of human concepts or semantics into robotic maps often assume the functionality expressed by observable objects as a strong indicator for the particular category (e.g., "office", "kitchen") of the given surroundings and base their reasoning or localisation strategies on these objects (e.g., Gálvez López *et al.*, 2008; Vasudevan *et al.*, 2007; Zender *et al.*, 2007). For the presented framework it was decided to focus on a more low-level representation that would allow to incorporate individual preferences according to the presentation strategies expressed by particular users rather than the autonomous categorisation of environments.

Another strong decision was made when (laser) range data sets were picked as the basis for the representation of the environment. The author is well aware of the fact that image based systems provide much richer and more human-like perceptions and are thus more intuitive (Booij *et al.*, 2006, 2007; Pronobis *et al.*, 2008; Tapus *et al.*, 2004a,b). However, a clear advantage of range data is their comparably low complexity, which made it attractive to exploit their range of descriptiveness useful to a Human Augmented Mapping approach. Additionally, range data usually have the advantage of reflecting the spatial properties of a given area somewhat more precisely than an arbitrary computer vision system can do.

The capturing of a complete area as one unit is suggested by Diosi *et al.* (2005), who use a watershed implementation after interactively labelling positions that are then related to the areas that include them respectively. Compared to the approach

presented here, a clear difference lies in the assumption implicitly understood from Diosi *et al.* that all rooms and other areas have to be specified in one complete tour to provide a correct representation of the given environment. This has to be considered a strong limitation since it was observed that potential users do not necessarily describe every room or area to a robot, but pick those that they personally consider important (see chapter 6 for details).

Mozos *et al.* show, how the *category* of a certain area ("room", "doorway", or "corridor") can be determined with the help of supervised learning (Martínez Mozos *et al.*, 2005). They generate a number of features from raw laser range data sets that were obtained at different locations corresponding to the named categories and use these features to form a training data base for the learning method. Kröse showed that it is possible to represent convex areas reliably by obtaining only one sample range data set and transform it to its centre point and bearing with the help of a principal component analysis to anticipate future scans (Kröse, 2000). The approach used for the presented work adopted in fact the idea of using a set of features computed with help of a principal component analysis to represent a laser range data set, that is obtained in a *region*, but uses an even more concise set of features than Mozos *et al.* did, as is described in the following.

Representing *regions*

In this section a very concise approach to representing *regions* with data obtained from a laser range finder is presented. It is assumed that the characteristics of an arbitrary *region* can be captured from a rather small data set (in this case a 360° laser range scan) obtained at one position.

The axes of the largest ellipse fitting the range data as two characterising features and the mass (area) of the complete space covered by the scan as a third feature are chosen as descriptive features for a *region*. This set of features proved quite useful in terms of the categorisation of different types of *regions*, but is less powerful in terms of recognition abilities of the system (see chapter 5, section 5.3 for a detailed discussion).

The ellipse itself allows to decide which geometrically defined area belongs to the *region*. This is obviously only a rough estimate, since not all parts of a rectangular room can be covered by just one ellipse and in some cases areas outside the actual *region* might be assigned to it. However, large parts of a *region* specified by the user are covered and previously specified items (*locations* and the respective parts of a navigation graph) can be hypothetically assigned to that *region*, keeping an uncertainty flag so that clarification dialogues can contribute to corrections if necessary.

The following features that characterise a laser range data set $\{X_i : 0 \leq i < N\}$, where N is the number of data points $X_i = (x_i, y_i)$ are investigated:

a) the area (or mass) m of the "visible" part of the represented region, and

b) the maximum range $l1$ and $l2$ along the two principle components of the data set (the axes of the "main" ellipse).

Locations according to the previously described model can be integrated into the region with their relative position to the centroid $\bar{X} = (\bar{x}, \bar{y})$ of the data set.

Due to the angular sampling in laser range finders the spatial representation is non-uniform[5]. To compensate for this effect the centroid is computed as a range weighted average

$$\bar{X} = (\bar{x}, \bar{y}),$$

with

$$\bar{x} = \frac{1}{\sum_{i=0}^{N-1} r_i} \sum_{i=0}^{N-1} r_i x_i$$

and

$$\bar{y} = \frac{1}{\sum_{i=0}^{N-1} r_i} \sum_{i=0}^{N-1} r_i y_i$$

where $r_i = \sqrt{x_i^2 + y_i^2}$ is the distance of the data point from the origin of the data set, i.e., the position of the laser range finder. The data set is then transformed to the set $\{X'_i = (x_i - \bar{x}, y_i - \bar{y}) : 0 \leq i < N\}$ relative to the centroid. To compute the mass of the region an ordered data set is assumed, i.e., each data point X'_i is required to represent a smaller bearing angle α'_i as its neighbour X'_{i+1}. This allows estimation of the area m bordered by the data set to

$$m = \left(\sum_{i=0}^{N-2} m_i \right) + m_{N-1},$$

with

$$m_i = \frac{1}{2} tan(\alpha'_{i+1} - \alpha'_i) {r'_i}^2$$

and

$$m_{N-1} = \frac{1}{2} tan(\alpha'_{N-1} - \alpha'_0)(r'_{N-1})^2$$

where r'_i is the distance of the transformed point from the centroid. Since this estimated covered area is depending on objects that are placed in the region it represents an index of clutter, which is helpful to differentiate between regions of the same basic layout, but with different furnishing.

[5] as a result of the equidistant angular resolution with which a laser range finder scans the environment objects in the direct vicinity of the laser range finder are represented with considerably more data points than objects that are further away

In order to obtain $l1$ and $l2$ a principal component analysis (PCA) has to be performed. The principal components correspond to the two eigenvectors E_1 and E_2 (to the corresponding eigenvalues λ_1 and λ_2) of the covariance matrix Q with

$$QE_i = \lambda_i E_i, \quad i = 1, 2$$

where

$$Q = \begin{bmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{bmatrix}$$

and

$$C_{XX} = \frac{N}{(N-1)\sum_{i=0}^{N-1} r_i} \sum_{i=0}^{N-1} r_i {x'_i}^2,$$

$$C_{YY} = \frac{N}{(N-1)\sum_{i=0}^{N-1} r_i} \sum_{i=0}^{N-1} r_i {y'_i}^2$$

and

$$C_{XY} = C_{YX} = \frac{N}{(N-1)\sum_{i=0}^{N-1} r_i} \sum_{i=0}^{N-1} r_i x'_i y'_i.$$

The covariances also have to be weighted due to the non-uniform sampling of the laser range data set[6]. Linear weights r_i are used, interpreting the original distances as the factor responsible for the *distribution* of the data samples around the laser range finder, which have to be compensated for. The two features $l1$ and $l2$ are now estimated as the maximum distances represented in the data set along the bearing angles of E_1 and E_2. To make sure that such a point is found, a tolerance threshold around the bearing angle is employed. The data set is now represented as $reg = (name, m, l1, l2, cX, cY, \beta)$, with cX and cY being the coordinates of the data sets centroid in the global coordinate frame and β the respective bearing angle of the main axis E_1, and is stored as a basis for comparisons. The very descriptive excentricity e can then easily be computed as

$$e = \sqrt{1 - \frac{l2^2}{l1^2}}$$

if necessary.

The feature based representation can be used for two types of comparisons, one of which is used to recognise or categorise a particular region. The results obtained led to the third conference publication this thesis is based on (Topp and Christensen, 2006) and are discussed in detail in chapter 5, section 5.3. The other

[6] the weighted variances are computed according to the National Institute of Standards and Technology's collection of formula at
http://www.itl.nist.gov/div898/software/dataplot/refman2/ch2/weighvar.pdf
(URL verified: June 10, 2008)

a) b)

Figure 4.5: *An illustration of a structural ambiguity. A corridor passes a hall, both are very different in their spatial/geometrical properties and probably also in their semantics, but there is no obvious "gateway" between them.*

comparison is a continuous check against a previously acquired representation to find transitions from one *region* to a neighbouring one. These transition detections and also what will be termed a "structural ambiguity" and was introduced as an "ambiguity of space" in chapter 3 are discussed in the context of the representation in the following.

Detecting structural ambiguities and transitions

This section focuses on structural ambiguities which are considered to be particular areas in the environment that appear quite differently regarding the used representation, that might or might not be parts of the same *region*. As an example can be named the combination of a rather narrow corridor with an entrance hall, where the transition between those two areas is not obviously marked by doors. Figure 4.5a) illustrates such a situation. A human user might want to distinguish between "corridor" and "hall" which needs to be reflected in the robot's representation of the environment. Obviously, a door detection would fail to segment the two *regions*, and the option of generating a "virtual door" for the map might be confusing to the user – it is probably very difficult to tell where exactly the delimitation between "hall" and "corridor" should be placed, and where the door passage should be assumed. The previously described approach of representing *regions* based on their spatial properties allows here to at least generate two different *regions* with maybe some unspecified or partly overlapping area in the middle, indicated by the question marks in figure 4.5b).

Given the approach to representing *regions* with the help of a small feature set it seemed quite natural to apply it to continuously compute a hypothetical representation while travelling through an environment and compare it to previously specified ones. This enables the system to detect transitions between areas with significantly different spatial properties, which can be used to generate clarification questions. Such questions are needed if, e.g., as illustrated by figure 4.5, the user

omits to mention the "corridor" during the "home tour", but has mentioned the "hall", which makes the robot wonder in the corridor, if this *region* should still be called "hall" since it appears fundamentally different.

Given the "home tour" scenario it has to be assumed that the range data sets will always contain the pattern of the human user's legs, more or less close to the sensor. This does not disturb the computation of the current representation if the user does not cover too wide an angular range completely. In situations where the robot is standing still though this happens every now and then. Thus, the system has to compensate for false alarms resulting from the distortion of the data sets generated by the user. Instead of comparing every available data set to a previously obtained representation it is assumed that the change has to be stable over a certain number of data cycles. Additionally it can be safely assumed that the robot cannot have entered a new (hypothetical or previously defined) *region* when it has not moved. Those two conditions allow to lower the computational effort and make the system more stable.

One question is to which previously generated representations the current – hypothesised one – should be compared. An option is to compare only to the representation that was last accepted as current one. In this case the system does not make use of previously acquired representations and cannot be used for the recognition of already actively specified regions. Comparing to all available *regions* to find the most likely current surrounding *region* is rather unnecessary and comparably expensive. Thus, a hybrid approach is used to deal with this situation. The currently hypothesised new *region* representation is compared to the previously accepted current one (which can be either a *region* specified by the user or an internal reference representation for the detection of changes that might or might not be handled in a respective dialogue).

In the case that a significant difference is detected, the representation is checked against all other available *region* representations, testing whether any of them matches sufficiently, to find out if a previously specified *region* has been re-entered. If this is the case, the matching *region* is hypothesised as current representation to compare to in further steps, otherwise a new representation is generated. When a *region* is specified actively by the user this *region* is assumed to be the current one immediately.

To decide if two *region* representations are sufficiently close to each other, a distance measure d is computed from the relative differences in each of the descriptive features:

$$d \quad = \quad \sqrt{\hat{m}^2 * \hat{l1}^2 * \hat{l2}^2 * \hat{e}^2}$$

with

$$\hat{f} \quad = \quad \left(1 - \frac{f_{hyp}}{f_{cur}}\right) \quad for \quad f \in \{m, l1, l2, e\},$$

with f_{cur} and f_{hyp} standing for the respective feature of the current and the hypothesised representation.

If the distance measure d exceeds a threshold a significant change in the environment representation is assumed and handled accordingly. Within the framework for Human Augmented Mapping this means, that the user is asked about the current situation and can thus help the system to proceed, as was described previously in chapter 3.

Two ways of updating a *region's* representation are considered, in case that a hypothesis for a detected transition is erroneous. One option is to compute a *region* representation as average of all available representations including the new hypothesised one, the other option is to have several different stored representations for each *region* to choose from. Since the classification performance test for the representation approach showed that this clustering method performs slightly better (see chapter 5, section 5.3), it seems the most useful way to proceed for continuous comparisons as well, assumed that not only subsequently generated representations are to be compared. The results that could be achieved with an experimental implementation of the transition detection led to the fourth conference publication this thesis is based on (Topp and Christensen, 2008). These and more interaction oriented results regarding the integration of the environment representation including the transition detection into a more complex interactive framework are discussed in detail in chapter 5, section 5.3.

The previous section focused on how *regions* can be represented and how the representations can be used to detect transitions and structural ambiguities in the environment. Another issue is in fact the representation of *locations*, which is described in the following.

Representing *locations*

The *location* is the second central concept to the HAM framework as it has been developed and implemented for this thesis. Nevertheless, the focus particularly of the implementation has been the representation of *regions* and their use as nodes in a topological graph structure. A *location* describes what in other works has been termed a "view" (Krieg-Brückner *et al.*, 1998), or a "snapshot". Within HAM a *location* usually incorporates a large object, e.g., a dinner table, which marks a particular work- or search space for the robot. This means that a useful representation (in terms of providing services, e.g., picking or placing a cup) of a *location* needs some kind of 3D- or at least $2\frac{1}{2}$D-information which can not be provided with a single (2D) laser range finder data set.

Corresponding to the representation of *regions* would be the idea of assuming the range data set obtained at the point of view when the *location* is specified as a descriptor for this *location*. Similar features as for the representation of *regions* can be extracted to allow for matching to decide, where with respect to a particular *location* the robot is located inside a *region*, given that angular (pose) information is

also observable. This would allow to represent also *locations* as an area independent from the exact pose they are specified at. Still, this would by no means enable a robotic system to actually perform a task, but it would allow to get close to the spot from where the task can be performed, given appropriate near navigation and mobile manipulation abilities.

However, for the work presented in this thesis, particularly regarding the implementation, *locations* are represented as a pose (x, y, θ) which gives the idea of the "view" but does not implement the flexibility of getting "somewhere close to the table" instead of going to a particular position.

4.5 Summary

Human Augmented Mapping (HAM) integrates human robot interaction with robotic mapping. This chapter proposed a hierarchical model that aims to reflect the theory of partially hierarchical representations of space in humans as connection between those two fields. Central (spatial) concepts of this model are assumed to be *floors*, *regions*, *locations* and *objects*. The applicability of the model, focusing on its two most central concepts *regions* and *locations* and its correspondence to the intentions of people interacting with a mobile robot was tested and supported in a user study setup that is described in detail in chapter 6. Consequently, the model could serve as the basis for the topological graph representation assumed as the main component of the mapping subsystem of the general architecture for a HAM system suggested in chapter 3, and a respective approach to the topological graph representation of arbitrary indoor environments with *regions* and *locations* was proposed and explained.

The mentioned user study also generated further questions regarding the correspondence of the proposed (spatial) concepts used in the proposed hierarchical model to observable human presentation strategies for particular items in a "guided tour" setting with a mobile (service) robot. These questions were discussed briefly in this chapter and informed the second user study design described in chapter 6.

A large part of the chapter dealt with the actual segmentation of indoor environments into the topological graph structure, using the concepts of *regions* and *locations* that were considered most central for an implemented system to represent. A concise laser range data feature based description for *regions* and its use for both the classification of *regions* and the detection of transitions between them was suggested. The actual implementation and empirical evaluation of the presented graph structure and environment segmentation based on the proposed *region* representation is subject of the following chapter.

Chapter 5

Empirical studies

The concept of HAM as presented in the previous chapter requires a number of functionalities and components which in itself can be compounds of several modules. In order to test the idea of HAM and explore requirements and limitations in a realistic context a prototypical system was implemented which links a mapping subsystem with functionalities for interaction. The mapping subsystem implements the graphical environment model as it was described in chapter 4. As far as the interaction abilities are concerned, the implementation focuses on the navigation related part of the interaction – the tracking and following component – and relies otherwise on a graphical user interface. The development of an adequate full dialogue management and processing component was considered beyond the scope of this thesis, however, in an integration effort within the project COGNIRON the mapping subsystem was transferred to an integrated interactive system to exploit the advantage of having natural language controlled interaction.

This chapter explains a standalone implementation with particular emphasis on both the tracking and following functionalities and the mapping subsystem. The evaluation of the particular components was also reported in several already published conference articles (Topp and Christensen, 2005, 2006, 2008) while in case of the integration reported in section 5.4 and a summarising report regarding the mapping subsystem respective articles are submitted for review.

An initial version of the implementation (using a simplified graph structure, focused on data recording, and combined with tracking and following) was used in the user studies described in chapter 6 which gives evidence of the applicability of the proposed models and implementation approach for the given context of a tour scenario.

5.1 An implementation for empirical studies

For an initial, experimental implementation of the HAM framework it was assumed that the modules responsible for the user tracking can be seen as driving components

Figure 5.1: *Minnie, the Performance PeopleBot as it was used in several tests*

for the system. Most of the robot's large scale tasks and actions are assumed to be initiated and controlled by interaction with a human user. Nevertheless, since the mapping process has to run concurrently to all other activities and the system needs to adhere to general principles of navigation as, for example, obstacle avoidance, a simple sequential processing of sensor readings and commands is not possible.

Chapter 3 of this thesis described the requirements for a human augmented mapping system on a high level of abstraction in terms of functionalities observable for the user. The rather implementation oriented requirements described here relate to low level control issues and the physically available robotic system.

The PeopleBot "Minnie"

The implementation work was conducted on the robot "Minnie", a Performance PeopleBot commercially available by MobileRobots (formerly ActivMedia). Figure 5.1 shows the robot as it was used for the studies described in chapter 6. Since no particular modifications were applied to the robot and the hardware is controlled with the help of the hardware abstraction software Player[1] the implementation can be assumed to be portable (as a whole or in parts) to other robotic systems, given that the following requirements are fulfilled.

- **Range data.** The tracking system as well as the mapping system rely on laser range data, provided in a plane at approximately knee height.
- **Position readings.** For the mapping process as well as for navigation tasks odometer readings need to be accessible (pose in 2D).
- **Interface.** Commands must be conveyed to the system as well as feedback to the user has to be provided. Typed input/output can be sufficient already to control the system as an operator.
- **Motor control.** Access to a motor controller is needed for a full "stand-alone" implementation.

[1] playerstage.sourceforge.net (URL verified June 23, 2008)

An integration of parts of the system into a complete interactive framework architecture was part of a project cooperation and is discussed later in this chapter in section 5.4. In the following the control system for a full implementation on the robot "Minnie" is presented.

Software packages

The implementation is based on a number of software packages that are either available as Open Source packages, test/research licences or developed as working group internal packages.

Player/Stage

Player/Stage is an Open Source project[2] providing hardware abstraction and basic robot control (Player) as well as a simulation tool (Stage). Recently the packages have been extended to also include basic services for navigation and map building to enable research groups using standard robotic systems to start out with some running system. The package is able to handle various types of robotic platforms and sensor configurations, which makes it attractive and easy to use.

The packages are used in the presented project for hardware abstraction and platform control. None of the standard methods provided by the package have been investigated so far.

Qt

Qt is a well established C++-library for the implementation of graphical user interfaces and visualisation tools[3]. Qt offers an easy to use signalling mechanism, which made it attractive for the implementation, since the communication with the user/operator is handled with the help of textual of graphical interfaces. Since Qt is only available freely for research institutions the functionality is kept separated from the essential modules as far as possible. Communication channels can be exchanged by other mechanisms if necessary.

CURE

The acronym CURE stands for "The CAS Unified Robotic Environment" and is the name of a C++-based software library providing utility algorithms for robot control. The library has been developed as a toolbox at the Centre for Autonomous Systems. Initially hardware abstraction was not integrated but has been included recently. For the presented implementation the integrated SLAM-packages and a number of navigation tools were used, together with the required data format classes.

[2] http://playerstage.sourceforge.net (URL verified June 23, 2008)
[3] http://trolltech.org/products/qt/ (URL verified June 23, 2008)

Architecture

Throughout the years a number of different design principles and architectural prototypes have been developed and postulated. More than 20 years ago an at that time revolutionary approach was presented by Brooks, who assumed for his "Subsumption architecture" layers of independently working, purely reactive behaviours (Brooks, 1986). This concept made a central planning component superfluous, each of the layers ideally was to maintain its own sense-plan-act (or rather sense-react) cycle. Such centrally controlled systems had been developed as deliberative architectures.

Brooks' approach worked nicely for the first three layers that were actually implemented, but turned out to be not as easy to expand to further layers as assumed. In addition it would be difficult to use any of the behaviours in different modes, which might be appropriate for complex, interactive systems.

A compromising solution was proposed by Arkin (1990). He combined the advantages of reactive behaviours and deliberation in his hybrid-deliberative architecture. Different - themselves reactive - behaviours (he called them motor schemas) were chosen deliberately, depending on the situation. A completely reactive component (a short cut connection from sensor system to motor control) was used as panic shunt.

Similar to Arkin's prototypical architecture the system presented for HAM implements a hybrid-deliberative design, also using a short cut connection that allows the motor control component to interpret sensor readings directly is established to enable "emergency" braking. The other system components represent a number of functionalities between which a central control module can switch according to the requested task. Since the system is not implementing different navigational strategies that have to be chosen autonomously, the deliberation is done by giving exclusive rights to the respective functionality. Figure 5.2 gives an overview of the system components and connections between them, which will be explained in detail in the following.

Support modules

A number of supporting modules that do not directly contribute to the functionality requirements for a Human Augmented Mapping approach, but which are needed to get a prototypical system running, will be described in the following.

Central control

The coordination between different software components is handled by a central control component. The CentralController connects supporting (resource) components to the receiving (interpreting) components. Also commands from the graphical interface or the console are passed on to respective modules and components. The central control component represents thus an event manager or basic planning component for the system. Using such a central component creates often

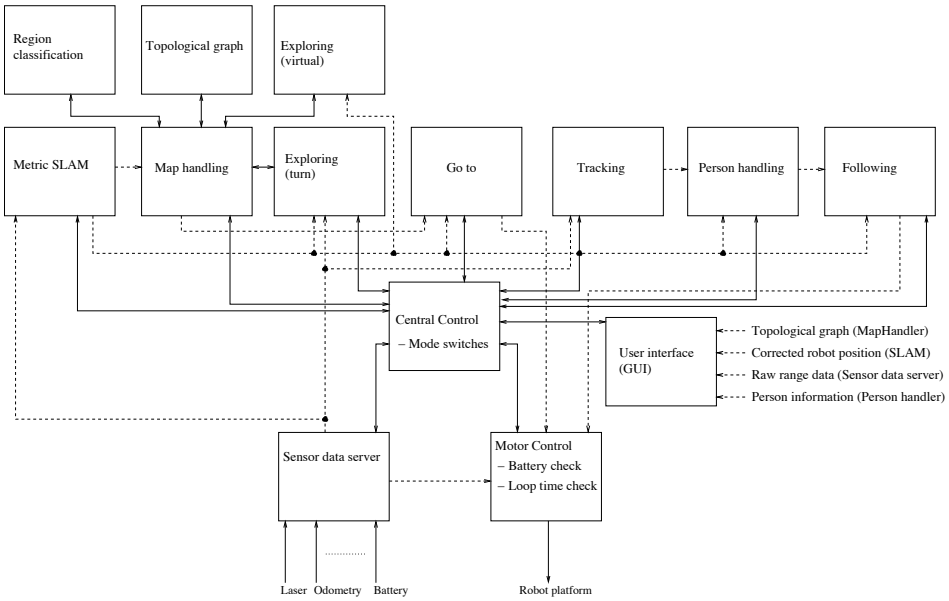


Figure 5.2: *Overview of the implemented system. Solid lines represent immediate connections, dashed lines represent communication links established by commonly used data containers (crossing lines are not connected, actual junctions are marked with black dots; the respective connections are still unidirectional). The connections from the different modules to the GUI are only sketched to keep the number of connections to a minimum.*

some sort of bottle neck for the event management, but since events do not come in too frequently it seemed an appropriate choice for a prototypical system. This could be assumed as the general data flow is not handled by this component but specific events (an incoming user command, a critical state message from another component) only.

The data server

The CURE library had no hardware abstraction incorporated when the implementation work was started. On the other hand the Player/Stage software provides hardware abstraction but produces data in a different format than the CURE library components expect. For this reason, and also to have the system working as well from recorded data, a general data server has been introduced as a mediating module between data collection and data consumption by the respective software modules. The central control component links the data server directly to all consuming components, which allows those to run in separate threads and collect data

in pull-connections whenever needed. The data server buffers sensor readings until a new round of data can be offered by the system and keeps track on time stamps to provide data as recent as possible.

Motor control

The motor control component is besides the data server the second tool that has access to the robot platform via Player. Direct connections are established to the system components that are responsible for goal setting and other navigation issues, e.g., following a person. To overcome timing issues that might cause problems when goal points or velocity settings are not updated properly, the component has a watchdog functionality. Whenever a critical time limit has been exceeded waiting for new instructions and the robot has been set to move, the speed is successively reduced to prevent dangerous situations. Additionally, an emergency obstacle avoidance is applied with a short cut connection to the data server. The platform is slowed down or stopped whenever sensory readings suggest that the robot is heading toward an obstacle closer than a certain safety threshold. Such mechanisms might prevent the robot from fulfilling a certain task but technical problems in other components do not cause danger to either the robot or people in its vicinity.

GUI

A simple graphical user interface is connected to the central control module to convey commands and information from the user to the system and to visualise internal processes for the user or an operator. The interface can be and was used to control the robot remotely in a "Wizard-of-Oz"-setup. In cases in which the graphical surveillance of the system is not necessary it is possible to switch to a purely text input based control tool. Parts of the commanding functionalities could be replaced by a speech recognition and processing system, as was done with the integration of larger parts of the system into a different communication and interaction framework (see section 5.4 for details).

Data containers

Some of the functional modules produce data to be displayed or sent to other modules. For the thread safe transport of these data a number of container classes has been implemented each of which provides recent data. The data are written by the respective modules and the container emits a signal that it has been updated, which can cause other modules to read the data. For example, the SLAM module writes its graphical map-information (lines) into an instance of a SlamData-Container and triggers the container to send out an "update" signal. Since the graphical interface is responsible for displaying relevant data, it is connected to the SlamData's update-signal and reacts by updating the display's properties according to

the changes in the container. Each module that writes or reads from a data container has to block access for all other modules. Since the data in the containers are copies, mostly needed for display- and diagnose purposes, no data are lost in case access cannot be granted once in a while. This method of communication is easy to replace by other communication tools, as could be shown with the transfer of the mapping subsystem to the integrated interactive framework for the project demonstration mentioned previously.

Specific components

The components relevant for the functionality provided by the system so far are described as specific components in the following. The components are roughly separated into groups dealing with low level data interpretation (feature computation), higher level situation interpretation (interaction monitoring, topological mapping), and robot control (following, navigating, exploring).

Data interpretation - Tracking

One of the contributing articles of this thesis relies on the results that could be achieved with the tracking module. These results are therefore described in more detail in section 5.2.

Data interpretation - SLAM

The SLAM-component is part of the CURE-library. It could be seen as a supporting component but since it is directly contributing to the complete mapping part of the system, it is named in this context. Based on raw laser range data and odometer readings delivered by the data server, the SLAM component extracts features (lines) and computes the actual position of the robot with respect to the starting point. This geometric framework is used later to provide the system with an accurate pose estimation to generate useful, geometrically defined links in the environment representation.

Situation interpretation - Person handling

The tracking module has no particular notion of its targets' relation to each other or to the robot. It purely delivers distinguished targets (numbered) to any consuming module. The interpretation in terms of, for example, which target might be the current "user" of the system, or whether a target is a person or some static, person-like object, is left to the person handling module. So far it classifies targets depending on their trajectories as

- STATIC_TARGET: This is the initial state, as all features that match our pattern classification are assumed to belong to a potential person of interest.

- WALKING_PERSON: As there are not any other means of classification for the user right now, more information is needed. Thus, the state for a target is set to WALKING_PERSON, whenever a certain distance was covered by it in relation to its initial point of detection.

- USER: To determine the user a simple rule is used: The closest WALKING_PERSON within a certain distance and angular area relative to the robot is assigned the USER flag. Only one user at a time can be present and once a person target gets the user flag, it will keep it, until it disappears from the scene.

- GONE: A target that has been removed from the set of targets is set to GONE in the person handler. This allows higher level processing of this state, for example, producing an error message when the user target is lost.

Situation interpretation - (Topological) mapping

For the topological mapping module a number of components are necessary. The central one is the MAPHANDLER, in which the topological graph is generated and maintained. The mapping subsystem is described in detail in section 5.3 together with the evaluation results obtained with particular components, which could be published in two of the conference articles contributing to this thesis. The mapping subsystem is also designed to be sufficiently self contained so that it could be easily decoupled from the complete framework to be transferred to a different robotic system, running with completely different communication and data access tools. This transfer is discussed later in this chapter.

Robot control - Following

When the command to follow is issued and the person handler can deliver a "user" target, the following component computes a desired goal point in a certain distance from the user that is passed on to the motor control component. Since the complete system is dynamic, i.e., both the robot's pose and the target destination are changing over time, the goal is computed (updated) continuously. The following component has no notion about the current configuration of the environment, the goal point is given as result from a straight interpolated connection with the target. The near navigation and obstacle avoidance are thus left to the motor control component.

Robot control - Exploring by turn

In order to capture a 360° range data set for the representation of a *region* the robot has to be turned around on the spot. When a respective specification is passed on to the map handler, it invokes the explore component's turn functionality. Other strategies of exploration and data collection could be implemented as well, but

the turning strategy is most convenient in terms of path planning and obstacle avoidance. The way of exploring could also be shown helpful for the interacting user, as will is discussed in the description of the pilot study conducted with the system (chapter 6, section 6.2).

Robot control - Virtual exploration

This module actually refrains from controlling the physical robot, but since it is closely related to the functionally of exploring by turn it is grouped into the robot control part of the system. For the continuous checks described in chapter 4 it is – at least with the currently available set of sensors – necessary to use virtual scans to obtain 360° range data sets. The VirtualExplorer maintains a local map and computes from this map an estimated set of data points in the back of the robot and matches this set with an actual scan from the laser range finder in the front. Thus, as soon as the robot has travelled at least a couple of meters or has turned around once, this technique provides a sufficiently precise estimate of the robot's current surroundings.

Robot control - GoTo

The current system can make use of the links between *regions* and the navigation graphs within them (see 5.3 for details) to navigate back to any known *location* or *region* (in the latter case the path is generated to a known – visited and thus probably reachable – position within the *region*). With the help of an A*[4] implementation provided by the CURE-library, the shortest path on the graph to the desired goal node is computed. This allows the robot with very basic methods to appear rather functional for user tests and studies (see also chapter 6).

In the following the particular parts of the system relevant to the tracking and following abilities and the mapping subsystem are discussed and evaluated.

5.2 Tracking for following – implementation and evaluation

In chapter 3 a number of issues regarding the abilities of a tracking approach for a Human Augmented Mapping system were discussed. Here the actual application used in the standalone implementation of the HAM framework is described and evaluated.

[4] A standard graph search algorithm (see also Russel and Norvig (2003, pp97–101))

Detecting people in laser range data

A common method to detect humans in laser data is to look for leg hypotheses, as done by Feyrer and Zell (2000), Kleinehagenbrock *et al.* (2002) and Schulz *et al.* (2001). The laser range data are analysed for leg sized convex patterns, either one of them or two at a reasonable distance from each other. Other systems rely on body shape as presented by Kluge (2002), or in previous work (Topp, 2003). In these cases a single "person sized" convex pattern is extracted from the data as a person hypothesis. The choice between the two approaches is often determined by the height the used laser range finder is mounted at. It seems that accepting leg patterns only is a rather strong constraint, as in this case a person wearing a skirt or baggy trousers would not be classified as person. Therefore three types of patterns are allowed in the implementation. These patterns can be classified as single leg, (SL), two legs appropriately separated, (TL) and person-wide blob, (PW). As accepting these patterns all the time would potentially generate a large number of false alarms, a rule based approach was adopted for the generation of new person hypotheses,

- TL and PW are accepted as features at any time they occur,
- SL are only accepted when they are close to an already detected and tracked target.

The latter constraint is based on the observation that a single leg pattern can only be seen for a short period of time when the leg of a moving person occludes the other. Therefore all other SL patterns are ignored, as they are unlikely to belong to a person. On the other hand the SL pattern is needed for a smooth tracking of the targets that have already been accepted.

Tracking and feature association

As outlined previously in chapter 3 SJPDAFs according to Schulz *et al.* (2001) are used to associate targets and features in a probabilistic framework. Each feature $z_j \in \{z_0, z_1, z_2, \dots z_n\}$ is assigned a posteriori probability β_{ij} that it was caused by the target $x_j \in \{x_1, x_2, \dots, x_m\}$. The feature z_0 represents the case that a target was not detected at all. The computation of the β_{ij} is based on a sample representation for the targets. Each target x_i has its own sample set for state prediction and is updated according to β_{ij}.

The sample space is composed of the state (x,y,v,θ) of their respective target, where (x,y) refers to the position, v is the translational velocity and θ the orientation relative to the robot. A first order Taylor expansion is used for the motion estimation.

This data association method is meant to handle a fixed number of targets. In the context of Human Augmented Mapping it can be expected that only very few new targets would enter or leave the scenery at exactly the same time, thus the method still seemed a valid way of solving the association problem.

Interpreting and using the tracker results

As mentioned above, the interpretation of the tracking results is handled by another software component, the PersonHandler. Depending on movements, covered distances and the pose relative to the robot the tracked targets are assigned flags such as "static", "moving" and "user". The user flag can only be assigned once at a time. Such a simple rule based decision does obviously not allow for sophisticated reasoning about people in the vicinity and their willingness to interact with the robot. Since the classification is done in the scope of this thesis work for a limited purpose - allow a particular person to draw the robot's attention, this lack of "natural" interaction abilities is not considered a problem.

Following and passing persons

In the currently implemented system the tracking system is used for following but not for passing persons. An attempt to integrate the tracking system with a method for the appropriate navigation in the vicinity of humans (Pacchierotti *et al.*, 2005) showed that also for this purpose in the initial phase the reliable tracking of one target was more important than the knowledge about different ones. Still, it seems possible to use the multiple target tracker when appropriate navigation becomes an issue. For the following the multiple target tracker can handle occlusions and crossing persons much better than a single target tracker would have been capable of.

Results from experiments

The tracking method was tested in three different scenarios. One test setting was a pure performance test for tracking of multiple targets in an artificially emptied "room". The other two reflected the behaviour of the tracker in a "real world" context, given the guided tour scenario. As differences in the quality of the results could be observed, the two test types are described separately, referred to as setup #1, #2, and #3 respectively.

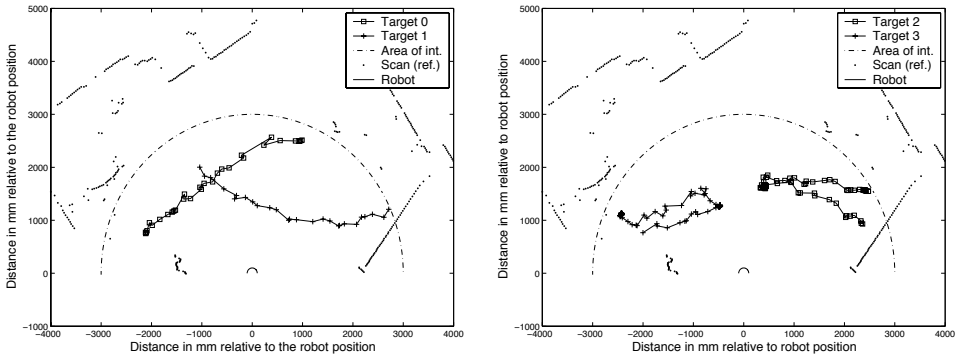
Experimental setup #1

In order to make sure that the number of persons present was controllable at any time during experiments, an empty area ("room") was generated by setting up a number of large plywood planks and cardboard pieces as walls for the experiments that involved a moving robot. A number of test cases was specified as follows:

1. robot not moving, one person present,
2. robot not moving, two persons present, occluding each other,
3. robot moving independently, up to three persons present, and
4. robot following one person.

With these tests it was aimed to test the tracker under different test conditions. Regarding the previously mentioned quality measurements the main interest was directed to problematic situations that might lead to confusions or the loss of a target. Therefore the participating test persons (co-workers who were willing to spend a couple of minutes walking in front of a robot) were asked to walk at different speeds, cross each other's trajectories in the field of view of the robot on purpose, "meet" in the middle of the room, "chat" and separate again, or perform unexpected changes in their moving direction. The laser range finder was set to a data transmission rate of 38400 baud to guarantee stable transmission and to determine, if this speed was enough for the purposes of tracking and following. During the tests all occlusions were handled correctly and no target was lost. This result could be confirmed by different tests under similar circumstances with the same models for movement and state prediction.

Robot still, one person: In this test scenario one person crossed the field of interest (in this case the area described by the laser range finder baseline (x=0 in the robot's coordinate system) and a radial distance of three meters) nine times, at varying speeds. The target was not lost at any time. It was always classified as a moving person and was assigned the user flag when entering the area where a user would be expected.



a) b)

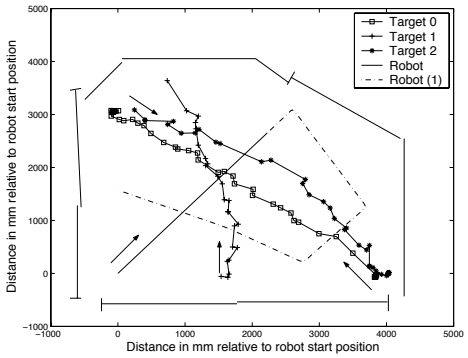
Figure 5.3: *The trajectories for a test with two persons moving in the field of interest, with the robot standing still. The dashed line marks the area of interest, the small half circle at position (0,0) represents the robot. The dots show a reference scan of the environment. a) The two persons cross the area of interest, with target 0 being occluded by target 1. b) The two persons walk into the middle of the area, stop at a comfortable "chatting" distance (about 80cm) and separate again.*

Robot still, two persons: Two persons crossed each other's paths in front of the standing robot, went out of the area of interest and came back. They met in front of the robot, "chatted" and separated. Figure 5.3 shows the resulting trajectories.

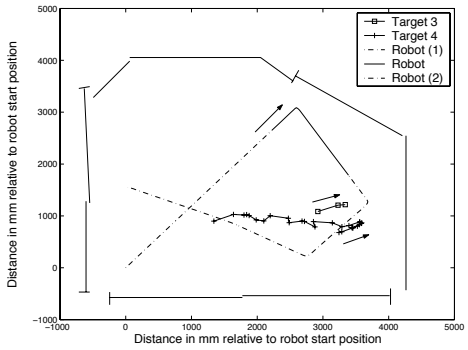
Again, the area of interest was set with a radial distance of three meters. In this case, the surrounding environment was the natural laboratory environment, but it was made sure that no disturbing objects were in the field of interest. This was possible, because the robot did not move. This test gives an example of the tracker being able to handle the short term occlusion of two persons passing each other. Both targets are classified as moving persons and the user flag is assigned to target 1, when entering the respective zone in front of the robot. The trajectory for target 0 seems to stop clearly within the area of interest, as the person gets occluded by some object indicated by the respective scan data points in the image. As the person does not come out of this hiding place for a while the system assumes the target as "gone". Even for the "chatting" scenario, the tracker could handle the situation, which shows that if two targets get close to each other, but are clearly distinguishable no confusions occur. Again, one of the targets (target 2) gets the user flag as it enters the respective zone first.

Robot moving, three persons: This test was the most relevant for the purpose of "following in the presence of bystanders", as it shows the abilities of the tracker running on the moving robot, together with the target classification that would make the robot follow one of the persons. Figure 5.4 shows the resulting trajectories from one of the tests covering this type of scenario, with three persons moving around while the robot is crossing the area. For this particular test the area of interest was set to a radial distance of eight meters. This means, that the whole "room" was in the field of interest. The robot moved straight across the area until it detected one of the walls at a certain distance. It turned then randomly to the left or to the right until it had enough free space in front to continue. In the first part of the scenario the user flag is assigned to target 1. This happens due to its proximity to the robot when it enters the "user zone". For the second part of the test target 3 did not remain visible long enough to be classified as a moving person, but target 4 is classified as moving person and user. For the last part target 5 is found as user and keeps the flag while it is present. The targets 6 and 7 are classified as moving persons, but not as user, as target 5 is still around. When target 5 steps out of the field of view, the user is lost, but immediately afterwards the newly arrived target 8 is classified as user.

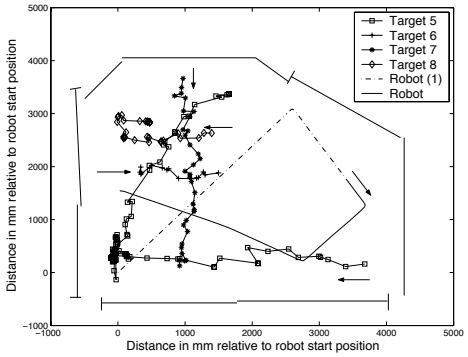
Robot following one person: To show the tracker's ability in a following scenario, the system was set in the respective mode and followed one person for about three minutes. During this time period the user changed her walking behaviour (speed and direction) frequently, sometimes came very close to the robot, so that it had to move backward, and stepped close to the walls of the empty room used in this experiment again. This test over a period of three minutes shows, that the



a)



b)



c)

Figure 5.4: *Person and robot trajectories in the x/y-plane of an experiment with up to three persons in the area at the same time and the robot moving independently. The resulting trajectories are presented in three images for clarity of presentation. The robot trajectory is shown as a solid line for the time period representing the events in the area. The rest of the robot's trajectory is shown dashed.*

a) Three persons cross the room in different directions. The "long steps" in the trajectories (in the starting steps for target 0, in the middle part of target 2's trajectory and in the last steps of target 1) occur due to occlusions. b) When the robot turns in the upper corner of its path, it loses the recently detected target 3 out of the field of view. Target 4 performs an indecisive behaviour by turning around and going back after a few steps. c) Target 5 remains in the scene for almost the whole time period shown in this graph, crossing the area from right to left, standing for a while in the bottom left corner and then continuing "up", being occluded by target 7 for a short moment.

first order motion model is able to handle arbitrary movements quite well, as the user was not lost at any time.

From these experiments it could be concluded, that under test conditions the approach can handle the situations of interest. Nevertheless, running the tracker with slight changes in the motion model on the same data sets for a number of times showed that there are situations in which the tracker fails, due to a seriously wrong

prediction of the further movement direction of a target in combination with a detection miss for the same target. This indicated that it might be useful to switch to a more sophisticated motion prediction model as derived, e.g., by Bennewitz *et al.* (2002) or Bruce and Gordon (2004).

Experimental setup #2

As it is impossible to assume clean test conditions for more general user studies and experiments, the tracker approach was tested on data collected during a comprehensive user study. The study was a Wizard-of-Oz experiment and is described in detail by Green *et al.* (2006a) and Hüttenrauch *et al.* (2006a,b). One important fact to note about this kind of experiment is that the robot was actually controlled remotely, while the test subject was told that the system performed autonomously. The scenario for the experiment was a guided tour through a "living room" (see figure 5.5 for an illustration). Subjects got the task to ask the robot to follow, present different locations and objects in the room and test the robot's understanding by sending it to learnt places and objects[5]. The study comprehends data from 22

Figure 5.5: *The experiment environment ("living room") seen from different perspectives*

trial runs. Laser range data were collected in all runs at a data transmission rate of 500000 baud, though due to a communication stability problem not all of the trials could be recorded completely. Still, a body of a couple of hours of experiment sequences could be collected, since every experiment lasted between 10 and 20 minutes. Figure 5.6 shows a raw scan taken from a typical start position during the tests.

[5]During this study the conceptual hierarchy for HAM proposed in chapter 4 was not yet available, thus the study subjects got instructions to present "places" and "objects" to the robot without any clear specification of what those terms actually meant. This turned out to confuse them in some cases and underlined the need for a clear terminology used in further work.

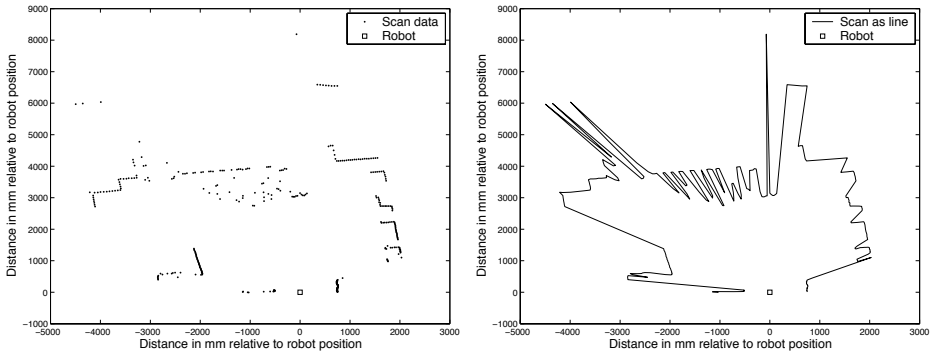


Figure 5.6: *The raw laser data (top) and the same data represented as polyline to show the data points in their angular order. The two peaks right in front of the robot are caused by the subject's legs, while the other peaks result from the table and chairs, that belonged to the experimental scenario.*

Running the tracking system on the data from the experiments showed, that performance in this kind of real world environment was significantly worse than expected after the results from the previously reported tests. The user target got confused with other targets rather frequently due to the problem outlined in the following paragraph.

As stated in section 5.2 static targets are allowed for the tracker, as this is reasonable in an interaction context. In fact, the experiments with the "inexperienced users" confirmed this assumption, as many of the subjects repeatedly stood still for quite a while (up to 50 seconds).

The images in figure 5.6 show a clear resemblance between some of the patterns and the subject's legs, even if some of them appear too pointy. Still, such patterns can fall under the classification thresholds for legs and a completely smooth representation for the targets' possible movement cannot be assumed (as this would conflict with the Sampling Theorem (Shannon, 1948) and the laser ranger finder's angular resolution). Therefore a target generated by a false positive (static) hypothesis detected in a number of data sets that is not detected in a consecutive data set due to robot movement and changing perspective might be incorrectly associated to a new false positive hypothesis that is close enough to the initial position of the erroneous target with respect to the motion assumptions of the tracker. Thus, the erroneously detected target(s) start to "move". The respective sample set picks up the motion estimation and predicts a new position. If the robot's viewpoint changes such that the "target" is not detected for a while, the predicted state gets more and more ambiguous. With the particles spreading toward the actual position of the user target and the appearance of a new (erroneous) target, the statistical approach is likely to confuse the feature–target association. As a consequence of

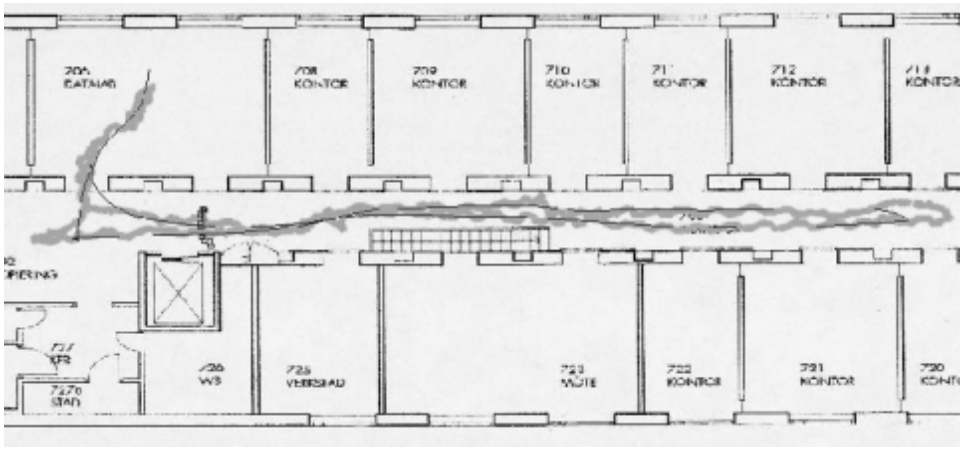


Figure 5.7: *The trajectories of the user (grey, thick line), the following robot (black, thin line) and a bystander crossing the way of the robot (small squares) between the robot and user.*

such a confusion, the tracker needs a few steps to recover, i.e. retrieve its certainty, which is even more difficult when the user stays close to distracting objects.

As the task for the subjects was to show the robot around in a furnished room, it is scenario immanent that the user moves around between objects in the room. On the other hand, it became obvious that in situations where the user was clearly distinguishable from disturbing objects, and those disturbing objects were detected reliably, the tracker and data association performed as expected. Occlusions were also handled properly in these situations.

Setup #3: Following through the office building

With this experiment it could be shown that the system is suitable for "real world" conditions, if the disturbances can be reduced to a still realistic minimum by the choice of environment. The robot followed the test person out of the laboratory and along the hallway, covering a distance of about 25 meters, and returned – still following – to the laboratory. On the way back a bystander was asked to cross the way between the robot and the user. Figure 5.7 shows the part of the office building together with the trajectories. The experiment took approximately four minutes and a distance of about 50 meters was covered, including two door passages. A total number of 26 targets was detected throughout the whole time period, one was accidentally classified as "moving", but did not get confused with the user. The user target was tracked reliably over the complete time period and one occlusion of the user by a crossing bystander was handled as expected. The bystander target

was classified as moving person correctly, so a respective person passing method could have handled the situation appropriately. In the test case the robot slowed down a bit, due to the influence of obstacles on the speed control.

Summarising these tests on "real world" data it was observed that

- the approach for tracking and data association is a valid method for tracking multiple targets in the context of following a user or passing persons.
- the approach is sensitive to motion models, but the choice of a good motion model does not seem to be as critical as the reliable detection of actual targets.
- problematic situations occur in "real world" scenarios, i.e., cluttered environments, when vicissitudinous false alarms lead to confusions.

These observations suggested to improve the system for following and passing persons by introducing other means for the detection of targets. Within the context of "showing the robot around" the system has to deal with an unknown, cluttered environment. From preliminary analysis of the user study can be alleged that persons in this context move differently compared to results from observations in long term experiments on a larger scale. Subjects tended to move to a certain position, stop and move around in a small area, to "explain" things to the robot. This type of movement seems rather stochastic, compared to the motion models that hold for long distance movements. Therefore improving the detection to eliminate confusing false alarms is a better way to improve the system for the given purposes.

An attempt to improve the detection method with the help of statistical data analysis is described briefly in the following.

Tracker improvements

Since the test results described above suggested to improve the tracking system by improving the reliability of the person detection, different methods of statistical data analysis were investigated in an undergraduate project (Platzek, 2005). The results showed, that with a supervised approach (in this case k-nearest neighbour) the classification of possible "leg"-patterns into actual legs and "non-legs" could be improved significantly. Using a rather small training set of human legs (with different types of trousers) and leg-like objects the ratio of false alarms to correctly detected legs on a test set of collected "real world" data could be clearly improved. However, due to technical issues, mainly the on-line- and real-time conditions the system for Human Augmented Mapping has to cope with, the improvement was not implemented so far as part of the complete system.

When later on a number of user studies were designed (see chapter 6) in which the tracking component was used for autonomous "following"-behaviour, it turned

out that even without the implementation of the suggested improvement the system's overall performance was reliable enough in the limited environment used for the study.

In the following the modules and components relevant to the mapping subsystem are discussed and evaluated.

5.3 Topological modelling – the mapping subsystem

In chapter 4 a model for the representation of arbitrary indoor environments was proposed. This model is used as basis for the implementation of a topological graph structure on top of a metric map obtained from a SLAM component. As mentioned previously in the description of the implemented software components this topological mapping component is a subsystem containing a number of components itself:

- Map handling: The component called MapHandler is the central one controlling the topological mapping process. This process includes regular checks whether a new *navigation node* (see chapter 4) has to be placed, including tests if a new / unknown *region* has been entered, but also the specification of new entries for the graph structure, when the user specifies new *regions* or *locations* according to the model in chapter 4.
- Region classification: The RegionClassifier[6] maintains a list of *region* representations that can be identified by name (thus it is possible to keep clusters of representations for one *region*) for comparisons to recognise a particular *region*. The component is also responsible for the "ad-hoc" comparisons between a current and a hypothesised *region* representation for the detection of transitions and structural ambiguities. It was decided to keep this extra list for easier access than would have been provided through the graph structure.
- Map graph: The actual graph representing the environment[7] with *region* nodes and links (*abstract edges*) between them.
- Node, Region, AbstractEdge, Location, NavNode, NavGraph, Edge: The elements (classes) of the complete graph structure as it was described in chapter 4. Both Region and NavNode are subclasses of Node which implements properties necessary for the A*-search (part of the CURE library, see above) used for path planning.

[6] due to "historical" reasons the actual software component is called "PlaceClassifier", and the representations generated and compared are called "Place", but they all refer to what is now termed a *region*.

[7] Also here the originally used term for the class description was "PlaceGraph", which was kept to avoid renaming issues

The MapHandler communicates with the help of "Data Containers" as they were described above, and explicit calls depending on the availability of new data from the SLAM component running concurrently. This made it very easy to decouple the subsystem from the overall framework to make it work with other communication and geometric mapping approaches. In the following the results that could be obtained with the different applications of the RegionClassifier are discussed, first the general segmentation of the *regions*, also in terms of the used approach's power of distinctiveness and determination, second the detection of transitions between two *regions*.

Region segmentation – implementation and evaluation

As mentioned previously in the description of the software components used for the full implementation of the HAM system, the mapping subsystem uses two different strategies to obtain the data sets for the computation of *region* representations according to the method proposed in chapter 4. For continuous comparisons (transition detections) virtual scans are used as described above. In the case of the *region* segmentation though it is assumed that data are provided in an explicit way, and thus the strategy for user specified *regions* is assumed in the following discussion.

For the acquisition of a respective data set for an explicitly specified *region* a turning strategy is used. The robot performs a 360° turn on the spot and gathers a specified number of range data sets (in this case two, one at 0° and one at approximately 180°), which are then merged according to the observed (corrected) pose at the time the data are collected. The descriptive features $(m, l1, l2)$ are computed according to chapter 4 and were evaluated as described in the following, which also led to the third conference publication relevant to this thesis (Topp and Christensen, 2006).

Evaluation

The method to represent *regions* proposed in chapter 4 can be evaluated in two different contexts. Firstly, one wants to know about the distinctiveness or the segmentation power of the features, i.e., it is necessary to know, how well the environment is described with *regions* that have been specified using the method described previously. The described feature sets can be used for a classification/categorisation approach to facilitate localisation. Here different issues have to be considered. One is that a metric SLAM method is integrated in the system that allows constant and exact localisation (in case that the system is not challenged with a "woken up" or "kidnapped" robot). What is thus interesting is the ability for the system to report its current position in the context of the graph it acquired, which means that also positions that are initially not included in the description of one *region* can be recognised as consistent with it. A second issue is in fact to facilitate localisation in a "waking up" or "kidnapped robot" scenario by reducing the search space for localisation on the basis of the metric map(s) or a topological exploration strategy.

Assumed that the system knows to be either in *region* A or *region* B, but definitely not in any other (known) *region*, the effort for exact localisation in large environments can be reduced significantly. The proposed method was thus evaluated in the two different contexts of distinctiveness and categorisation.

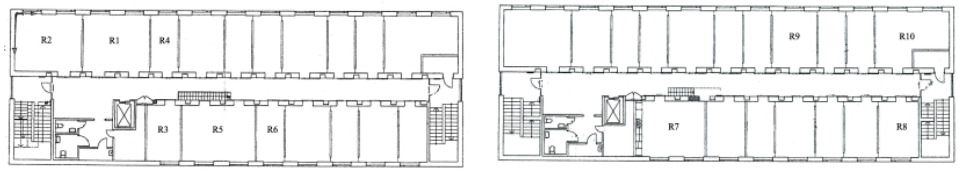


Figure 5.8: *The ten rooms of the office environment, that were used for the tests. Above is shown the plan of the floor that contained six of the ten tested rooms. The bottom plan shows the floor that contained the rest of the used rooms.*

Categorisation The approach to represent *regions* was evaluated in the context of classification and recognition of specific *regions* (in this case correspondent to rooms, which will be the term used in the context of this evaluation) for a number of rooms in the office and laboratory environment of the CVAP group at KTH. Figure 5.8 shows a schematic drawing.

In each of the rooms R1 to R10 at least four 360° range data sets were obtained at different positions in the rooms. Those sets of representations for each room were used in different comparison setups to find the best way of finally representing the rooms and indicate a useful strategy to update the representation if necessary. Such updates become interesting in the context of the interactive "home tour" scenario, given the previously described idea for the detection of structural ambiguities and transitions and the need to be able to accept different representations (spatial properties) for one *region*.

Looking at the picture it becomes obvious that certain groups of rooms can be identified considering the size. Within the groups the rooms are quite similar to each other as far as their size and shape are concerned. Since additionally the larger rooms correspond to robotic or vision laboratories and the kitchen, where the smaller rooms are offices and a workshop, they are also quite similarly furnished. Thus, it is not surprising that the results for the classification and recognition of particular rooms are not convincing. Table 5.1 shows the confusion matrix for one test setup with clusters. In this setup the obtained and stored *region* representations were compared to all other feature sets available. The nearest neighbour according to the description in the previous chapter was picked as the recognised *region* representation (room).

Test \ Answer	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
R1	50%	25%			25%					
R2					25%		75%			
R3			50%			25%				25%
R4			25%	75%						
R5	25%	25%					50%			
R6						75%			25%	
R7		30%			10%		60%			
R8								20%		80%
R9									60%	40%
R10			17%	33%		17%		33%		

Table 5.1: *Confusion matrix: Tests with clusters*

The overall recognition rate for this test was 40% which is clearly not sufficient for classification. For other test setups (using the average of the feature descriptions for each region or a one-shot presentation) similar low rates were observed. Table 5.2 shows the confusion matrix for an initial test setup where the test representation was compared to a single representation of each of the rooms R1-R7.

Test \ Answer	R1	R2	R3	R4	R5	R6	R7
R1	25%				41%		33%
R2		41%			16%		41%
R3			75%	25%			
R4			58%	41%			
R5	25%	25%			33%		16%
R6						100%	
R7	8%	25%			33%		33%
entries truncated, therefore not always a sum of 100% is displayed							

Table 5.2: *Confusion matrix: Tests with "one-shot" presentation*

Thus, it became obvious that a classification (recognition) of particular *regions* would not be feasible with the very simple approach proposed. Still, it was interesting to see if it would be applicable for a *categorisation* that could be used to facilitate localisation by reducing the search space.

Looking at the rooms in three groups or categories (i.e., "large open spaces" = R1, R2, R5, R7, "medium size cluttered/odd offices" = R8, R9, R10, and "small cluttered offices" = R3, R4, R6), a recognition rate of 88% within the groups could be observed. Here it has to be noted that the categories were chosen only from the roughly estimated size of the complete room, given the architectural sketch. It becomes obvious that recognition errors mostly occurred for one of the medium size rooms (R10). Now, this office has a considerably different shape (L-shape) than all other rooms available for the tests and is heavily cluttered with office furniture. Figure 5.9 shows a number of laser range scans matched (manually) to represent the perceivable area of R10. It becomes clear that this office can easily be perceived as several small, cluttered offices which were in fact the ones it got confused with.

Since the overall HAM framework assumes no prior knowledge of categories, a grouping of defined *regions* would have to be done according to a similarity measure. Such a measure could be the likelihood of confusing a particular feature set with another that belongs to a differently labelled *region*.

Grouping according to this measure (i.e., "large open spaces" = R1, R2, R5, R7, "medium size offices" = R8, R9, and "small/odd, very cluttered offices" = R3, R4, R6, R10) would result in a recognition rate of 94%. The remaining errors are mainly due to the fact that a previously correct "in group" recognition for R10 becomes an error by regrouping. These rates suggest, that it is in fact possible to give a rather strong estimate for the validity of a hypothesis for global localisation in terms of *categories* of rooms or *regions*. The author believes that this holds for most indoor environments in which at least two types of rooms can be found. The uncertainty for a global localisation in a "waking up" scenario could thus be reduced significantly before invoking either a metric localisation method or an exploration strategy to disambiguate the situation. Such strategies have been proposed already by Kuipers and Byun (1988) and have been investigated later also by Seiz *et al.* (2000).

Distinctiveness The other issue to be evaluated is the distinctiveness of the method. Given that a particular *region* is presented to the system the question is, how dependent the acquired representation is on the current position of the robot. Intuitively and along the argumentation of Kröse (2000) one would assume, that the data obtained in a simply structured (but not empty) convex room with only one door will be rather similar for different positions. Figure 5.10 shows such a room (R7) with the positions from where the 360° range data sets were taken (P1–P6). Additionally the positions where the system calculated the corresponding centre of the obtained laser range data are marked with grey dots and numbers 1–6. The thinner (blue) lines represent the line features extracted for the metric

Figure 5.9: *Several scans matched to fit the area of R10. Blue dots represent the laser range finder data points, the black lines correspond to the architectural sketch of this room.*

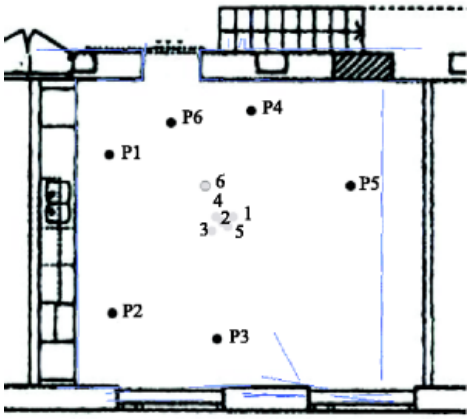


Figure 5.10: *One of the rooms (the kitchen – R7) with the positions (P1-P6) from where the data sets were obtained and the corresponding centroids (1-6).*

SLAM. The line along the upper wall is caused from a long sofa/bench. Other furniture in the room (tables and chairs) cause only scattered data points and are thus not relevant for the line extraction. Not surprisingly they all fall into an area of about 35cm radius, but one (no.6). This particular data set was obtained very close to and in the line of the doorway, where a significant portion of the corridor could be perceived already. Table 5.3 shows the mean and variance of the

Feature	Mean	Variance
"Mass" / area (m) [m^2]	21.23	2.73
Length 1 (major axis) ($l1$) [m]	8.45	4.37
Length 2 (minor axis) ($l2$) [m]	5.10	0.13
Excentricity (e)	0.71	0.04
Distance between centroids (D) [m]	0.34	0.05
Angular difference between major axes (A) [rad]	1.11	1.04

Table 5.3: *Statistical values for R7*

differences between features (or measures) calculated for each pair of consecutively taken positions on the path in R7. From those values it becomes obvious that for the

major part of nearly convex areas the position to acquire the features for this region is arbitrary. In the immediate proximity of doorways though the representation becomes slightly unstable. This is still acceptable when interpreted in the sense of a human environment representation, where a door passage might be a transition not only in the spatial sense and thus is difficult to describe in a binary way as strictly "inside" or "outside". In fact, for the generation of clarification questions in the context of transition detections it feels quite sensible to have an area in which the system's uncertainty gradually reaches and exceeds a threshold instead of having to deal with strong hypotheses and thus unnecessary questions frequently.

More interesting than the convex and nearly convex *regions* or rooms are actually those that are of particular shape or have a very distinct type of furnishing. This is in the given set of rooms the case for R8 and R9 (furniture) and R10 (shape and furniture). Figure 5.11 shows similar to figure 5.10 the positions (P1–P5) from

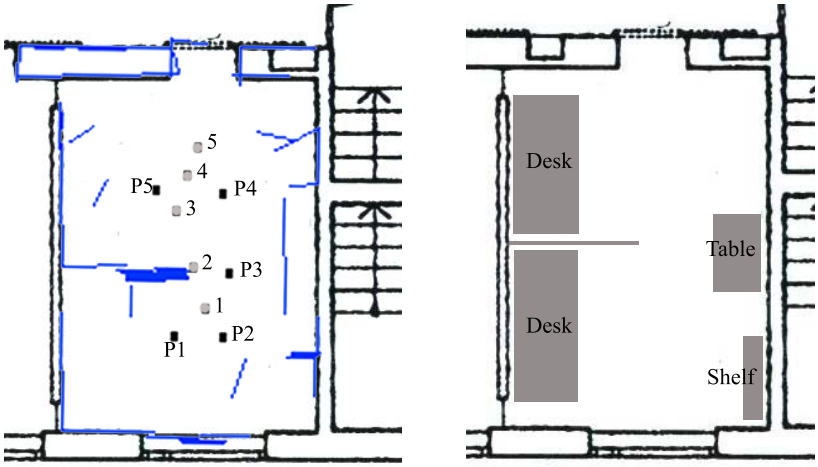


Figure 5.11: *Positions from where data sets were taken in R8 together with the laser range data features (thinner (blue) lines) (left) and a schematic drawing of the furnishing in R8 (right).*

where data sets were taken together with the corresponding centroids (1–5) and the laser range data features. Additionally an illustration of the furnishing that makes R8 look like two cubicles connected by a corridor is shown. Still, since the room is only of medium size and thus the "cubicles" are not too deep, a large portion of the room can be perceived from at least positions P2, P3, and P4. Accordingly the feature sets are altered gradually along the path from P1 to P5. Table 5.4 shows the variation over a number of measures for rooms R8, R9, and R10.

For those cluttered or heavily structured rooms the higher variances (compared to table 5.3) for the initial features $(m, l1, l2)$ indicate an unstable geometric representation of the perceived area along the paths the robot took. Apart from those

the distance of the centroids from each other can give quite good an indication for the change of the area perceived when used while travelling. The most significant change for R10 though can be observed in the area. The other features do not change as significantly as each part of the room represents an area quite similar to the others as far as the shape is concerned, but different in direction and size.

	R8		R9		R10	
Feature	Mean	Var	Mean	Var	Mean	Var
m	18.04	10.71	13.63	9.93	23.45	364.62
l1	7.67	9.17	7.03	5.68	8.56	3.22
l2	2.76	0.53	3.42	0.44	2.39	0.38
E	0.82	0.05	0.84	0.01	0.95	0.00
D	1.15	0.33	1.02	0.26	1.47	0.47
A	1.33	1.57	1.02	0.63	1.14	0.73

Table 5.4: *Statistical values for R8, R9, R10*

The angular distances can be interpreted as follows: In case of a generally low excentricity of the main ellipse an angular distance close to any multiple of $\pi/2$ does not represent a significant change since the ellipse is almost circular. In the case of a high excentricity an angular distance close to any odd multiple of π would indicate a significant change of the shape of the perceived environment. This means for the proposed representation method, that measurements for "similarity" can be based on those features. The features displayed in the tables above can all be derived from the originally calculated features $m, l1$ and $l2$ together with the global position of the region represented.

Summary From these results it can be concluded that the previously described method to represent distinct *regions* works well as a categorisation approach for global localisation. More important, the distinctiveness for the segmentation of an environment is very good for simply structured *regions* as almost convex rooms. In strongly structured areas the representation is altered depending on the position the data set was obtained from. Still, since the observed changes occur gradually a similarity measure can be used here to identify ambiguities that can be resolved by the interaction with the user. The investigation, in how far the concise feature based representation could be used with continuously obtained data samples to identify transitions from one topologically consistent *region* into the next one, is discussed in the following section.

Detection of structural ambiguities and *region* transitions – implementation and evaluation

Previously the approach to the representation of *regions* was discussed purely in the context of distinctiveness and categorisation applicability. Another aspect was to apply it to continuous comparisons to detect transitions from one *region* to a neighbouring one to allow for different interaction strategies ("curiosity") of the robot. Since the robot used for the implementation of the complete system has only one laser range finder and a differential drive, it is not possible to obtain 360° laser range data sets directly. For the acquisition of a *region* representation in an explicit specification case this problem can be solved by turning the robot, as previously described. For continuous comparisons this is obviously not possible, given the interactive context – who would want to interact with a robot that stops frequently and turns on the spot just to get a notion of the surroundings?

Thus, it was decided to use virtual scans to estimate a range data set covering the robot's backside. The respective module, the VirtualExplorer was discussed together with the other software components previously in this chapter.

The following discussion focuses on the results that could be obtained with the transition detection implemented according to the ideas presented in chapter 4. The evaluation resulted also in the fourth conference publication directly relevant to this thesis (Topp and Christensen, 2008).

Evaluation

Since the interesting issue for the complete framework lies in actually using the information that can be obtained from human users guiding the robot, data sets acquired during user studies were used for the evaluation of the environment representation, i.e., in this case the transition detection. The user studies though were in fact conducted to understand, how users guide around a robot and present an environment, and in how far the robot's representation of the environment corresponds to the user's understanding. Consequently, the data collected during the studies is exactly what is needed for the (off-line) evaluation, but the full implementation of the system was not yet available to be evaluated in the context of the studies. Thus, the approach for the detection of transitions is discussed here in the context of a number of different data sets, obtained

a) as part of a public data set[8] acquired in domestic settings with respect to the "home tour" scenario (post-hoc run through recorded data)

b) during experiments in domestic environments (post-hoc run through recorded data),

c) during experiments in a laboratory environment (post-hoc run through recorded data),

[8] http://staff.science.uva.nl/~zivkovic/FS2HSC/dataset.htm (URL verified June 18, 2008)

d) during test runs ("simulated tour") in the laboratory / office environment with the full system running (ad-hoc, on-line),

Several runs (guided tours) from and in those different environments were evaluated with respect to the following criteria:

- Consistency of the generated separation of regions in the environment with the "common understanding" of this separation.
- Loop closing ability on the conceptual / semantic level when coming back to a previously specified region through a new entry point
- Overall number n of detected ambiguities / transitions (and clarification questions asked by the system for the fully implemented system), with $nCorr$ being the number of expected transition detections between structurally different areas given the path of the robot.
- Number $nSens$ of ambiguities detected in a sensible range (approximately 1 to 2 meters in a standard indoor / domestic environment) from an obvious transition in the environment (e.g., a doorway), or in situations that can appear as structurally ambiguous (e.g., at a hallway junction or door opening).
- Number $nSpurious$ of obviously spurious (erroneous) detections of ambiguities (e.g., in the middle of an open area)
- Number $nMiss$ of obviously missed transitions into a structurally different area

The value $nCorr$ (and thus also $nMiss$) is of course a subjectively set number, estimated by counting actual transitions between *regions* and structurally unstable areas (hallway/doorway junctions). An idea here would be to conduct a short evaluation with different people to get an idea of a sensible segmentation to support a quantitative evaluation. However, since an interactive context is assumed, the author believes that not only the quantifiable results are of importance, but the interaction flow that is generated or disturbed by the transition detection. Thus, only one subjectively chosen ground truth was used for these quantitative tests.

In cases were the data sets were collected in laboratory settings of actual user experiments, the generation of a new, explicitly specified, *region* was not considered as a detected change. When on the other hand a specified *region* was obviously left a detection should have occurred, otherwise a miss is counted.

a) Domestic environment, data collection run The first domestic environment considered for these tests was a rather small apartment (approximately 50m^2) with narrow passages and doorways. In the apartment the living room, the bedroom and the kitchen were presented to the robot on a path assumed suitable for a guided tour scenario. The "home tour" in this case was however simulated for

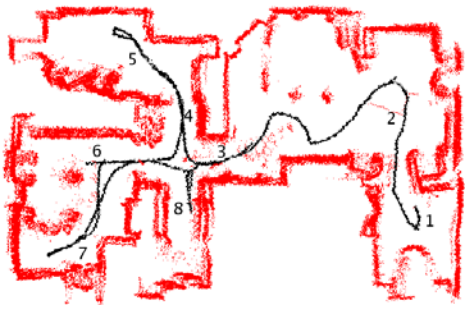


Figure 5.12: *The run through the small apartment used for the quantitative test of the transition detection. The red dots represent laser range data points of several scans (roughly matched), giving an illustration of the apartment's layout. The black line marks the robot's trajectory (starting at "1", following the numbers through "8" and coming back approximately to "3"), the numbers indicate areas where changes (transitions) should be detected.*

the purpose of collecting the data, so that spurious detections of transitions are unlikely to occur due to interaction related situations.

Figure 5.12 illustrates one of the runs and the "ground truth" for this apartment, indicated by the numbers. Approximately around those points one would expect the system to react to changes in the representation according to chapter 4. For the small apartment the numbers are very convincing, in two runs (the one depicted and a similar one) an overall number of $n = 18$ (of $nCorr = 18$, given that the robot passed several of the interesting points more than once, summing up to 10 estimated detections for the depicted run) ambiguous situations / transitions were detected, all of which appeared sensible given the current positioning of the robot. As expected the fact that no spurious detections (i.e., in the middle of an open area) occurred can be explained with the fact that no human user was actually interacting with the robot. On the other hand it appears that in three situations a change in the environment should have been more obvious and thus should have been detected earlier than this was the case. This delay might be due to the movement condition, implicating that the robot had to move at least one meter from the last point where a transition was detected.

b) Domestic environment, user study The second apartment considered for the tests was a medium sized flat of approximately 85m^2 with partially rather wide passages and open spaces. In this apartment the hallway is opening directly into the living room without any door or other obvious separator between them. The hallway, living room, one bedroom and the kitchen were presented to the robot, in this case in an actual "home tour" scenario, observed during the second user study reported in chapter 6. Two runs were considered for the quantitative evaluation.

Figure 5.13 illustrates one of the runs and the transition detections that were expected marked as numbers 1–7 with the help of a layout sketch. Due to some

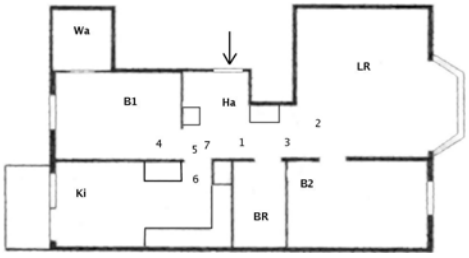


Figure 5.13: *The run through the larger apartment (layout sketch; not exactly proportional). The robot was guided from hallway (Ha) to living room (LR), back through the hallway into bedroom 1 (B1) and then into the kitchen (Ki). The tour was concluded in front of the bathroom (BR).*

navigation issues[9] the continuously running SLAM module was not able to provide corrected pose estimations through the complete run, which makes the use of a scan matching based illustration impossible in this case. For the detection of transitions as used in this experiment though the correct position estimation is not crucial, thus the data could still be used for this purpose.

In the two evaluated runs $n = 22$ ($nCorr = 24$) transitions were detected, with $nSens = 20$ (91% wrt n, 83% wrt $nCorr$) of them appearing sensible regarding the surroundings ($nSpurious = 2$, 9% wrt n). In this apartment in $nMiss = 4$ (17% wrt $nCorr$) occasions an obvious change in the environment was not detected. An analysis of the similarity values showed, that differences between region representations seem generally slightly smaller in domestic settings than in a laboratory environment. Adaptive setting of the threshold values or the application of a more sophisticated change detection filter can be an option to cover such cases more appropriately.

c) Laboratory environment, user study In the laboratory environment two user study experiment runs were evaluated, which covered a large of the corridor and some of the rooms (one office, a meeting room and the kitchen/lunchroom). Figure 5.14 illustrates the environment and the results for continuous checking in one of the collected data sets. The similarity distance threshold was (empirically) set to 1.5, the robot had to travel at least 1.0 meter before a new comparison was started and changes were accepted with only one data cycle of occurrence.

In both runs together $n = 45$ transitions or structural ambiguities were detected. Of these can $nSens = 35$ be classified as sensible while $nSpurious = 10$ have to be considered as spurious. Most of those spurious detections were actually due to the user blocking the field of view of the robot significantly, so that the data sets appeared quite differently during short periods of time. Still, with a rate of roughly 78% sensibly detected changes in the environment representation the approach seems helpful in the interaction context it is intended for. No miss had to be counted.

[9] The robot got stuck on a threshold and collected highly erroneous odometer readings due to wheel slip

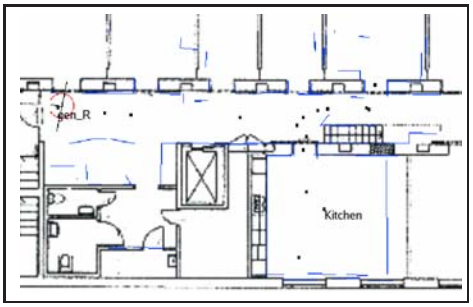


Figure 5.14: *The part of the laboratory travelled by the robot and its guide. The black dots without label mark the positions where the system decided to generate a new (hypothesised) region. The labelled dots refer to an explicitly specified region's centre ("Kitchen") and the starting point ("gen_R", being the origin of the "generic region").*

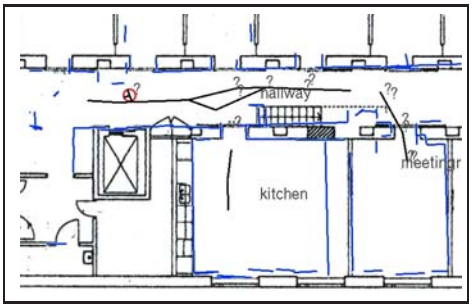


Figure 5.15: *The laboratory run on one of the office floors (visualisation fit into schematic drawing), starting from the left hand side of the environment, entering the kitchen and the meeting room and travelling back to the starting point; question marks are indicating the positions where the robot asked for clarification.*

d) Test runs in the laboratory / office environment, full system In the office environment two runs were evaluated, one of which covered a large part of the corridor of one of the floors and two of the rooms (a meeting room and the kitchen/lunchroom). With the other run the applicability for loop closing was tested by specifying the "living room" (one large laboratory room) and the connected hallway, where the robot was guided back into the living room through a different door than was used when leaving it. One important issue for these runs is, that a strong assumption was based on the idea of personal preferences expressed in the discussion for the environment model (chapter 4) and the first user study (pilot study, chapter 6, section 6.2): As long as no specific *region* was presented to the system, it assumed to be located in the "generic *region* and ignored observed changes in the environment representation. Thus, the system would not pose frequent questions on a level of the hierarchy that is not (or at least seems not up to a certain point) relevant to the user.

The complete system as described previously in this chapter was running on the robot's internal PC, while the graphical user interface was exported via a wireless network connection to a laptop to enter commands and take screen dumps.

Figure 5.15 illustrates the environment and the results visualised as screen dumps of the GUI, taken on-line during the runs. In the first run three *regions* were specified, ("kitchen", "hallway" and "meeting room") one of which (the "hallway") was specified as a correction of the assumption the system made when an ambiguity was detected. The navigation graphs (black solid lines) indicate a clear,

sensible cut between "generic *region*" / "hallway" and "kitchen", while the transition between "meeting room" and "hallway" was initially placed quite far out of the "meeting room" which led to clarification questions posed by the system so that the initial, inconsistent assumptions could be corrected. Overall $n = 13$ times the system asked for clarification (the respective points are shown with question marks). Of these can all $nSens = 13$ ($nCorr = 13$) be classified as sensible while none have to be considered as spurious or missing ($nSpurious = nMiss = 0$). Note that the system asked twice at points where it passed through twice. For the purpose of verifying the ability to find transitions and ambiguities the confirmations or corrections given by the user had no persisting impact on the stored representation of the environment, except when a new *region* was explicitly specified in a clarification situation. This means simply that the system forgets whether it has already asked about a particular transition. No sensible viable loop occurs on this floor of the office environment, thus the applicability to loop closing (re-entering through a previously unknown gateway) had to be tested in a second laboratory run.

During the second laboratory environment run (see figure 5.16) the robot was placed in a rather large room (the "living room") which was specified actively right away (figure 5.16a)). The robot was guided out of the room, detected a transition in the doorway (here it was confirmed, that this was still the "living room") and it got the specification of the "hallway" before it could ask a second time (figure 5.16b)+c)). After this specification the robot was guided back into the living room (detecting an ambiguity as soon as it reached the crossing of doorway and hallway), and hypothesised being back in the "living room" as soon as this was entered (figure 5.16d)). After travelling a bit more inside the living room, checking the viability between loose ends of the graph the two separate graph sections got linked together (figure 5.16e)). Overall $n = 4$ ($nCorr = 4$) ambiguities are detected at spots related to doorways – one is detected twice again due to the fact that corrections or confirmations (i.e., the positions where they were given) are not persistent in the system. The one transition detected when coming back into the living room is marked with two question marks due to synchronisation issues with the GUI, thus only one hypothesised transition is actually reported.

Summary The results from the eight evaluated runs show, that most of the obvious transitions (e.g., junctions of hallways, entering a room, hallways opening into a room) in as well a lab environment as in two different domestic environments are quite well detected. Some missed detections in the domestic settings however suggest to consider the application of a more adaptive decision process. In the particular case of the "home tour" scenario though it can be safely assumed, that the human user would take care of such situations. If he or she thinks that the robot should be aware of a spatial distinction, a respective region representation would be specified actively according to the information that the user would give. Errors in the space representation the robot built during the tour would be revealed when the representation is to be used and can be corrected according to the ideas

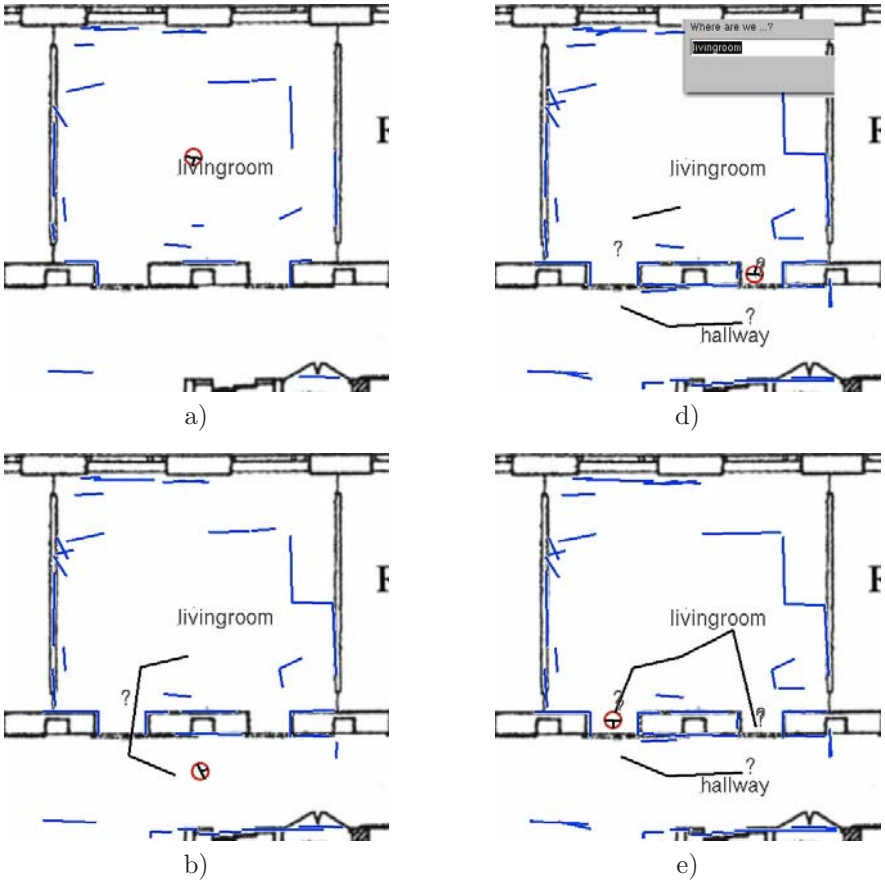


Figure 5.16: *The laboratory run with the robot starting in the "living room", leaving it and coming back in through another door, guided by a user. Question marks indicate positions where the system asked for clarification and solid black lines represent the navigation graphs generated within the regions. a) At the starting position, b) leaving the "living room" through one door, c) specification of the "hallway", d) getting back into the "living room", e) the final graphs*

discussed in chapter 3.

A number of spurious detections in one of the domestic settings can be explained with the user being very close to the robot (due to the interactive scenario) and thus covering larger parts of the laser range finder's "field of view". Since for this particular evaluation a change of the current region representation was accepted immediately after only one cycle of occurrence, such spurious detections can easily be avoided by applying a higher threshold, e.g., three cycles of actually updated laser range data. This was done for the laboratory runs, where it seemed to have immediate impact in the sense that spurious detections did not occur as frequently.

As an overall result the approach to separating *regions* and the detection of transitions between them is considered a useful tool to support the acquisition of conceptual understanding of the environment in the context of the Human Augmented Mapping framework.

5.4 Transfer of the mapping subsystem to "BIRON"

The described framework incorporates dialogue and interaction management as a central part of a system for Human Augmented Mapping. However, the work for this thesis was rather focused around the environment representation and issues connected to the navigation abilities of the robot in an interactive context, which were investigated with the tracking and following approach. A full dialogue management system was thus considered out of scope and replaced by a graphical user interface / text input component and text-to-speech output, that allowed the developer and experiment leaders to control the system for experiments and studies. Without the transition detection most of the time the *user* would have the initiative, which allowed quite well to work with this rather simple interface implementation.

Still, it was of considerable interest both for the author and the COGNIRON project partner, i.e., the Applied Computer Science group at the University of Bielefeld, to see a fully integrated HAM system working in an interactive setting, particularly considering the idea of the *robot* taking initiative and asking the *user* for clarification in particular situations, when a transition or structural ambiguity was detected.

The "**Bi**elefeld **Ro**bot Companio**n**" BIRON fulfilled a number of requirements as integration platform (Haasch *et al.*, 2004). BIRON's base is, just as was previously used for the tests, a Performance PeopleBot, thus providing approximately the same properties in terms of mobility and basic appearance[10], but also offering a much more sophisticated interactive interface. It is customised with a number of additional sensors (visual and acoustic) and runs under a complex interaction framework including a multi-modal "person attention" module and following abilities. The interaction framework is connected to a dialogue management and pro-

[10] In fact, BIRON is an earlier model of the Performance PeopleBot by MobileRobots, inc. (former ActivMedia, inc.) than the PeopleBot "Minnie" that was used for most of the described tests and experiments, but it clearly is the same type of platform

cessing module, further referred to as "the dialogue" (DLG) as a simplification (Li, 2007).

BIRON's internal communication framework was developed as a memory based architecture with data from and for particular components being stored and made available as XCF/XML-chunks (Wrede *et al.*, 2004; Spexard *et al.*, 2008), which allowed to integrate the mapping subsystem of the HAM-framework with a rather low number of minor modifications. For the support component that requires access to motor control though (i.e., the "TurnExplorer") those modifications were slightly more drastic but still easy to accomplish.

Adapting the communication interface

In the full implementation the CentralController module is responsible for the communication between components and user interface (here the GUI). To encapsulate the mapping subsystem's functionality as far possible, the idea of the "Central Control" component was kept and the respective implementation was cut down to the communication between MapHandler, TurnExplorer, and the actual interface to BIRON's communication framework implemented as the "biHAM"-main program ("biHam" standing for "Bielefeld-HAM"). One drastic change involved the TurnExplorer: In the original implementation the MapHandler has direct access to the explorer, which itself has direct access to the motor control's velocity setting function. Velocities can then be altered by obstacle avoidance or explicit "stop"-commands. In the BIRON-framework this direct access would have caused problems given the different interaction modalities and modules that might ask for motor control access. Thus, the component had to be changed to a) be accessed *outside* the MapHandler (the *region* representation is generated in the explorer and then fed into the MapHandler by the control component) and b) specify goal poses instead of a velocity.

The "biHAM" program transforms the XCF/XML data structures to and from the data containers used in the HAM framework. Thus, the mapping subsystem communicates internally as before and data structures are transformed at one specified location in the module. Figure 5.17 illustrates the modified communication structure between the components and to the main program interfacing BIRON's framework.

In the following section the actual data flow for particular situations is described.

Data flow for HAM on BIRON

"biHAM" is running within BIRON's communication framework as an independent component in the sense that other components (i.e., the dialogue) provide information that might or might not be used and vice versa also "biHAM" produces information that might or might not be used by other components, without causing the interaction system to act incoherently. The information to and from "biHAM"

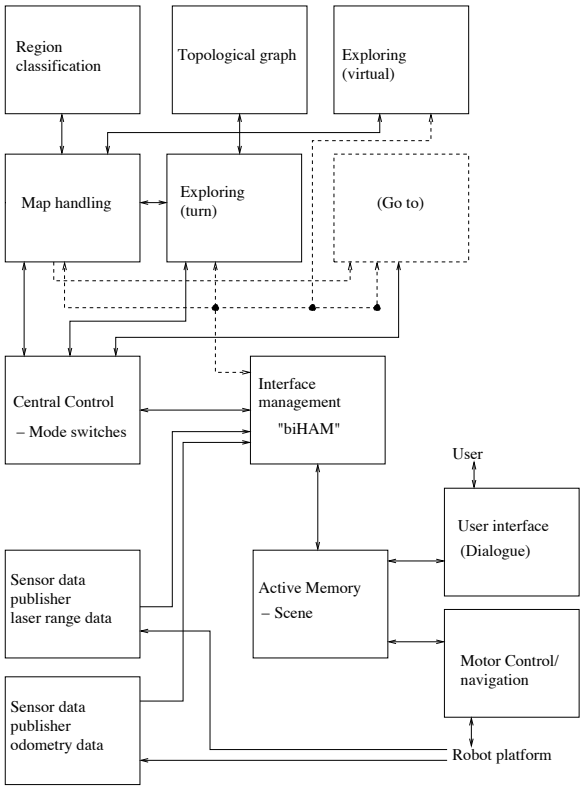


Figure 5.17: *The structure of the mapping subsystem software, adapted to the communication framework of BIRON*

is communicated via the "Scene" memory component of the Active Memory implemented on "BIRON". "biHAM" subscribes itself to relevant entries in those memories and receives a respective notification as soon as a relevant entry comes in. The program provides information as soon as something interesting has happened (i.e., a transition was detected) by writing a respective entry into the memory. In order to reduce the amount of data being sent to the memories, DLG keeps an internal representation of the current whereabouts of the robot that is updated only when changes occur. The following situations according to chapter 3 are considered so far.

Specification of a new *region* or *location*

These two situations can be considered the "standard case" for the system, corresponding to the situation termed "explicit (initial) information – user driven, concept obvious" on page 34; DLG has the category knowledge to decide if a *region*

or *location* is specified and provides an entry in the "Scene" memory with a new *region* or *location* label and a flag indicating that this is a specification (as opposed to a correction or confirmation), "biHAM" gets notified and feeds the label into the CentralController and the TurnExplorer is triggered. Accordingly (in case of a *region* being specified), "biHAM" sends a pose request into the memory to gain motor control. When the respective rotation has taken place and the explorer has gathered the relevant data, "biHAM" sends a confirmation to the memory and feeds label and data set (*region* representation) into the MapHandler. For a *location* the confirming message is sent immediately, since no representation has to be obtained by turning. The current labels for the *region* and if available the closest *location* inside this *region* are maintained and updated internally to answer appropriately on localisation requests.

Change of the "closest location"

In case the "current" or closest *location* changes with the robot moving about, "biHAM" provides a respective entry in the "Scene" memory, which is stored but not used by DLG until an explicit request is made by the user ("BIRON, where are you" results in "I am in the X, and there is also the Y", as long as both are available; being in the "generic *region*" the robot states that it does not know the name of the surroundings). This corresponds to the situation "Implicit information – data driven, *Location*" mentioned on page 38.

Transition detection, robot taking initiative

The MapHandler runs the previously described continuous checks on new data sets being available. It keeps track on observed changes with the help of two flags: REGION UNCERTAIN and REGION CHANGED. This corresponds to the situations described under "Implicit information – data driven, *Region*" on page 37 and under "Implicit localisation" on page 38.

REGION UNCERTAIN If a transition is detected, but the best estimate for the current representation is still the same or the robot is even still inside the initial ellipse that was generated when the *region* was specified, the REGION UNCERTAIN flag is set to "true". "biHAM" sends then out *the label of the current region* to the "Scene" memory, which is interpreted by DLG as a request for confirmation ("Are we still in the X?"). Depending on the user's answer DLG sends out a confirmation (same label with respective flag) or a correction (new label with respective flag). A confirmation leads thus to a reset of the REGION UNCERTAIN flag, while a correction is handled in two steps, due to the dialogue management. Firstly the dialogue indicates that the hypothesis was wrong by sending a respective message with "generic_region" as the current label. In a second step DLG asks the user for the actual whereabouts ("What room is it?") and passes the answer as a correction to "biHam" where it is handled either as an internal correcting update of the now

current *region* including the representation (from the hypothesised representation available) or "biHAM" triggers a new *region* specification, in case the user specified a previously unknown *region*. If no confirmation or correction comes in within a certain time threshold, "biHAM" sets back its internal waiting state and can thus generate a new request for clarification, if not the flags have been set back internally due to a completely new situation being reached.

REGION CHANGED If a transition is detected and the best estimate indicates that the robot either has entered a previously known, neighbouring *region* or has left a specified one and is now in the "generic *region*", a respective flag is set and "biHAM" sends out *the label of the hypothesised region* to the "Scene" memory. DLG gets notified and can ask the user depending on the label that was sent for clarification ("We just entered X, right?" in case a previously known *region* was – hypothetically – entered, "We just left Y, right" in the case of the "generic *region*" being entered). Depending on the type of question and the user's answer a confirmation or correction is sent (as described above) and interpreted by "biHAM". Here as well it is generally assumed that no answer within a certain time threshold means a problem with communication for the time being and the "biHAM" program continues by sending out a new request for clarification if necessary.

Since the flag setting is handled in a way that assumes the most likely hypothesis as the default, the system would keep a consistent representation in most of the cases, even if the user does not answer, possibly due to other issues requesting her reaction during interaction with the robot. This implicit anticipation of (non-existent) user reactions might still lead to a slightly deviating environment representation (compared to what the user has in mind) but it would not disturb the general flow of interaction. In case the uncertainty remains, a new transition detection will trigger a new clarification question that can then be handled appropriately.

Asking the robot where it is

This corresponds to the situations described under "Explicit query – localisation" on page 40. In the case the user asks the robot where it is, DLG reacts based on its own internally kept representation of the "current *region*" and "closest *location*"; it is assumed that uncertainties are either handled immediately or at least lead to appropriate updates according to the *currently hypothesised region and location*, so that no explicit query has to be propagated into the "biHam" program.

The fully integrated system was shown as a demonstrator for the Key Experiment 1 – the "home tour" – of the COGNIRON project. The results that could be demonstrated are discussed in context of an experimental run in the following.

Evaluation of HAM on BIRON

The system running within the fully interactive framework was evaluated regarding the transition detection and resulting robot initiated communication with the user. BIRON was guided around a part of a laboratory of the cooperating group at the University of Bielefeld that can be compared to a part of an apartment including living room, kitchen, and a part of a hallway, which were the labels chosen for the tour. The "kitchen" can be reached both from the "hallway" and the "living room" which integrates a loop in the environment, allowing to investigate the ability of the transition detection and region representation to recognise a particular, already known *region* and react appropriately. Figure 5.18 shows a simplified illustration of the environment.

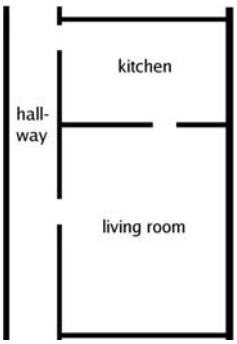


Figure 5.18: *The three rooms used for the run with BIRON, introduced during the tour as "living room", "kitchen", and "hallway" (illustration based on architectural sketch). The "living room" has a size of approximately 40m^2. An opening in the "hallway" to the left was ignored to reduce the illustration to the three relevant regions.*

One interesting aspect of this experimental run was that no correction of the internal pose estimations coming from the respective sensors on the robot platform was available. An initial decision not to transfer the SLAM module from the original full implementation was made in an early stage of the project cooperation, due to intellectual property right considerations. Some technical problems occurred during attempts to implement other mapping / SLAM methods on BIRON through the time of the project cooperation, thus the final decision was made to run the fully integrated system without any SLAM approach and investigate its capabilities running on the original pose estimations delivered from the sensory system. Consequently, the results achieved in the presented experimental run have to be discussed before this background.

The "guided tour" with BIRON

Figure 5.19 illustrates the guided tour with BIRON through the laboratory environment. The position estimation was corrected in a post-hoc run through the data stored during the experiment. From additional logs generated during the run the positions where the robot took the initiative and asked about a hypothesised transition to a new or known *region* are available and are thus marked in the il-

lustration with question marks. The hypothesis for the "current *region*" at the respective point in time is shown in the upper left corner of each of the frames. In the following the tour is described according to the most interesting events relevant to the *region* representation and transition detection process and the respective communication between the robot and the user. The user specified also a number of *locations* inside the "living room" and the "kitchen", which are omitted in the following description and illustration.

In the "living room" The "tour" started in front of the "living room" which was entered and presented to the robot. The tour continued towards the "kitchen", and the system got uncertain about still being in the "living room" when the robot came into the line of the open door, as shown in figure 5.19 a). The robot asked: "Are we still in the living room?" and received the confirmation "Yes", which led to a respective confirmation of the hypothesis "living room".

Leaving the "living room", getting to see the "kitchen" When the "living room" was left the system detected this clear transition, hypothesising to be in the "generic *region*", since the "kitchen" was not known at this point. The robot uttered: "We just left the living room, right?", this was confirmed by the user ("Yes").

In the "kitchen" Following the respective rules for the dialogue the robot then asked "What room is it?" and received the answer "This is the kitchen!", which was treated like a user initiated specification of the *region* "kitchen". After getting the information about the "kitchen" a respective representation was generated. Moving towards the door to the hallway the robot got uncertain and hypothesised "Are we still in the kitchen?", which was confirmed by the user ("Yes") (figure 5.19 c)).

Leaving the "kitchen", getting to see the "hallway" Still in the kitchen the robot got – just as in the "living room" – uncertain when it came into the line of the doorway and asked again "Are we still in the kitchen?", which was still confirmed. Shortly after that the robot left the "kitchen" (see figure 5.19 d)), the system noticed that fact and hypothesised to be in the "generic *region*". Consequently, the robot asked "We just left the kitchen, right?". This was confirmed ("Yes") and led to the same short exchange ("What room is it" – "This is the hallway") as before in the "kitchen".

In the "hallway" Travelling along the now specified "hallway", the system noticed one potential transition when the robot reached the door to the "living room", which led again to a short confirming exchange ("Are we still in the hallway?" – "Yes"), as shown in figure 5.19 e).

Leaving the "hallway" – getting back into the "living room" The robot was guided back into the "living room". The transition was noticed by the system, which hypothesised "We just left the hallway, right?", assuming to be in the "generic *region*", as shown in figure 5.19 f). The user answered "Yes", which led to a second question "What room is it?", answered by the user with "This is the living room". The hypothesised "current *region*" was corrected accordingly.

Leaving the "living room" – getting back into the "kitchen" Crossing the "living room" the robot was guided towards the "kitchen" again. On the way it got uncertain twice, once getting closer to the door, asking "We just left the living room, right?", and receiving an implicit correction ("No"). The other time a transition was detected right in the doorway to the "kitchen" ("Are we still in the living room?"). This hypothesis was corrected again, the user answered "No", the dialogue generated a follow-up question "What room is it?" and the system received the statement "This is the kitchen", which led to a respective correction of the hypothesised "current *region*". In the "kitchen" the robot got uncertain twice, both times, however, hypothesising correctly to be in the "kitchen". This led to two respective exchanges ("Are we still in the kitchen" – "Yes"), (see figure 5.19 g)).

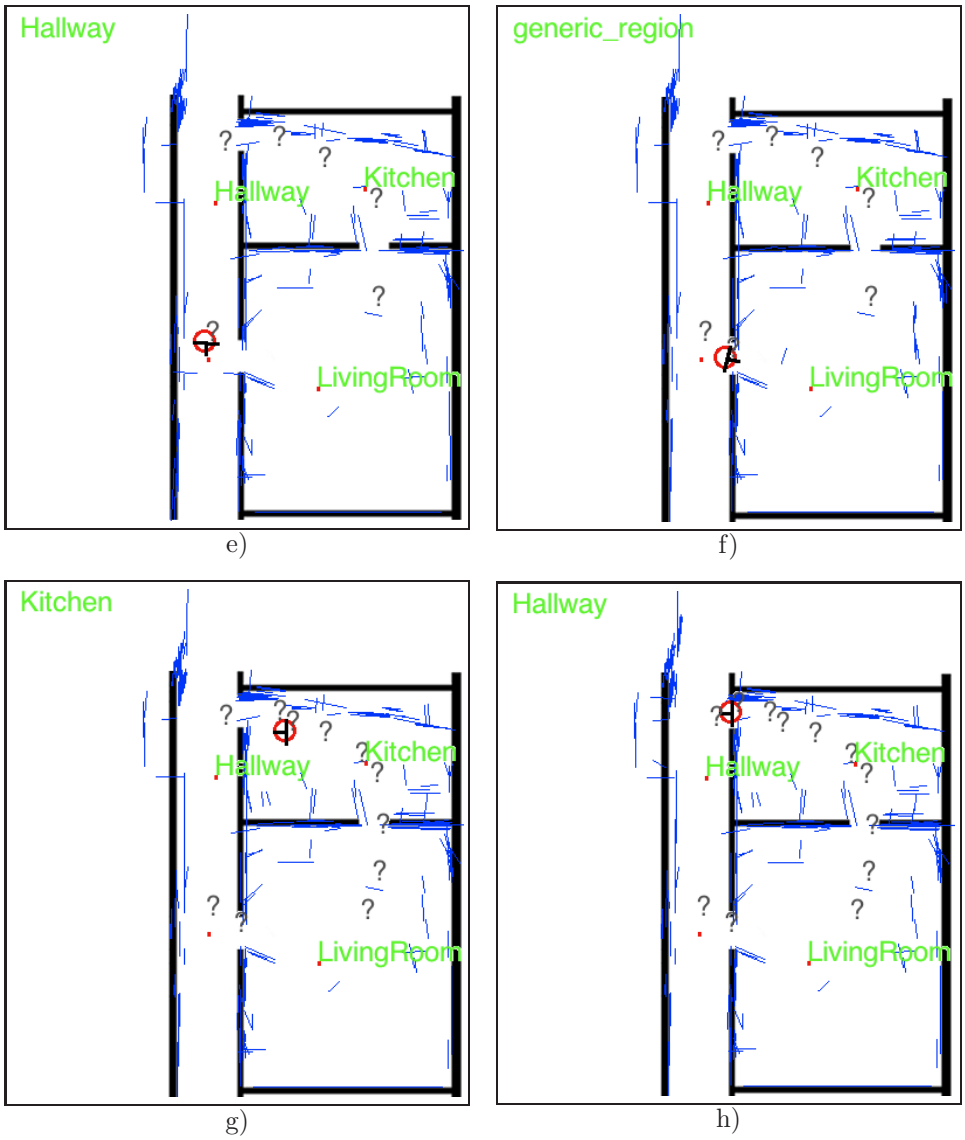
Leaving the "kitchen" – getting back into the "hallway" When reaching the door to the "hallway" for the second time, the system noticed the transition and hypothesised correctly to have entered the "hallway" again, stating "We just entered the hallway, right?" (see figure 5.19 h)). This assumption was confirmed and the tour was terminated by the user without further discourse in the "hallway".

Given the significantly different layouts of the three rooms used for the experiments the author started to wonder about the seemingly odd discourse happening during the second encounter of the "living room". Taking a closer look into the uncorrected data though revealed, that the position estimation had accumulated an error large enough to make the system refrain from accepting the "living room" as current hypothesis when the room was re-entered. Figure 5.20 a), b), c), and d) illustrate four significant points of the tour corresponding to the figures 5.19 a), c), e), and h) with uncorrected position estimation. Since the error is obviously mostly depending on rotations of the robot platform, the position estimation error is kept on a level that allows to hypothesise the "hallway" correctly as "current *region*" when it is re-entered, since no significant turning movements "on the spot"[11] had been made after its specification.

Overall the following results can be stated in terms of the criteria listed above:

[11] Also in the implementation on BIRON the robot turns around once to gather a 360° laser range data set.

Figure 5.19: *The experimental run with BIRON, visualised with a corrected pose estimation in a post-hoc run using the SLAM method of the full implementation version. The blue lines indicate the line features generated and used by the SLAM module.*



The region labels are marked at the positions where they were given to the robot by the user. a) In the "living room", b) leaving the "living room", c) in the "kitchen", d) leaving the "kitchen", e) in the "hallway", f) back into the "living room", g) back into the "kitchen", h) back into the "hallway"

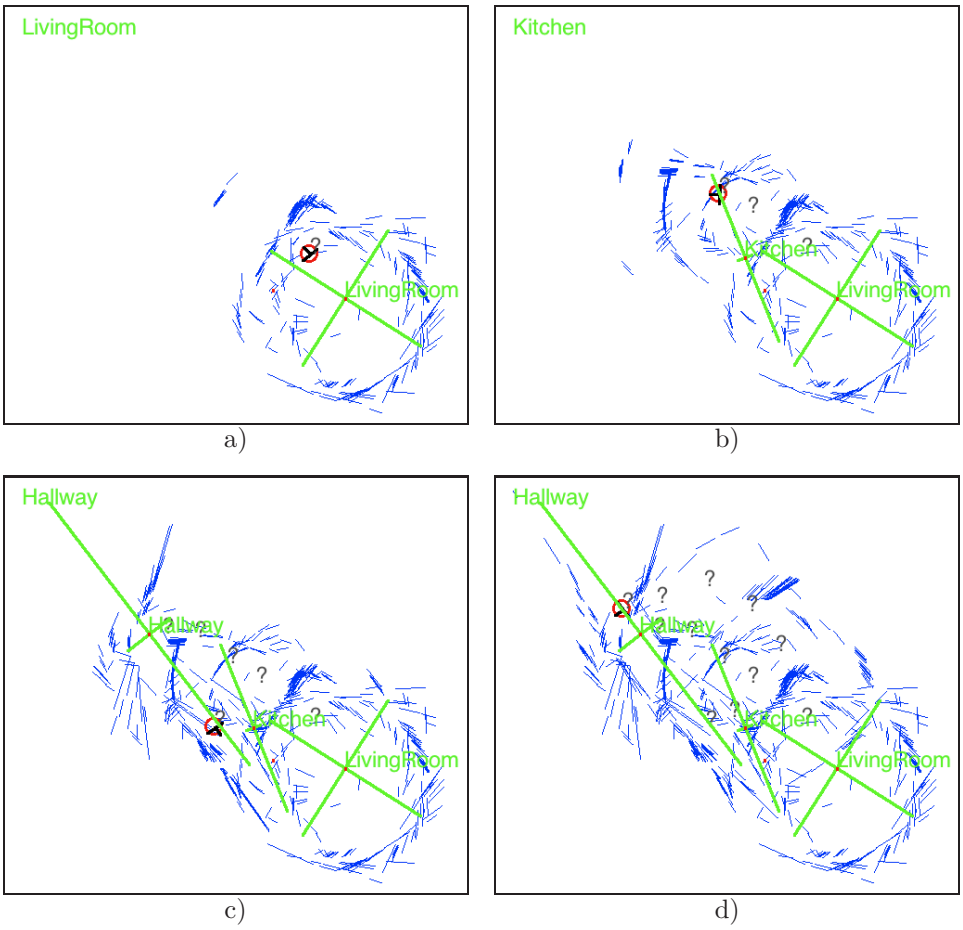


Figure 5.20: *The experimental run with BIRON, visualised with the original uncorrected pose estimation in a post-hoc run. The blue lines indicate the line features generated and used by the SLAM module, that ran under the assumption of receiving perfect pose estimation data. The region labels are marked at the positions where the system computed their centre points, showing also the axes of the describing ellipse.*

- Overall $n = 12$ ($nCorr = 5$) transitions were detected (including four detections that were plausible due to the robot being close to the respective door).
- None of the expected detections (those where in fact a *region* was left) were missed ($nMiss = 0$).
- The $nSpurious = 3$ somewhat questionable detections (also in the context of the surprising hypothesis of having left the "living room" in the middle of that *region*) that occurred in the "living room" and "kitchen" can be explained partially with the uncertainty the system had gathered due to the error in the position estimation. For the spurious detection occurring in the first round through the rooms in the "kitchen" a possible explanation might be that the user was blocking the robot's "view" for a significant amount of time. Taking a look into the run itself it becomes obvious that the user and the robot spent a couple of minutes in the "kitchen", being relatively static, while the user specified a *location* to the robot and had some problems with the general interaction flow, not related to "biHam". Figure 5.21 shows the respective "spurious" detections highlighted and numbered in the order of the occurrence of all transition detections.

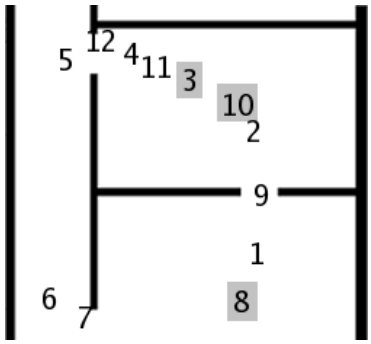


Figure 5.21: *All transition detections marked with their number in chronological order of occurrence. The spurious detections are highlighted with grey rectangles around them. One (no. 3) occurred during the firs round through the environment, the other two (no. 8 and 10) occurred in the second round when the system had accumulated a rather high error in its pose estimation.*

Discussion

The aim of the integration of the mapping subsystem with the fully interactive framework on BIRON was to see if a meaningful interaction in and about the surroundings can be achieved with the proposed models and used representations. Since the technical integration of dialogue and mapping subsystems itself is a complex task, it was decided to limit the functionality of the resulting overall system to the rather basic situations described above, including the specification of *regions* and *locations*, and the detection of transitions together with the resulting requests for confirmation. Within this limited context the question mentioned above can

be positively answered at least for the discussed environment. The robot detected all expected transitions and produced only a very limited amount of surprising questions. Despite the problems due to a large accumulated error in the pose estimation the tour could be concluded in a consistent state of the robot's and its user's understanding of the environment.

More complex scenarios of the "home tour" remain to be considered as future ideas. With the mapping subsystem and its mechanisms to generate requests, thus being able to take the initiative in the interaction, a number of aspects are possible to investigate, for example, a situation in which a user specifies a new *region* with a previously used and thus ambiguous label. For the time being this would lead to an erroneous "correction" of the existing representation, due to issues of the communication between dialogue and mapping system that could not be solved within the scope of the cooperation.

5.5 Summary

This chapter presented the implementation and evaluation of a system for Human Augmented Mapping. A full standalone implementation of the proposed general architectural framework includes a mapping subsystem and an interaction subsystem, each of which subsumes a number of different components implementing the proposed environment model and representation for *regions*. The evaluation of particular components and modules was done with different simplified versions of the complete standalone implementation and considered the module for tracking as well as the mapping subsystem consisting of the *region* representation, the topological graph generation and the detection of *region* transitions or structural ambiguities. Additionally the mapping subsystem was modified and transfered to a fully integrated interactive communication framework on the robot BIRON, the "**Bi**elefeld **Ro**bot Companio**n**" within a project based cooperation, resulting in an evaluation particularly of the transition detection in the context of the interaction with the user.

Chapter 6

User studies

In this chapter three user studies are described. All three of them were conducted with the robot "Minnie", the Performance PeopleBot which was also used as platform for empirical studies regarding the environment representation and topological graph modelling discussed in chapters 4 and 5. The user studies were carried out by the author in cooperation with Helge Hüttenrauch[1], and for the third study also a master's project (the Swedish "examensarbete") was assigned. The author's immediate contributions to design, setup and analysis that exceed the technical involvement – i.e., providing the software system the studies were conducted with – are mentioned in the respective sections for each study. The results of the first study were reported in conference articles relevant to this thesis (Topp *et al.*, 2006a,b), while results of the second study contributed to a journal article submitted for review.

A similarity of all the studies is their exploratory character. In all 27 sessions of the three studies only one robot was used, thus it is by no means possible to transfer the observations to general human-robot interaction, but it is possible to make statements for the interaction that could be observed in the given scenario with a mobile robot with the appearance of a Performance PeopleBot. The respective findings can thus in all cases be used as a basis for a qualitative discussion and inspire further studies.

Despite the general similarities the basic ideas for the studies and thus also their setups were quite different. The first study, so far referred to as the "pilot study" was designed to test the proposed environment representation for applicability and investigate people's behaviour and strategies when guiding a robot through a known (office) environment. The setup was exploratory; instructions given to the subjects were rather clear in terms of what they should do, but did not specify how they

[1]At the time of writing, Helge Hüttenrauch is affiliated with the Human-Computer Interaction group of CSC at KTH (http://hci.csc.kth.se/personView.jsp?userName=hehu, URL verified August 27, 2008).

should handle the task. However, a number of working hypotheses were formulated to guide the study design and realisation.

For the second study, which was designed as a follow up study to the pilot study, a very strong decision regarding the setup was made. The robot was moved out of the laboratory and to people's homes to guarantee familiarity of the subjects with the surroundings. This decision in itself of course gave options to observe a lot of interesting and fascinating situations. Still, this study had some clear questions as a basis, which made the instructions for the subjects slightly more canonical than for the pilot study.

The third study actually appears thematically as the first, but the questions that led to the setup idea arose only during the process of developing the environment model, implementing the system and using it for the previous studies. A general issue that had not been investigated before was to observe a human presenting an environment to another *person* and a *robot* and thus find out about differences or similarities between human-human and human-robot interaction in the context of the "home tour" scenario. For the human-robot interaction part the setup was similar to the pilot study, though instructions given to the subjects differed due to the fact that the tour also had to be given to another person, which required a little more sophisticated background story to motivate the "home tour".

6.1 An implementation for user studies

For the studies a rather early version of the full implementation of the HAM system (described in chapter 5) was used to guarantee a certain robustness and stability. This version of the system does not implement the full graph model proposed in chapter 4, but allows to store labels for the items subjects should present to the robot together with a time stamp and the corrected pose the robot has, so that the stored data can be used for technical evaluations later. This was, for example, done for the evaluations on recorded data discussed in the previous chapter, section 5.3. Nevertheless, it offers the operator to distinguish between items of the different conceptual categories proposed in chapters 3 and 4, which means also, that a proper *region* representation is gathered and stored, when a *region* entry is made. Despite the fact that the actual implementation of the graph model only handles *regions* and *locations*, this "tool"-version for studies was extended so that also *object* entries can be made. Those entries have no impact on the graph structure but the event (timestamp, current pose of the robot, label, item type) is stored for potential further analysis.

The operator, i.e., the experiment leader runs the software on the robot and exports its graphical user interface (GUI) to a laptop computer that can be carried around to be in the vicinity of the robot. Thus, the experiment leader can take control over the robot whenever this seems necessary. The graphical user interface provides a number of menus and short cut keys to feed commands and labels into the system, e.g., pressing the key "r" for "specifying a *region*" opens a dialogue

that allows the operator to choose the needed label from a list, or (in case it is not assumed a frequent label) to type it. The lists (one for *regions*, one for *locations* and later also one for *objects*) were extended gradually.

This "tool"-version of the implementation was iteratively improved after the first (pilot) study and offers also a couple of convenient tools for the operator, including a key to take a snapshot with the on–board camera of the robot to get an impression of "the robot's view" in a certain situation (the camera is otherwise not used) and the option of making the robot say a precoded utterance ("Please move on, you are too close to me") to overcome a particular type of "deadlock" problem (discussed in context with the pilot study described in the following section) that might not be detected by the system.

Equipped with the "tool"-version of the HAM system it was possible to control the robot in a semi–autonomous mode with the dialogue being simulated in a Wizard-of-Oz manner by the experiment leader. In all three studies the subjects were told to talk to the robot, while the experiment leader and robot operator interpreted their utterances and fed the respective commands into the system via the GUI. The implementation allowed to switch to direct remote control immediately to guarantee the participants' safety and to overcome possible technical problems, e.g., if robot and user "got stuck" in a very narrow passage.

All three studies are presented in detail in the following sections.

6.2 System and model in use - the Pilot Study

The study described in this section was set up and initiated by the author to find out about different strategies users might show when presenting a known environment to a robot. This was assumed to contribute to the design of the environment model proposed in chapter 4. Presumably such a process of a "home tour" or "guided tour" is influenced by personal preferences of the individuals and expresses to some extent the attempt to personalise the robot's environment representation. Personalisation along the taxonomy of Blom (2000) 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).

The study served also as an exploratory tool – i.e., in fact a pilot study – to the extension of a previously conducted, similar study by Green *et al.* (2006a), that was limited to one single room. The extension of that study should comprise several rooms, as did the pilot study described here, and thus the methodology could be investigated for a more comprehensive follow-up experiment. Those two goals – investigating presentation strategies and the applicability of the proposed environment model, as well as the more general investigations of the study setup and interaction aspects – gave the opportunity to the mentioned collaboration mostly in realisation and general analysis of the study. The initial idea for the setup,

the working hypotheses, and the analysis and discussion with respect to the environment model as they are presented in the following are the contributions of the author herself to pilot study.

In the following the study setup and hypotheses are described together with the observations made during five experiment sessions or trials. Some observations are directly related to the implementation of the tracking and following system (see chapter 5) and the interaction, while others are more related to the environment model to be investigated.

Scenario

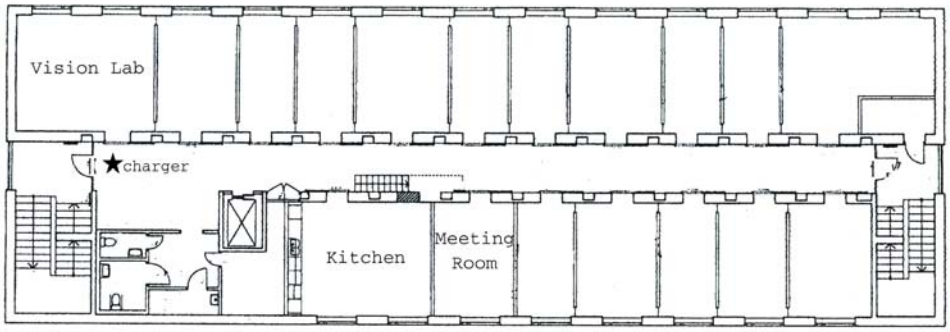


Figure 6.1: *The floor plan of the office environment on which the experiments took place. The star marks the starting point, where subjects encountered the robot*

The scenario of the study was a "guided tour" through a portion of an office building, namely one of the floors that hosts a part of the Centre for Autonomous Systems[2] the author was affiliated with[3]. Figure 6.1 shows the floor plan with offices (not marked), the kitchen, the meeting room and the computer vision laboratory of the office floor where the trials were conducted. Subjects were instructed to show the robot ("Minnie") around in the environment so that it later could perform general, not particularly specified, service tasks. This could only be achieved if the robot had "seen" the respective *locations* (a more detailed description of the instructions and the technical realisation is given in the following paragraphs).

[2]The Centre for Autonomous Systems (CAS) is a research centre with participation from different groups and laboratories at the Royal Institute of Technology (KTH), among them the Computational Vision and Active Perception (CVAP) group, whose offices were in this case used for the trial runs

[3]At the time of writing the author is still affiliated with the CVAP / CAS group.

Method

The pilot study was designed as an exploratory (case) study (Denscombe, 1998) with some guiding assumptions that will be discussed later in this chapter. The analysis of the collected data was thus limited to qualitative considerations, but could be based on a number of quantifiable observations. A significant effort was put into the selection of subjects and the instructions, which gave the study a somewhat controlled character despite its generally exploratory design. The following section explains this selection of subjects, the instructions given to them, and the methods used for data collection.

Subjects

As important precondition to the study subjects were assumed 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 experiment sessions took place in. To require familiarity with the robot's operation area is a design choice that differs from other human-robot interaction studies, where subjects often are invited into an unfamiliar or even a "simulated" environment. The deliberate choice came at a price however: since the used office environment hosts a research group working with various approaches to robotics and computational vision systems some subjects of the pilot study had to be expected to be familiar with the internals of robotic systems. As a consequence the use of a different environment for further user studies was considered to make sure that the familiarity with the robotic system was balanced by subjects without experience in robotics research. This was achieved with the follow-up study in which the robot in fact travelled to subjects' places, as will be discussed later in this chapter.

To assure at least some variety in familiarity with robotic systems the five subjects were selected actively among the members of the Computational Vision and Active Perception Laboratory on the KTH campus. The group of 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.

Instructions

Subjects were given an instruction sheet (see appendix A) that explained the task and the functionalities and abilities of the robot. To give the subjects a context for their experiment session they were introduced to the service robot "Minnie" that

should be made familiar with the office environment to be able to move around and provide its services, which could be, for example, fetch-and-carry type tasks or welcoming a visitor and guiding him or her to a particular room. The task was to use a number of spoken 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 experiment sessions took place on, given the possible tasks it should be able to perform later. 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" by sending it to previously specified items). In the instruction none of the words *region, location, position* or *place* was named. References were made 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 were any examples given (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 small objects like cups or phones were not of any interest, since the robot had no object recognition abilities; it just would need to know "where" to go to perform its tasks. This strong decision on not giving any potentially priming information was made since one goal of the study was to observe people's strategies for presenting an environment, particularly to find out if there was any correspondence to the concepts and hierarchy observable that were proposed by the author and presented in chapter 4, and consequently determine, how an implementation of the model would have to be designed.

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 at the used configuration setting only offers a forward field of view of 180° with a range of eight meters (for the detection of users it was actually reduced further to three meters). The subjects got thus some idea about 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. They got the explanation 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 actually start following them after uttering the command "follow me", the user had to gain a distance to the 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 spoken commands and utterances were interpreted by an experiment leader and fed manually into the system. Since object recognition was not incorporated it was suggested 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: it clearly stated which of the functionalities of the robot were in fact simulated or remotely controlled (see also the paragraph on "technical realisation", page 117) by an experiment leader that followed the (subject and robot) pair. This information was given to the subjects due to two considerations. Firstly, it had to be assumed that some of them were sufficiently familiar with the work to know that the robot had no speech recognition abilities available (at least not without any headset needed). Secondly, the purpose of the study was to observe the overall strategies for the tour with respect to the environment presentation, thus it simply did not matter if they knew about the robot's inabilities and thus it seemed unnecessary to mislead them initially and explain about that afterwards.

Subjects also were informed on what they 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. The participants were offered to ask for help before and during the actual trial, and knew that they could abort the experiment at any time.

Technical realisation

The robot "Minnie" was used – as previously described – with the "tool"-version of the Human Augmented Mapping system. In contrast to a previous user study performed with this robot (Green *et al.*, 2004) the robot thus navigated autonomously when it was following a user or moved toward a specified goal.

As mentioned previously the dialogue system was simulated by the experiment leaders. User utterances were interpreted into commands and labels for *regions* and *locations* and fed manually into the graphical user interface with pre-defined command codes and the labels the subjects uttered.

The robot was additionally provided 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 the environment model and the respective definitions of *regions* and *locations*. To determine what the respective participant intended to present, common sense knowledge about the labelling in indoor environments seemed a sufficient base for this process. The subjects were asked in the short interview if the behaviour switches seemed to correspond to their intentions, i.e., whether the robot turned to "have a look at the surroundings" when they intended to present a surrounding room or *region* and did not turn when they intended to present something particular "to look at". The strategy used by the experiment leader was thus supported *post-hoc* (see the observations discussed on page 119ff. for details). The turning behaviour was kept

also for the full implementation of the graph structure, to capture a full 360° data set for the *region* representation according to chapters 4 and 5.

Observation methods and data collection

By storing the data provided by the robot's sensory systems a full "real time" (graphical) representation of each of the trials could be obtained, since the complete HAM system implementation can be run on stored data sets as well as on-line. Additionally the sessions were recorded 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 trial runs the participants were asked to answer a number of questions (see appendix B) on their "tour" 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. It was of particular interest 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.

Hypotheses

The study was set up mainly to investigate the overall strategies of different individuals when they present a known environment to a mobile robot and test the relation between the resulting information and the environment model presented previously. The term "strategy" refers in this case to the choice and order of items presented and particular ways of presentation for particular items, e.g., find out if subjects always would enter a *region* to present it to the robot. A number of working hypotheses (WH) were used 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 entered *regions* in the environment" (e.g., they are probably moving through the hallway, but they will not necessarily present it),

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

WH3: "Users point out *regions* without entering them" (specifying, e.g., and office by just pointing through its door from the hallway).

Those hypotheses were used to test whether the observations from the pilot study can be related to the graphical environment model and particularly the assumption about the implementation of a partial hierarchy, as was described in chapter 4. The investigation whether familiarity with robotic systems had any particular influence on the "tour" was not in the focus of the study. Nevertheless the participating robotic researcher particularly familiar with map representations was expected to

be more precise in specifying items than subjects not familiar with robotic environment representations. Another assumption was, along the argumentation of Sidner *et al.* (2005), 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*.

Observations and results from the study

In this section the results from the pilot study are summarised. It is obvious that the data set is small and not entirely representative. However, it is possible to analyse the outcome of the experiment sessions in terms of *occurrence* of different phenomena, which makes the study an exploratory, or qualitative study according to, for example, Denscombe (1998). Additionally, the observations and the subjective answers obtained in the short interviews allowed to investigate how subjects reasoned about their strategy to show *regions* and *locations* and to improve the implemented system for the following studies.

As one outcome the methodology for conducting the pilot study was confirmed to show the validity of the approach in getting information on individually different ways of building map representations in an interactive, joint process. Furthermore the soundness of the environment model described above seems to be supported by its ability to handle the diverse situations and strategies observed. In table 6.1 quantifiable results are summarised to give an overview over observations and statements from the interviews.

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 6.1: *Quantifiable results from the pilot study.*

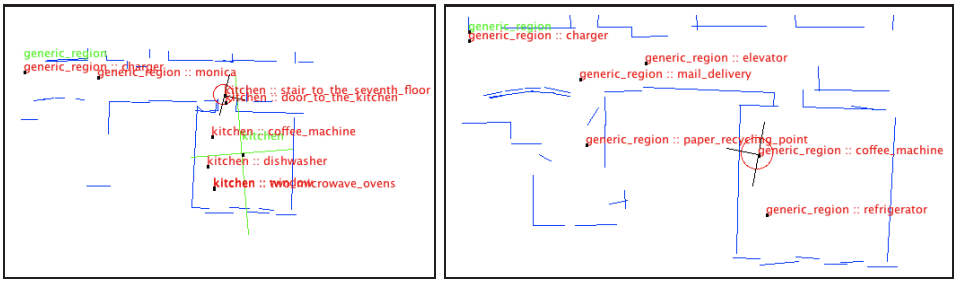


Figure 6.2: *Two different representations of the same environment generated after two runs with subjects of the pilot study. Regions are marked in green with the label, also showing the representation of the spatial properties with the two axes of the ellipse describing a respective data set and locations are marked in red together with the label of the region they are assigned to.*

Observations

All subjects but one used the full time frame to present the environment to the robot. The "tour" started for each experiment session at one end of the corridor (see Figure 6.1, page 114), 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 this automatically generated *location* was not taken 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 tour.

All participants took the robot into the kitchen, probably because this is a central room in the used 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*. Figure 6.2 shows the resulting "maps" for the same area generated by two different subjects. Already from the small sample of data it is thus possible to 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 presentation strategies observed for the "kitchen" and its *locations* correspond to expectations expressed in hypotheses `WH1` and `WH2`.

It was also noted that none of the subjects named the corridor or hallway – leading toward 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 further support for 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 expectations 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. 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 sessions with the same subject.

All subjects that had presented a mixture of *regions* and *locations* (four out of five) were asked if they had perceived the difference in reaction of the robot (turning by 360° for a *region* vs. not turning for a *location*). At this time they were also informed about the two different concepts the experiment leaders used. 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". This was stated despite the fact that written information had been given to all subjects, stating clearly that all dialogue features were to be translated from spoken commands to typed command codes by the experiment leader.

Despite some technical problems (see section 6.2 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.

This can be seen as confirmation that the implementation described in chapter 5 (particularly the tracking system) is sufficiently robust and stable in its performance to be used in a study setup with unexperienced participants. Still, a couple of implementation improvements were employed for the second user study to overcome at least some of the technical problems observed.

Particular situations

Even with the limited number of subjects some interesting strategies for the presentation of the environment were observed. The observations are related to the graphical environment model in the following. Observations that are more general are not considered in this analysis but were used to inform further studies.

As of the time the study was conducted, the feature set based *region* representation discussed previously had not yet been investigated fully. Still, it was assumed that the system had some ability to generate hypotheses on delimited *regions* from range data sets, to establish the link from the observations made during the trials to the environment model.

It was also assumed that a general knowledge model distinguished between *regions* and *locations* and a dialogue model that uses this knowledge base was included in the system.

From the trials evidence was collected 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 located in they stated only 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 was assumed to give a signal on the actual intention and was in fact one of the issues leading to the questions for the second user study that is described later in this chapter.

Departing from detailed observations some key situations can be specified that need to be handled by the robotic environment model (see figure 4.2) together with possible solutions to cope with them. Those suggestions were taken into account when the actual *region* segmentation was integrated into the full implementation of the mapping subsystem as it was described in chapter 5.

Presenting persons During the trial sessions, persons were pointed out twice to the robot. In one case the respective person was walking by and thus the presentation was ignored by the experiment leader. In the second case the participant actually intended to present the office in which the presented person was sitting, by pointing through the door. This was not entirely clear to the experiment leader who nevertheless decided spontaneously to feed a *location* as link to the office into the system. Later the intention of the participant was confirmed in the interview. 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).

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 the environment model. 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 are now associated to this new branch in the hierarchy (see also chapters 4 and 5 for details).

Links to *regions*/rooms With the internal "connector nodes" of the model links to rooms (e.g., 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

a) that a *region* is presented, and

b) that it is not the one the robot is currently located in.

This type of differentiation was one of the main issues investigated with the second user study.

Pointing out doorways explicitly The environment model could cope with explicitly pointed out doorways (as observed in the trial run with the robotics researcher) by generating a *location* with the respective name. However, there are several possibilities to represent such an entry in the hierarchy. 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 this respective strategy could only be observed with the robotics researcher, it was assumed to be rarely observable with a differently structured sample of users to test with and thus not explicitly taken care of in the implementation. Overlaps between *regions* are handled implicitly by overwriting (see also page 54).

Summarising it seems that the model holds at least for the variety of strategies to present a known environment to a robot observed in this pilot study.

Particular interaction styles

A few observations were made that seemed particular in the sense that they only occurred rarely. In one trial the subject presented items without any article (stating "this is table" instead of "this is the table"). In the same trial the subject used a very explicit strategy to make the robot follow through doors: the robot was guided up to the door along the hallway where it should stop. Then, it was made to turn to "face" the door, the subject walked into the respective room and asked the robot to follow again. Those observations are discussed further in the context of the third study reported in this chapter.

Interaction issues

During the pilot trials some issues with the technical implementation were observed, that had consequences for the actual interaction between subjects and the robot. Those issues led to a number of adjustments in the implementation to facilitate further studies.

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 start to move. As a consequence, user and robot ended up in a "deadlock"

situation, both waiting completely static for the other one to act. From carefully studying the interactions recorded on video it was concluded that the robot would need to indicate with a body (movement) gesture that it is ready and able to follow. An improvement of the system to incorporate a better signalling and feedback strategy to resolve this kind of "deadlock" was investigated in a master's project conducted by Mahani (2006). As a result the system was extended with a turn towards the user before it stated "I will follow you" and an operator triggered pre-coded utterance ("Please move on, you are too close to me"). Tests with this utterance being triggered by certain time limits or other technical measures did not lead to a satisfying conclusion within the (time) scope of the mentioned master's project.

A similar problem occurred, when subjects made the robot face something to "look at it" and wanted to continue the tour afterwards. This could also be resolved by making the robot turn back toward the user to indicate, that it is ready to continue after storing a presented item. However, these situations did not occur as often and were thus not explicitly handled.

Other problematic situations occurred due to failures of the tracking system (see section 5.2), but those were rare enough (they only occurred about once per eight minutes of interaction on average, not regarding the time when the robot was sent to previously shown places) to not cause severe interruptions in the interaction between user and robot.

With the mentioned improvements the general study setup with a user and the mobile robot interacting in a partially Wizard-of-Oz controlled, partially autonomous mode was considered successful and thus could build the basis for the following multiple room study.

6.3 Investigating presentation patterns – the Multiple Room Study

The previously discussed pilot study was initially designed mainly to find out about strategies for the presentation of an environment to a robot and the correspondence to the proposed environment model. When the study was designed it turned out though that the setup matched the plans for a study on human-robot interaction in a multiple room setting that should be a follow-up study to a previously conducted experiment using only one room (Green *et al.*, 2006a; Hüttenrauch *et al.*, 2006a,b). Thus, the pilot study served not only as it was originally intended as "proof-of-concept" for the model but actually as a feasibility test and basis for improvements in the setup for a more comprehensive study, in the following referred to as the "multiple room study (MRS)". Since some particular observations on presentation techniques for particular items were made both in the initial "one room" study and

in the previously described pilot study (see also a brief discussion of these observations in chapter 4, page 54), but could not be used as basis for an in-depth analysis due to the differing study setups, the author contributed in the planning phase of the MRS with a set of questions and hypotheses that also influenced significantly the study design and the instructions given to the subjects. The purpose of the study from the author's point of view was thus to investigate a possible correspondence of people's strategies to present particular items to the robot with three levels of the conceptual hierarchy (considering *regions*, *locations*, and *objects*) proposed in chapter 4. In the following discussion of the study the focus is thus set on this issue and the results related to it, more or less disregarding general human-robot interaction aspects that were of particular interest to the collaborating researcher, Helge Hüttenrauch. The analysis and result discussion presented here are thus the contribution of the author herself. Otherwise the author contributed with the study ("tool")-version of the HAM system implemented on the robot "Minnie", that was improved and customised after the previous pilot study, and participated in the realisation of the study and the general analysis of the observations.

Scenario

The scenario of the Multiple Room Study (MRS) was again the "home tour" idea. A user should guide the mobile robot through an indoor-environment by using spoken dialogue / commands and the robot's functionality to follow a person autonomously. A significant change however was made in the choice of environment. One of the prerequisites of the pilot study was that subjects should be familiar with the environment to at least an extent that allowed them to guide the robot without being guided themselves before. Thus, they would not be primed in their behaviour by the tour given to them. Further, this being more related to the general purpose of the study, a domestic setting rather than an office environment was preferred, and subjects should be as unfamiliar with robots and robotics research as possible, to counterbalance the choice of subjects in the pilot study. Those conditions were kept for the MRS, resulting in the problem of finding a domestic environment, familiar to subjects not being involved in robotics research. The only environments that would fulfil these requirements were assumed to be people's (study subjects') homes. Consequently, the robot "Minnie" was transported around the Stockholm area, visiting 7 different apartments or houses and 8 study subjects. This choice, however, came at the price of never being absolutely sure about the feasibility of a run in a given apartment or house, since the experiment leaders did not want to spend more than roughly one hour with each subject. Despite a couple of limitations due to the encountered environments' properties (see the paragraph on "technical issues", page 143ff. for details), it was technically possible for the subjects to show at least three different *regions* by entering them (including a "hall" or "hallway") and one more where the robot was not supposed to enter (e.g., the bathroom), even if not all of them actually did that.

Method

The study design can be summarised as "semi-controlled", since the instructions given to all subjects were the same, they all had the same task to fulfil and they received a clear list of suggestions, what should be presented, that aimed to make the conditions somewhat similar for each of the subjects. However, since the study took place in different environments those similarities could not exceed a certain threshold and thus the study has aspects of a field or case study with an exploratory design that allows at best to analyse the gathered data in a qualitative sense. This somewhat ambiguous situation allows to quantify the data as far as this is possible, but it does not support any statistical analysis or hypothesis testing. The immediate methodology applied for the trials is described in the following.

In general the procedure was to make an appointment with the subject, transport the robot to the respective home, and conduct the trial. For each of the trips the robot was wrapped and secured for transport in a regular car and taken to the subject's place by one of the experiment leaders, where both then teamed up. One informed the subject about the details of the trial and gave her the instruction sheets to read, while the other experiment leader checked the area for potential navigation issues (doorsteps, carpets, etc.) and prepared the robot. In a short discussion with the subject it was decided which area(s) should be avoided and doorsteps were prepared if necessary in the way described later in this section. The subject was given a short demonstration on how the system could be used and then had about 15 minutes to present the environment to the robot. A short, unscripted interview was conducted to wrap up the session, while the robot was again prepared for travel. The overall time spent in people's apartments and houses was tried to be kept around one hour. Since the robot was used by other groups in the laboratory and sessions had to be scheduled individually with the subjects, the robot was taken back to its "home" (the laboratory) after each trial.

Three experiment sessions though were carried out in the apartments / houses of the experiment leaders, which made it possible to have the robot available a little longer to have a more thorough feasibility test. Thus, the first two sessions were also considered pilot trials to ensure that the general setup worked and instructions were comprehensible. No severe changes had to be applied after those runs though, so that the respective observations are considered fully integrated into the study. The third subject was very familiar with one of the apartments – though not living there, which allowed to avoid having the robot travel to this particular subject, still meeting the requirement of familiarity with the surroundings.

Subjects

The subjects were chosen through "social snowballing", by asking around among friends, neighbours and co-workers and their families, so that the subjects were in most cases only known superficially to the experiment leaders and had definitely no deep insight in the work before being confronted with the task. They were

informed thoroughly about the intentions and process before they were to agree on participating. The subjects ranged in age from 27 to 68 years, with mean 39,88 and median 36. None of them were particularly familiar with robots, most of them had encountered robotic characters in movies or other media. Thus, it was assured for the sessions that none of the subjects had any other information of the robots perceptive abilities and functionalities than those given to them by the experiment leaders and it could also be assumed that media caused images and ideas of robotic systems did not interfere with reality in this context.

Instructions

For the pilot study the general idea was to allow the participants of the study to be completely free in their decisions what to present in the indoor environment and how to do that, given some information on possible commands and abilities of the robot. This resulted in strongly differing observations regarding the spatial concepts that were presented (some subject presented no *regions* at all, while others mixed both *regions* and *locations*, for example). For this MRS though it was important that each subject would at least present one, preferably several entities of each considered level of the hierarchy discussed in chapter 4, i.e., *region*, *location* and *object*. This was important for the analysis regarding different presentation strategies for different (spatial) concepts, that could not be performed on data collected in previous studies, since these never covered the range over the three concept-levels (*region*, *location* and *object*) in one trial, i.e., presented by one subject. Additionally it was of interest to observe a sufficient number of presentation strategies for *regions* that were not to be entered by the robot itself (e.g., a bathroom), and thus were likely not to be entered *with* the robot during the tour either. Consequently, the experiment leaders decided to generate clear instructions and a list with suggestions of what could be presented, to make sure that each subject would cover those three levels. A strict list was not possible to produce, given the different environments the study took part in. However, a number of spatial entities and objects could be assumed to be located somewhere in each of the homes (e.g., living room, kitchen, refrigerator, table, book, cup), to give at least an idea of what could and should be presented to the robot, including at least one room or area that the robot should avoid (e.g., the bathroom); the complete list given to the subjects can be found in appendix C. Thus, it was possible to generate data for the analysis regarding different presentation strategies for different (spatial) concepts.

The instructions given to the subjects included also some information on how to communicate with the robot, how the robot perceived its environment, difficult situations (e.g., narrow passages, thresholds) and how they could move the robot, by either having it follow or use direct navigation commands. Additionally they were informed about the roles of the experiment leaders, one recording the interaction on digital video, the other one monitoring and if necessary controlling the robot, including the interpretation of spoken commands.

Subjects were also informed that they could ask for help at any time or could abort the trial without giving any reasons. They were informed and agreed upon the further use of the data in research contexts and were rewarded with a lunch ticket. The instruction sheet was very similar to the one for the pilot study apart from the explicit list with suggestions on what should be presented.

Another significant change in the instructions given to the subjects was that they got a short demonstration by one experiment leader, described in the following.

Demonstration

For the pilot study it was decided not to give any demonstration of the functionalities of the robot – again in order to avoid any priming. For the multiple room study it was initially decided to continue in the same way. However, after a first trial without demonstration it turned out that the instructions were a bit too exhaustive and a short demonstration would allow to shorten the instruction sheet to an amount of text that was easier comprehensible given the rather strict time constraints for the trials. Thus, it was decided that one experiment leader (the one otherwise running the video camera) should demonstrate in a short sequence how to make the robot follow and briefly present one item in each concept-level of the hierarchy. Such a demonstration took about two to three minutes. This gave the subjects the opportunity to listen to most of the standard utterances the robot would make and also observe the robot's reaction to different commands including its movement, to avoid surprising situations in their own homes. The experiment leader giving the demonstration – in fact the author of this thesis – decided spontaneously which items to present, most often picking items that had a high likelihood of being presented anyway (e.g., the *region* living room, since that was one of the examples on the list) or something that was likely not to be relevant for the subject at all (e.g., the robot's own wireless keyboard or the video camera, which were part of the experiment leader equipment). This demonstration was also used as a final test for the system setup, before the actual trial with the subject was started.

Technical realisation

The implementation run on the robot "Minnie" was, as described before, the "tool"-version of the HAM system described in chapter 5. As mentioned before, this version of the implementation did in fact not implement the full graph structure, but allowed to store all the data required to run off-line experiments with the more technical oriented mapping subsystem, as described in chapter 5. As in the pilot study speech recognition and interpretation were simulated in a "Wizard-of-Oz" like setup, and the subjects were informed about that fact. The version of the HAM implementation was improved with a number of tools (direct motion commands, image storage, experiment leader triggered warning utterance "Please move on, you are too close to me", etc.) and a turning strategy before the robot started to follow

to avoid a particular type of "deadlock" situation as discussed previously in this chapter on page 124.

Another helpful improvement of the GUI was the introduction of a visible clock that showed the system time stamp for a certain period of time to allow synchronisation with the recorded video by filming the clock in the beginning of the trial runs.

Observation methods and data collection

A complete trial of the MRS included a short – mostly demographic – questionnaire to learn about subjects' age, professional background, and their experiences with robots, the latter also including exposure to media robots such as R2D2[4]. The trial was then recorded on digital video by one of the experiment leaders following the user–robot pair during interaction. Additionally the laser range data and odometer readings available from the robot were stored, and also every entry the subject initiated regarding *regions*, *locations* or *objects* was stored in a list together with the estimated pose information given at the time the information was uttered / entered into the system. For a *region* in fact also the 360° range data set was created and stored for further analysis and off-line experiments.

Hypotheses

The semi–controlled study design for the MRS corresponds mostly to an exploratory case study (Denscombe, 1998), since a) the environment was changing and thus not fully controllable and b) subjects only got a list of suggestions of what to present, while the items on this list might or might not be available in the particular surroundings. Thus, the observations made cannot be analysed quantitatively as in a controlled experiment that would confirm or reject particular hypotheses. However, from observations made during previous studies, including the pilot study, the author formulated the following working hypotheses or assumptions to guide the study:

- (WH1) There is a general pattern observable across subjects that allows to derive the conceptual category (using the three categories *region*, *location*, and *object*) of a presented item from the way it is shown to the robot.
- (WH2) There is at least a hierarchy observable *within-subject* with respect to the presentation strategy for entities of three particular concepts from the proposed conceptual hierarchy, namely *region*, *location*, and *object*.

More specifically, a number of assumptions on the hypothesised patterns were made regarding the concepts *region*, *location*, and *object*, for the latter those can be

[4]The beeping mobile "trash bin" from the "StarWars"-movies that can probably be seen as the most famous robotic individual created for the screen, offering social interaction by beeping, blinking and turning its upper part

compared to findings reported by Foster *et al.* (2008) on the role of "haptic-ostensive references" as they term the phenomena they observed in multi-modal dialogues between humans and a robotic system.

- (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.

The analysis of the video data was then performed by relating the observations made during each item presentation sequence (a SHOW-phase according to Hüttenrauch *et al.* (2006b)) to the conceptual categories used in the proposed environment model (chapter 4), as described in the following.

Observations, analysis and results

In a previous study it was found that the "home tour" scenario could be segmented in different phases (SHOW, FOLLOW, and VERIFY) (Hüttenrauch *et al.*, 2006b), while there were certain TRANSITIONS observable between these phases. Particularly the transitions between a FOLLOW and SHOW phase or two SHOW phases (which both are very likely to be observed in this order) could be seen as a preparation of the actual SHOW phase. Thus, the presentation of an item to a robot cannot be seen as a static event but is rather a dynamic process, which was taken into account for the analysis with respect to particular interaction patterns according to the assumptions listed above.

Observations and analysis

The SHOW phase for each presented item 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 focus of the analysis was for both preparation and show event on "how" the subjects conducted the two segments, and on "what" was observable, but not so much on "when" certain poses or gestures were performed, e.g., whether a pointing gesture was performed before, after or during a label was

uttered. The target of the observations was the subject who performed or initiated the respective actions. Thus, all preparations, even if they involved the robot performing a particular movement, are attributed to the user.

The following categories for observed presentation patterns were used (categories were specified heuristically after a preliminary viewing of the video data and from observations made during previous studies).

- Preparations that the subjects conducted, sometimes also making the robot move to a specific pose (several of them can actually occur in a sequence):
  - Move robot to particular position
  - Make robot turn to "face" item
  - Carefully position self to allow robot "see" item
  - Leave robot, fetch item from inside a storage place (e.g., get a glass out of a cupboard)
  - Leave robot, fetch item from an open surface (e.g., get a book from a shelf)
  - Stay with robot, take item out of a storage
  - Stay with robot, pick item from nearby open surface
  - Arrange item on surface (either after fetching it or after picking it)
  - Remove (optical) obstacle (e.g., open a door)
  - No explicit preparation
- Gestures:
  - Hold item in front of robot's camera, possibly shake or wave it
  - Fine specific deictic gesture ("fingertip pointing"), touching or nearly touching item with one fingertip
  - Hold item in one hand, fine specific deictic gesture with finger of the other hand
  - Coarse specific deictic gesture ("hand touch"), touching item with more than one finger or whole palm
  - Coarse directed deictic gesture ("hand pointing"), directing robot's attention towards item with full arm/hand gesture, not touching item, using more than one finger or whole hand for pointing
  - Coarse undirected deictic gesture ("hand waving" or "sweep"), not touching, sloppy gesture (e.g., sweep arm around body half)
  - No deictic gesture

a) b) c)

Figure 6.3: *Two often observed gestures for the presentation of a) an object and b) a location, seen from the robot's perspective and c) one way of showing a room without entering it.*

The preparations, particularly those involving the robot or subject take a specific pose can be interpreted also in the context of "spatial prompting" (Green and Hüttenrauch, 2006) and the general spatial management of robot and user (Hüttenrauch *et al.*, 2006a), as the preparation of the actual *show event* obviously is an act of adaptation to the given situation, with which the user aims to find an optimal pose with respect to robot and item to be presented.

Due to the semi–controlled setup, a quantitative analysis can only be (and was) used as basis for a qualitative discussion of the findings of this study. However, it is appropriate to quantify the observations and relate them to the assumptions and working hypotheses to identify patterns in preparations and gestures performed by the subjects. There is of course in some cases a close relation expected between the type of preparation and the following (if any) gesture. For example, the "arrange item on surface" preparation is most likely followed by some kind of pointing gesture, rather than by picking up the item again and holding it in front of the robot.

In some cases it was also observed, that the robot was moved as little as possible, trying to present as many items as possible at one position, which of course involved preparation phases of the type "leave robot, fetch item ..." or "position self", rather than "move robot to particular position".

Figure 6.3 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*, and c) illustrates one way of showing a "forbidden" room to the robot, in this case by using the preparations "moving and turning robot to particular pose" and "position self" after "removing visual obstacle" (opening the bathroom door).

Overall 129 SHOW episodes with the presentation of 34 *objects*, 61 *locations* and 34 *regions* were observed, counting repeated presentation of the same item[5]

[5]In some cases the subject tried to present an item to the robot while it was still processing

as separate episodes/items, but a sequence of preparations and gestures within the same episode[6]. The numbers, summarised in table 6.2, include some special cases that were taken into account: One of the *regions* was never explicitly shown to the robot, but its entrances were pointed out to delimit the *region*, thus the number of *regions* actually shown to the robot is reduced to 33. Instead the four entrances that were used to describe the room in question are counted as *locations*, already included in the overall number of 61. Furthermore, 9 of the *regions* were shown but not entered, which of course changed the SHOW sequence significantly. Thus, they are counted separately as 9 *non-entered regions*, with 24 *regions* remaining. One *region* was entered by the subject, while the robot had difficulties to follow inside due to a doorstep. This *region* is however counted into the 24 *regions* presented being inside, since the respective room was very small and the robot was about as close to the subject as possible when it was still standing in the doorway. Additionally when asked the subject confirmed that the intention was to show the room being inside it, despite the fact that it was one of the *regions* suggested as "areas that should be off-limits for the robot" (the bathroom).

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 6.2: *Summarised numbers for observations regarding the presentation episodes with explicit preparations and gestures.*

One of the objects shown to the robot was a stroller – the experiment leader decided to mark this as *object* rather than a *location* since it is very likely to move or be occasionally removed from the scene. Chairs on the other hand were initially

a previously given input. When the robot then was done with this previous task and stated that, the subject would realise this miscommunication situation and repeat the new specification.

[6]In several cases a sequence of preparations could be observed for one item.

counted as *location* despite the fact that they can be moved around – but in general they would be likely to be found at approximately the same place. In fact, chairs are difficult to conceptualise in the sense of the used conceptual hierarchy, and the observations regarding how subjects handled them suggest that there should be a differentiation between chairs (as *objects*) and armchairs (as *locations*). Tables

Preparation / Item category	All	Region	Location	Object	Non-entered region
all (excl. "none")	122	6 (5%)	61 (50%)	39 (32%)	16 (13%)
Move rob. to pos.	32	2 (6%)	21 (66%)	3 (9%)	6 (19%)
Turn robot	35	2 (6%)	22 (63%)	6 (17%)	5 (14%)
Position self	15	–	10 (67%)	3 (20%)	2 (13%)
Go fetch item f. cont.	2	–	–	2	–
Go grab item f. surf.	5	–	–	5	–
Take item f. cont.	6	–	–	6	–
Grab item f. surf.	13	–	–	13	–
Arrange item	1	–	–	1	–
Remove obstacle	13	2 (15%)	8 (62%)	–	3 (23%)
None	44	18 (40%)	22 (50%)	2 (5%)	2 (5%)

Table 6.3: *Categories of presented items in relation to observed preparations*

6.3, 6.4, 6.5, and 6.6 summarise these observations, listing the presented item type (concept) in relation to the observed behaviour and vice versa.

Table 6.3 shows that most of the observed preparations are connected to the presentation of a *location* or an *object*, while a large number of unprepared SHOW episodes falls under the category *region*. It has to be noted that also a high number of *locations* is shown without particular preparation of either environment or robot,

but this is due to the fact that 61 *locations* were shown compared to 24 *regions*. In general it is observed that a large number of preparations relate to the preparation of the robot (67 cases), or the preparation of the item if possible (27 cases), but not so many preparations in terms of the conscious positioning of the own body can be counted (15). In general, 61 (50%) of all explicit preparations are observed in the context of a *location* presentation, and 39 (32%) of them are seen in connection to *objects*. Obstacles are removed every now and then, but never in relation to an *object* that is to be shown – in cases where this would be appropriate it is easier and more intuitive to grasp the *object* and hold it in front of the robot. As far as the gestures are concerned a similar pattern is noticed.

Gesture \ Item category	all	Region	Location	Object	Non-entered region
all (excl. "none")	100	6 (6%)	55 (55%)	34 (34%)	5 (5%)
Hold item	28	–	1^{I} (4%)	27 (96%)	–
Fingertip point	4	–	1 (25%)	2 (50%)	1^{II}(25%)
Hold and fingertip	–	–	–	–	–
Hand touch	35	–	31 (89%)	3 (8%)	1^{II}(3%)
Hand point	23	–	18 (78%)	2 (9%)	3 (13%)
Hand wave / sweep	10	6 (60%)	4^{III}(40%)	–	–
None	29	18 (62%)	7^{IV}(24%)	–	4 (14%)

I: a chair
II: pointing on/to the respective door
III: "waving" with the instruction sheet
IV: incl. the four "delimiting" *locations* to specify the *region*

Table 6.4: *Categories of presented items in relation to gestures*

Table 6.4 shows that most observed gestures are of the type "hold item" (for *objects*) or "hand touch" / "hand point" for *locations*. As expected, not that many "waving" gestures were observed, most of them related to the presentation of

regions. Four "waving" gestures were observed in relation to *locations*, where the subjects would not release the instruction sheet and use this to "wave" (trying to make the sheet "point") in the direction of the item to be shown. It can obviously be assumed that there would have been a "hand point" gesture if the subjects did not have the rather instable piece of paper in their hands.

Item \ Gesture	all	Hold item	Fingertip point	Hand touch	Hand point	Hand wave / sweep	None
all	129[I]	28	4	35	23	10	29
Region	24	–	–	–	–	6 (25%)	18 (75%)
Location	62[I]	1[II] (2%)	1 (2%)	31 (50%)	18 (29%)	4[IV] (6%)	7 (11%)
Object	34	27 (79%)	2 (6%)	3 (9%)	2 (6%)	–	–
n-e R.	9	–	1[III] (11%)	1[III] (11%)	3 (33%)	–	4 (44%)

I: incl. two differing gestures for one *location* used sequentially
II: a chair
III: pointing on/to the respective door
IV: "waving" with the instruction sheet

Table 6.5: *Gestures in relation to presented items*

"No gesture" was observed mostly for *regions*, as expected, but also in rare occasions for *locations* and *non-entered regions*. In most of those cases actually a particular preparation was observed before – the subject(s) positioned themselves very carefully, thus using their complete body to "show" the item, e.g., by standing in a doorway to indicate an "exit" or "entrance". These cases were observed during the very particular sequence in which the subject presented the surrounding *region* by showing its four entrances / exits. Since this was in fact a single observation it

can safely be assumed that respective sequences are not representative and unlikely to be observed frequently.

Tables 6.5 and 6.6 relate the observed preparations and gestures to the type (concept) of the presented item. Thus, these tables give a better understanding of the quantitative relationship between certain types of preparations or gestures and the presented item type. A large number of *regions* (18, making 75% of the respective episodes) are shown without preparation, only in a few cases the robot is moved to a particular pose or an obstacle is removed.

Regions are also shown in most cases without any explicit deictic gesture (75%), or with a very vague "waving" gesture only (25%). Those gesture types on the other hand are only observed for a small number of *locations* (11, 18% of the respective episodes), while other gestures and preparations are much more often observed for them.

Locations involve in most cases a preparation of the robot or a positioning of the user (53, 64% of the respective episodes). According to the numbers, *objects* are most often shown after a "fetch" preparation (27, 71%, including all four listed particular "fetch"/"grasp" types) and by holding the item in front of the robot's perceptual system (27, 79% of the respective cases). For the *non-entered regions* that were shown without being entered it can be noted that the respective episodes include more likely a preparation of some kind than the cases of an entered *region*, (only 2, 22% of the non-entered *regions* vs 18, 75% of the entered *regions* are shown without preparation), mostly those preparations involve moving the robot (often to the respective door and turning it to "look" inside) and removing a visual obstacle – most often opening the door to the respective room. Further regarding the gestures used when a *region* is presented without entering, these are more often of a specific type (pointing) than vague or non-existent, though the overall number of observations here is rather small and maybe not entirely representative. However, it can clearly be stated that a *region* is only presented with a specific gesture when it is not entered, otherwise the gesture is vague or non-existent.

Result

From the quantifiable observations it is possible to conclude the following regarding the hypotheses proposed previously on page 129.

For the working hypothesis WH1 some support can be found in the data, since there are patterns observable across subjects, but a generalisation of this statement is not appropriate due to the limited number of subjects and the fact that only *one* robot was used for the trials.

The working hypothesis WH2 seems supported with respect to the data available, as there are clear patterns visible in the data, showing that differing presentation strategies are used for the different conceptual categories by each subject. Again, the observations were made only with *one* particular robot, thus a full generalisation is not possible.

Item / Preparation	all	Move robot	Turn robot	Position self	Go fetch item	Go grab item	Fetch item	Grab item	Arrange item	Remove obstacle	None
all	166	32	35	15	2	5	6	13	1	13	44
Region	24	2 (8%)	2 (8%)	–	–	–	–	–	–	2 (8%)	18 (75%)
Location	83	21 (25%)	22 (27%)	10 (11%)	–	–	–	–	–	8 (10%)	22 (27%)
Object	41	3 (7%)	6 (15%)	3 (7%)	2 (5%)	5 (12%)	6 (15%)	13 (32%)	1 (2%)	–	2 (5%)
n-e R.	18	6 (33%)	5 (28%)	2 (11%)	–	–	–	–	–	3 (17%)	2 (11%)

Table 6.6: *Preparations in relation to presented item categories*

Assumption 1, *objects*: For the presentation of 34 *objects* the following summarised observations are noted.

- For 94% (32) of all shown *objects* a preparation is used, of which 71% (27) of the cases involve picking up and carrying or arranging the *object*.
- In 16% (6) of the cases the robot was prepared by being turned towards the *object*, one of the cases corresponds to the one *object* that was arranged and pointed to.
- In only 3% (1) of the cases an object was arranged on a surface and then presented with a fingertip point. No other item was arranged.
- In only 5% (2) of the cases a fingertip point was used to present an *object*, while these two cases on the other hand make 50% of all fingertip points.
- In 79% (27) cases of an observed gesture in context with an *object* it is held in front of the robot. All but one of the "hold item" gestures overall observed refer to an *object*[7].

Hence, the data seem to support 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 it seems safe to assume that *if an object is fetched or picked up it is unlikely to be put down on a surface to be shown.* However, it can not be assumed, that an *object* is shown with a "fingertip pointing" gesture, but if such a gesture is observed, it is likely that an *object* is presented.

Assumption 2, *locations*: For the presentation of 61 *locations* the following can be noted.

- For 64% (39) of all *locations* a preparation is used, of which 52% (43) of the observations relate to movement (position and turn) of the robot.
- In 69% (43) cases of a robot movement preparation a *location* is shown, 18% (11) of those cases refer to a *non-entered region* and 15% (9) refer to an *object*.
- In 50% (31) cases a *location* is presented with a "hand touch" or "hand point" gesture, while these are 89% of the cases in which those gestures are used at all.
- In only 11% (7) of the cases no gesture was used to present a *location*, but this occurred in most cases after an explicit preparation (move and turn or positioning of the user).

[7]This is due to the fact that chairs were initially categorised by the author as "location" as are sofas and armchairs. Given the observations this categorisation seems less useful and should be changed.

From those numbers also assumption 2 seems supported 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, *regions*: For the presentation of 33 *regions* the following can be stated.

- 75% (18) of the entered, but only 22% (2) of the non-entered *regions* are shown without explicit preparation (moving inside the *region* not counted).
- All of the entered, but only 44% (4) of the non-entered *regions* are shown without any gesture or a "sweep".
- 60% (6) of the "sweep" gestures refer to *regions*[8].
- The distribution of observations made for *non-entered regions* is more similar to that for *locations* than to that for *regions*, though it is not exactly the same.

Summarising, also assumption 3 seems to be 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.*

Those results 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 results still can only be used as a basis for the generation of hypotheses that should be confirmed in further, rather controlled experiments.

For the conceptual hierarchy proposed in this thesis these results actually confirm the usefulness of the three levels used so far. Those conceptual categories or levels of the hierarchy correspond quite well to the presentation strategies that were observed for different item categories in the "home tour" scenario and it seems even

[8] the remaining 4 referring to *locations* are probably due to the fact that the subjects could not point more clearly, since they had the instruction sheets in their hands

possible to reassign a conceptual level to particular items according to the observations made in the study. This is the case for the particular object type "chair". Chairs were initially considered as *locations*, similar to their functional siblings "sofa" and "armchair", but it turned out that they were much more likely to be treated as *object* (moved – as "removed obstacle", arranged, even held up in front of the robot when explicitly presented). Thus, it seems that the used definition for *objects* (see page 49) should be revised in the sense that an *object* is not necessarily "small".

In the following a number of particular observations and technical issues are discussed that contribute to the general results of the study, but were not considered in the (quantitative) analysis presented previously.

Particular observations

Particular observations are considered situations that were observed only once (or with only one subject) or remarkable in the sense that they occurred frequently in a specific context.

Showing a room by presenting its entrances / exits

One of the subjects was very precise with the robot and presented a number of *locations* before remembering that also the "room" itself (in this case the living room) should be presented. Despite instructions and a short demonstration the subject hesitated for quite a while, discussing the issue of "how should I present the room...?" The conclusion for this subject was to make the robot move to each of the entrances / exits (in this case there were four, one to the hallway, one to the balcony, and two to extra rooms), stand in the respective doorway and present those *locations* as, for instance, "This is the exit to the balcony from the living room". The subject realised later when the robot (i.e., the experiment leader) registered the "hall" instead of "the entrance to the hall"[9] and turned around according to its rules, that there would have been the easier option of just specifying the "living room". This episode is certainly a very rare case but it is an example for the adaptation of the user to an *a priori* image of the robot's capabilities. The observation is another indicator for possible differences between the interaction with a robot and a human, based on the different images the respective person has of her interaction partner, which was one of the guiding questions for the third study reported in this chapter. For this study it was assumed that this was a single observation and that in general the instructions were sufficient for the subjects to get an understanding of the robot's capabilities that was sufficient to guide their

[9]This "mistake" of the experiment leader happened due to the subject being a bit hesitant, stating initially (already in the hallway) "And this is the hall ..." – which made the experiment leader react and feed the robot a *region* entry labelled "hall" – and continued "... er ... er .. the entrance to the living room from the hall" when it was too late to stop the robot, that already stated "I will have a look at it ..." and started its turning movement for *regions*.

interaction, but for following sessions it was made sure that the demonstration was clear and did not omit anything that might be needed by the subjects.

Praising the robot

The robot that was used for the study has a completely mechanical appearance, it definitely is a "rolling machine". The subjects were informed that not the robot but one of the experiment leaders would be listening to them, who would then translate their utterances to commands that would be given to the robot by typing / via a GUI. Despite all these factors, the experiment leaders observed in a number of cases that the subjects started to praise the robot ("Very good!") or say "Thank You" politely when a certain task was completed. This phenomenon seems to go along with the findings of Reeves and Nass (1996, 2003), who stated that people tend to treat machines as some sort of social agent or individual and was one of the underlying ideas for the third and last study discussed in this thesis.

Leaning on furniture

It was observed with several subjects that they would not just touch a *location* – or rather the "large object" that defined it – but that they would actually *lean* on, e.g., a table, to specify the *location* (see also figure 6.3). This phenomenon was only noted so far and not analysed in detail, but it seems to go along with the concept of *affordances* of furniture as described for example by Norman (2002).

Protecting the robot – or the furniture

In one case it was observed that a subject held a hand around a corner of a table when the robot was moving very close to this piece of furniture and seemingly had difficulties to find its way out of the very narrow passage. This gesture could either have been to protect the robot or the furniture, or both, but it looked exactly like the protective gesture that parents or other caregivers use to more or less unconsciously protect toddlers from bumping their head onto furniture corners when they are focused on playing and would not notice the "danger" above them. This phenomenon would fit as well into the category "individualising the robot", as the other ones discussed above.

Presenting locations by naming the objects in it

In several cases subjects presented *locations*, i.e., shelves, by labelling them according to their contents, for instance, "Here are the DVDs" or "Here are some books and magazines". This seems to be a similar case than those observed in the pilot study, where subjects did not present a *region* but the *locations* in them, since those were the important items to them. Here, it is not important to know that the furniture is called a "shelf", but it is important to know that the DVDs can be found "there". Similar to the "generic *region*" concept proposed for the implementation

of the environment representation discussed in chapter 4 a "generic *location*" entry could solve a respective situation.

Entering particular "forbidden" areas

The instructions for the subjects included suggestions for *regions* that should be shown to the robot, but as an area that would not be allowed to be entered in the future, for example, the bathroom. In most observed runs this issue was solved by presenting a non-entered *region* and adding the information "do not go there". In one case though the subject entered the bathroom with the robot – which was difficult because of a doorstep and a narrow passage, and presented it. In this case it seems that the subject had the notion that the robot could only know about *regions* that were presented "properly" and did not consider to just present the link to the room. It is not entirely clear, if this would have been different with different instructions / demonstrations, but since also this mismatch between instruction and execution / interpretation occurred only once, it was not taken into account for detailed analysis so far.

Technical issues

Some technical issues had to be dealt with, most of them related to the environments the study was conducted in. Those issues are summarised here to give an idea about improvements that would have to be considered for further setups of this kind.

Doorsteps

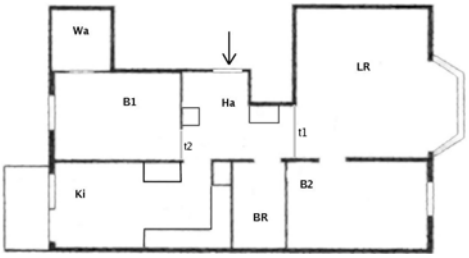


Figure 6.4: *Navigation issues the robot encountered. Two changes in flooring material (resulting in small, "one-directional" thresholds) are marked as "t1" and "t2", leading from the hallway (Ha) into living room (LR) and bedroom (B1). Also the passage out of the bedroom (B1) and directly into the kitchen (Ki) appeared difficult for the unexperienced user with the robot following.*

As an example of feasibility issues encountered, figure 6.4 shows the architectural sketch of one of the apartments with thresholds / changing in the flooring material and particular passages, that made it difficult for the robot to navigate and follow the user. All but one of the environments used for trial runs had at least one doorstep that had to be prepared with a small "ramp" to make it possible for the

robot to drive from one room to another, as illustrated with one example in figure 6.5. This also led to the decision to limit the trial runs to accessible areas of the

Figure 6.5: *One example of how the customised "ramps" over doorsteps were built, in this particular case two "ramps" (most often planks of plywood) on either side of the doorstep had to be used.*

given apartment or house, in the latter case of course the "tour" was already limited to one floor in the first place. None of the apartments or houses was particularly designed or equipped to have a mobile robotic platform moving in it. One of the apartments used for a trial was actually equipped with some particular ramps due to one of the inhabitants being bound to a wheelchair. Still, the small "thresholds" between different types of flooring material (about 2cm high) that were no obstacle for the wheelchair actually made it necessary for the experiment leaders to build an additional "ramp" for the robot. Such "ramps" had to be applied in most of the apartments and houses used for the study, and "construction time" and available material most often reduced the number of viable areas to the absolute minimum considered useful for the study (three *regions* including, e.g., a hallway, plus at least one *non-entered region*). These issues need to be considered when such a study setup is planned for, also considering the material the experimenting team needs to provide.

Narrow, angled passages

In at least two cases subjects experienced significant problems when they tried to make the robot follow into the kitchen, in one case in the apartment depicted in figure 6.4, coming from the bedroom, in the other case in a house where a very narrow aisle led to the kitchen over a doorstep. Inside the kitchen a cupboard and a counter formed a kind of second doorway leading to the left right after entering. Figure 6.6 illustrates this situation. Apart from this passage this particular house had very open spaces and wide doorways, which led to a kind of "break" in the flow of the run, when all of a sudden a problem was encountered with the navigation.

In both cases the subjects had to struggle to find a position where the robot a) would perceive them, b) was not blocked by them and c) could actually generate a useful path through the passage. In both cases the experiment leaders had to

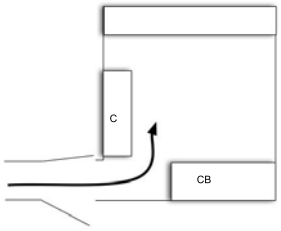


Figure 6.6: *The passage into the kitchen that caused some difficulties for the participant to get the robot into the room and later out of it again (illustration). The doorway had a width of approximately 80cm, as had the passage between counter (C) and cupboard (CB).*

help and remind the subjects that the robot could also be controlled with direct commands like "move forward" or "turn". In general this shows, that such direct navigation commands are absolutely crucial when a robotic system is to be moved out of the usually quite spacious office or laboratory environment. Furthermore the adaptation of the follow routines, for instance in terms of the distance thresholds, to the given environment is still an interesting issue to be investigated before the background of such differing environments.

Observations in comparison to the pilot study

The multiple room study (MRS) was designed as a follow up study to the presented pilot study. Nevertheless a number of aspects were changed and the focus was not so much the confirmation of the usefulness of the proposed environment model but the investigation of possible correspondences between the model and the interaction strategies of the users. Thus, also the instructions were much more detailed in the sense that the participants received a list of suggested items that additionally was organised corresponding to the three investigated levels of the conceptual hierarchy. It is hence an open question whether those explicit instructions and also the demonstration that the subjects observed could have influenced the behaviour of the subjects significantly. The author assumes that this was the case for the overall strategy for the tour, since the subjects tried to make sure to present all items on the list in a certain order (first the *region*, then the *locations* and *objects*, the latter sometimes in context with a *location*). Thus, they followed a more hierarchical structure than the participants of the pilot study did. On the other hand the range of observed strategies for the actual presentation of the particular items (preparation and gestures) seems to reflect the range of personal problem solving strategies given the task of showing items to a robot. Since both studies differ a little in their instructions and information given to participants, results from both seem to be relevant for the correspondence between human space representation, the proposed model and observable interaction strategies in the "home tour" scenario.

In a third study it was investigated, in how far it actually would be possible to observe human-human interaction in the particular scenario of the "home tour" to draw conclusions for the respective human-robot interaction scenario.

6.4 Humans guiding a robot and a person – the Comparison Study

Kirsh (1995) stated that the observation of human activities – the interaction with the environment – allows to draw conclusions for the design of robotic or in general technical applications. Consequently, it might be considered sufficient to observe a human interacting with a human in a "home tour" scenario, to gather information for the design and implementation of respective environment models and interaction frameworks for a service robot. The previously conducted pilot study though was designed under the assumption that there might be significant differences in the behaviour of human "guides", depending on whether they interacted with another human or a robot, and thus the human should be observed in the interaction with a robot. Reeves and Nass (1996, 2003) found that people tend to communicate with machines as if those were individuals, but not necessarily human. This trend is dependent on the type of machine and how it is perceived by the human user in a particular context. Observations made during the pilot study and the MRS suggested in fact, that there are some particular strategies in the interaction with a robot that would seem surprising in the interaction with another human. For instance, a particular way to make the robot follow through a door was observed in the pilot study (see page 123) and another particular case was discussed previously in context with the MRS, in which the subject presented a *region* by pointing out all its entrances or exits. To find out more about the observable similarities and differences in the interaction strategies people use when they show around another *person* or a *robot* in an environment in the same scenario, the third user study – in the following referred to as "comparison study" – contributing to this thesis was designed and conducted.

The author contributed mainly to the idea and overall design of the study, and was involved in the procedure as assisting experiment leader, controlling the robot with the previously described "tool" version of the HAM system implementation. The detailed design, realisation and analysis was assigned as a master's project (Swedish "examensarbete") to Farah Hassan Ibrahim, supervised jointly by Helge Hüttenrauch and the author. Since at the time of writing of this thesis the respective analysis and report were not yet available, the study is described and discussed by the author herself in the following.

Scenario

As for the pilot study the scenario was a "guided tour" in the office environment of the CVAP group (illustrated again in figure 6.7) where in this case a number of particular *regions* should be presented (hallway, computer vision laboratory, bathroom, office and kitchen), and in each of those a number of *locations* and *objects*. The significant differences in the scenario are thus the explicit list of items to be presented and the fact that subjects should guide both the robot and a person in two runs through the environment. To make the level of detail plausible for the

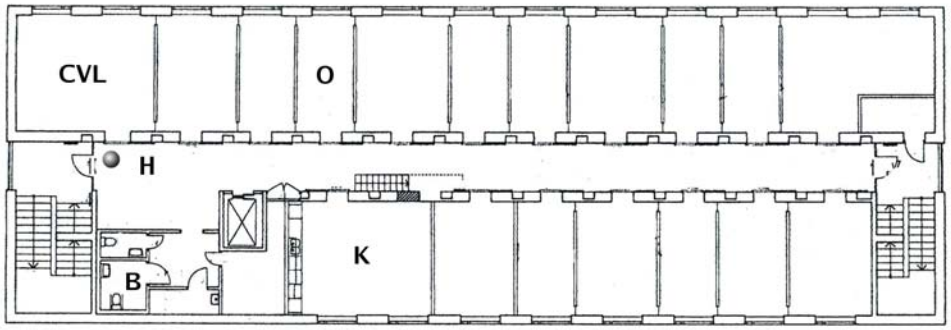


Figure 6.7: *The office environment as it was used for the comparison study. Depicted are the hallway (H), the computer vision laboratory (CVL), the bathroom (B), one office (O) and the kitchen (K). The tour started in all cases at the "dot" in the left corner of the hallway.*

tour with the person, the participants received a rather sophisticated background story, or script, in which the robot and the person appeared as "housekeeper" or "cleaning help" that needed to be instructed carefully. This particular scenario was used to make subjects present or discuss items for the human "cleaning help" that usually can be assumed familiar to other adults and thus would not need to be presented or discussed. Hence, it could be assumed that the participants would use the same level of detail in terms of the proposed conceptual hierarchy when they presented the environment, and it was possible to observe similarities and differences for the actual presentation of the particular items. As for the other two study setups it was initially assumed, that subjects should be familiar with the environment. It turned out though that it was not possible to find enough participants in the used office environment that had not participated in the pilot study and would not have too much insight in robotics research. Taking the robot out of the laboratory again was not an option since this would have required resources (time, a car and subjects with apartments large enough to be considered useful) the involved researchers did not have available at that time. Thus, it was decided to use the environment the robot was easy to use in, that would not pose any problems for the robot to navigate, and that appeared at least as a typical university laboratory environment to the subjects (students from the campus, not involved too deeply in robotics). It was also decided to give the participants the opportunity to make themselves familiar with the environment according to the list of items they should present. In all cases the tour started at the same point in the hallway, close to the entrance to the office environment from the stair case.

Method

The study was again of an exploratory character, but aimed to have some controlled aspects, so that a number of assumptions could guide the trials. Consequently, the subjects received the same instructions and task and it was in general aimed to keep the controllable conditions equal for all trials (i.e., the used implementation of the robot control system was not supposed to be altered and the instructions were to be kept on the same level of detail for all participants). Regarding the analysis this had as for the other two studies the consequence that any quantifiable data could only be interpreted as trend indicators for a qualitative investigation of human-human compared to human-robot interaction in a "guided tour" scenario.

The study was designed as a "semi-controlled within-subject" setup, i.e., each of the trials consisted of two runs, one in which a person was shown around and one with a robot. To balance potential habituation effects the order was altered so that half of the group started with the person and the other half started with the robot. A couple of spontaneous informal test runs performed by the main experiment leader suggested that people who had been guided around in an environment and were asked to guide someone themselves would more or less copy the actions of their guide, even a couple of hours after they had been shown around. Thus, it was assumed that none of the persons being shown around should act as "guide", so that the "guides" should only be prepared with a list of items and a sketch of the environment, instead of being shown around initially by an experiment leader. To keep the overall number of participants per trial as low as possible, the participants acting as "cleaning help" were asked to do that twice so that the following methodical setup was used: Participant A shows around the robot - participant A shows around participant B - participant A is interviewed briefly and done - participant C shows around participant B - participant B is done - participant C shows around the robot - participant C is interviewed briefly and done. Thus, the "participant B" group could serve for both groups A and C as receiver of the "tour" and 16 trials (including two pilot trials) could be conducted with only 24 participants that had to be recruited.

Subjects

The 14 + 7 study subjects were students of the Royal Institute of Technology, recruited in different places around the campus, the 2 (+1) participants of the pilot trials were students as well, recruited from a group of friends of the experiment leader, since they were expected to spend more time than the regular subjects to improve the setup if necessary. In the following only the 14 participants that acted as "guides" in two runs each are referred to as "subjects", since the 7 participants acting as "cleaning help" could more or less be considered as "guinea pigs", not taking part in the action directly. The group of subjects was rather homogeneous in factors like age and education, as can be expected with a group of undergraduate students, and comprised 7 males and 7 females. In most of the cases they

were invited in (peer) groups of three participants (2 subjects + 1 "cleaning help actor") which resulted in mostly gender homogeneous pairings for the trials. The students were recruited from the fields of Mechanical Engineering, Biotechnology, Vehicle Engineering, Chemical Engineering, Computer Science, Industrial Economy, Physics, Microelectronics, and Material Design and Engineering. None of them had particular experience or knowledge in the field of robotics.

Instructions

The subjects received an instruction sheet in general similar to the one used for the pilot study that explained the task and the environment. Regarding the environment however the instructions were more specific than those used for the pilot study. The sheet included a sketch ("map") of the floor, depicting the rooms or *regions* that should be included in the tour. The subjects got the task to guide around both a robot and a person in the role of a "cleaning help" or "housekeeper" and to explain for particular *locations* and *objects* what they should do or not do with them. Five *regions*, ten *locations* and nine *objects* – according to the definitions proposed in chapter 4 – were mentioned as items to be discussed during the tour. The instruction sheet suggested, for example, that occasionally there would be *dishes* spread on the *tables* in the *kitchen* that should be put in the *sink* before 6pm. The *flowerpot* in the *office* should always be located on the *desk* near the *window*, there should not be any *papers* lying around on the *printer desk* and the *printer* should not be moved during cleaning. With this suggested background stories and the particular instructions it was made plausible for the subjects to present and mention items in the environment that they might not have considered important to point out for a person that should be guided around.

They were informed that the robot could be controlled with spoken commands and that it would also accept spoken explanations about the environment. Since the purpose of the study was to compare what the subjects did in the two runs according to their own understanding of the situation they did not get an exact list of commands or utterances to use.

After exploring the surroundings on their own with the sketch and the instructions, the subjects had the chance to ask any question about the trial they had in mind. In fact, the instructions did not tell too much about the robot's functionalities and responses, since it was assumed that it was sufficient to give them approximately the same information about possible commands as was given to the subjects in the pilot study, but explanations of potential problems had to be reduced significantly to keep the instruction sheet readable and comprehensible in a rather short time despite the background story.

The subjects were also informed that the robot was moving autonomously but that it was monitored constantly by the assisting experiment leader so that inconvenient situations (tracking problems, navigation close to obstacles) could be resolved. Each subject was given 15 minutes to interact with the robot. For the interaction with the other person the same time limit would have been applied, but no subject

needed as much time – the runs with persons took about two minutes on average (ranging from 1 to 4 minutes, with the median at 2 minutes, 50 seconds). The time limit for the robot trials was applied since it was observed in the pilot trials that this time was clearly sufficient to conduct the complete tour. It had thus to be assumed, that trials exceeding this time suffered from technical problems with the robot, which was considered an inconvenience subjects should not be forced to deal with for a longer time. Overall in four trials the time limit actually had an effect on the tour, since one room had to be omitted, otherwise the subjects were able to complete their runs, needing from about five to 15 minutes, which includes the time the robot needed to move around in the environment without any immediate interaction taking place. Excluding this time the subjects spent on average about six minutes interacting with the robot (ranging from 2 minutes, 20 seconds to 9 minutes, 50 seconds, the median being at approximately 5min, 20sec). Including preparation and a short interview after the two runs each subject spent between 45 and 60 minutes in the laboratory. All participants received a cinema ticket as compensation for their time.

Technical realisation

As for the MRS the "tool" version of the HAM implementation was used for the runs with the robot. Again, the speech recognition / speech processing was simulated in a "Wizard-of-Oz"-like setup by the assisting experiment leader who controlled the robot via the GUI. Since the Royal Institute of Technology offers quite a lot of international programmes a large number of participants were not native Swedish speakers, which made the choice of Swedish as interaction language not obvious. Additionally the used tool version of the HAM system implementation assumes English as interaction language, and choosing Swedish would have required quite some changes. Thus, English was chosen as the overall most neutral language for all participants and used in the sessions.

For each group consisting of two "active" subjects and one participant "to be shown around" the robot was initially placed at the starting position for the tour for the first run, then it was brought to the kitchen where the HAM system was started to collect the tracker data to observe the interaction between the two persons in the following two runs. After that the robot was moved back to the starting position for the last run in the block of sessions. Thus, the waiting time for the participants who where all invited to their particular runs directly could be reduced to a minimum. As mentioned before the runs were organised as follows within each of the groups:

- subject A and participant B arrive and receive instructions and prepare while
- the robot is prepared for the first run
- subject A shows around the robot (first run)
- the robot is prepared to collect sensor data of the second and third run

- subject A shows around participant B (second run)
- subject C arrives and receives instructions and prepares while
- subject A is interviewed, receives the cinema ticket and leaves
- subject C shows around participant B (third run)
- participant B receives the cinema ticket and leaves while
- the robot is prepared for the last (fourth) run
- subject C shows around the robot (fourth run)
- subject C is interviewed, receives the cinema ticket and leaves.

To verify the feasibility and validity of the study setup, two pilot trials were conducted after which the instructions were adjusted to make the complete tour possible within the time limits, particularly regarding the robot's rather slow movements in cluttered areas and narrow passages.

Observation methods and data collection

Each of the trials consisted of the preparation phase, the runs with the robot and the other participant and a short interview. The preparation phase included a short questionnaire on the background of the participants and their knowledge or experiences about robots and robotics. The trial runs were recorded on digital video with one camera operated by the experiment leader, as were the interviews. Additionally the data available from the sensory system of the robot were stored during the "robot-runs" and during the "person-runs" the robot was placed in one of the rooms of the tour (the kitchen), so that data from the tracking system were available and stored to find out more about the spatial relationship between guide and recipient. It turned out though that the time the subjects spent in the kitchen together with the human interaction partner was too short to collect significant data.

Hypotheses

Two working hypotheses were worked out that guided the study design:

- WH1: People compare a mobile robot to a human and treat it similar to a human when they interact with it, but they do not see an actual "human" in the robot.
- WH2: People use more time and are much more precise when they explain the environment and a task to a robot than they do with another human in the same context.

WH1 is meant in the sense that the subjects would use communication patterns that are usually observed in human-human communication (e.g., use a narrative style rather than verbalised "computer command input", or utter polite phrases like "thank you"), but there are still differences observable in the communication, e.g., gestures might be applied differently. From the particular observations of the previous studies it even seemed that there could be fundamental differences in the strategies for, e.g., navigation through narrow passages and the presentation of a room.

Observations, analysis and results

The digitalised video tapes from the 14 trials were analysed in terms of the occurrence of particular observations and phenomena. Compared to the particular strategies that were observed for the navigation through narrow passages in both the pilot study and the MRS none of the subjects of the comparison study used such an explicit strategy with the robot. The subjects seemed in general a bit hesitant to use near navigation commands as "turn left" or "move forward" if they were not explicitly told by the experiment leaders that this was an option to get them out of a "deadlock" situation. Overall, the subjects seemed to suppose similar "navigation" capabilities for the robot as for their human interaction partners and would, for example, just walk into a room with the robot following them, obviously assuming (correctly) that it was capable of doing that. Thus, no general differences in the movement strategies of the subjects were observed, given the fact that of course the resulting coordinated movement appears different, since the robot moved significantly slower than a person. In the following the observations are discussed according to the conceptual category (*region*, *location*, or *object*) that was to be presented, also considering the particular order that was most often used by the subjects, probably due to the background story and the placement of the robot.

Regions

Of 70 (14x5) possible *regions* 63 were shown to the robot and 66 to a person, where all *regions* are counted that were entered at some point during the tour, even if they were not explicitly labelled or mentioned, and those that were explicitly presented but not entered. This excludes only *regions* that were ignored completely by the subjects, either due to the time limit being exceeded or because they deliberately chose to leave them out from the tour. Thus an event of "entering but not presenting" is still an interesting aspect that is taken into account for the analysis.

Grouping the *regions* according to their positions in the environment, it can be stated that 12 of 14 subjects (85%) presented first the group "hallway, laboratory, bathroom" and then the group "kitchen, office" in some relative order, both to the robot and the person and only two subjects had different orders in at least one tour, while only one chose to move along the larger part of the hallway three times. Furthermore 12 of 14 subjects (85%) used exactly the same order of *regions* both

for the tour with the robot and the one with the person, one had a switch within one group and one had a different starting position with the person than with the robot. Summarising it can thus be stated that the overall strategy was the same for the tour with the robot and the one with the person. Since the subjects received an instruction sheet that listed the *regions* and their particular properties in the most often observed order, these observations are not surprising. More interesting are the strategies used to actually present the listed *regions*. Six different strategies were observed, where a differentiation is made between "mentioning" a *region's* name (e.g., saying "now we'll proceed to the kitchen") and "presenting" a *region* (e.g., saying "this is the kitchen"):

1. presenting / labelling the *region* when being inside it, e.g, entering the kitchen and saying "this is the kitchen"
2. mentioning the *region's* name before going there, e.g., saying "we are going to the kitchen now" when leaving the previously presented place
3. mentioning the name before entering and labelling the *region* when being inside it
4. presenting / labelling the *region* in front of it and entering it then, e.g., saying "... and this is the kitchen" immediately before entering the kitchen
5. presenting / labelling the *region* from outside, i.e., saying "this is the office" standing outside and pointing through the door
6. not mentioning the name at all, but being inside it

Table 6.7 summarises the observations according to the particular rooms / *regions*. Looking at the numbers considering the six different strategies there is a tendency to be noted, that *regions* are more often *explicitly labelled being inside the respective room* to the robot and more often *mentioned being outside of the room* to a person. In general *regions* are more often presented (labelled) explicitly to the robot than to the person (61% vs 56%), though the numbers do not indicate a significant difference. An overall surprising observation is that the subjects tended to be more precise with the person and put less effort in labelling *regions* to the robot than expected. The fact that rooms like the office (labelled sometimes as "Elin's office") or the computer vision laboratory were labelled quite often to a person is not very surprising since they needed a particular specification (Elin's office, opposed to "the" or "an" office, and a "computer vision laboratory" is obviously a room existing in only a few indoor environments and can not be expected to be encountered on a regular basis). However, the subjects labelled the "hallway" explicitly quite often to a person, in fact more often than to the robot, which is somewhat surprising. One explanation could be that they picked it up from the instruction sheet and uttered it in the flow of the interaction. On page 157ff. some reflections on the influence of the instruction sheet are presented within the paragraph on "general observations".

Receiver	Region / Observation	Presented inside	Mentioned before entering	Mentioned before and presented inside	Presented before entering	Presented outside	Not mentioned / presented, but entered	Sum presented inside	All presented outside or mentioned only
Robot	All	18	9	14	6	1	15	32 (51%)	
	CV Lab	5	1	2	1	-	4	7 (54%)	
	Hallway	3	4	1	-	-	5	4 (29%)	
	Bathroom	2	-	5	2	1	2	7 (58%)	7
	Kitchen	5	1	4	2	-	2	9 (64%)	
	Office	3	3	2	1	-	2	5 (45%)	
Person	All	15	15	6	12	4	14	20 (30%)	
	CV Lab	8	1	-	2	-	2	7 (54%)	
	Hallway	5	3	2	1	-	3	7 (50%)	
	Bathroom	-	3	1	4	3	1	1 (8%)	16
	Kitchen	1	5	1	3	-	4	2 (14%)	
	Office	1	3	2	2	1	4	3 (23%)	

Table 6.7: *Observations regarding the presentation of regions to either a robot or a person, listed according to the particular region (CV Lab = Computer Vision Laboratory).*

Locations

Ten *locations* were listed as suggestions of what to present or discuss within the background story of the instructions in the *regions*: computer desk or computer (in two rooms), printer desk, printer, toilet, sink (in two rooms) or sink counter, table (in two rooms), and television or TV. In most cases the subjects only discussed items that were on the list, but since a number of not listed items were presented as well, all observations are considered in the following as overall numbers. Thus, 121 *locations* were presented or mentioned to the robot, which gives an average of 1.92 per *region* and 122 *locations* were presented or mentioned to the person (1.85 per *region*). Hence, the overall numbers do not show any significant difference, possibly also due to the instructions, since only in four cases a *location* was named to the robot or the person, that had not been on the list (dishwasher, whiteboard, cupboard, and sofa for the robot; dishwasher (twice), whiteboard, and coffeemaker for the person). More interesting is again the differentiation of strategies that could be observed in the context of *locations* being presented. Three different strategies for the presentation or introduction were observed:

1. presenting the *location* explicitly (e.g., saying: "This is the printer desk" or "Here we have the printer desk")
2. mentioning the *location* as part of the instructions for the particular *region* only (e.g., saying "Make sure that the computer desk is cleaned every other day")
3. presenting the item first and giving instructions later.

Since presenting or mentioning a *location* was often accompanied by some type of gesture a second list of observations was used to classify those:

- pointing to the item, more or less precisely
- point in a "sweeping" move around
- touch a *location* or the object defining it
- not using any gesture

Observation \ Receiver	Robot	Person
Presented explicitly	30	19
Mentioned in instruction	44	68
Presented and then mentioned	47	36
Presented (line 1+3)	77	55
Pointed to *location*	44	71
"Sweep"	13	17
Touched	30	15
No gesture	34	19

Table 6.8: *Observations regarding the presentation of locations to either a robot or a person*

Table 6.8 summarises the observations regarding the presented *locations*. A clear difference can be noticed here in the strategies; the subjects tended to present

locations to the robot more often than to persons, while mentioning them only in the instructions was observed more often with persons than with the robot. This is not surprising, given the fact that a human conversation partner with a comparable social and cultural background can be assumed to have a similar world knowledge base to start with. For the robot this is not the case, the subjects had to assume a different or no knowledge base and had to find out during interaction what they could expect from the robot. An interesting observation in this respect is, that the presentation or explanation was strikingly more often accompanied by a pointing gesture, when the interaction partner was a person, while the robot was instructed comparably often by touching the item or without any gesture. Overall these observations seem to correspond to the idea of "being more precise and explicit" with the robot and use more of a conversational style with the person, with the exception of the fact that less gestures were used with the robot. In some cases this was probably due to the fact that the subjects positioned the robot to face the item very carefully and seemed to assume, that an explicit gesture was thus superfluous.

Objects

Nine *objects* were listed in the instructions, spread over the five different *regions* relevant to the tour: paper (sheets), chair(s) (four times in three rooms, differentiating between a regular chair and the office chair in the office), a clothes-hanger, dishes, a flower, and the TV-remote control. An additional *object* that was named in one case (both to robot and person) was a paper bin. Chairs were counted as *objects* due to the considerations resulting from the multiple room study that were previously discussed, and also during the trial runs of the comparison study subjects showed a tendency to touch and move chairs when they were presenting or mentioning them.

Overall 90 *object* presentations or mentions to the robot were observed (1.43 per presented *region*), and 96 *object* presentations or mentions (1.45 per *region* on average) were counted for persons. Similar to the *locations* different presentation strategies could be noted, again differentiating between "mentioning in instructions" and "presenting explicitly". Three different strategies were observed and counted in terms of occurrence in the interaction:

1. presenting the *object* explicitly (e.g., saying "this is the remote control")

2. mentioning in instructions only (e.g., saying "make sure that the flower is watered every week")

3. presenting first and mentioning later again.

Observation \ Receiver	Robot	Person
Presented explicitly	20	13
Mentioned in instruction	46 (51%)	56 (58%)
Presented and then mentioned	24	27
Presented (line 1+3)	44 (48%)	40 (41%)
Pointed to *object*	38	62
Touched / held	21	13
No gesture	31	21

Table 6.9: *Observations regarding the presentation of objects to either a robot or a person*

As for the *locations* different gestures could be observed that accompanied the presentation of the *objects*:

- pointing to the *object* (any kind of pointing gesture is subsumed here)
- touching / holding the *object*
- no explicit gesture

Table 6.9 summarises the observations regarding the presentation of *objects*. There is a trend to be seen that *objects* just as *locations* are more often presented explicitly to a robot and more often mentioned only to a person. However, the numbers do not indicate a striking difference. An interesting observation is though – as it was for the *locations* – that the subjects tended to point to the relevant item significantly more often when they interacted with a person than with a robot, while they did not compensate for this fully by being *more explicit* (touching or holding the item) with the robot, but were in fact even *less explicit*, in the sense that they more often did not use any gesture when they interacted with the robot than with the person.

General observations

Some general observations were made that were even more difficult to quantify than the particular strategies used for the presentation of items. Those qualitative

observations are discussed in the following, also considering the statements uttered by the participants regarding their experiences during the session and their own reflections about their behaviour.

In a large number of trial runs the subjects obviously tried (and managed to a large extent) to treat both the person and the robot in the same way, which they also confirmed in the interviews. The instructions however did not mention anything about "trying to act in the same way", thus this phenomenon needs to be discussed more thoroughly, leaving an open question for further studies. It also seems questionable whether the quantitative interpretation of the observations remains an entirely valid answer to the question about the differences and similarities between human-human and human-robot interaction before this background. Nevertheless, the observations can be regarded from a slightly different point of view, stating that *despite* their willingness to act in the same way with both interaction partners the subjects actually *succeeded with this intention only to a certain degree* (not considering differences in the interaction flow that were caused by the robot's abilities or lack of those).

The intention to give the same tour to both the person and the robot actually led to interesting particular observations. E.g., in one case the subject guided the robot first, using very clear and simple presentation techniques. The subject only labelled the items, did not use any articles or explanations along the line of the instruction script and pointed to or touched the named item very thoroughly. This resulted in a very precise but simple interaction flow, i.e., entering one of the rooms, pointing around and uttering for example "kitchen – chair – table – sink", giving the robot the opportunity to give feedback – "stored information about..." – after each item, and leave the room. Given the observations from previous studies this was not a very surprising style to interact with the robot, though it had been a rare observation (see also "particular interaction styles" on page 123). More surprising was the fact, that this particular subject interacted with the person in the following run *exactly in the same way* which obviously confused or amused the second participant who expected to be shown around in the role of an – obviously human – adult "housekeeper" or "cleaning help".

In other cases the robot (i.e., the assisting experiment leader) had difficulties to cope with the flood of information that was given by the subject in one sentence, using entirely natural communication strategies, language, and modulation in the same way this was done earlier or later with the person, e.g., entering the kitchen, stating "... this is the computer vision lab and here I want you to make sure that all computer desks are cleaned every other day and that none of those chairs is moved out of the room - oh and ...". In those cases the subjects were in the beginning surprised by the rather slow interaction offered by the robot, that gave feedback on each of the named items, but with a certain delay caused by the time needed to process the utterances. Still, they kept their narrative style during the rest of the tour with the robot, just slowing down a bit in their frequency of the utterances.

Another interesting observation was made in a number of cases, both with the robot and the person, when subjects seemed to *present* all items listed within one

region by just reading them aloud, and then start to give the instructions about the "cleaning" and handling of them. Since this strategy was observed a couple of times and in one case the subject even started to read a part of the (meta) instructions about the scenario aloud to the robot, it seemed a questionable decision to have subjects carry around their instruction sheets, instead of forcing them to remember the instructions or improvise. During the study, however, it was decided to make the conditions comparable for all trials and allow all subjects to keep their instruction sheet to feel more comfortable.

Along with these observations some issues regarding the use of English as trial language were encountered. In a couple cases it turned out that the participating students were not as familiar with the English language as expected in the given international context of the Royal Institute of Technology (KTH). As a consequence those subjects felt obviously rather uncertain about the task and used the instructions not only as suggestions for a background story but as script to read from during the interaction. These issues certainly have to be taken into account for the interpretation of the overall observations. Furthermore some misunderstandings occurred that led to problems with the interaction and possibly altered the participant's original choice of presentation strategies. The robot uttered "stored information about <item>" when an item was presented or mentioned. In a couple of cases the subjects did not understand the word "stored" properly, assuming the robot asked for more information ("start information about ..."), and repeated their utterance. This again forced the assisting experiment leader to have the robot utter its confirmation "stored information about ..." and the interaction ended up in a loop that could only be resolved by the experiment leaders, using both English and Swedish in attempts to explain the robot's utterance. In some cases this led to even more confusion, probably due to the uncertainty of some subjects with both the English and the Swedish language and an additional uncertainty due to their lack of experience with technical systems of the given kind. However, during the study this particular problem caused by the word "stored" was discussed but not solved, since this would have forced the experiment leaders to allow different conditions for different subjects, which would have been against the initial decision to keep the conditions as comparable as possible.

From the interviews and the observations it is possible to state that there is a trend to anthropomorphise the robot (e.g., assuming similar abilities to receive and process complex instructions for both a person and the robot), but still see it as a machine during interaction. Unfortunately not all subjects answered the question "did you see an individual or a machine in the robot" explicitly, thus it is difficult to quantify this observation in an appropriate way. Additionally the author is aware of the fact, that this or a similar question should have been formulated differently, since the term "individual" was difficult to consider for the subjects in the short time used for the interviews (about two minutes). As the interview was only roughly scripted, the particular terminology used during it was not discussed in detail beforehand. Most subjects however used – spontaneously – a personal pronoun ("he") for the robot instead of a neutral one ("it") in the discussion, while

the interviewer referred to "the robot", not suggesting any grammatical genus. The fact that the subjects tended to use a male pronoun is interesting in itself, since the instruction sheet introduced the robot as a female ("Minnie"), while due to technical reasons the voice used for the text-to-speech system on the robot appears male. Obviously, the male voice together with the technical appearance of the robot had stronger influence on the subjects' impression than the name.

Result and discussion

Summarising the following result can be extracted from the observations made during the sessions.

Considering the hypotheses that were formulated in the design phase of the study it does not seem possible to either reject or support them fully. However, WH1 ("People compare a mobile robot to a human...") seems to be supported with the statements from the interviews and the observation, that a large number of subjects used a rather conversational style (combining "mentioning" with "presenting" items) when they interacted with the robot. This is not surprising given the general ability and willingness of humans to anthropomorphise a presumably somewhat intelligent agent or a machine when they interact with it (e.g., Reeves and Nass, 1996, 2003).

WH2 ("People use more time and are much more precise when they explain the environment and a task to a robot than they do with another human in the same context") has to be revised against the background of the observation that subjects tended to use more specific gestures and explanation when they interacted with a person than with a robot. In terms of time it is difficult to draw a conclusion due to a rather slow response frequency the robot offered, which of course made subjects slow down the interaction. However, the assumptions cannot be rejected entirely either, since there are obvious differences observable in the interaction style and presentation strategies, *despite* the fact that subjects *explicitly* tried to give the same tour to robot and person and "treat them equally". The author's personal conclusion is thus that it is necessary to consider experiments in a realistic scenario with a prototypical robot to inform a potential design process and decisions for underlying models and techniques, though there is a certain similarity to human-human interaction that can probably be exploited as initial background.

Regarding the methods and decisions made in the design process of this study it seems though that despite the two pilot trials a number of problems – regarding the understanding of the instructions and obvious uncertainty about the task – only manifested themselves during the actual sessions, possibly even setting the validity of the setup in question. Given these issues and the generally exploratory character of the study, the author refrained from any attempt to a statistical analysis of the quantifiable data, and only interpreted the numbers as trend indicators. For further studies with similar within-subject setups it seems necessary to consider more thorough pilot trials, also regarding the possible priming effect of the first run for the second. Potential language issues need to be considered more thoroughly,

since they turned out to have too strong an impact on the observed presentation strategies, as well as the instructions seemed to have.

Results in comparison to previous studies

Results from the previously discussed MRS were used to inform the analysis of this comparison study, particularly for grouping the items and observed strategies. Since the character of the study was different from the multiple room study the data were not considered as additional material for the respective analysis, but in general the data do not contradict the results from the MRS. A comparison with the first study (the pilot study) does not give any contradictions either, while a more detailed discussion seems not necessary due to the much more complex and comparably controlled setup applied for the comparison study.

6.5 Summary

In this chapter three user studies were presented that all were conducted with the robot "Minnie" in a setup of the "home tour" scenario. The first study is considered a pilot study that confirmed the usability of the previously proposed environment model as well as the applicability of a tool version of the HAM system implementation to study setups of this kind. The second study was designed as a follow up study to both a previous Wizard-of-Oz experiment and the pilot study. An interesting aspect with this study was that it was conducted in subjects' homes instead of in the laboratory. The question relevant to this thesis was whether the conceptual hierarchy used for the proposed environment model and implementation could be confirmed in people's interaction styles and strategies. This question could be answered positively to a certain extent. The third study investigated differences and similarities in human-human vs. human-robot interaction in a "home tour" scenario in a within-subject setup. The analysis of the study indicates subtle but not ignorable differences in the interaction. This supports the author's assumption that the observation of a human interacting with a real robot gives the best information on this interaction – as opposed to observing a human interacting with a human and transfer the conclusions to robot system design immediately.

Chapter 7

Summary and concluding discussion

Human Augmented Mapping (HAM) is an approach to robotic mapping that integrates information given interactively by a human user into the mapping process. The concept does not aim to propose a new and sophisticated method of robotic mapping or simultaneous localisation and mapping (SLAM). Neither does the concept in itself provide new ways of dealing with interaction. Nevertheless both, the mapping process and the interaction of a human user with a service robot can be facilitated and supported by integrating these two fields in one framework.

Based on a broad variety of work presented in the fields of a) robotic mapping, particularly considering approaches to topological mapping and segmentation of space, and b) human-robot interaction, communication and cognitive modelling this thesis proposed a conceptual framework for Human Augmented Mapping. A generic environment model based on a conceptual hierarchy employed in a partially hierarchical graph structure was suggested. This model was discussed in the context of as well an implemented system as a number of user studies. The framework offers advantages for the robotic mapping process by integrating information given by the user in the interactive setting of a "guided tour" or "home tour" which was assumed as the background scenario for the discussion, not at least to limit the technical efforts of the underlying work to a manageable level. This information, either given actively by the user or asked for by the robotic system can help to disambiguate certain situations. On the other hand, the interaction with the user about the given environment is possible in a natural way only if the environment representation built up by the robot corresponds to the human concepts. Therefore, the hierarchical environment representation based on psychological findings and common spatial concepts is used as a link between robotic mapping and interaction.

7.1 Summary

The thesis presented an architectural framework for HAM and picked three main aspects for a system design for implementation and both experimental and user related evaluation: A tracking and following mechanism that facilitates the "guided tour", the conceptual environment model and its implementation in a topological graph structure, and the mapping subsystem that subsumes the building of the graph structure and the underlying segmentation of the environment.

A partially hierarchical structure for the environment model was proposed, using the concepts of *regions* and *locations* for the actual implementation work and including *objects* for the more conceptual, user related investigations. The *regions* (corresponding to architecturally or functionally delimited areas in indoor environments, typically rooms) form the nodes of a topological graph structure.

To achieve a meaningful segmentation of a presumably arbitrary indoor environment into *regions* that build the nodes of this graph, a laser range data based approach to the representation of such *regions* was suggested. This representation was also used and discussed in the context of the detection of transitions between *regions* that allowed the robot to take the initiative in the interaction during the assumed "guided tour". The author found that in the given interactive setting the suggested representation of *regions* can be applied as a very concise, computationally relatively inexpensive method for the detections of transitions and – within certain limitations – also for the classification and thus recognition of particular *regions*.

Three user studies were conducted in the context of the thesis work. The first one served mainly as a pilot study investigating the usability of an initial implementation of the HAM system in a user study context and setup. Additionally a couple of assumptions for the proposed environment model were supported by the observations made during this pilot study. Already with a small number of subjects a rather broad variety of presentation strategies, particularly regarding the order of shown items, could be observed.

A second study was designed as a follow-up study to investigate, in how far the proposed conceptual hierarchy is supported by interaction strategies of human users in the scenario of a "guided tour" with a robot. The findings of this study supported the hypothesis that there would be particular presentation strategies observable depending on the conceptual category of the item to be shown to the robot.

To investigate more general aspects of and potential differences and similarities between human-human and human-robot interaction in the scenario of a "guided tour" a comparison study was designed. The subjects of this study were asked to present the same environment both to a robot and to a person, to observe their strategies for showing particular items to the different interaction partners. Despite some issues regarding the general setup and realisation of this study some very interesting observations could be made that revealed subjects' tendency to

treat robot and person differently even while they explicitly tried to handle the two tours in the same way.

7.2 Concluding discussion

The general idea of Human Augmented Mapping offers many issues to discuss and investigate and thus, many lessons were learnt from the work conducted within this framework. First of all, the overall conclusion from the investigations done for this thesis can only be that there is no "final" concluding statement; the overall problem of Human Augmented Mapping is not yet considered solved. Nevertheless, a number of reflections can be made, starting with some general remarks on important decisions taken in the beginning of the doctoral thesis project.

Breadth vs. depth

The idea of Human Augmented Mapping involves two main aspects, namely robotic mapping and human-robot interaction, both limited to the context of an indoor environment in which a service robot has to adapt to its working area that is to be shared with (a) human user(s). Both these aspects themselves cover a lot of interesting questions, which could each be used as a starting point for a complete thesis project. Realising that building and investigating a full system for Human Augmented Mapping from scratch was an impossible project, the author was forced to decide, whether only one particular aspect, e.g., the mapping part, or a broader variety of aspects should be investigated.

The decision fell on the latter strategy, coming thus at the price of not being able to investigate any of the respective aspects as deeply and thoroughly as this could be the case with very focused research regarding one particular problem. However, the author would not have made this decision, if she had not seen the opportunity in the combination of different fields and aspects, trying to find connections between human mental models, the way these models are expressed and their possible implementation in a robotic system that facilitates communication. Overall one outcome is that the complete system will always be more than the sum of its parts, considering the rather simple approaches to environment segmentation and representation that were used as components in a rather complex communication framework, allowing meaningful interaction with a human user in the cooperative integration work on the robot "BIRON", discussed in chapter 5.

The real world and simulations

A decision had to be made regarding the possible simplification of the overall problem by using simulations or a robot simulator. Not at least due to the influence of the common statement often heard in the research group that "it is not working until I have seen it on the robot", the author considers simulators as a tool for a quick test in an implementation process, offering the opportunity to test algorithms

on the robot without risking any damage with the real one. However, as soon as real-world conditions and human users come into the picture, no simulator will be sufficient to observe only nearly as many interesting aspects as this can be done with a real robot and a person *in situ*, given a background scenario to investigate. The evaluation of the mapping subsystem, particularly the transition detection, in different "real-world" environments, and the user studies, particularly the multiple room study in subjects' homes, clearly supported this standpoint.

The mapping subsystem

Despite the rather broad approach to the problem of Human Augmented Mapping, the implementation of the overall system had a strong focus on the mapping subsystem, where in particular the *region* representation and its application for transition detection were evaluated in a set of experiments. Those experiments made use of a public data set and other data sets available from data recordings in office and domestic environments as well as they were run on-line within an interactive framework. Thus, the focus of the evaluation was much more to investigate the applicability of the proposed approaches in the assumed context of an interactive "guided tour" through an arbitrary but limited indoor environment rather than to compare the work to other mapping or space segmentation approaches.

The author does not claim to build a complete and consistent categorised model of an environment, as other related approaches do (e.g., Friedman *et al.*, 2007). Nevertheless, the proposed approach has to be considered a fast and easy to apply complement to such more complex ways of representing an environment. It proved fully applicable in the integrated interactive communication framework on the robot BIRON, where it supported a meaningful, yet limited, discourse about an indoor environment.

An aspect that would have exceeded the scope of the thesis and also that of the integration work is the investigation of ambiguities in the environment representation that are caused either by different users and their possibly different (functional) understanding of an environment or by the actual existence of more than one item of the same kind (e.g., two "bathrooms", or two "printer rooms"). With the current state of the integrated system the mapping subsystem is capable of generating the necessary requests for clarification and could also handle the anticipated input accordingly, but the currently implemented framework is not.

Another issue not yet incorporated in the actual handling of the topological graph structure is a correction of the underlying pose estimation in the SLAM module according to a correction of the hypothesised current whereabouts generated by the *region* classification and transition detection.

Even without those issues that remain ideas for future work the author concludes that the proposed representation for *regions* based on property features computed from laser range data sets is a concise and fast approach that can be applied standalone to generate a representation of the environment already meaningful to the user. Combined with other methods to the segmentation of space, e.g., a door

detector or an off-line categorisation of a complete indoor environment it would be a source of different information cues to improve a fully integrated interactive system.

Semantics and concepts

In the literature, the idea of Human Augmented Mapping is also discussed as "semantic mapping" (Zender *et al.*, 2007, e.g.,), since the general idea is to provide the assumed service robot with the same (semantic) understanding of the environment the human user has. For the work reported in this thesis the term "semantics" was avoided, since its use seemed somewhat controversial. Often, "semantics" as a term is related to the functional understanding and appearance of particular delimited areas in indoor environments – e.g., perceiving a refrigerator and a coffee–maker in a spatial context allows to conclude that the surroundings most probably would be called a "kitchen", or vice versa, being in the "kitchen" should start a query for respective items probably available. This level of semantics was considered to exceed the scope of the thesis, as well as assuming *a priori* knowledge on the most likely disposition of a particular indoor environment, allowing to reason that an apartment most likely would contain at least a "kitchen", a "bathroom" and one "room".

This type of semantic knowledge was considered very limiting given that arbitrary indoor environments were assumed, including, e.g., studios, where it might be difficult to find delimited "rooms" assigned to the type of semantics listed above. Thus, the author decided to assume only the *a priori* knowledge needed to derive the conceptual category of an item to be presented with respect to the proposed hierarchy and definitions of spatial concepts given in chapters 3 and 4.

The resulting accumulated knowledge about the environment can be described as "semantic" in a way similar to the understanding used for the Spatial Semantic Hierarchy (Kuipers, 2000): The semantics of each spatial concept used in the hierarchy proposed in this thesis relate to the options of "perception and use of space" (affordances, e.g., Norman (2002)) it holds for the user and the robot. The *region* describes the surroundings where it is possible to *be inside* and *move around in*, the *location* is a spatial entity to *work at* and the *object* can be manipulated and *worked with*.

Tracking and interaction

An obvious issue of the "guided tour" idea is that the robot in question needs to be able to track its user, both to interact in a meaningful way and to follow the user around through the environment. This thesis presented an implementation and evaluation of a tracking approach earlier proposed by Schulz *et al.* (2001), used in the context of the overall system for Human Augmented Mapping. Additionally to the experimental evaluation the tracking approach was used for all the 27 trial runs conducted in the three reported user studies. As discussed in the respective

context (chapters 5 and 6) a number of tracker failures occurred, but never to an extent that would have made interaction impossible. However, in all cases the experiment leaders were able to a) monitor the robot and the system's interpretation of the situation and b) control the robot remotely to resolve any inconvenient situation. Problems occurred most often when the tracking system erroneously assigned the "user" tag to a static, "person-like" object, resulting in a kind of deadlock situation with the robot focusing on the object and the user not knowing about this problem. Consequently, the author concludes that not only the absolute quality of the tracking approach, but even more its integration into an interaction monitoring framework is crucial, when a fully autonomous interactive system is to be designed. Again, this supports the idea of getting more than the summed quality of simple components by combining several of them.

User studies

Three user studies of rather different character were designed and conducted around the "home tour" scenario. Two of them were set in a laboratory/office environment while one, named the multiple room study, was conducted in subjects' homes. In all three studies, users were asked to present the given environment to the "service robot Minnie", more or less specifically instructed about the details of the task. In one of the studies in the office environment this tour was also given to a person to investigate similarities and differences in human-human and human-robot interaction in the limiting context of the "home tour" scenario.

In general, user studies in human-robot interaction can be conducted under different paradigms, e.g., either being a controlled experiment or a rather exploratory (field) study. In all three cases it was decided to work under the paradigm of exploration, but still trying to control as many factors of the respective setup as possible without losing the exploratory character. Consequently, a term that seems to be appropriate for this type of study design would be a "semi-controlled experiment", with a number of controlling elements for the task given to subjects or the environment, but leaving the subjects free to choose their own strategies and ways of solving the task.

The first study setup left the subjects in a known office environment completely free regarding the items they would present to the robot and the order they would do that in. Consequently, it was not possible to draw any conclusion about interaction patterns or particular strategies assigned to conceptual categories of items to be presented to the robot. This had to be investigated with the second study, in which subjects got a much more specific instruction on what to present to the robot – in this case in their own homes. The respective study delivered rather specific observations with respect to this particular issue, leading to the resulting statement that there are interaction patterns and strategies observable across subjects and often some correspondence to the proposed hierarchy for the environment model can be established within-subjects. This allows to draw the conclusion that the underlying robotic mapping process of an assumed HAM system might be in-

formed regarding the category of the item to be presented by interpreting the action sequence the user performs to present the item, in case that *a priori* knowledge is not available. On the other hand the results of the study are still of exploratory character so that the author would not go so far to declare *a priori* knowledge for items potentially occurring during a "home tour" scenario as obsolete.

Regarding the third user study observations concerning both the study's guiding question and the general study design and conduction could be made. It is possible to declare as a study result that study subjects tended to treat a robot differently from a person in the same context of a "guided tour", even if they *explicitly did not intend to do that*. However, since a number of issues regarding the instructions and some problems with the used language (English) seemed to influence subjects' attitude and assumptions significantly, it has to be concluded that further studies regarding the similarities and differences in human-human and human-robot interaction in particular contexts remain to be conducted, making use of the general observations from the study described in this thesis. A respective study setup needs obviously to consider more carefully the influence of language choice and use than this has been done with the presented study.

A priori knowledge vs. learning from scratch

An interesting issue the author encountered during the thesis project work was the question regarding the necessary amount of *a priori* knowledge a general purpose service robot would have to be provided with. With the second user study a correspondence between interaction patterns and conceptual categories of (spatial) items could be established to a certain extent, that might suggest to omit precoded knowledge about the conceptual categories. On the other hand, the results of the study can not be interpreted as entirely representative or generally applicable. There are also particular situations observable in the "home tour" scenario that suggest to rely on precoded knowledge together with the observation of a – seemingly inappropriate – interaction pattern to solve a particular case of ambiguity. This can be the presentation of a *region* from "outside", where the interaction patterns correspond rather to the presentation of a *location*, while precoded knowledge of the conceptual category of the given label might suggest a *region*. Thus, the mentioned question remains unanswered as to how much precoded knowledge exactly is needed, but it seems obvious that some of it is definitely crucial to have to avoid inconsistent interpretations of particular situations.

The main lesson learnt

An overall lesson learnt from the decisions made in the beginning and the proposed approach to Human Augment Mapping is thus, that at the price of giving up depth and at the risk of losing focus, a great opportunity to find interesting aspects in each involved research area was won. Many interesting observations leading to new questions for further investigations both regarding the system implementation and the

interaction aspects would not have been made without continuously switching from system development to user studies and back. The author considers it therefore absolutely necessary to see robotic research fields and human(-robot) interaction as strongly intertwined, producing the need of considering both in parallel when systems that are supposed to work in the proximity of humans are to be designed.

7.3 Future ideas

As mentioned in the previous section a number of particular issues remain to be investigated within the framework of Human Augmented Mapping.

An open issue is the handling of the presentation of *regions* from "outside", e.g., by pointing towards the door and labelling the room behind it. As a consequence, a respective system needs to create a link to a *region* of unknown spatial properties, which is also an issue to be more thoroughly investigated. This way of showing a *region* is in fact a strategy that could be observed in a number of cases, and it is obvious, that any interactive system that is not made aware of the possibility of a respective situation to happen is bound to fail due to an inconsistent representation of the surroundings being acquired. The present thesis discussed this issue as an interesting case and approached possible solution strategies by investigating potential indications in human presentation styles. Further, more controlled investigations of that matter seem in order, as well as an experimental implementation of a system that makes use of the recognition of particular action and gesture sequences to distinguish between a *region* presented being "inside" and one presented being "outside".

In the context of the integration with the interactive framework on the robot BIRON another type of ambiguity was named that cannot be solved with the current integrated implementation. In case that an already known *region* label is mentioned this is always handled as a correction or confirmation, not as a specification of a new *region* that happens to have the same label. This would be an issue to be solved with appropriate handling through a dialogue and interaction system so that potential ambiguities in the mapping subsystem can be handled as this was the case for transition detections.

On a more general level the overall framework and idea of Human Augmented Mapping is a topic that this thesis only started to investigate. With the exploratory user studies a number of hypotheses were formulated and supported, but a more thorough investigation, potentially with increasingly controlled setups, seems adequate to learn more about the actual situations a future service robot system would have to deal with.

Appendix A

Instructions for the Pilot Study

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

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

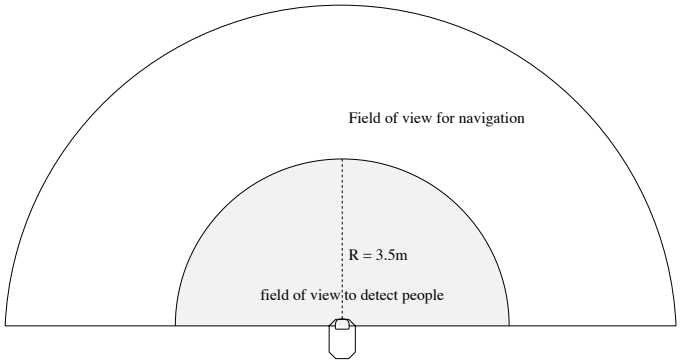
Elin A. Topp and Helge Hüttenrauch,
School of Computer Science and Communication (CSC),
Royal Institute of Technology (KTH), Stockholm, Sweden

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.

a)



b)

What is this all about?

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 <u>not</u> 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.

Appendix B

Interview questions for the Pilot Study

The following questions formed a loose guideline for the interviews in the pilot study.

HRI Pilot study: Interview questions

Elin A. Topp and Helge Hüttenrauch, Subject:

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?

Appendix C

Instructions for the MRS

The instructions for the subjects in the MRS were a somewhat shorter version of the instructions used in the pilot study, listing not so many particular issues regarding problematic situations. The subjects were informed that the experiment leaders would help them out whenever they would ask for it. Due to the similarity the full instruction sheet is not shown again, but the "new" particular instructions regarding what to present is illustrated in the following.

```
Your task

Please take the robot on a tour around your home and teach it

1) four different rooms,
                              the kitchen
                              the living room
                              the hallway/entrance/corridor.


   The last room to teach to the robot is "special" - as the
   robot could be damaged if used in rooms like the bathroom
   or staircases, please show the robot either
                              a bathroom  or  a staircase

   without entering or coming too close!
```

... continued on the next page

2) in the kitchen, the living room and the hallway the robot needs to learn where to perform tasks or where to find things, please show it

in the kitchen:	the kitchen sink
	the fridge (microwave or dishwasher)
	a table where you eat
in the living room:	where to get/place books (or magazines)
	your TV (or video or other elec. equipment)
	a chair (or similar) where you can sit
in the hallway:	where you place your shoes
	the entrance to your home
	a light switch (or cloth hanger)

3) the robot also needs to learn a few examples of possibly important items, please show and teach the robot the following:

kitchen:	a cup (or a glass)
	a container of milk from your fridge
	a salt shaker (or other spice container)
living room:	a remote control
	a book (newspaper or magazine)
	a video tape (or CD or similar)
hallway:	a pair of shoes
	your mobile phone
	your keys to your home

Bibliography

Althaus, P. and Christensen, H.I. *Automatic Map Acquisition for Navigation in Domestic Environments.* In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Taipei, Taiwan.* 2003

Althaus, P., Ishiguro, H., Kanda, T., Miyashita, T., and Christensen, H.I. *Navigation for human-robot interaction tasks.* In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), New Orleans, LA, USA.* 2004

Anderson, J.R. and Lebiere, C. *The Atomic Components of Thought.* Lawrence Erlbaum Associates, Mahwah, NJ, USA / London, United Kingdom. ISBN 0-8058-2817-6. 1998

Arkin, R.C. *Integrating behavioural, perceptual and world knowledge in reactive navigation.* Robotics and Autonomous Systems (RAS), 6:pp. 105–122, Elsevier. 1990

Arulampalam, S., Maskell, S., Gordon, N., and Clapp, T. *A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking.* IEEE Transactions on Signal Processing, 50(2):pp. 174–188. 2002

Bakker, B., Kröse, B., Christensen, H.I., Belpaeme, T., Paquier, W., and Chatila, R. *Cognitive Models of Space and Objects.* Project Deliverable for EU Integrated Project COGNIRON, FP6-IST-002020, University of Amsterdam, The Netherlands. www.cogniron.org. 2005

Beeson, P., Jong, N.K., and Kuipers, B. *Towards Autonomous Topological Place Detection Using the Extended Voronoi Graph.* In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Barcelona, Spain.* 2005

Bennewitz, M., Burgard, W., and Thrun, S. *Using EM to learn Motion Behaviors of Persons with Mobile Robots.* In *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS), Lausanne, Switzerland.* 2002

Blisard, S., Skubic, M., Luke, R.H. III, and Keller, J.M. *3-D Modeling of Spatial Referencing Language for Human-Robot Interaction.* In *Proceedings of the ACM*

Conference on Human-Robot Interaction (HRI), Salt Lake City, UT, USA. Mar. 2006

Blom, J. *Personalization: A Taxonomy*. In *CHI'00 extended abstracts on Human factors in computing systems*, SIGCHI. ACM. 2000

Booij, O., Terwijn, B., Zivkovic, Z., and Kröse, B. *Navigation using an Appearance Based Topological Map*. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Rome, Italy*. 2007

Booij, O., Zivkovic, Z., and Kröse, B. *Sparse Appearance Based Modeling for Robot Localization*. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Beijing, China*. 2006

Bosse, M., Newman, P., Leonard, J., and Teller, S. *SLAM in Large-scale Cyclic Environments using the* Atlas *Framework*. International Journal on Robotics Research, 23:pp. 1113–1139, Sage publications. 2004

Breazeal, C. *Sociable Machines: Expressive Social Exchange Between Humans and Robots*. Doctoral dissertation, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, USA. 2000

Breazeal, C. *Designing Sociable Robots*. MIT Press, Cambridge, MA 02142, USA. ISBN 9-780262-025102. 2002

Brennan, S.E. and Clark, H.H. *Conceptual Pacts and Lexical Choice in Conversation*. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(6):pp. 1482–1493, American Psychological Association. 1996

Brooks, A.G. and Breazeal, C. *Working with Robots and Objects: Revisiting Deictic Reference for Achieving Spatial Common Ground*. In *Proceedings of the ACM Conference on Human-Robot Interaction (HRI), Salt Lake City, UT, USA*. 2006

Brooks, R. *A robust layered control system for a mobile robot*. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), San Francisco, CA, USA*. 1986

Bruce, A. and Gordon, G. *Better Motion Prediction for People-Tracking*. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), New Orleans, LA, USA*. 2004

Buschka, P. and Saffiotti, A. *Some Notes on the Use of Hybrid Maps for Mobile Robots*. In *Proceedings of the International Conference on Intelligent Autonomous Systems (IAS), Amsterdam, The Netherlands*. IOS Press. 2004

Castellanos, J.A., Montiel, J.M.M, Neira, J., and Tardós, J.D. *The SPmap: A Probabilistic Framework for Simultaneous Localisation and Mapping*. IEEE Transactions on Robotics and Automation, 15(5):pp. 948–953. 1999

Chong, K.S. and Kleeman, L. *Large Scale Sonarray Mapping using Multiple Connected Local Maps*. In *Proceedings of the International Conference on Field and Service Robotics (FSR), Canberra, Australia*. Australian Robot Association. 1997

Choset, H. and Nagatani, K. *Topological Simultaneous Localization and Mapping (SLAM): Toward Exact Localization Without Explicit Localization*. IEEE Transactions on Robotics and Automation, 17(2):pp. 125–137. 2001

Clark, H.H. and Brennan, S. *Grounding in Communication*. In Resnick, L.B., Levine, R.M., and Teasley, S.D., eds., *Perspectives on socially shared cognition*, pp. 127–149. American Psychological Association, Washington, DC, USA. 1991

Denscombe, M. *The Good Research Guide – for small-scale social research projects (Swedish version: Forskningshandboken – för småskaliga forskningsprojekt inom samhällsvetenskaperna)*. Open University Press, Buckingham and Philadelphia (Swedish version: Studentlitteratur, Lund, 2000). ISBN (Swedish version) 91-44-01280-2. 1998

Diosi, A., Taylor, G., and Kleeman, L. *Interactive SLAM using Laser and Advanced Sonar*. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Barcelona, Spain*. 2005

Feyrer, S. and Zell, A. *Robust Real-Time Pursuit of Persons with a Mobile Robot Using Multisensor Fusion*. In *Proceedings of the International Conference on Intelligent Autonomous Systems (IAS), Venice, Italy*. IOS Press. 2000

Fletcher, L., Apostoloff, N., Petersson, L., and Zelinsky, A. *Vision in and out of Vehicles*. IEEE Intelligent Systems, 18(3):pp. 12–17. 2003

Folkesson, J. *Simultaneous Localization and Mapping with Robots*. Doctoral Thesis. ISBN 91–7178-145-5, TRITA-NA-0528, KTH School of Computer Science and Communication (CSC), Stockholm, Sweden. 2005

Foster, M.E., Bard, E.G., Guhe, M., Hill, R.L., Oberlander, J., and Knoll, A. *The Roles of Haptic–Ostensive Referring Expressions in Cooperative, Task-based Human-Robot Dialogue*. In *Proceedings of the ACM Conference on Human-Robot Interaction (HRI), Amsterdam, The Netherlands*. 2008

Friedman, S., Pasula, H., and Fox, D. *Voronoi Random Fields: Extracting the Topological Structure of Indoor Environments via Place Labeling*. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Hyderabad, India*. Springer. 2007

Galindo, C., Saffiotti, A., Coradeschi, S., Buschka, P., and Fernández-Madrigal, J.A. *Multi-Hierarchical Semantic Maps for Mobile Robotics*. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Edmonton, AB, Canada*. 2005

Gálvez López, D., Sjöö, K., Paul, C., and Jensfelt, P. *Hybrid Laser and Vision Based Object Search and Localisation.* In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Pasadena, CA, USA.* 2008

Gockley, R., Bruce, A., Forlizzi, J., Michalowski, M., Mundell, A., Rosentahl, S., Sellner, B., Simmons, R., Snipes, K., Schultz, A.C., and Wang, J. *Designing Robots for Long-Term Social Interaction.* In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Edmonton, AB, Canada.* 2005

Green, A. and Hüttenrauch, H. *Making a Case for Spatial Prompting in Human-Robot Communication.* In *Multimodal Corpora: From Multimodal Behaviour theories to usable models, workshop at the Fifth international conference on Language Resources and Evaluation (LREC2006), Genova, Italy.* 2006

Green, A., Hüttenrauch, H., and Severinson Eklundh, K. *Applying the Wizard-of-Oz Framework to Cooperative Service Discovery and Configuration.* In *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication (Ro-Man), Kurashiki, Okayama, Japan.* 2004

Green, A., Hüttenrauch, H., and Topp, E.A. *Measuring Up as an Intelligent Robot: On the Use of High-Fidelity Simulations for Human-Robot Interaction Research.* In *Workshop on Performance Metrics for Intelligent Systems (PerMIS), Washington, DC, USA.* 2006a

Green, A., Hüttenrauch, H., Topp, E.A., and Severinson Eklundh, K. *Developing a Contextualized Multimodal Corpus for Human-Robot Interaction.* In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC2006), Genova, Italy.* 2006b

Gustafsson, F. *Adaptive Filtering and Change Detection.* Jon Wiley & Sons, Ltd., Chichester, England. ISBN 0-471-49287-6. 2000

Haasch, A., Hohenner, S., Hüwel, S., Kleinehagenbrock, M., Lang, S., Toptsis, I., Fink, G.A., Fritsch, J., Wrede, B., and Sagerer, G. *BIRON – The Bielefeld Robot Companion.* In *Proceedings of the International Workshop on Advances in Service Robotics (IWASR), Stuttgart, Germany.* Fraunhofer, IRB. 2004

Holsanova, J. *Spatial language and dialogue: A multimodal perspective.* In *Workshop on Spatial Language and Dialogue.* Delmenhorst, Germany. http://www.lucs.lu.se/people/jana.holsanova/. 2005

Hüttenrauch, H., Green, A., and Severinson Eklundh, K. *Investigating Spatial Relationships in Human-Robot Interaction.* In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Beijing, China.* 2006a

Hüttenrauch, H. and Severinson Eklundh, K. *Fetch-and-carry with CERO: Observations from a long term user study with a service robot.* In *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication (Ro-Man), Berlin, Germany.* 2002

Hüttenrauch, H., Severinson Eklundh, K., Green, A., Topp, E.A., and Christensen, H.I. *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.* 2006b

Kirsh, D. *The intelligent use of space.* Artificial Intelligence, 73:pp. 31–68, Elsevier. 1995

Kleinehagenbrock, M., Lang, S., Fritsch, J., Lömker, F., Fink, G.A., and Sagerer, G. *Person Tracking with a Mobile Robot based on Multi-Modal Anchoring.* In *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication (Ro-Man), Berlin, Germany.* 2002

Kluge, B. *Tracking Multiple Moving Objects in Populated Environments.* In *Sensor based Intelligent Robots*, vol. 2238 of *Lecture Notes in Computer Science (LNCS).* Springer, Heidelberg, Germany. 2002

Knoop, S., Vacek, S., and Dillmann, R. *Sensor Fusion for 3D Human Body Tracking with an Articulated 3D Body Model.* In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Orlando, Florida.* 2006

Krieg-Brückner, B., Röfer, T., Carmesin, H.-O., and Müller, R. *A taxonomy of spatial knowledge for navigation and its application to the Bremen Autonomous Wheelchair.* In *Spatial Cognition I – An interdisciplinary approach to representing and processing spatial knowledge*, pp. 373–397. Springer, Berlin, Germany. 1998

Kröse, B. *An Efficient Representation of the Robot's Environment.* In *Proceedings of the International Conference on Intelligent Autonomous Systems (IAS), Venice, Italy.* IOS Press. 2000

Kruijff, G.-J.M., Zender, H., Jensfelt, P., and Christensen, H.I. *Clarification Dialogues in Human-Augmented Mapping.* In *Proceedings of the ACM Conference on Human-Robot Interaction (HRI), Salt Lake City, UT, USA.* 2006

Kuipers, B. *The "Map in the Head" Metaphor.* Environment and Behavior, 14(2):pp. 202–220, SAGE. 1982

Kuipers, B. *The Spatial Semantic Hierarchy.* Artificial Intelligence, 119:pp. 191–233, Elsevier. 2000

Kuipers, B. and Byun, Y.-T. *A robust qualitative method for spatial learning in unknown environments.* In *Proceedings of the National Conference on Artificial Intelligence (AAAI), Los Altos, CA, USA.* AAAI Press. 1988

Kuipers, B. and Byun, Y.-T. *A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations.* Technical report, University of Austin, Texas, USA. 1990

Kuipers, B., Modayil, J., Beeson, P., MacMahon, M., and Savelli, F. *Local Metrical and Global Topological Maps in the Hybrid Spatial Semantic Hierarchy.* In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), New Orleans, LA, USA.* 2004

Kuipers, B.J. and Byun, Y.-T. *A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations.* Robotics and Autonomous Systems, 8:pp. 47–63, Elsevier. 1991

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

Li, S. *Multi-model Interaction Management for a Robot Companion.* Doctoral thesis, Applied Computer Science, Faculty of Technology, Bielefeld University, Bielefeld, Germany. 2007

Lisien, B., Morales, D., Silver, D., Kantor, G., Rekleitis, I.M., and Choset, H. *The Hierarchical Atlas.* IEEE Transactions on Robotics, 21(3):pp. 473–481. 2005

Lowe, D. G. *Object recognition from local scale-invariant features.* In *Proceedings of the IEEE International Conference on Computer Vision (ICCV), Kerkyra, Greece.* 1999

Lu, F. and Milios, E. *Robot Pose Estimation in Unknown Environments by Matching 2D Range Scans.* In *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference (CVPR).* 1994

Mahani, Maryamossadat Nematollahi. *Towards Resolving Ambiguities in Service Robots' Behaviour.* Master's thesis, Department of Computer and Systems Sciences (Stockholms University and Royal Institute of Technology), KTH School of Information and Communication Technology, Royal Institute of Technology (KTH), Stockholm, Sweden. 2006

Martínez Mozos, Ó., Stachniss, C., and Burgard, W. *Supervised Learning of Places from Range Data using AdaBoost.* In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Barcelona, Spain.* 2005

Martínez Mozos, Ó, Triebel, R., Jensfelt, P., Rottmann, A., and Burgard, W. *Supervised Semantic Labeling of Places using Information Extracted from Sensor Data.* Robotics and Autonomous Systems, 55(5):pp. 391–402, Elsevier. 2007

McNamara, T.P. *Mental Representations of Spatial Relations.* Cognitive Psychology, 18:pp. 87–121, Elsevier. 1986

McNamara, T.P. and Shelton, A.L. *Cognitive maps and the hippocampus*. TRENDS in Cognitive Science, 7(8):pp. 333–335, Cell Press. 2003

Montemerlo, M., Pineau, J., Roy, N., Thrun, S., and Verma, V. *Experiences with a Mobile Robotic Guide for the Elderly*. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*. AAAI Press. 2002

Norman, D. *The Design of Everyday Things*. Basic Books (Perseus). Reprint of the original book "The Psychology of Everyday Things", 1988. 2002

Nourbakhsh, I., Powers, A., and Birchfield, S. *DERVISH An Office-Navigating Robot*. AI Magazine, 16(2):pp. 53–60, AAAI Press. 1995

Pacchierotti, E., Christensen, H. I., and Jensfelt, P. *Embodied social interaction for service robots in hallway environments*. In *Proceedings of International Conference on Field and Service Robotics (FSR), Brisbane, Australia*. Australian Robot Association. 2005

Platzek, S. *Statistical Analysis of Human Patterns in 2D Laser Range Data*. Project Report (Studienarbeit), Department of Numerical Analysis and Computer Science, Royal Institute of Technology, Stockholm and Institut für Technische Informatik, University of Karlsruhe, Germany. 2005

Powers, A. and Kiesler, S. *The Advisor Robot: Tracing People's Mental Model from a Robot's Physical Attributes*. In *Proceedings of the ACM Conference on Human-Robot Interaction (HRI), Salt Lake City, UT, USA*. 2006

Prassler, E., Bank, D., and Kluge, B. *Motion Coordination between a Human and a Mobile Robot*. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Lausanne, Switzerland*. 2002

Pronobis, A., Martínez Mozos, Ó, and Caputo, B. *SVM-based discriminative accumulation scheme for place recognition*. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Pasadena, CA, USA*. 2008

Reeves, B. and Nass, C. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. The University of Chicago Press, Chicago, IL 60637 USA. ISBN 9-781575-860534. 1996, 2003

Russel, S. and Norvig, P. *Artificial Intelligence – A Modern approach*. Prentice Hall Series in Artificial Intelligence, 2 edn. Pearson Education, Inc., Upper Saddle River, NJ, USA. ISBN 0-13-080302-2. 2003

Schulz, D., Burgard, W., Fox, D., and Cremers, A.B. *Tracking multiple moving targets with a mobile robot using particle filters and statistical data association*. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA), Seoul, Korea*. 2001

Seiz, M., Jensfelt, P., and Christensen, H.I. *Active exploration for feature based global localization.* In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Takamatsu, Japan.* 2000

Shannon, C.E. *A Mathematical Theory of Communication.* Bell System Technical Journal, 27:pp. 379–423. 1948

Siciliano, B. and Khatib, O., eds. *Springer Handbook of Robotics.* Springer Reference. Springer Verlag, Heidelberg, Germany. ISBN 978-3-540-30301-5. 2008

Sidner, C.L., Lee, C., Kidd, C.D., Lesh, N., and Rich, C. *Explorations in engagement for humans and robots.* Artificial Intelligence, 166:pp. 140–164, Elsevier. 2005

Spexard, T.P., Li, S., Wrede, B., Hanheide, M., Topp, E.A., and Hüttenrauch, H. *Interaction Awareness for Joint Environment Exploration.* In *Proceedings of the Special Session on Situation Awareness in Social Robots at the IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man), Jeju Island, Korea.* 2007

Spexard, T.P., Siepmann, F.H.K, and Sagerer, G. *A Memory-based Software Integration for Development in Autonomous robotics.* In *Proceedings of the International Conference on Intelligent Autonomous Systems (IAS), Baden-Baden, Germany.* IOS Press. 2008

Tapus, A., Ramel, G., Dobler, L., and Siegwart, R. *Topology Learning and Place Recognition using Bayesian Programming for Mobile Robot Navigation.* In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Sendai, Japan.* 2004a

Tapus, A., Tomatis, N., and Siegwart, R. *Topological Global Localization and Mapping with Fingerprints and Uncertainty.* In *Proceedings of the International Symposium on Experimental Robotics (ISER), Singapore*, STAR. Springer. 2004b

Tellex, S. and Roy, D. *Spatial Routines for a Simulated Speech-Controlled Vehicle.* In *Proceedings of the ACM Conference on Human-Robot Interaction (HRI), Salt Lake City, UT, USA.* 2006

Thomaz, A. L., Hoffman, G., and Breazeal, C. *Experiments in Socially Guided Machine Learning: Understanding Human Intent of Reward/Punishment.* In *Proceedings of the ACM Conference on Human-Robot Interaction (HRI), Salt Lake City, UT, USA.* 2006

Thrun, S. *Robotic Mapping: A Survey.* In *Exploring Artificial Intelligence in the New Millenium.* Morgan Kaufmann. 2002

Thrun, S. and Bücken, A. *Integrating Grid-Based and Topological Maps for Mobile Robot Navigation*. In *Proceedings of the National Conference on Artificial Intelligence (AAAI), Portland, OR, USA*, vol. 2. Portland, OR, USA. 1996

Thrun, S., Burgard, W., and Fox, D. *Probabilistic Robotics*. MIT Press, Cambridge, MA 02142, USA. ISBN 978-0-262-20162-9. 2005

Tolman, E.C. *Cognitive Maps in Rats and Men*. Psychological Review, 55:pp. 189–208, American Psychological Association. 1948

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

Topp, E.A. *An Interactive Interface for Service Robots – Design and Experimental Implementation*. Diplomarbeit / Master's Thesis, Faculty of Computer Science, University of Karlsruhe, Karlsruhe, Germany. 2003

Topp, E.A. *Evaluation of a multiple target tracking approach for following and passing persons*. Technical Report TRITA-NA-P0502, ISSN 1101-2250, CVAP-296, School of Computer Science and Communication, Royal Institute of Technology (KTH), Stockholm, Sweden. 2005

Topp, E.A. *Initial Steps Toward Human Augmented Mapping*. Licentiate Thesis. ISBN 91-7178-452-7, TRITA-CSC-A 2006:13, CVAP-303, KTH School of Computer Science and Communication (CSC), Stockholm, Sweden. 2006

Topp, E.A. and Christensen, H.I. *Tracking for Following and Passing Persons*. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS), Edmonton, AB, Canada*. 2005

Topp, E.A. and Christensen, H.I. *Topological Modelling for Human Augmented Mapping*. In *Proceedings of the IEEE/RSJ International Conference of Intelligent Robots and Systems (IROS), Beijing, China*. 2006

Topp, E.A. and Christensen, H.I. *Detecting Structural Ambiguities and Transitions during a Guided Tour*. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Pasadena, CA, USA*. 2008

Topp, E.A. and Hüttenrauch, H. *Human Augmented Mapping – A Pilot Study*. Technical Report TRITA-CSC-CV 2006:1, ISSN 1653-6622, CVAP-301, School of Computer Science and Communication (CSC), Royal Institute of Technology (KTH), Stockholm, Sweden. 2006

Topp, E.A., Hüttenrauch, H., Christensen, H.I., and Severinson Eklundh, K. *Acquiring a Shared Environment Representation*. In *Proceedings of the ACM Conference on Human-Robot Interaction (HRI), Salt Lake City, UT, USA*. 2006a

Topp, E.A., Hüttenrauch, H., Christensen, H.I., and 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.* 2006b

Topp, E.A., Kragic, D., Jensfelt, P., and Christensen, H.I. *An Interactive Interface for a Service Robot.* In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), New Orleans, LA, USA.* 2004

Trafton, J.G., Schultz, A.C., Perzanowski, D., Bugajska, M.D., Adams, W., Cassimatis, N.L., and Brock, D.P. *Children and Robots Learning to Play Hide and Seek.* In *Proceedings of the ACM Conference on Human-Robot Interaction (HRI), Salt Lake City, UT, USA.* 2006

Vasudevan, S., Gächter, S., Nguyen, V., and Siegwart, R. *Cognitive Maps for Mobile Robots – An Object Based Approach.* Robotics and Autonomous Systems, 55(5):pp. 359–371, Elsevier. 2007

Wada, K., Shibata, T., Musha, T., and Kimura, S. *Effects of Robot Therapy for Demented Patients Evaluated by EEG.* In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Edmonton, AB, Canada.* 2005

Wang, E., Lignos, C., Vatsal, A., and Scasselati, B. *Effects of Head Movement on Perceptions of Humanoid Robot Behavior.* In *Proceedings of the ACM Conference on Human-Robot Interaction (HRI), Salt Lake City, UT, USA.* 2006

Wang, R.F. and Spelke, E.S. *Human spatial representation: insights from animals.* TRENDS in Cognitive Science, 6(9):pp. 376–382, Cell Press. 2002

Werner, S., Krieg-Brückner, B., and Herrmann, T. *Modelling Navigational Knowledge by Route Graphs.* In Freksa, C., Brauer, W., Habel, C., and Wender, K., eds., *Spatial Cognition II - Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications*, pp. 295–316. Springer, Berlin, Germany. 2000

Wrede, S., Fritsch, J., Bauckhage, C., and Sagerer, G. *An XML Based Framework for Cognitive Vision Architectures.* In *Proceedings of the IEEE International Conference on Pattern Recognition (CVPR).* 2004

Zender, H., Jensfelt, P., Mozos, Ó.M., Kruijff, G.-J.M., and 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

Zivkovic, Z., Booij, O., Kröse, B., Topp, E.A., and Christensen, H.I. *From Sensors to Human Spatial Concepts: An Annotated Data Set.* IEEE Transactions on Robotics, 24(2):pp. 501–505. 2008