2008 IEEE International Conference on
Robotics and Automation
Pasadena, CA, USA, May 19-23, 2008

Detecting Structural Ambiguities and Transitions during a Guided Tour

Elin A. Topp
School of Computer Science and Communication
Royal Institute of Technology (KTH)
10044 Stockholm, Sweden
topp@csc.kth.se

Henrik I. Christensen
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0760, USA
hic@cc.gatech.edu

Abstract— **Service robots designed for domestic settings need to navigate in an environment that they have to share with their users. Thus, they have to be able to report their current state and whereabouts in a way that is comprehensible for the user. Pure metric maps do not usually correspond to the understanding of the environment a user would provide. Thus, the robotic map needs to be integrated with the human representation. With our framework for Human Augmented Mapping we aim to deal with this issue and assume a guided tour as basis for an initial mapping process. During such a tour the robotic system needs to be able to detect significant changes in its environment representation – structural ambiguities – to be able to invoke a clarification discourse with the user. In this paper we present our approach to the detection of such ambiguities, that is independent from prior specification and training of particular spatial categories. We evaluate our method on data sets obtained during several runs in indoor environments in the context of a guided tour scenario.**

I. INTRODUCTION

Service robots or mobile robot companions need to navigate in an environment that they share with their users. They also have to be able to report their current state and whereabouts in a way that is comprehensible for the user. Robotic localization and mapping is most often performed using geometric features being derived from sensory measurements. Those feature based representations are typically different from the spatial models that humans use to define and reason about the same environment. This poses a challenge in particular if the system is to be operated by novice users without any robotics background.

There is thus a need to reconcile the user's models of the environment with the robot's representation of the same space – a *shared* model that can be *personalized* is needed. We consider our framework of *Human Augmented Mapping* [12] a possible way to approach this issue in general. The framework subsumes different aspects of robotic mapping, spatial representation and human robot interaction.

Within the context of Human Augmented Mapping we assume an interactive scenario – a "home tour" – as the most natural way of providing the robot with the needed information about the environment as it is seen by the user. The human user guides the robot and gives names to things and places according to her personal preferences, while the robot builds a suitable (hybrid) map that is augmented with this information. More particular, we assume the robot to use a representation for "regions" (typically rooms or distinct areas in which the robot can navigate) that is based on statistics extracted from laser range data [13]. Those regions form the nodes of a topological graph. In order to enable the communication with the human user in a useful way the separation of such regions needs to correspond quite closely to the user's understanding of the distinct regions in an environment. The general assumption is that a "region node" is generated when the user shows a particular region to the robot. During a guided tour though a number of different ambiguities can occur, one of which we term a "structural ambiguity". This is considered to be a situation, in which the robot detects a significant change – a transition from one region into another – in the environment, while the user did not (yet) introduce any new region actively. It might be necessary for the robot to know in such a situation if it is still *inside* or already *outside* a previously defined region. In this article we describe, how our statistical representation can be used to detect such transitions and how the region distinction (and the corresponding map) can then be updated with the help of the user.

A. Overview

In the following we give a short overview over our framework of Human Augmented Mapping and the system context (section II). Reflections on related approaches are given in section III. In section IV we describe what we mean by structural ambiguities and how they can be detected and used to separate an environment into particular regions. We present our implementation in section V and evaluate the approach in section VI. Section VII contains our conclusions and ideas for future works.

II. SYSTEM CONTEXT

We propose an approach to detect changes in the representation of an environment that can be used for negotiations about the current location with a human user. Since this is part of a conceptual approach to building a human comprehensible environment representation, the overall concept of Human Augmented Mapping is outlined to provide a context for this paper.

A. Human Augmented Mapping - concept

We assume a service robot designed to work in a domestic environment populated by humans - potential users. We also suppose that for appropriate communication about the robot's workspace a graph representation is needed, that

978-1-4244-1647-9/08/$25.00 ©2008 IEEE.

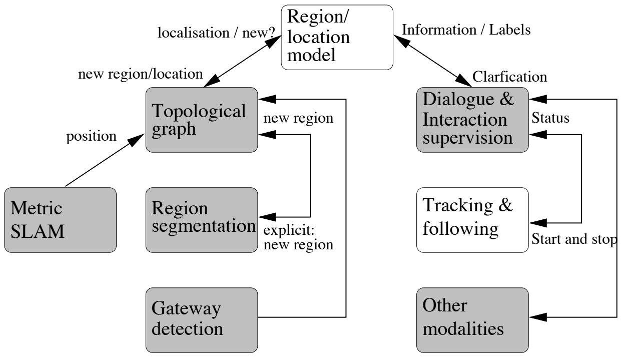

Fig. 1. Human Augmented Mapping (HAM) overview.

can incorporate concepts into its nodes. On the other hand we assume that even an underlying metric representation is needed for the robot for exact localization and navigation to perform its services. By taking the human knowledge and abilities into account and controlling the mapping process interactively, it is possible to integrate the human concepts and understanding of the environment into the resulting environment representation. This helps the human to communicate with the robot about its tasks and whereabouts according to the semantic and conceptual understanding the user has. For the robotic mapping process it is helpful to consider the user's information also for building a topological representation that forms the link between the conceptual graph representation and the metric map. It becomes possible to resolve ambiguities and answer questions related to different levels of the overall representation, e.g., "you mentioned one bedroom already, is this the same or a second one?" (conceptual/semantic level) or "was this a door we passed?" (topological level). This implies that a two way communication has to be made possible. Different types of events, i.e., external conceptual input from the user and internal detection of topologically significant structures, have to be considered. Figure 1 shows a possible system in a schematic way. A form of topological node distinction is needed, which is accomplished by the region segmentation module. This module is used both for the explicit specification of regions and the continuous detection of significant changes in the environment. The latter is the main focus of this paper.

B. Regions

To classify spatial entities (e.g., rooms) it is necessary to separate them as spatial unit from the rest of the environment representation. We use the term *region* for such a spatial unit. Topologically speaking, a *region* can contain several distinct places (e.g., different in terms of the perceivable appearance of the environment at the respective location) that form the nodes of a topological graph representation.

C. Events

We consider two types of events that can trigger the system to segment the environment in the internal representation. One is to receive external input that annotates a certain spatial entity with a label (e.g, "... this is Elin's office..."). The other type of event is the data driven detection of a "new area". One possible solution is a gateway detection (e.g., for doors), but we believe that our statistics based descriptor [13], so far used for the representation and classification of specific *regions*, is more flexible when applied to the detection of structural ambiguities and transitions in general that might not necessarily involve doors. Those ambiguities and the use of our region representation are explained in detail in section IV

III. RELATED WORK

Since our presented work concentrates on obtaining a topological partitioning of a given environment but is embedded in an interactive framework we give an overview of related work in both areas.

A. Topological representations

Approaches to topological environment representation, or map building, have been reported in the context of different presentation strategies or learning. One strategy is to predefine the topological structure of an environment and use this map for localization and navigation purposes. Nourbakhsh *et al.* used this strategy for their implementation of a path planning system for "Dervish" [10]. The limitations of such an approach in the context of our Human Augmented Mapping framework and the arbitrary environment we assume are obvious: the complete possible working environment for the robot needs to be known in advance, including all possible transitions along the edges and measurements to describe doors and hallway intersections.

Other, more adaptive methods that assume the robot to acquire a topological representation of its environment are based on (sensory) data obtained while traveling. Those can be subdivided in unsupervised and supervised approaches.

An unsupervised/autonomous method for the detection of *places* is suggested by Beeson *et al.* [2]. The authors propose to use the extended Voronoi graph for the segmentation of the environment which was initially investigated by Choset *et al.* [3] and under a paradigm for an exploration strategy by Kuipers and Byun [8]. Their definition of a place suits the requirements and abilities of an autonomous system, but does not necessarily correspond to a personalized representation of a human user. This limitation can be observed also for other completely unsupervised methods of topology learning, as for instance proposed by Tapus *et al.* [11]. Here a method based on a combination of images and laser range data is used. The appearance of a certain area is captured as a "fingerprint", a string that represents different types of observed features (color occurrences, vertical lines, corners, etc.) in the angular order they were perceived in. Such fingerprints are actually a rather concise description of a certain area that can be obtained on-line and can probably be triggered by both external or internal events. A disadvantage of the purely sequential representation though is that not the area (*region*) itself is captured as a spatial entity.

For the representation of convex areas Kröse showed that it is possible to represent such regions reliably by obtaining only one sample range data set and transform it to its center point and bearing with the help of a principal component analysis to anticipate future scans [6]. This method alone has the limitation of working robustly only in convex spaces but we believe that it is usable also for other areas as one method in a more complex framework. Our representation for *regions* is thus related to this proposed approach.

Mozos *et al.* show, how the *category* of a certain area (room, doorway, or corridor) can be determined with the help of supervised learning [9]. They generate a number of features from raw laser range data sets that were obtained at different locations corresponding to the named categories and use these features to form a training data base for the learning method. We adopt the idea of using a set of features to represent a laser range data set, that we obtain in *regions*, but use an even more concise set of features [13]. Further Mozos *et al.* label places in the complete environment into a fixed number of categories, while we do not rely on any previously defined categories for the regions that can be specified by the user. This allows us to concentrate on the transition from one region into the other, not regarding what category (in the sense of the mentioned work) the regions or the transition itself belong to. This is also the main difference in comparison to work recently presented by Friedman *et al.* [5] who use Voronoi Random Fields to label points in the environment according to their spatial category (room, corridor, doorway, or junction) and build a respective topological map on top of an occupancy grid map.

B. Mapping and interaction

An earlier approach to supervised learning of topology was reported by Althaus and Christensen [1]. They had a user guiding a robot through an office environment and assumed an explicit external trigger given by the user, when a new node in the topological graph had to be created. They assumed nodes as rooms that allowed for activities and doorways / gateways as edges. The triggering event had to occur exactly (metrically) where the link between nodes should be placed. The approach presented in this article offers a more lax strategy to supervised map annotation.

The capturing of a complete area as one unit is suggested by Diosi *et al.* [4], who use a watershed implementation after interactively labeling positions that are then related to the areas that include them respectively. Compared to our approach, a clear difference lies in the assumption implicitly understood from Diosi *et al.* that all rooms and other areas have to be specified in one complete tour to avoid merging of too many unlabeled regions into one "room". We consider this a strong limitation. In a pilot study [12] it was observed that potential users do not necessarily describe every room or area to a robot, but pick those that they personally consider important.

A very similar approach to an integrated system for "Human Augmented Mapping" is reported by Zender *et al.* [14]. Their system concentrates though more on the conceptual/semantic level of the mapping process as far as the interaction and higher level functionalities are concerned.

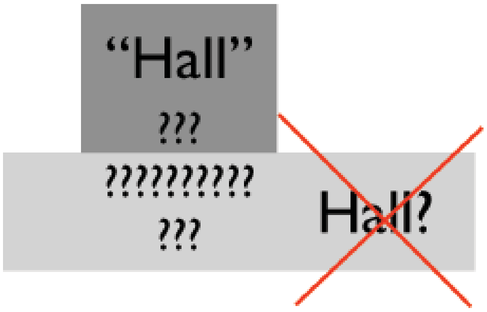

Fig. 2. A illustration of two regions, not clearly separated by a door or other common border, but with distinct meaning to the human user.

On the mapping side the reported system uses the already mentioned method proposed by Mozos *et al.* [9] to categorize the environment into two types of spatial entity: "room" or "corridor", which makes it then possible for the conceptual levels to distinguish between particular rooms. We believe that in a domestic setting it is not always possible to specify areas of interest as particular room. E.g., recently built apartments tend to have rather open layouts, where it still is possible to specify "the kitchen" or the "living room" - possibly with rather vague specifications of the border between them, while those regions do not necessarily correspond to the idea of a closed *room* with a door.

IV. HANDLING STRUCTURAL AMBIGUITIES

As mentioned before we assume the scenario of a guided tour as basis for our work. A human user guides a robot through a well-known environment and specifies particular *regions* (typically rooms and larger distinct areas) and *locations* (particular places where the robot can and should go to perform tasks). Different types of ambiguities can occur during such a tour. In this article we focus on what we term a "structural ambiguity" [12]: parts of the environment that appear quite differently regarding the representation we currently use, and might or might not be parts of the same *region*. As an example can be named the combination of a rather narrow corridor with an entrance hall, where the transition between those two areas is not obviously marked by doors. Fig. 2 illustrates such a situation. A human user might distinguish between "corridor" and "hall" which needs to be reflected in the robot's representation of the environment. Thus, it is necessary to provide the robotic system with means to separate the two relevant regions from each other disregarding the user's mentioning both *actively*. This is also applicable to more obvious region boundaries, where the system still might need a confirmation to the hypothesis of having left a particular – actively specified – region.

A. The generic region

Users do not necessarily mention all possible regions in the environment explicitly during an initial tour. It also happens that particular *locations* are presented to the robot without having presented the surrounding region [12]. It might though still be necessary for the robot to have some representation of the "unspecified" space to be able to handle such locations. We introduce thus a "generic region" into our topological graph representation of the environment that works as a container for everything that does not belong to any specified region. This generic region is different from the other – specified – regions with respect to its topological representation: it does not have any, which means that whenever the robot is not in any of the known and specified regions in the graph it is expected to be located in the generic region.

B. The region representation

In order to detect the transition between two regions, or between a specified region and the generic region a continuous comparison of the – hypothesized – representation of the current surroundings with the representation of the specified region has to be made. We represent specified regions with the help of a number of statistical features computed from a 360° laser range data set [13, for details]:

- Mass m: The free space surrounding the robot
- Length $l1$ and $l2$: The length along the two principal components of the data set
- Angle θ: The angle of the first principal component relative to the origin of the map / the starting point of the tour
- Excentricity e: The excentricity of the ellipse described by the two principal components

As soon as a region is specified it is both made available as a region in the topological graph that models the environment and added to a list of region representations for comparisons. The part of the environment (so far specified locations and navigation nodes, see section V for details) that is covered by the ellipse described by the two principal component vectors is hypothesized as part of the new region and handled accordingly. Within a region all relations between entries are specified relative to the center of the region (the center point of the data set), so that it can be metrically decoupled from the rest of the map. To ensure the link to other regions (including the generic regions), we use a navigation graph with connecting nodes for region transitions. We are aware that the ellipse specified by the principal components is only a very coarse estimate, but proves sufficient as an initial hypothesis, given that our main focus currently is the distinction of regions from each other.

C. Detecting changes while traveling

While traveling the environment the available data sets are continuously used to generate a representation of the current surroundings. This representation is compared to previously acquired ones to decide, if the environment has changed significantly so that it is likely to have entered a new region.

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

One question is to which previously generated representations the current – hypothesized one – should be compared. One option is to compare only to the representation that was last accepted as current one. In this case the system does not make use of previously acquired representations and cannot be used for the recognition of already actively specified regions. Comparing to all available regions to find the most likely current surrounding region is rather expensive. Thus, we use a hybrid approach to deal with this situation. The currently hypothesized new region representation is compared to the previously accepted current one. In the case that a significant difference is detected, the representation is checked against all other available representations if any of them matches sufficiently. If this is the case, the matching region is hypothesized as current representation to compare to in further steps, otherwise a new representation is generated. When a region is specified actively by the user this region is assumed to be the current one immediately.

2) Comparing representations: To decide if two region representations are sufficiently close to each other, we compute a distance measure d from the relative differences in each of our descriptive features:

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

with

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

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

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

D. Updating regions

We consider two ways of updating a region representation, in case that a hypothesis for a new region is erroneous. One option is to compute the region representation as average of all available representations including the new hypothesized one, the other option is to have several different stored representations for each region to choose from. Since a classification performance test for the approach showed that this clustering method performs slightly better [13], it seems the most useful way to proceed for continuous comparisons as well, assumed that not only subsequently generated representations are to be compared.

V. IMPLEMENTATION

Our implementation of the suggested approach to continuous detection of structural changes in an environment is part of an initial implementation of the framework for Human Augmented Mapping [12]. The complete system is able to run as well on recorded data sets (odometry readings and corresponding laser range data sets) as on our robot (a Performance PeopleBot by MobileRobots[1]) in an on-line fashion. While traveling the system incorporates the information obtained from the user into the map representation of the environment. At the same time the continuous detection of structural changes in the environment is run to generate requests for clarification if necessary. We assume thus always the presence of a human user in a guided tour scenario and refer to this situation accordingly in the following.

We consider two types of event relevant to generate a new region representation. One is the – user initiated – specification of a new region, the other is the – data driven and robot initiated – detection of a structural change in the environment that forces the generation of a new representation.

1) Specifying a new region: In this particular case the user triggers the generation of a new region representation. Since our approach to build such a representation relies on a 360° range data set, but our robot only is equipped with one (forward oriented) laser range finder, the robot performs a complete turn to obtain the required data set. This has actually proven useful for the interaction, since it allows the human user to understand that the robot is actually considering the surroundings for its representation of the mentioned region [12].

2) Continuous generation of region hypotheses: It is obviously not possible to constantly rotate the robot while it is traveling. Neither is it comprehensible or helpful for the user if the robot stops at arbitrary positions during the guided tour to obtain a 360° degree data set. Thus, we overcome the problem of only having one laser range finder available on the robot by using virtual scans generated from a local map. This can actually lead to quite high a number of structural changes detected in the starting phase of a tour, since the hypotheses for the virtual backward oriented scan data can only be sensible when the robot has left parts of the measured environment behind. Thus, we only start the comparing process when the robot has traveled a certain distance.

VI. EVALUATION

Since we are interested in actually using the information that can be obtained from human users guiding the robot, we use data sets acquired during user studies [12, and ongoing] to understand, how users guide a robot around and present an environment, and in how far the robot's representation – in this case the structural ambiguity detection – corresponds to the user's understanding. Thus, we discuss our approach here in the context of a number of different data sets, obtained

a) during a user study in our lab environment,
b) as part of a public data set[2] acquired in domestic settings with respect to the "home tour" scenario[3]
c) during an ongoing user study in domestic environments.

In case a) we picked two runs out of five that covered both a large part of the used office floor and a useful number of rooms (*regions*) to get enough transitions. In case b) one of the data sets from a real apartment was picked arbitrarily while for case c) the set was picked from the available ones as a contrast to case b) in terms of room size and layout.

It has to be noted that all those environments are small compared to data sets commonly used for SLAM experiments and comparisons. Since we are in fact interested in our approach as part of an interactive framework, applicable in a limited – mostly domestic – setting, we do not consider this as a limitation.

We evaluated several runs (guided tours) in those different environments with respect to the following criteria:

- Overall number n of detected ambiguities / transitions
- Number $nSens$ of ambiguities detected in a sensible range from an obvious transition in the environment
- Number $nSpurious$ of obviously spurious (erroneous) detections of ambiguities
- Number $nMiss$ of obviously missed transitions into a structurally different area

The values are obviously based on somewhat subjective estimates (is a transition detected correctly when the system reacts about one meter after passing a door or not?), but might give an idea of the validity of our approach. The "sensible range" for $nSens$ (and also for $nMiss$) was in fact estimated to about one meter, and "obviously spurious" detections refer to those occurring in the middle of an empty area inside a *region*. We do not consider the generation of a new specified region as a detected change, but when this specified region was obviously left a detection should occur, otherwise a miss is counted.

A. Lab environment

In the lab environment we evaluated two runs which covered a large part of the corridor and some of the rooms (one office, a meeting room and the kitchen/lunchroom). Fig. 3 illustrates the environment and the results for continuous checking in one of the runs. The part of the corridor covered has a length of approximately 16 meters, the size of the rooms shown to the robot varies from approximately 17m^2 to 25m^2. The robot traveled from the starting point ("SP") along the corridor to the kitchen, was shown around in there and then sent back to a previously specified *location* (marked with "L1"). The robot then was sent back to the kitchen door and afterwards guided past this spot ("KD") into the office (upper right corner of the image), which it then left to return

[1] former ActivMedia Robotics

[2] http://staff.science.uva.nl/∼zivkovic/FS2HSC/dataset.htm

[3] Key experiment 1 of the European Integrated Project "Cogniron", http://www.cogniron.org

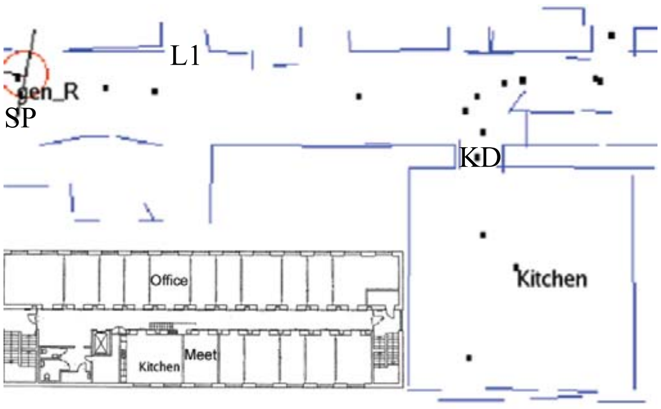

Fig. 3. The part of the lab environment (schematic drawing in the left lower corner) traveled by the robot together with one guiding person. The black dots without label mark the positions where the system decided to generate a new (hypothesized) region. The labeled dots refer to an explicitly specified region's center (Kitchen) and the starting point (gen_R). Already stored regions were not considered – each dot marks a point where the previously accepted representation does not match the current (hypothesized) one sufficiently anymore. A movement threshold of one meter was assumed, the similarity distance threshold was set to 1.5, changes were accepted immediately.

to the starting point ("SP"). Numerous transition detections occurred around the entry to the kitchen, due to the fact that here the layout of the corridor changes, and the robot passed the area four times. Thus, some of those transitions can be seen as confirmations of earlier detected ones, particularly when the robot was reaching the relevant area from different directions. In general most transitions are detected in the vicinity of door passages and other openings. The left most "spurious" detection occurred right after the system started the continuous checking and due to the fact that before the gathered data were not sufficient to form a useful virtual scan to cover the robot's rear "view".

The similarity distance threshold for the runs was (empirically) set to 1.5, the robot had to travel at least 1.0 meter before a new comparison was started and changes were accepted with only one data cycle of occurrence. This last condition resulted in a number of spurious detection but allowed to get a better idea on the general robustness of the approach.

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

B. Domestic environment

Two different domestic environments were considered, one being a rather small apartment (approximately 45m^2) with narrow passages and doorways, the second being a medium sized flat (83m^2) with partially rather wide passages and open spaces. In the latter the hallway (approximately 1.6 meters wide) is opening directly into the living room (ca 25m^2) without any door or other obvious separator between them. In both apartments the living room, the (or one) bedroom and the kitchen were presented to the robot. In the first case (the small apartment) though the tour was simulated just for the purpose of collecting the data, so that spurious detections of transitions are unlikely to occur due to interaction related situations.

The thresholds were set to the same values as for the lab environment, i.e., the similarity distance threshold was set to 1.5, the minimum travel distance was set to 1.0 meter and change detections were accepted after one cycle.

For the small apartment the numbers are very convincing, in two runs an overall number of $n = 18$ ambiguous situations were detected, all of which appeared sensible given the current positioning of the robot. As expected the fact that no spurious detections occurred can be explained with the fact that no human user was actually interacting with the robot. On the other hand it appears that in three situations a change in the environment should have been more obvious and thus should have been detected earlier than this was the case.

In the larger apartment also two runs were conducted, both being "real" guided tours. In both runs $n = 22$ transitions were detected, with $nSens = 20$ (91%) of them appearing sensible regarding the surroundings. Again, also in this apartment in $nMiss = 4$ occasions an obvious change in the environment was not detected. An analysis of the similarity values showed, that differences between region representations seem generally slightly smaller in domestic settings than in a lab environment. Adaptive setting of the threshold values or the application of a more sophisticated change detection filter can be an option to cover such cases more appropriately.

C. Summary

The results from the six evaluated runs show, that most of the obvious transitions (e.g., junctions of hallways, entering a room, hallways opening into a room) in as well a lab environment as in two different domestic environments are quite well detected. Some missing detections in the domestic settings however suggest to consider the application of a more adaptive decision process. We assume though in the general case, that the human user would take care of such situations. If he or she thinks that the robot should be aware of a spatial distinction, a respective region representation would be specified actively according to the information that the user would give.

A number of spurious detections in the lab environment as well as those occurring in one of the domestic settings can be explained with the user being (due to the interactive scenario) very close to the robot and thus covering larger parts of the laser range finder's "field of view". Since we chose for this evaluation to allow a change of the current region representation to be accepted immediately after only one cycle of occurrence, such spurious detections can probably be avoided by applying a higher threshold, e.g., three cycles, here. Another reason for such spurious detections can be the fact that we use virtual range data sets to cover the environment in the robot's back. Those data sets do of course only correspond sufficiently to the actual situation, when the robot has been turned at least slightly to generate an appropriate, roughly consistent local map.

As an overall result we consider our approach a useful tool to support the interaction between robot and user during a guided tour.

VII. CONCLUSIONS AND FUTURE WORKS

In this paper we presented our approach to the detection of structural ambiguities, or region transitions, in indoor environments. We assume an interactive guided tour in which a human user presents and explains a known environment to our robot. We explained, how this detection can be used within our interactive framework for Human Augmented Mapping to disambiguate the situation for the robot.

A. Conclusions

We tested our approach on data sets recorded in different indoor environments on a regular PC. Still, since it is part of our complete implementation of the interactive framework for Human Augmented Mapping [12] it is possible to run the system directly on the robot during a guided tour. This makes our approach to the detection of transitions in indoor environments very flexible. No prior knowledge of spatial categories is needed to use our method. Our tests showed sufficiently good results in both lab and domestic environments. We consider it thus a helpful tool within an interactive setting. Other mentioned approaches aim to label an environment with spatial categories [5], [9], while our method can rather be considered to deal with transitions between any type of spatial categories. This makes it more flexible in situations where the spatial category is difficult to determine even for a human user. Thus, we consider our approach as a fast and easy-to-apply complement to such categorizing methods.

B. Future Works

As already mentioned in section VI it would seem natural to investigate a more adaptive method to decide if in fact a transition has occurred. This should make the method better suitable to different types of environments (generally narrow or more open) without needing to adjust parameters manually. An open issue is to consider the interaction state more directly in the mapping (and transition detection) process to tell user induced ambiguities from structural ones. Further the method has to be integrated fully into our framework to actually investigate the impact on the interaction with a user. This is already ongoing work.

VIII. ACKNOWLEDGMENTS

The work described in this paper was conducted within the EU Integrated Project COGNIRON ('The Cognitive Robot Companion' - www.cogniron.org) and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020. We also thank Patric Jensfelt and John Folkesson for their technical help with parts of the implementation of the complete framework for Human Augmented Mapping.

REFERENCES

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

[2] P. Beeson, N.K. Jong, and B. Kuipers. Towards Autonomous Topological Place Detection Using the Extended Voronoi Graph. *in Proc. of the IEEE International Conference on Robotics and Automation,* April 2005, Barcelona, Spain.

[3] H. Choset and K. Nagatani. Topological Simultaneous Localization and Mapping (SLAM): Toward Exact Localization Without Explicit Localization. *IEEE Transactions on Robotics and Automation,* 17(2), April 2001.

[4] A. Diosi, G. Taylor and L. Kleeman. Interactive SLAM using Laser and Advanced Sonar. *in Proc. of the IEEE International Conference on Robotics and Automation,* April 2005, Barcelona, Spain.

[5] S. Friedman, H. Pasula, and D. Fox. Voronoi Random Fields: Extracting the Topological Structure of Indoor Environments via Place Labeling *in Proc. of the International Joint Conference on Artificial Intelligence (IJCAI),* January 2007, Hyderabad, India.

[6] B. Kröse An Efficient Representation of the Robot's Environment. *in Proc. of Intelligent Autonomous Systems 6,* 2000, Venice, Italy.

[7] G.-J. Kruijff, H. Zender, P. Jensfelt, and H.I. Christensen. Clarification dialogues in human-augmented mapping. *in Proc. of the ACM Conference on Human-Robot Interaction,* March 2006, Salt Lake City, UT, USA.

[8] B. Kuipers and Y.-T. Byun. A robust qualitative method for spatial learning in unknown environments. *in Proc. of the National Conference on Artificial Intelligence (AAAI-88),* 1988, Los Altos, CA, USA.

[9] O.M. Mozos, C. Stachniss, and W. Burgard. Supervised Learning of Places from Range Data using AdaBoost. *in Proc. of the IEEE International Conference on Robotics and Automation,* April 2005, Barcelona, Spain.

[10] I. Nourbakhsh, R. Powers, and S. Birchfield. DERVISH An Office-Navigating Robot. *AI Magazine,* 16(2):53–60, AAAI, 1995.

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

[12] E.A. Topp. Initial Steps Toward Human Augmented Mapping. *Licentiate Thesis,* School of Computer Science and Communication, Royal Institute of Technology (KTH), October 2006, Stockholm, Sweden.

[13] E.A. Topp, and H.I. Christensen. Topological Modelling for Human Augmented Mapping. *in Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* October 2006, Beijing, China.

[14] H. Zender, P. Jensfelt, Ó. M. Mozos, G.-J.M. Kruijff, and W. Burgard. An Integrated Robotic System for Spatial Understanding and Situated Interaction in Indoor Environments. *in Proc. of the 22nd Conference on Artificial Intelligence (AAAI-07), pp. 1584–1589.* July 2007, Vancouver, BC, Canada.