



Initial Steps Torward Human Augmented Mapping

ELIN ANNA TOPP

Licentiate Thesis
Stockholm, Sweden 2006

TRITA-CSC-A 2006:13

ISSN-1653-5723

ISRN-KTH/CSC/A--06/13--SE

ISBN 91-7178-452-7

KTH

School of Computer Science and Communication

SE-100 44 Stockholm

SWEDEN

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framläggas till offentlig granskning för avläggande av teknologie licentiatexamen i datalogi fredagen den 6 oktober 2006 klockan 13.00 i sal D35, Kungl Tekniska Högskolan, Lindstedtsvägen 17, Stockholm.

© Elin Anna Topp, september 2006

Tryck: Universitetservice US AB

Abstract

With the progress in research and product development humans and robots get more and more close to each other and the idea of a personalised general service robot is not too far fetched. Crucial for such a service robot is the ability to navigate in its working environment. The environment has to be assumed an arbitrary domestic or office-like environment that has to be 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 could 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 the robot acquires. Advantages become also obvious for the mapping process itself, since in an interactive setting the robot could 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.

Based on results from robotics research, psychology, human-robot interaction and cognitive science a general architecture for a system for Human Augmented Mapping is presented. This architecture combines a hierarchically organised robotic mapping approach with interaction abilities with the help of a high-level environment model. An initial system design and implementation that combines a tracking and following approach with a mapping system is described. Observations from a pilot study in which this initial system was used successfully are reported and support the assumptions about the usefulness of the environment model that is used as the link between robotic and human representation.

Acknowledgements

There are many people who had their part in getting together this thesis. I want to thank my advisers Henrik Christensen and Kerstin Severinson Eklundh who facilitated the work with their ideas and questions.

And there are many more: Helge for inspiring discussions and his willingness to cooperate. Anders for many more inspiring discussions.

Patric for teaching me a lot about the internals of robots and always having some piece of code somewhere that could be the answer to my question. Mattias who had to live with a lot of rather odd questions about wiring and power consumption of robots. John for writing code with which this work was made possible. Danica for spreading a lot of energy and sharing shopping tips as willingly as work-related advice.

Jeanna and Mariann for dealing with loads of travel budgets, reimbursements and strange payment policies.

Eric, Josephine, Gareth, Stefan, Hugo, Oskar and all the other “lunching people” for a lot of interesting discussions on football and other stuff. All the CVAP-members for a more or less technical chat here and there in the kitchen: Jan-Olof, Mårten, Babak, Staffan, Daniel, Calle, Klas, Frank, Simone, Elena, Christian and many more ...

But still there is one left: Ludwig Seitz, who had to put up with all my moods and who got many drafts to read and a lot of seemingly strange questions at the breakfast table. Thanks!

The work described in this thesis was funded through the EU Integrated Project COGNIRON (‘The Cognitive Robot Companion’ - www.cogniron.org) by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020. The funding is gratefully acknowledged.

Contents

Contents	VII
List of Figures	IX
1 Introduction	1
1.1 Motivation	1
1.2 Initial steps of HAM	6
1.3 Organisation of the thesis	6
1.4 Contributions	8
2 Maps and Interaction	11
2.1 Cognitive models and space representations	12
2.2 Hybrid mapping	14
2.3 Topological mapping and space segmentation	17
2.4 Language and communication	19
2.5 Human-Robot Interaction	20
2.6 Summary	24
3 Human Augmented Mapping	25
3.1 Context, advantages, and limitations	25
3.2 Spatial concepts, situations, and requirements	27
3.3 Overall architecture	39
3.4 Main aspects for the thesis	40
3.5 Summary	40
4 Technical setup - initial system	41
4.1 The system	41
4.2 Tracking for following	49
4.3 Summary	59
5 Hierarchical environment representation	61
5.1 A partially hierarchical graph model	62
5.2 System and model in use - a pilot study	64

5.3	Observations and results from the study	68
5.4	Summary	73
6	Summary and discussion	75
6.1	Open issues and problems	76
6.2	Conclusion	78
7	Future ideas and plans	79
7.1	User studies	79
7.2	Topological modelling	80
7.3	Disambiguation in ambiguous environments	82
A	The instruction sheet	83
B	Interview questions	87
	Bibliography	89

List of Figures

1.1	A robotic (line based) map	2
1.2	The robotic map filled with human concepts	3
3.1	Hierarchy of situations and tasks	29
3.2	Interaction flow: user driven	32
3.3	Interaction flow: updates	33
3.4	Interaction flow: robot initiative	34
3.5	Interaction flow: continuous localisation	36
3.6	Interaction flow: continuous localisation	37
3.7	Interaction flow: continuous localisation	37
3.8	The overall concept for HAM	39
4.1	The PeopleBot Minnie	42
4.2	Overview of the implemented system	48
4.3	Trajectories of two test persons	53
4.4	Trajectories of a moving robot and three persons	54
4.5	Environment for test setup #3	56
4.6	Confusion of the tracker	57
4.7	Robot and user trajectory in the corridor	58
5.1	Spatial relations between objects	62
5.2	A partially hierarchical graph model	63
5.3	Environment for the study setup	64

Chapter 1

Introduction

The present licentiate thesis consolidates results of two years work within the integrated EU-project “Cogniron - The cognitive robot companion” and serves also as a proposal for further investigations.

A conceptual design for an approach to hierarchical, interactively controlled robotic mapping and localisation will be discussed together with results that could be achieved with an initial system implementation. Those results include also the observations made in a pilot study conducted with five subjects to test the assumptions underlying the models used for the work.

“Human Augmented Mapping” is a central term which is used to subsume aspects of robotic mapping and human robot interaction. In the following the general concept of “Human Augmented Mapping” (HAM) is introduced and motivated to allow the comprehension of the assumed scenarios, their constraints and limitations as well as the advantages of an interactive approach to robotic mapping.

1.1 Motivation

Since the work underlying this thesis is closely related to the ideas and plans of the project “Cogniron” mostly scenarios involving a companion/service robot can be assumed as a starting point. The following scenario should illustrate a 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 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. In the descriptions they get from the shop they also say 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

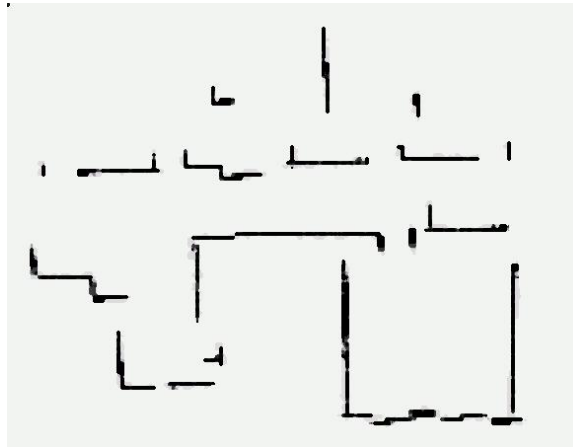


Figure 1.1: How your “ButlerBot” sees the environment with the help of its laser range finder

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 –

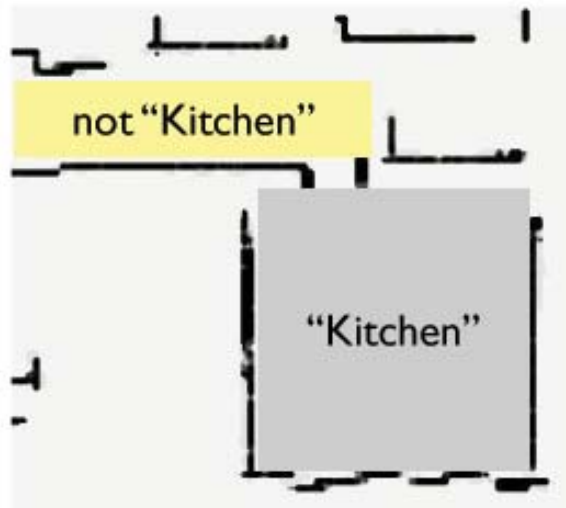


Figure 1.2: How your “ButlerBot” can see the environment if you help it to find its way

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? Or what happens if they have their house redecorated? What if one of the granddaughters really 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. In the following the background concepts for 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) (Folkesson, 2005; Thrun, 2002), 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 ex-

traction 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. 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. Already 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. Still, 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 have to be assumed 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 – well 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” 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 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, but 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.

Human Augmented Mapping

The term “Human Augmented Mapping” (HAM) describes the conceptual framework suggested 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 have never interacted with a robot before.

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 more or less two ways. One is to ignore the overall context and concentrate only on one single area, e.g., the interaction and dialogue components for an 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 seemingly different fields of robotics research. Still, all those areas have in common that they contribute to the overall, integrating concept. Even more, it becomes obvious that also in this field of robotics research the integrated system is more powerful than the sum of its components, regardless if a full implemented system or the design is considered.

1.2 Initial steps of HAM

The previous sections describe a motivating scenario and the background for a doctoral thesis project on “Human Augmented Mapping”. The present licentiate thesis deals with the initial steps toward a system for Human Augmented Mapping. These steps are the design of a conceptual framework for an overall system, the implementation of a subsystem that facilitates joint exploration of an environment for robot and user and the conduction of a pilot study to find out about how users would actually present an environment to a robot. It was assumed that observing users guiding around a robot in a more or less domestic or office-like environment is a good way of gathering more information on how a joint representation of this environment needs to be built, and what the open issues for the mapping process are. Thus, the initial system that was implemented in the framework of HAM concentrated on tracking and following persons, which is a crucial ability for the robotic system. Using this implemented subsystem the conducted pilot study investigated in fact a model that can link robotic and human environment representations. The thesis aims to give an overview about the different areas that are later on considered for the integration into the framework and hence describes one possible solution to the problem of integrating a human user into the process of robotic mapping.

1.3 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 over a first implementation to a small user study. All these steps build the base for further investigations in the context of HAM.

Chapter 2 - Background

Chapter 2 gives an overview over existing work in the research areas contributing to the concept of HAM. A large part of the technological background has its roots in the field of robotic mapping, which involves both geometric, topological, and hybrid approaches. The design of the environment representations used has its background in psychology and cognitive science, and also a number of references on 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 - Conceptual design

In chapter 3 the overall conceptual design for HAM is presented. 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 - Initial implementation

A design and an implementation for an initial HAM-system are discussed. The system is the result of the first phase of implementation work for the doctoral thesis project, which runs along the lines of the EU project COGNIRON. One important component of the system is the tracking and following mechanism that represents a central part of the interaction subsystem.

Chapter 5 - Environment representation

Chapter 5 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. To motivate a partially hierarchical graph model a small user study was conducted.

Chapter 6 - Summary and Discussion

In chapter 6 the results that could be achieved with the presented initial system and the study are summarised and discussed with respect to open issues.

Chapter 7 - Future work

Chapter 7 contains a discussion of the open issues and a strategy for how they could be addressed in future work. Those issues comprise the improvement of

components, the further development of spatial (topological) node representation and the development of disambiguating cues.

1.4 Contributions

The main contributions of this thesis can be summarised as

- A conceptual design for a human augmented approach to robotic mapping that considers aspects from autonomous robotic mapping and human-robot interaction;
- Results from an initial system implementation concentrating on the subsystem for tracking and following a user; and
- A model for a flexible high level environment representation suitable to incorporate individual information and motivated with results from a pilot user study.

The concept and results have been presented at international conferences and have been published accordingly in the following articles:

- Elin A. Topp and Henrik I. Christensen, “Tracking for Following and Passing Persons”, in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Edmonton, AB, Canada, August 2005
- Elin A. Topp, Helge Hüttenrauch, Henrik I. Christensen, and Kerstin Severinson Eklundh, “Acquiring a Shared Environment Representation”, in *Proc. of the ACM Conference on Human-Robot Interaction (HRI)*, Salt Lake City, UT, USA, March 2006

Other publications that are related, but did not directly contribute to the thesis:

- 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 the IEEE Workshop on Robots and Human Interactive Communications*, Hatfield, UK, September 2006, *to appear*

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, Royal Institute of Technology, Department of Numerical Analysis and Computing Science, TRITA-NA-P0502, ISSN 1101-2250, ISRN KTH/NA/P-05/02-SE, CVAP-296, March 2005
- Elin A. Topp and Helge Hüttenrauch, “Human Augmented Mapping - a pilot study”, Technical Report, Royal Institute of Technology, School of Computer Science and Communication, TRITA-CSC-CV 2006:1, ISSN 1653-6622, CVAP301, March 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

Chapter 2

Maps and Interaction

The idea of HAM is influenced by a number of different research areas which will be discussed in this chapter.

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 according to what can be observed and assumed in human behaviour for autonomous robotic mapping. The concept for HAM also relies on findings from psychological studies in terms of the models used for the communication (link) to the user.

Many approaches in robotic mapping deal with the problem of handling large maps. A solution that reduces the complexity of pure geometric maps is to build hierarchical representations that use local, geometric, often feature based maps that are then representing the nodes of a global topological map. Those approaches relate closely to the cognitive models mentioned above. Since such a map hierarchy consequently is corresponding quite well to the conceptual hierarchies that are used by humans the relation to HAM is obvious. Thus, a number of relevant contributions to hybrid solutions of the SLAM problem are presented.

A fundamental question is how to build a topological layer that can establish the link between geometry and higher level conceptual (“human”) representations. A number of approaches to pure topological mapping and space segmentation are explained to motivate the decisions for the system design presented later.

The idea to obtain a shared model relates closely to the problem of establishing common ground in communication. A lot of work has been done in the resolution of deictic (spatial) references and the use of spatial language. Since HAM can be seen as an approach to obtain common ground with a shared environment model, the principles and ideas are sketched with respective references.

It is possible to interpret the whole idea of HAM as well in the context of human-robot interaction as in robotic mapping. Human-robot interaction is a very broad field of research, ranging from cognitive models for the investigation of human learning abilities to social interaction with entertainment robots. Since HAM relates

to in fact both, a number of relevant publications in these areas are discussed.

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 (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).

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. He 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*. An strongly hierarchical model would suggest that 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) suggests that knowledge is encoded in two ways, as *production rules* and *chunks* and models it in ACT-R accordingly. The two types of representation differentiate in fact between declarative (or explicitly encoded) and procedural knowledge. A chunk describes thus 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 a number of hypotheses for the environment model proposed in this thesis.

Exploration and the use of maps

Kuipers captures in his “Spatial Semantic Hierarchy” (SSH) (Kuipers, 2000) 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) 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 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 Milius, 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 builds topological maps on top of in this case grid based metric maps was presented by Thrun and Bücken (1996). The authors 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 graph¹ 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.

Also 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 information on neither 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, 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. With the hybrid approach these control laws for localisation are replaced by a metric localisation in *Local Perceptual Maps*. The local maps thus do in this case not represent an absolutely specified and delimited region with a spatial extent, they are limited by the perceptual

¹The Voronoi graph of the free space is the set of all points that are equidistant to two or more obstacles (Russel and Norvig, 2003)

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).

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. Also 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 modelling of space in humans.

Tomatis *et al.* propose a framework, where features describing a certain area are explicitly grouped and linked in a topological graph (Tomatis *et al.*, 2003). They focus on a 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 the SSH are in fact exceeding the requirements for the presented concepts in terms of 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”.

HAM allows for a more open setting as a starting point, being independent from precoded knowledge 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 concentrate often 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.

Topological maps

Nourbakhsh *et al.* proposed with DERVISH (Nourbakhsh *et al.*, 1995) 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.* extend the idea of using the Voronoi graph for the autonomous learning of places (Beeson *et al.*, 2005). Still no relation to human verbalised concepts is given.

Althaus and Christensen suggested an approach to interactive mapping in which a user could take a robot on a tour and present an indoor environment (Althaus and Christensen, 2003). 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. Additionally, the system did not try to model the observable area according to human concepts other than the gateways.

Segmenting space

Another 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 offline 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 algorithm². 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.

With a biologically inspired approach to topological mapping Tapus *et al.* (2004) 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.

Two suggestions to actually capture the spatial layout of a given environment were made by Kröse (2000) and later by Mozos *et al.* (2005). Kröse 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

²A watershed approach delimits convex areas from each other with borders that are similar to the critical lines. In an indoor environment most of the borders are in fact doorways, but also parts of L-shaped rooms can be separated

such a method only holds for convex areas. Mozos uses 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 scans taken by the robot are continuously classified with the respective type-label. His method then uses AdaBoost for learning and classification of new examples.

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

2.4 Language and communication

Another aspect to the modelling of space and spatial relations is the use of (spatial) language. The presented work aspires to establish common ground (Clark and Brennan, 1991), using an environment model, which is a well investigated area in communication. The term *spatial language* often refers to deictic references (“... over there ...”, “This ...”), that might be accompanied by a pointing gesture, but also relations (“left”, “right of”) and concepts (“room”, “building”) have to be considered. Common ground for communication of spatial issues can be achieved, for instance, with the use of graphical maps or drawings (Holsanova, 2005). Nevertheless, such communicative aids 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.* in a study in which subjects were told to send a toy robot around on a table top “map” that represented an urban scene (Kyriakou *et al.*, 2005). 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 – 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 on the dialogue and communication. 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.* delivers interesting aspects to look into, not only in terms of communication and clarification but also for methods to generate space segmentation hypotheses other than a gateway detection.

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 human augmented mapping.

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. Robotic systems are developed and used in studies to learn about the effects of emotions, personality traits, or behaviour changes (Gockley *et al.*, 2005, as a recent example). Fundamental studies on social aspects of human-robot interaction have been presented throughout the years by, for example, Breazeal *et al.* (Breazeal, 2000; 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. 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, 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.* presented an approach to navigation in an interactive context, where a robot enters a group of people and adapts its dynamic behaviour to the configuration changes of this group (Althaus *et al.*, 2004). Starting from such observations and systems, the coordination of user and robot in an interactive scenario is an issue currently investigated in a number of user studies (Green *et al.*, 2004; Hüttenrauch *et al.*, 2006b).

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 conclude 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 in 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 publication 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 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. 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.

As mentioned above embodied interaction 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 example). Also for activity recognition the tracking of a user is necessary, 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. Opposed to the mostly computer vision based approaches mentioned above, a pure position tracker can be realised 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 (Prassler *et al.*, 2002, for example).

The work presented in this thesis relates strongly to the idea of a “home tour” scenario which requires such an ability. Since for initial steps only the rather coarse coordination of the robot’s movement with the user has to be assumed, 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 one 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.

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 on the other hand the target 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 Kleinhagenbrock *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.* suggest to use Expectation Maximisation to predict directions that possible targets might choose given the spatial context they are in (Bennewitz *et al.*, 2002). 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 details of the tracking approach will thus be described in chapter 4 together with the system design for the initial approach to Human Augmented Mapping.

2.6 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 framework.

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 tags¹, 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 has to be seen as 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 base 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 closing² 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” scenario³ the uncertainty in the system can be significantly reduced with the help

¹Radio Frequency Identification (RFID) is an automatic identification method, relying on storing and remotely retrieving data using devices called RFID tags or transponders. An RFID tag is a small object that can be attached to or incorporated into a product, animal, or person. RFID tags contain silicon chips and antennas to enable them to receive and respond to radio-frequency queries from an RFID transceiver. Passive tags require no internal power source, whereas active tags require a power source. (From Wikipedia, <http://en.wikipedia.org/wiki/Rfid>)

²Loop 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.

³With 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

of a clarification dialogue with a human user. Hypotheses generated on the current location and position can be confirmed or rejected interactively. Thus, the 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 can be 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 list explains a number of spatial concepts or entities and their relations to each other as well as to HAM and the present thesis.

Object Small objects that can be manipulated (cup, plate, remote control). (not incorporated in the framework so far)

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"⁴ (one of the central concepts for the presented approach to HAM)

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

map.

⁴The 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)

hierarchy and deliver a local description. Here, the term is used in a similar way and represents the second central concept for the space representation in HAM)

Floor A collection of regions, distinct by the level (in height). Requires artificial separations of the representation of each floor to deal with similar structures if no altitude information is available. (not incorporated in the framework so far)

The common concept of “room” is explicitly not used in this hierarchy 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 somehow separated 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.

With the main concepts of *regions* and *locations* a hierarchical space representation can be established, as will be explained further in chapter 5. 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.

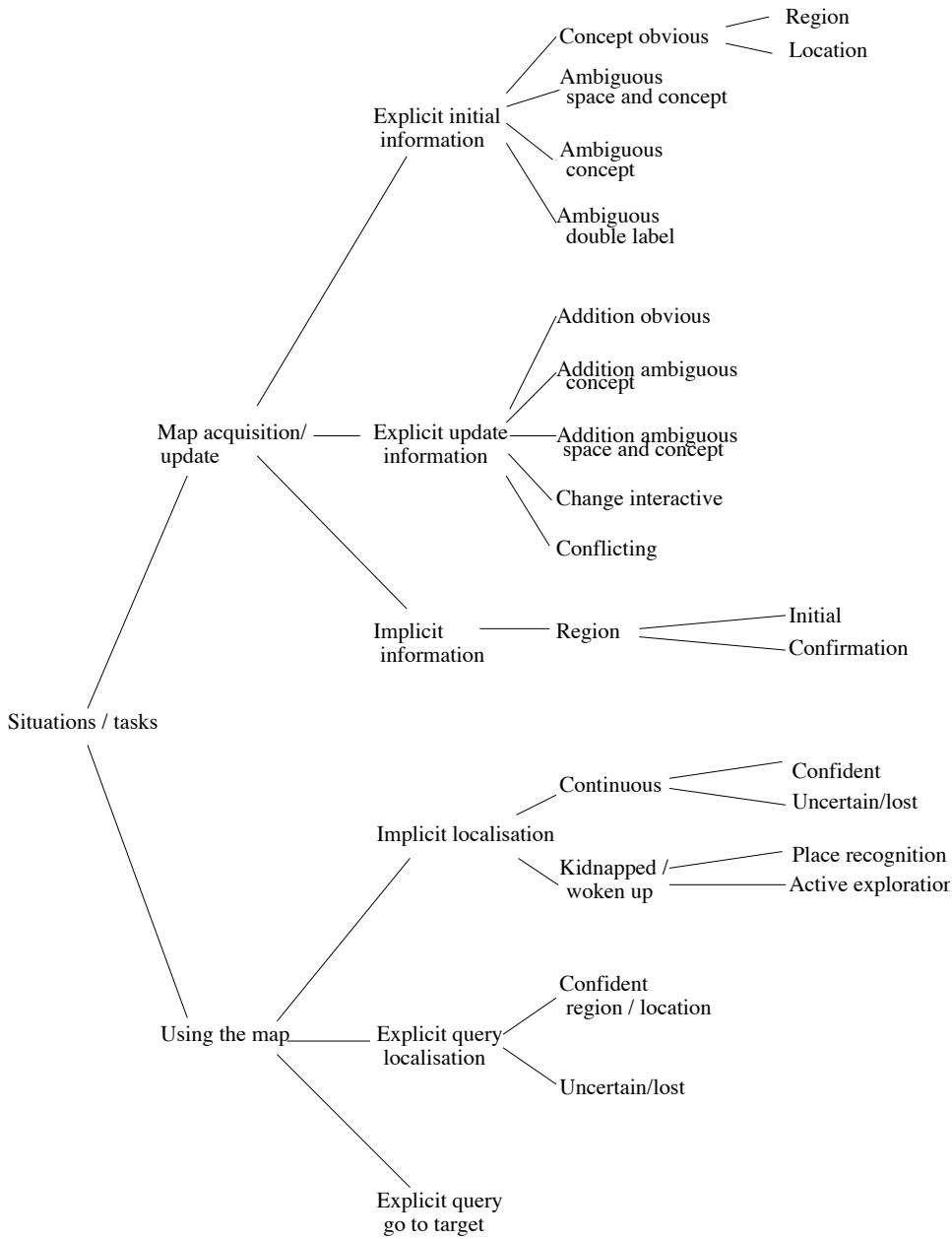


Figure 3.1: The relevant interaction situations and tasks of a guided tour scenario

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	loc confidence	(double measurement,
confidence		classification
		confidence)
Closest location:		
Label:	name	(string)
Description:	location descriptor	(position relative to
		region, pose)
Localisation	metric confidence	(double measurement,
confidence		metric localisation
		confidence)
Overall	f(loc 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 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 – 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 – 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 discussed in more detail in chapters 6 and 7. Here it can only be mentioned as a challenging situation for which it has to be investigated, if it can be identified with the help of the available sensory input and interpretation tools.

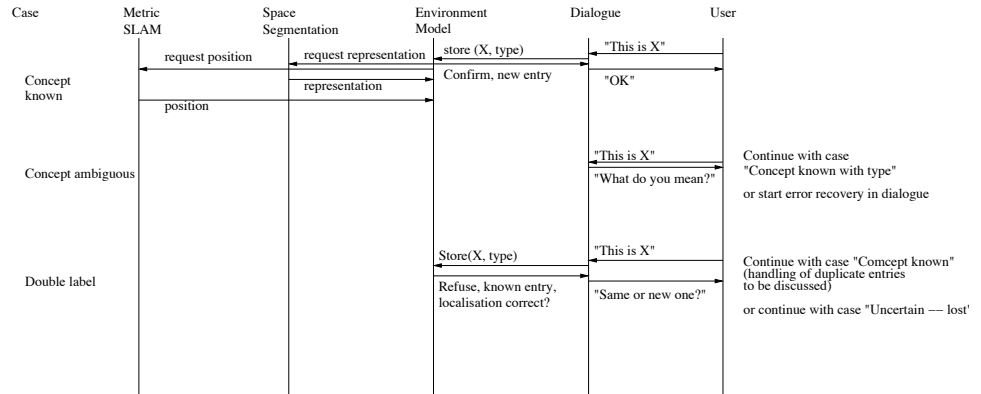


Figure 3.2: The interaction flow for the initial user utterance for user driven annotation. Depending on user responses (in case those are necessary), the interaction flow corresponds to other, initial situations

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 situation “Ambiguous – space and concept” can be detected, it remains as a challenging situation to be resolved.

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.

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

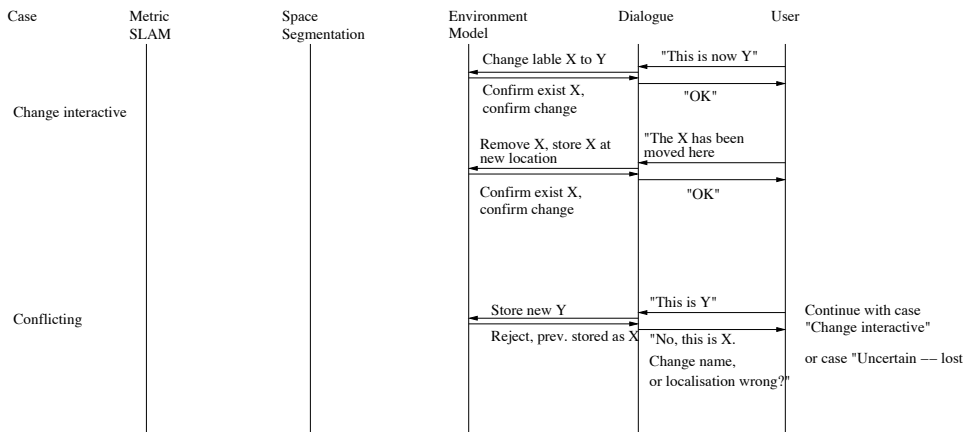


Figure 3.3: *The interaction flow for updates – those cases that correspond to the initial cases are omitted here*

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.

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.

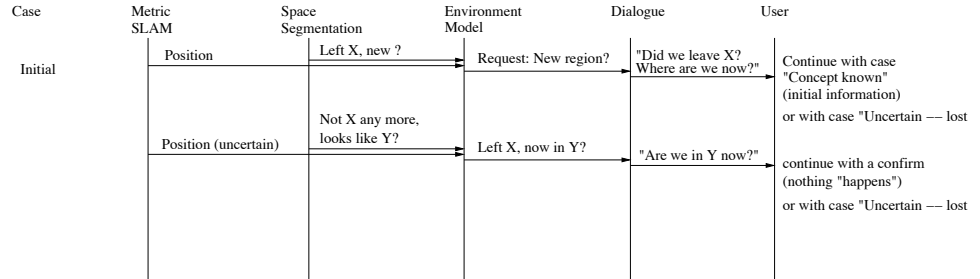


Figure 3.4: *The interaction flow for cases in which the system detects something and needs to get more information about it*

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. Thus, the system assumes to have left a certain *region* and entered a new one when a gateway was passed (Kruijff *et al.* (2006)). Another option is to detect significant changes with the help of a space segmentation component based on geometric features and generate a hypothesis on “being still in the same *region*” vs. “having left the *region*”.

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

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/topological representation 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 system 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 confidence	LOW	(double measurement, classification confidence)
OR		
Description:	location descriptor	(position relative to region, pose)
Localisation confidence	LOW	(double measurement, metric localisation confidence)
AND		
Overall confidence	LOW	(summarised 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

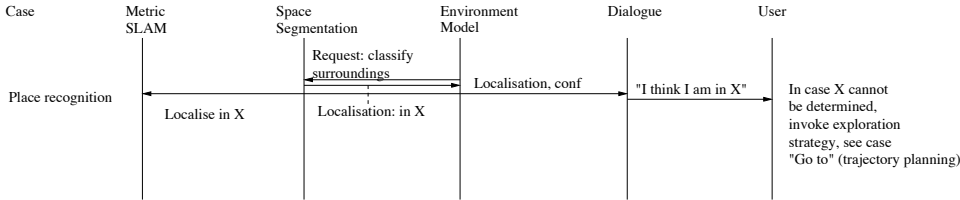


Figure 3.6: The interaction flow for explicitly invoked localisation, after waking up the system or kidnapping it

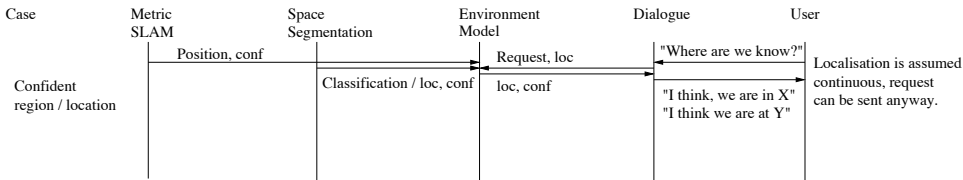


Figure 3.7: The interaction flow for explicit user queries regarding the current whereabouts

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, waypoint 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 plan-

ning 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 necessary⁵.
- 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.
- Topological representation. In order to relate spatial entities to each other a topological representation (a graph) is needed.
- 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 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.
- 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.
- 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.

⁵Approaches 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.

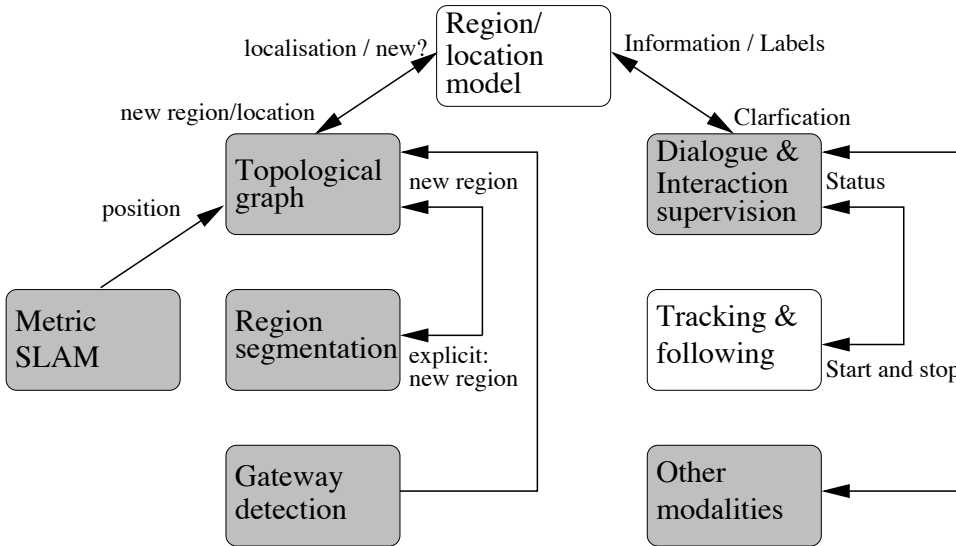


Figure 3.8: A concept for integrating all the necessary parts for Human Augmented Mapping. The white modules are discussed in detail in the following chapters, grey ones represent parts of the concept that contribute to the discussion but are not of main interest

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 Overall architecture

The components for HAM can be grouped functionally into two bundles, one representing the models 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. 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 are considered possible, one of which is a

gateway detection (Kruijff *et al.*, 2006). Another option, discussed in more detail in chapter 7 is a space segmentation based on geometric properties (Mozos *et al.*, 2005) or appearance (Tapus *et al.*, 2004) of the environment.

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 together with the pilot study presented in chapter 5.

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. An approach to a tracking and following system for HAM is presented together with an implementation related design of an initial system in chapter 4 of this thesis.

3.4 Main aspects for the thesis

The design and implementation of a complete system for Human Augmented Mapping would exceed the scope of the thesis. Starting from the general concept two main aspects will be discussed in the following chapters. One is the initial implementation of a tracking and following system integrated with a geometric mapping approach. The second is the development of a partial hierarchical environment model described in the context of a pilot study for HAM.

These two aspects form already a significant part of the whole concept, but the link through the topological graph is not fully established. The integration of a dialogue model into the approach is another open issue for future work and collaborations. Those issues as well as the aspects of the space segmentation are subject to future work as will be presented in detail in chapter 7.

3.5 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. On the other hand can interaction (communication) about the robot's workspace be facilitated with the integrating concept, as information can be retrieved and used in a human comprehensible way. Two aspects of the general concept have been picked as the main issues to be discussed in the remainder of this thesis.

Chapter 4

Technical setup - initial system

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. This chapter explains this initial implementation with particular emphasis on the tracking and following functionalities. The work was also reported in a conference article (Topp and Christensen, 2005).

The implementation was used in the pilot study described in chapter 5 which gives evidence of the usability of such an approach for the given context of a tour scenario.

4.1 The system

The modules responsible for the user tracking can be seen as driving components 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.

The previous chapter 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



Figure 4.1: *Minnie, the Performance PeopleBot as it was used in several tests*

4.1 shows the robot as it was used for the study described in chapter 5. Since no particular modifications were applied to the robot and the hardware is controlled with the help of the hardware abstraction software Player¹ 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 odometry 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.

An integration of parts of the system into a different control architecture is part of the future work within the Cogniron Project. The present thesis chapter though presents the control system for a full implementation on the robot “Minnie.

¹playerstage.sourceforge.net

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 project² 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 tools³. Qt offers an easy to use signalling mechanism, which made it attractive for the implementation, since the communication with the user is handled with the help of textual or 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 control⁴. The library has been developed during the last years at the Centre for Autonomous Systems and is aimed to be made an Open Source package. 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>

³<http://www.qt.org>

⁴<http://www.CURE.org>

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 component delivers 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 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

Despite the fact that the Player/Stage software itself provides hardware abstraction, the data format used in the CURE library and to have the system working also 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 speed 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 realised 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. The interface can be 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.

Data containers

Some of the functional module 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 Slam-Data’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 only needed for display- and diagnose purposes, no data gets lost in case access cannot be granted once in a while. This method of communication is easy to replace by different communication tools.

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 main contributing articles this thesis relies on deals with the results that could be achieved with the tracking module. These results and the respective article will therefore be described in more detail in section 4.2.

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 odometry 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 describe the relation between entries (nodes) in the topological hierarchy.

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 if 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 map handler, in which the topological graph is generated and maintained. Whenever information about a new *region* is given by the user the map handler generates a new region-node in a graph structure. In the current implementation those nodes are represented by the position where they were specified and a raw 360° range data set. Likewise if a new *location* is specified a new node

is generated, in this case only represented by the pose of the robot at the specification time. For navigational purpose the map handler also creates supporting “navigation”-nodes whenever the robot has covered a certain distance from the last generated node. These supporting nodes are to be removed successively with the mapping system getting more sophisticated (see chapter 7 for more detailed ideas).

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 make 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

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 be described in chapter 5.

Robot control - GoTo

The current system can make use of the topological graph to navigate back to any known *location* or *region* (in the latter case the path is generated to the exact position where the *region* was specified). With the help of an A* 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 5).

Architecture

Throughout the years a number of different design principles and architectural prototypes have been developed and postulated. 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

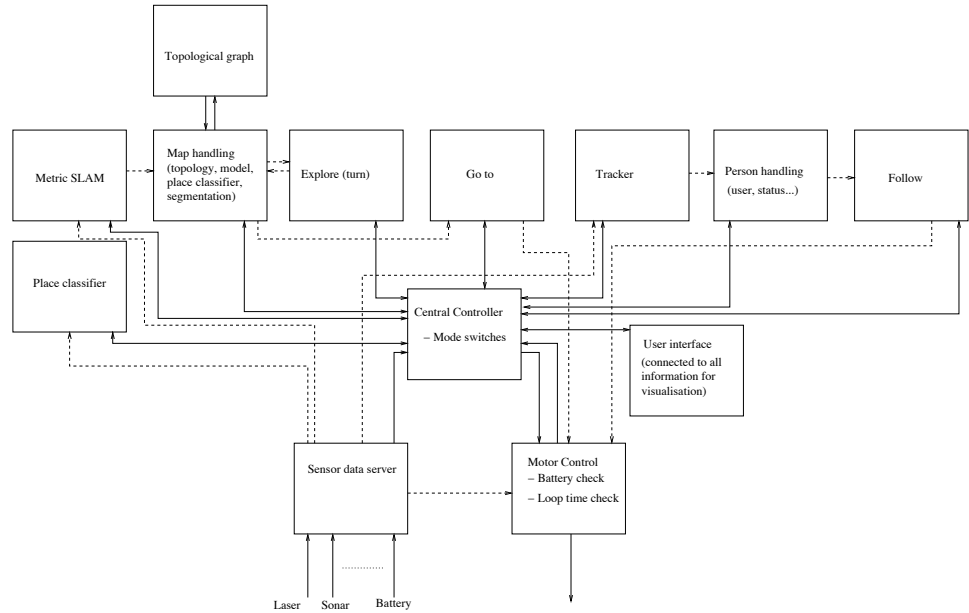


Figure 4.2: Overview of the implemented system XXX description/change in the image required!!! XXX

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 as a hybrid-deliberative design. Also here 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 4.2 gives an overview of the system components and connections between them.

Since the tracking component is central to the use of the complete system in an interactive context, the remainder of this chapter deals with the detailed investigation of the system's tracking component.

4.2 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. 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.

The tracking method

The laser range data based multiple tracking system uses the approach presented by Schulz *et al.* (2001). Their system 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 (SJPDF). Using this they can track multiple targets from a mobile robot, while motion compensation is done by scan matching. The work presented here adopts the idea of using the SJPDF approach for tracking and associating, but in contrast to Schulz *et al.* the detection and tracking methods allow handling of people standing still, which is needed for interaction.

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), Kleinhagenbrock *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 this case 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 been already accepted.

Tracking and feature association

As mentioned above SJPDFAs 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 person handler. 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 had the possibility to help) were asked to walk at different speeds, cross each other 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

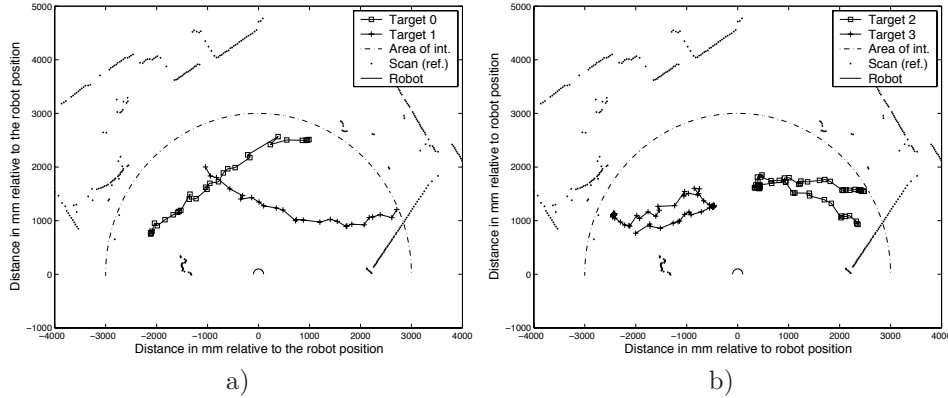


Figure 4.3: *The trajectories for a test with two persons moving in the field of interest, with the robot standing still. The dashed line marks the area of interest, the small half circle at position $(0,0)$ represents the robot. The dots show a reference scan of the environment. a) The two persons cross the area of interest, with target 0 being occluded by target 1. The trajectory for target 1 starts rather late after walking into the area. This can be explained by the accepting rule for features. Taking a look at the data at the corresponding time steps shows that the person in question was represented by a single leg pattern for a couple of steps. b) The two persons walk into the middle of the area, stop at a comfortable “chatting” distance (about 80cm) and separate again.*

moving person and was assigned the user flag when entering the area where a user would be expected.

Robot still, two persons: Two persons crossed each other 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 4.3 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 lab 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”. As well for the “chatting” scenario the tracker could handle the situation, which

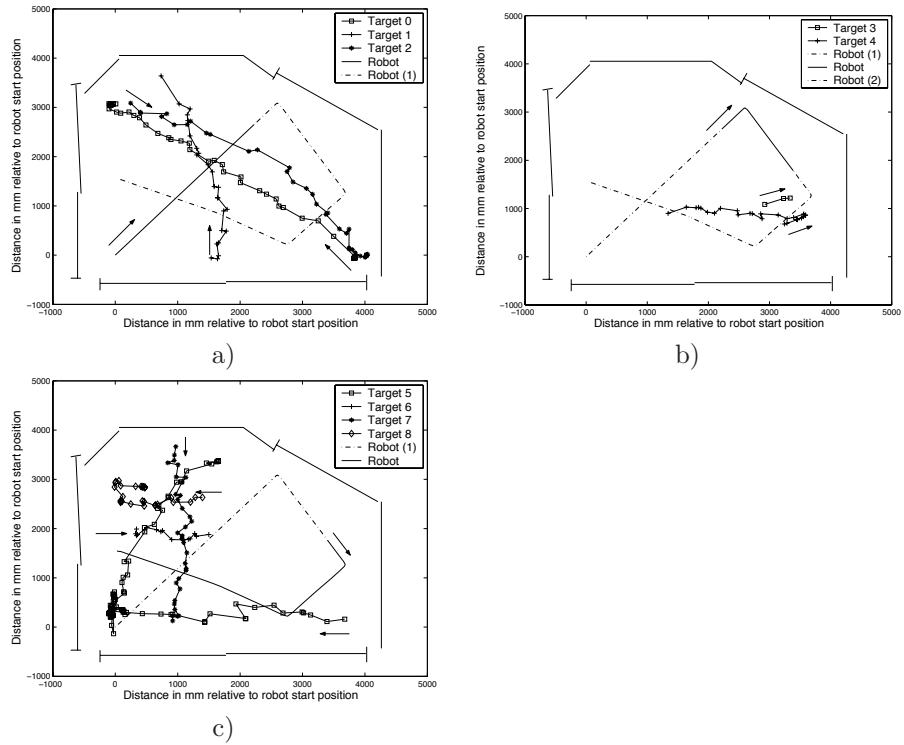


Figure 4.4: Person and robot trajectories in the x/y -plane of an experiment with up to three persons in the area at the same time and the robot moving independently. The resulting trajectories are presented in three images for clarity of presentation. The robot trajectory is shown as a solid line for the time period representing the events in the area. The rest of the robot's trajectory is shown dashed for convenience. 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.

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 4.4 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. Then it turned 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 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 conducted at our laboratory. The user study was a Wizard-of-Oz experiment and is described in more detail by Green *et al.* (2006) 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. 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 locations and objects. The study comprehends data from 22



Figure 4.5: *The experiment environment (“living room”) seen from different perspectives*

experiments. Laser range data was collected in all experiments at a data transmission rate of 500000 baud, though due to a communication stability problem not all of the experiments 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 4.6 shows a raw scan taken from a typical start position during the tests.

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 4.2 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).

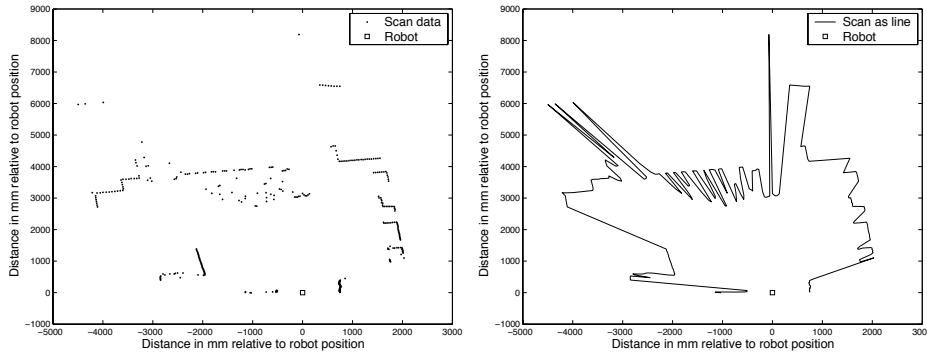


Figure 4.6: *The raw laser data (top) and the same data represented as polyline to show the data points in their angular order. The two peaks right in front of the robot are caused by the subject’s legs, while the other peaks result from the table and chairs, that belonged to the experimental scenario.*

The images in figure 4.6 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 we cannot assume a completely smooth representation for the targets’ possible movement (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 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.

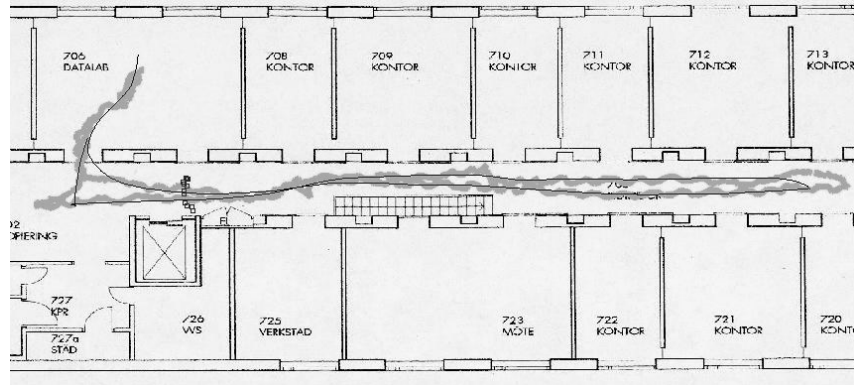


Figure 4.7: The trajectories of the user (grey, thick line), the following robot (black, thin line) and a bystander crossing the way of the robot (small squares) between the robot and user.

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 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 4.7 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 location, 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 second user study was designed (see chapter 5) 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.

4.3 Summary

The present chapter described the general design and implementation of an initial system for Human Augmented Mapping. Different types of architectures were considered, leading to a kind of hybrid-deliberative system, in which the planning component (or arbitration) is done by an interactively controlled central component. Reactive data interpretation components and a motor control module the different system components to use the data according to their demands.

One module of the system, consisting of the tracking and person handling components has been tested extensively in different experimental setups. Results of these tests led to one of the main contributing publications for this thesis (Topp and Christensen, 2005) and have been presented together with reflections on improvements and the use of the tracking system in further test scenarios.

Chapter 5

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

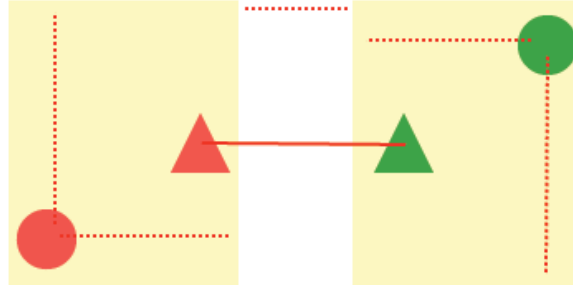


Figure 5.1: *The spatial relationship between two objects far from each other (circles) are rather estimated with the help of the relation to the region and the relation between the regions. For close objects (triangles) the relation is estimated directly, ignoring the “border” between the regions*

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.

5.1 A partially hierarchical graph 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 5.1 shows the two different ways of relating spatial entities with each other.

Given that a partial 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). (not incorporated in the framework so far)
- **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 previous 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”.

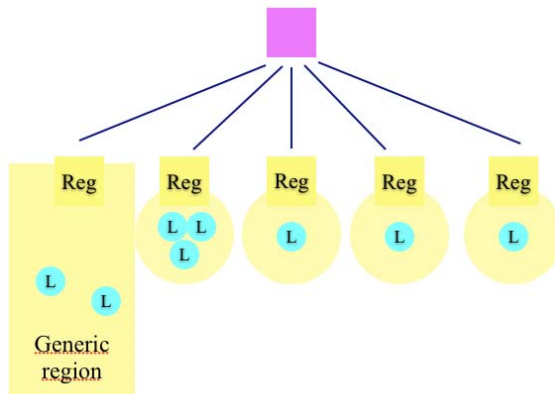


Figure 5.2: *The two-level model with “regions” (R) and “locations” (L). On a higher level “floors” could be integrated, while on a lower level “objects” would be appropriate*

- **Region** A container for one or several locations. Offers enough space to navigate (rooms, corridors, delimited areas in hallways).
- **Floor** A collection of rooms, distinct by the level (in height). Requires start of a new map to deal with similar structures if no altitude information available.

Partiality can 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 strictly 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 5.2 illustrates this simple two-level hierarchy with the “generic” entity. 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.

Both, the use of this graphical, high-level environment model and the implemented system described in chapter 4 could be tested in a user study setup. A

pilot study with five subjects was conducted which showed the implementation to be sufficiently robust and the environment model adequate for an interactive setting. The pilot study resulted in the second conference article mainly contributing to his thesis.

5.2 System and model in use - a pilot study

The study described in this section was set up to find out about different strategies users might show when presenting a known environment to a robot. Presumably such a process is guided 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).

In the following the study setup and hypotheses are described together with the observations made during five experiments. Some observations are directly related to the implementation of the tracking and following system (see chapter 4) and the interaction, while others are more related to the environment model to be investigated.

Scenario

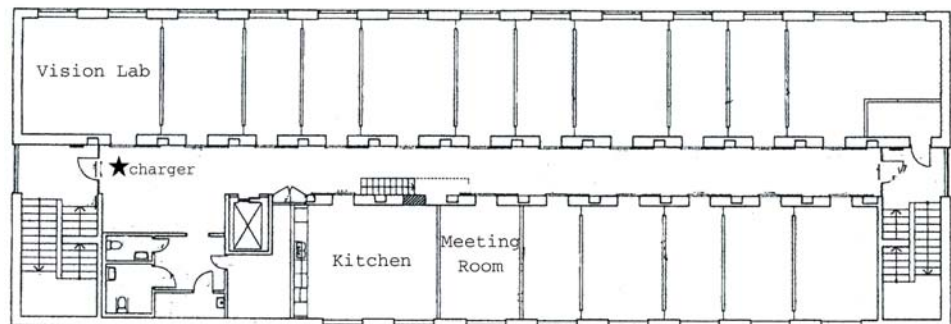


Figure 5.3: *The floor plan of the office environment on which the experiments took place. The star marks the starting point, where subjects encountered the robot*

The scenario of the study was a "guided tour" through a portion of an office building. Figure 5.3 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 around in

the environment so that it later could perform non-specified service tasks and in order to do this needed to have “seen” the respective *locations* (a more detailed description of the instructions and the technical realisation is given in sections 5.2 and 4).

Method

The following section explains the selection of subjects, the instructions given to them, and the methods used for data collection.

Subjects

As important precondition to the pilot experiments 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 experiments took place in. For further studies the question arises, if such requirements can be held up for technical reasons. Nevertheless, it is planned to assure as much familiarity with the environment as possible for further similarly experimental study setups. 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 “simulated” environment. The deliberate choice comes at a price however: in the used office environment 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 future user studies needs to be considered to make sure that the familiarity with the robotic system is counterbalanced by subjects without experience in robotics research.

To assure at least some variety in familiarity with robotic systems the five subjects were selected actively among the members of the Computer Vision and Active Perception Laboratory¹ that hosts a part of the Centre for Autonomous Systems² on the KTH campus. The group of pilot subjects included one secretary (familiar with robots from films, presentations and frequent encounters in the office environment, but not familiar with their internals), three computer vision researchers, one of them somewhat familiar with the internals of robotic systems, and one robotics researcher from the field of robotic mapping. Thus, the participants represented the full range of robot expertise available at the laboratory. All subjects had been working in this particular office environment for about two years.

¹<http://www.nada.kth.se/cvap>

²<http://www.nada.kth.se/cas>

Instructions

Subjects were given an instruction sheet (see appendix A) that explained the task and the functionalities and abilities of the robot. The task was to use a number of commands (`follow me`, `go to <target>`, `stop`, `turn left`, `turn right`) and explanations (`this is <item>`) to make the robot follow and to point out everything that the subject considered important for the robot to know on the floor the experiments took place on. The time frame given to the subjects for the completion of their task was about 20 minutes (15 minutes for the guided tour and five minutes to test the robots “memory”). In the instruction none of the words *region*, *location*, *position* or *place* was named. 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 were not of any interest, since the robot had no object recognition abilities; it just would need to know “where” to go. The instruction sheet included a drawing that showed, how the field of view of the robot looked like, and explained that the robot used a laser range finder to detect the subject for following and “looking around”. This information was important to the users’ understanding of the robot’s capabilities and behaviour, since the laser range finder at the used configuration setting only offers a forward field of view of 180° with a range of 8 meters (for the detection of users it was actually reduced further to 3 meters). The subjects could thus understand, how the robot perceived its environment.

A particular instruction given to the subjects regarded the approach to the robot and initiation of the robot’s following functionality. 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 start following them, 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 commands 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 “Technical realisation”, page 67, for details) by an experiment leader that followed the (subject and robot) pair. Subjects also were informed one 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 experiment, and knew that they could abort the experiment at any time.

Technical realisation

The robot “Minnie” was used as described in chapter 4 with the complete initial implementation of the Human Augmented Mapping system. In contrast to a previous user study performed with this robot (Green *et al.*, 2004) the robot navigated thus autonomously when it was following a user or moving toward a specified goal. Nevertheless the implementation allowed to switch to remote control immediately to guarantee the participants’ safety and to overcome possible technical problems.

Since the system so far is only controllable by a graphical user interface and the use of a speech recognition system seemed to rise too much instability when the study was planned, 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 interface.

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 afterwards if the behaviour switches seemed to correspond to their intentions, which could be confirmed (see section 5.3).

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 experiments could be obtained. Additionally the experiments 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 experiments the participants were asked to answer a number of questions (see appendix B) on the experiment in a short interview. This interview was scripted with a list of prepared questions on the motivation for naming or not naming certain *locations* or *regions* and for the handling of the tour scenario. 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 to investigate how different individuals present a known environment to a mobile robot and test the relation between the resulting information and the environment model presented above. 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 *regions* in the environment”,

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

WH3: “users point out *regions* without entering them”.

Those hypotheses were used to test whether the observations from the pilot study can be related to the graphical environment model. No specific hypothesis was given for the dependency “familiarity with robotic systems vs. way of explaining the environment to a robot” to explore this issue. Nevertheless the participating robotic researcher particularly familiar with map representations was expected to be more explicit 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*.

5.3 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 consequently not entirely representative. However, it is possible to analyse the outcome of the experiments in terms of *occurrence* of different phenomena. 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 further studies (see chapter 7).

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 observed. In table 5.1 quantifiable results are summarised to give an overview over observations and statements from the interview.

Observations

All subjects but one used the full time frame to present the environment to the robot. The “tour” started for each experiment at one end of the corridor (see Figure 5.3), where the robot awaited its user. An initial *location* (the “charger”) was

Observation	Subject	VR	VR	VR	SE	RR
Interaction time		22min	19min	11min	25min	24min
# <i>regions</i>		4	2	–	2	2
# <i>locations</i> ^I		4	4	5	4 ^{II}	8 ^{III}
# <i>regions</i> w o loc.		3	2	–	1	1
# loc. w o <i>region</i>		3	4	5	2	3 ^{IV}
# <i>regions</i> 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 <i>regions</i> 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 5.1: Quantifiable results from the pilot study

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 trial.

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*. 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 naming 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 evidence 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*). 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 simulated by the experiment leader.

Despite some technical problems (see section 5.3 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 4 (particularly the tracking system) is sufficiently robust and stable in its performance to be used in a study setup with unexperienced participants.

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.

A general requirement for this relation is that the system has the ability to perceive *regions* that are delimited from other *regions* autonomously. Such a delimiting process is subject to future investigations (see chapter 7). It also is assumed that a general knowledge model distinguishes between *regions* and *locations* and a dialogue model that uses this knowledge base is included in the system. Also these issues are subject to current and future investigations.

From the experiments 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 in they just stated that this was “the < *name* >” without positioning the robot with “turn commands”. The detection of such differences in the user’s behaviour and spoken utterances could give a signal on the actual intention and is subject to further analysis of this and other user studies to be conducted.

Departing from detailed observations some key situations can be specified that need to be handled by the robotic environment model (see figure 5.2) together with possible solutions to cope with them. Those suggestions can be taken into account when the actual topological *region* delimiting components are to be integrated into the implementation.

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 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 (e.g., walls, doorways) are now associated to this new branch in the hierarchy.

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.

Such a differentiation might be possible to obtain by carefully analysing the user's strategy of positioning the robot, which could be observed as described above.

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

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.

Interaction issues

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

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 come after the user. From carefully studying the recorded interaction on video it was concluded that the robot actually needs to indicate with a body (movement) gesture like turning toward the user that it has seen the user and is ready to follow. An improvement of the system to incorporate a better signalling and feedback strategy here is subject to current investigations.

A similar problem occurred, when subjects made the robot face something to "look at it" and wanted to continue the tour afterwards. It is planned to make the robot turn back toward the user to indicate, that it is ready to continue after storing a presented item.

Other inconvenient situations occurred due to failures of the tracking system (see section 4.2), 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.

5.4 Summary

Human Augmented Mapping integrates human robot interaction with robotic mapping. The connection between those two fields is a hierarchical graph model that aims to reflect the theory of partially hierarchical representations of space in humans. This chapter presented the background for this model and a user study setup which was specified to find out about the usability of this model in an actual interaction context. Five pilot experiments were conducted with this study setup, in which the initial implementation of a Human Augmented Mapping System described in the previous chapter was used and thus tested under realistic conditions. Working hypotheses of the study and the applicability of the implementation in a user study context could both be shown.

Chapter 6

Summary and discussion

Human Augmented Mapping (HAM) is an approach to robotic mapping that integrates information given by a human interactively 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 of research.

This thesis presented the overall concept for Human Augmented Mapping as well as results that could be achieved with experiments and a user study. The work was conducted within the European Integrated Project “Cogniron – The Cognitive Robot Companion” and was thus guided by aims and scenarios defined for the project. As a starting point a “guided tour” scenario is used, in which a user is supposed to teach a service robot about an indoor domestic or office environment. This scenario is used also in the work presented in this thesis to limit the context for experiments to a manageable level.

HAM offers advantages for the robotic mapping process by integrating information given by the user in the interactive setting of the “guided tour” or “home tour”. 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, an environment representation based on psychological findings and common spatial concepts is used as a link between robotic mapping and interaction.

An initial implementation of a system for HAM was tested with a strong focus on the tracking system that is necessary to enable the user to take the robot on a tour. The overall system and the tracking module were described together with the results that could be achieved. Apart from the conclusion that the used multiple target tracker is sufficiently robust for the “follow” task the robot has to deal

with, it seems possible to use it as well for the detection and adequate handling of bystanders, which was not incorporated into the system, but could be an interesting extension. The implementation was also used in a number of pilot experiments in a user study setup (see below) which again confirmed the applicability of the system.

In a separate chapter the background and design for the environment model that forms the link between user and robot was described together with the setup of a user study. This study aimed to confirm the soundness and validity of the environment model by observing users who presented a known environment to a service robot. A large variety of strategies for such a “tour” could be observed, which suggest a number of situations the environment and a fully implemented system would have to cope with.

6.1 Open issues and problems

The described implementation has to be seen as an initial implementation that did not incorporate the full hierarchy of mapping approaches or the environment model. Neither was any dialogue system included so far, which means that spoken dialogue based interaction still has to be mediated by an operator. Including such enhancements as part of the implemented system is subject to ongoing work. Additionally there are still open questions about some conceptual issues to be answered in test settings. One is how to measure the quality of an implemented system for Human Augmented Mapping in the context of a complete, interactive service robot setup. Another question is how much predefined knowledge a complete system has to have in terms of spatial “common” concepts and semantics.

Quality measurements

The breadth of Human Augmented Mapping makes it difficult to compare the performance of an implemented system to other systems, or just measure the quality of one system in itself. The fact that, for example, the system components dealing with more or less traditional robotic mapping measure up to state of the art mapping systems does not cause automatically an “easy-to-control” experience for an arbitrary user, when the system is tested in a real world context. A perfectly working tracking system does not guarantee appropriate feedback to the user in any possible situation in the interaction flow. Since understandability is difficult to measure in numbers setting up user studies seems to be the most appropriate way to get feedback on an integrated systems performance in “reality”.

One possibility to learn more about the requirements for highly interactive systems is in fact to conduct user studies to observe in a setting as realistic as possible the interaction between a user and a robot. Often those studies are limited to a very narrow context to learn more about human preferences that could possibly be met with the help of certain configurations. A more open setting is an exploratory study setup which has been described in this thesis. Such setups allow to identify and

investigate interesting issues for the modelling behind the system, not necessarily allowing to directly draw a conclusion on the best configuration for further systems. The combination of both methodologies, exploratory and clearly specified together with traditional performance measures for the particular system components seem a useful combination of measurements for Human Augmented Mapping.

Given the quality measurements and “ribbons” in traditional fields of robotics, such as mapping, the quality of the single components in an integrated system might not seem to be stunningly high, but one has to keep in mind that the concept builds on the integration of various complex tasks that have to be performed synchronised and in an on-line setting. Thus, the quality of the components should be measured against traditional, pure methods, but always before the background of the specific context they need to be used in.

Knowledge

A fundamental question in the context of Human Augmented Mapping is “Who should have the knowledge?”. Knowledge in this context can be precoded concept categories, e.g., a “kitchen” is a “room” (is a *region*) and even the semantics of such categories (“If this is a “kitchen”, then there should be a fridge” and vice versa, “If this is the coffee machine, and we are in a home, then we are most likely in the kitchen”). How much of such knowledge needs to be represented up front and where? Since architectures change and often modern domestic settings do not have clear delimitations (“rooms”) for many areas, it seems (to the author at least) almost impossible to come up with a clear, general categorisation and reasoning scheme for the semantics. Precoded knowledge implies here a very careful design of hypotheses handling, allowing to deviate easily from the scheme. If such knowledge is to be handled up front, this is clearly not to be done in the “mapping part” of a system. On this side of the system information can only be stored and retrieved, together with a level of uncertainty, that would allow higher level interpretation mechanisms to decide on what to do next.

This leads to a tentative answer for the other issue. How much of the general categorisation along the environment model should be known in advance and where should it be specified? A limited knowledge base provided by the dialogue system can be of great help to easily and quickly define the type of information given to the system. In ambiguous situations it seems also appropriate to leave the disambiguation (Do you mean the room we are standing in or the object you are pointing at?) to a dialogue system. This would leave the mapping part of a HAM-system without any responsibility for reasoning. There might be one case though, where both dialogue/interaction supervision and mapping process have to cooperate to disambiguate a situation. This is the case when a room is presented by the user from the outside. Here a combination of different cues can actually make it possible to “understand” the situation and integrate the information properly into the environment representation. From the dialogue model it has to be known, that a “room” (*region*) is presented. The interaction/gesture supervision identifies a

presenting behaviour that is normally not used to present the surrounding *region*. From the environment representation it is either known, that the surrounding *region* has been specified already and it is unlikely to get a new label now, or it is obvious that the system is positioned to face an opening into an unknown area. Here the already gathered knowledge of the mapping and representation part is needed, but it seems that this subsystem itself can be kept free of precoded knowledge and reasoning processes. This is a design choice which will be expressed in the further extensions of the initial system, especially with respect to the integration of a dialogue.

6.2 Conclusion

This thesis presented an approach to Human Augmented Mapping, a concept to facilitate robotic mapping and human robot interaction by integrating both into an interactively controlled mapping system. The initial implementation presented could be used in a user study setup with which certain issues and individual strategies for an interactive mapping process were investigated. Overall the approach of integrating the different parts seems to be promising while numerous interesting questions still have to be answered by future improvements and studies. Thus, the work presented here can be concluded to provide a good basis for further investigations in Human Augmented Mapping, both in terms of user studies as well as in terms of environment representation and mapping experiments.

Chapter 7

Future ideas and plans

The implemented system so far allows to be used for both, further experiments and developments in the topological representation for disambiguation purposes, and the setup for further user studies. This allows to investigate different aspects of the robotic mapping process as well as those evolving from the interaction. The following open issues are planned to be part of the future work on the doctoral thesis as well as they are to be investigated within the Cogniron project. Since some of the questions require to conduct user studies that closely relate to the project work done at the IPLab of CSC/KTH and to cooperations with other partners, they are only roughly sketched here. The planning thus needs to involve further discussion.

7.1 User studies

One possible study setup is the extension of the described pilot study to be more comprehensive and thus also more representative. Other study related issues include the investigation of interaction strategies in different situations.

Enhancing the pilot study

Such a setup would involve a basic study that possibly requires a larger amount of subjects (>10) to investigate in a within-subject design the possible differences for the presentation of an environment to a human and to a robot. According to the literature it seems evident, that humans tend to treat computers as individuals, but it is not entirely clear if this applies to service robots in the same way. Also it is not obvious, if the treatment as “individual” implies a treatment “as a human”. Such an investigation might give evidence about the mental model users have of a particular robot and its abilities which can lead to design implications for underlying models and the design of interaction capabilities.

Interaction strategies

In the pilot experiment differing strategies for the presentation of objects (*locations*) and complete rooms (*regions*) could be observed. This difference manifested itself in a “preparation phase” for the presentation of an object in which the robot was explicitly turned (with near navigation commands) to face the respective object. Such preparation was omitted for the presentation of rooms in case that robot and user were currently in the room to be presented. The issue to investigate is in how far the detection of such preparation commands together with a certain knowledge about the current configuration in space can be used to determine the category of the item to be presented. Possibly the hierarchy can be extended to incorporate objects, for which also a particular strategy of presentation seems to be evident from previous studies. This would allow to reduce the requirements for a (spoken) dialogue model knowledge base that otherwise would have to provide the categorisation. This issue is to be investigated on conceptual level in cooperation with the IPLab group at KTH and the Cogniron project partner group at the University of Bielefeld.

7.2 Topological modelling

So far the described system concentrates on the integration of both a mapping system with the interaction. Another central part of the system is in fact the topological modelling that forms a link between the used metric mapping approach and the high level graph model for the environment representation. This topological link is needed to enable the robotic system to “understand” what is meant with *regions* and links between them. Initial experiments have already been conducted, but were not considered part of this thesis, since more sound results are to be expected.

Representing *regions*

Based on the idea to have a simple and concise set of features that can be acquired in an on-line fashion a method to represent *regions* has been implemented and tested. The approach is based on the assumption that differing areas in an environment can be represented by three features: The area they cover and their maximal length and width along the principal components of a laser range data set obtained in them. The applicability of this approach is going to be tested in different and more complicated data sets.

Separating *regions* and building a topological map

Given a metric map, the feature based representation for *regions*, and the high level environment model, the link needed is a topological map that ties all together. Initial tests showed so far, that the representation used for specified rooms (*regions*)

can also be used for continuous updates. Areas with little changes in the feature representation can thus be grouped to topological nodes. The high level map can then - according to users' preferences - group those areas as a labelled node. For the initial tests only a 180 degree laser range data set was used which showed to be highly sensitive to rotations.

A dataset for space representation in a home like scenario has been generated in cooperation with the University of Amsterdam's project group. On these data an analysis of the proposed region segmentation method is to be performed. Initial tests show, along the preliminary tests on data collected in our own group's office building, that the method performs quite well. Further tests will involve a virtual range data set covering 360 degrees to test the applicability in the context of constant updates.

Resolving structural ambiguities

The topological segmentation and representation can be considered under two perspectives. Making use of the categorisation that could be shown possible so far for a localisation strategy is one of them. More interesting is actually the investigation in how far areas that cannot be classified clearly even by a human can be detected as areas of interest where respective clarifying questions can be posed. In this context different experiments will be run and analysed to see, if the earlier proposed geometric feature based representation can help to solve at least some of the problems. As a first step the method will be applied on virtual 360 degree scans instead of capturing 360 degrees by making the robot rotate. This allows to become more independent from the sensor configuration. Further it will be tested, if based on those virtual scans the method can be used in an on-line fashion to detect areas of change while travelling and in how far those can be used to segment the environment topologically, still with a possible link to high level map representation as a requirement. Given that a useful segmentation can be achieved, the method will be compared and/or combined with a door detector to show assumed improvements. It seems possible that the combination of both methods reduces the number of both false positives and false negatives for the detection of spatial delimiters significantly. Both the implementation and the possible test setup (finding/generating a respective data set) will have to be investigated.

Separation of metric maps for the topological structure

The geometric features used to describe the topology of the environment currently rely on metric relations. This requires very accurate positioning throughout the whole process of acquiring a representation of the environment. In this step it should be investigated, in how far global metric positioning can be reduced to small scale local metric relations within what is termed a "region" throughout the work. Localisation should then be reduced to a pure geometric descriptor based recognition process.

7.3 Disambiguation in ambiguous environments

Several types of ambiguities might have to be resolved in an interactive mapping process. One type is caused by individual preferences that are expressed by different users. Other types have been previously mentioned as the general ambiguity in a localisation process (the system needs to decide under uncertainty “where the robot is”) and structural/conceptual ambiguities evolving when the environment does not provide a clear cut between two areas, for example between a narrow corridor and an open entrance hall.

Personalising and updating maps

A question in the context of a service robot that can be assigned to different users is the personalisation of the map information given to the user. It is possible to think about two basic approaches to handle personalised information. One is to incorporate everything in one map and resolve ambiguities by adding more and more information that might be assigned to specific users. The other way is to generate separate maps for every user of the system. Advantages and disadvantages of both options have to be investigated and respective applicability checks for the map generation need to be considered. This work is to be expected to provide results useful for the disambiguation of semantic/conceptual ambiguities.

Those issues will be investigated as the basis for the continued work on a doctoral thesis, which this licentiate thesis will be part of.

Appendix A

The instruction sheet

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**

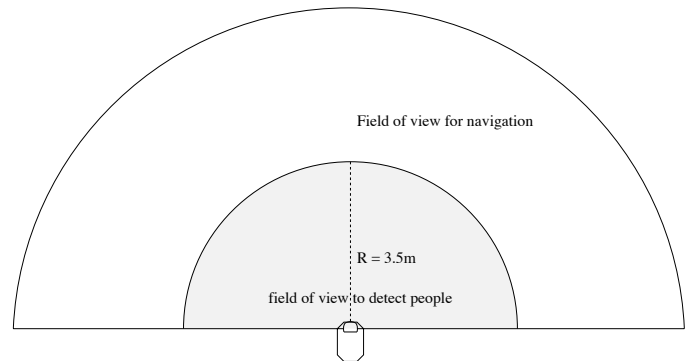
Hello and welcome to this pilot study. This document should give you the information you need to participate in our study on human robot interaction and interactive (robotic) mapping.

**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 not try are to:

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

Some notes on technical issues

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

Control of the robot The experiment leaders - we - will follow you and the robot for several reasons. One is, that we want to observe and videotape the experiment not only from the robot's point of view, but also from a more general point of view.

The other reason is, that we want to assure your safety and comfort. We can interrupt the robot's automated control at any time and switch to manual control. Or abort the experiment completely. That is one reason for a laptop being carried around. The second is, that we do not want to rely on speech recognition. The contents of your utterances are translated into the respective commands and informations "by hand" and given to the system by typing. That is the second reason for the laptop.

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

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

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

Your privacy

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

Appendix B

Interview questions

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, Testperson:

1. Did you notice the difference in the reactions of the robot to regions/rooms and places/locations?
2. Do you think this difference was appropriate?
3. Why do you think the robot had these differences in its behaviour?
4. Did this give you the impression that the robot was “thinking” of the same thing as you were?
5. Why did you not show the ...?
6. Why did you show the ..., but not the ...?

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

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*. Taipei, Taiwan. Sep. 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. 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. Feb. 2002
- Bakker, B., Kröse, B., Christensen, H.I., Belpaeme, T., Paquier, W., and Chatila, R. *Cognitive Models of Space and Objects*. Project deliverable, University of Amsterdam, The Netherlands for EU-integrated Project COGNIRON, FP6-IST-002020. 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*. Barcelona, Spain. Apr. 2005
- Bennewitz, M., Burgard, W., and Thrun, S. *Using EM to learn Motion Behaviors of Persons with Mobile Robots*. In *Proceedings of the 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*. Salt Lake City, UT, USA. Mar. 2006

- Blom, J. *Personalization: A Taxonomy*. In *CHI'00 extended abstracts on Human factors in computing systems*. 2000
- Bosse, M., Newman, P., Leonard, J., and Teller, S. *SLAM in Large-scale Cyclic Environments using the Atlas Framework*. *International Journal on Robotics Research (IJRR)*, 23:pp. 1113–1139. Jul. 2004
- Breazeal, C. *Sociable Machines: Expressive Social Exchange Between Humans and Robots*. Doctoral dissertation, Department of Electrical Engineering and Computer Science, Cambridge, MA, USA. 2000
- 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. 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*. Salt Lake City, UT, USA. Mar. 2006
- Brooks, R. *A robust layered control system for a mobile robot*. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, vol. RA-2. San Francisco, CA, USA. 1986
- Bruce, A. and Gordon, G. *Better Motion Prediction for People-Tracking*. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*. New Orleans, LA, USA. Apr. 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*. Amsterdam, The Netherlands. Mar. 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). Oct. 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*. Canberra, Australia. 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). Apr. 2001
- Clark, H.H. and Brennan, S. *Grounding in Communication*. In Resnick, L.B., Levine, R.M., and Teasley, S.D. (Eds.), eds., *Perspectives on socially shared cognition*, pp. 127–149. American Psychological Association, Washington, DC, USA. 1991

- 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*. Barcelona, Spain. Apr. 2005
- Feyrer, S. and Zell, A. *Robust Real-Time Pursuit of Persons with a Mobile Robot Using Multisensor Fusion*. In *Proceedings of the 6th International Conference on Autonomous Systems*. 2000
- Fletcher, L., Apostoloff, N., Petersson, L., and Zelinsky, A. *Vision in and out of Vehicles*. *IEEE Intelligent Systems*, 18(3). May/June 2003
- Folkesson, J. *Simultaneous Localization and Mapping with Robots*. Ph.D. thesis, KTH School of Computer Science and Communication (CSC), Royal Institute of Technology (KTH), Stockholm, Sweden. Sep. 2005
- Galindo, C., Saffiotti, A., Coradeschi, S., Buschka, P., and Fernández-Madriral, J.A. *Multi-Hierarchical Semantic Maps for Mobile Robotics*. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. Edmonton, AB, Canada. Aug. 2005
- 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*. Edmonton, AB, Canada. Aug. 2005
- 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 *Proceedings of the 2006 Workshop on Performance Metrics for Intelligent Systems (PerMIS)*. Washington, DC, USA. 2006
- Gustafsson, F. *Adaptive Filtering and Change Detection*. Jon Wiley & Sons, Ltd., Chichester, England. ISBN 0-471-49287-6. 2000
- 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/>. Oct. 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*. Beijing, China. to appear. 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*. 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*. Hatfield, UK. to appear. 2006b
- Kirsh, D. *The intelligent use of space*. *Artificial Intelligence*, 73:pp. 31–68. 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 (ROMAN)*. Berlin, Germany. 2002
- Kluge, B. *Tracking Multiple Moving Objects in Populated Environments*. In Hager, Christensen, Bunke, and Klein, eds., *Lecture Notes in Computer Science (LNCS)*, vol. 2238. Springer Verlag, 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*. Orlando, FL, USA. May 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 Freksa, C., Habel, C., and Wender, K., eds., *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*. Venice, Italy. 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*. Salt Lake City, UT, USA. Mar. 2006
- Kuipers, B. *The "Map in the Head" Metaphor*. *Environment and Behavior*, 14(2):pp. 202–220. 1982
- Kuipers, B. *The Spatial Semantic Hierarchy*. *Artificial Intelligence*, 119:pp. 191–233. 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*. Los Altos, CA, USA. 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*. New Orleans, LA, USA. Apr. 2004
- Kyriakou, T., Bugmann, G., and Lauria, S. *Vision-based urban navigation procedures for verbally instructed robots*. *Robotics and Autonomous Systems (RAS)*, 51:pp. 69–80. Jan. 2005
- 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
- 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
- McNamara, T.P. *Mental Representations of Spatial Relations*. *Cognitive Psychology*, 18:pp. 87–121. 1986
- McNamara, T.P. and Shelton, A.L. *Cognitive maps and the hippocampus*. *TRENDS in Cognitive Science*, 7(8):pp. 333–335. Aug. 2003
- Mozos, O.M., 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*. Barcelona, Spain. 2005
- Nourbakhsh, I., Powers, A., and Birchfield, S. *DERVISH An Office-Navigating Robot*. *AI Magazine*, 16(2):pp. 53–60. 1995
- Pacchierotti, E., Christensen, H. I., and Jensfelt, P. *Embodied social interaction for service robots in hallway environments*. In *Proceedings of IEEE Field and Service Robotics*. Brisbane, Australia. Jul. 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*. Salt Lake City, UT, USA. Mar. 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

- 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*, pp. 1665–1670. Seoul, Korea. May 2001
- Shannon, C.E. *A Mathematical Theory of Communication*. Bell System Technical Journal, 27:pp. 379–423. 1948
- 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. 2005
- 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. Jun. 2004
- Tellex, S. and Roy, D. *Spatial Routines for a Simulated Speech-Controlled Vehicle*. In *Proceedings of the ACM Conference on Human-Robot Interaction*. Salt Lake City, UT, USA. Mar. 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*. Salt Lake City, UT, USA. Mar. 2006
- Thrun, S. *Robotic Mapping: A Survey*. In Lakemeyer, G. and Nebel, B., eds., *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 AAAI National Conference on Artificial Intelligence*, vol. 2, pp. 944–950. Portland, OR, USA. 1996
- Tolman, E.C. *Cognitive Maps in Rats and Men*. Psychological Review, 55:pp. 189–208. 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. 2003
- Topp, E.A. *An Interactive Interface for Service Robots – Design and Experimental Implementation*. Diplomarbeit (written in english), Faculty of Computer Science, University of Karlsruhe, Karlsruhe, Germany. Oct. 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. Mar. 2005
- 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*. Edmonton, AB, Canada. Aug. 2005
- 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. Mar. 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*. Salt Lake City, UT, USA. Mar. 2006
- 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*. New Orleans, LA, USA. Apr. 2004
- Trafton, J.G., Schultz, A.C., Perzanowski, D., Bugajska, M.D., Adams, W., Casimatis, 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*. Salt Lake City, UT, USA. Mar. 2006
- 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*. Salt Lake City, UT, USA. Mar. 2006
- Wang, R.F. and Spelke, E.S. *Human spatial representation: insights from animals*. *TRENDS in Cognitive Science*, 6(9):pp. 376–382. 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