Moving from augmented to interactive mapping

O. Booij, B. Kröse
Intelligent Systems Lab Amsterdam, University of Amsterdam
1098 SM Amsterdam, The Netherlands
Email: obooij@science.uva.nl

J. Peltason, T. Spexard, M. Hanheide
Applied Computer Science, Bielefeld University
33594 Bielefeld, Germany
Email: jpeltaso@techfak.uni-bielefeld.de

I. Introduction and problem statement

Recently[1] there has been a growing interest in human augmented mapping[1, 2]. That is: a mobile robot builds a low level spatial representation of the environment based on its sensor readings while a human provides labels for human concepts, such as rooms, which are then augmented or anchored to this representation or map [3]. Given such an augmented map the robot has the ability to communicate with the human about spatial concepts using the labels that the human understand. For instance, the robot could report it is in the "kitchen", instead of a set Cartesian coordinates which are probably meaningless to the human.

Even if the underlying mapping method is perfect, two main problems occur in the approach of augmented mapping. When guiding a robot through a number of rooms, humans tend to not provide labels for every visited room [4]. The result is that the robot has difficulty to model where one room ends and the other room starts. This problem could be solved by detecting room transitions through the sensor data. Although good attempts using such an approach have been made in office environments [5, 6], applying these to other environments such as real homes is nontrivial. Another problem is that the generalization of the labeled map to newly acquired sensor data can be much different from the humans ideas. That is: there is a mismatch between the human representation and the robots representation. In our case the robots generalizes labels using visual similarities, while humans could use the function of the room. Even among humans there are differences between spatial representations. Think of a living room with an open kitchen. Where does the living room end and the kitchen begin?

Our solution to both of these problems is to use pro-active human robot interaction. We briefly describe how the robot learns a map of the environment using a vision sensor and active dialog with a human guide. The method is implemented on Biron (the Bielefeld Robot Companion) which supports an integrated human robot interaction system based on XCF (XML Communication Framework) complete with person attention, spoken dialog, person following, gesture recognition and localization components [7].

[1]The work described in this paper was conducted within the EU FP6-002020 COGNIRON ("The Cognitive Companion") project.

Fig. 2. Biron and human guide in a home environment.

II. Augmented mapping

A. Appearance based topological mapping

To map the environment we use images taken by an omni-directional vision system. From each image SIFT features are extracted which are used to find image point correspondences between pairs of images by matching their SIFT descriptors. False point correspondences are then removed by imposing the epipolar constraint. By dividing the minimal number of SIFT features of two images i and j by the number of correspondences, one finds a measure for the distance of the two images in appearance space:

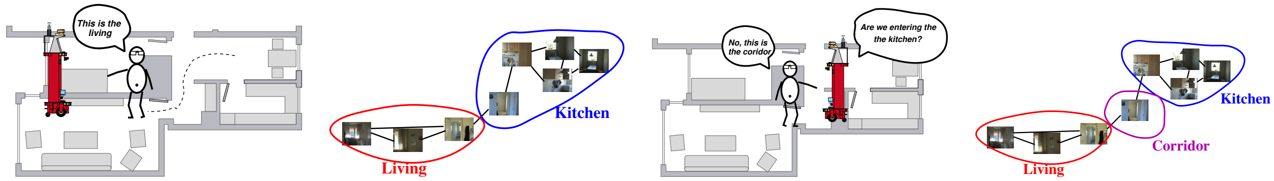
$$d_{ij} = \frac{\min(\#\text{SIFTS}_i, \#\text{SIFTS}_j)}{\#\text{correspondences}_{ij}}$$

These computed distances are put in a graph representation in which the nodes denote the images and distances are put on the links, effectively creating a topological map of the environment. If the distance is above a certain threshold, which was set to 10 in our experiments then no link was created.

The complete map building system is running in real time on one of the robot-laptops, processing around one image per second. To keep the number of comparisons limited we used the Connected Dominating Set method to pick key images from the previous image set. For an in depth treatment of this map building scheme see [8].

B. Human augmentation of room labels

While the robot is driving through the environment following the human guide and building a topological map, room-



(a) (b) (c) (d)

Fig. 1. A sketch of the proposed method. (a) The human guide provides a label. (b) After a second label is provided the map consists of two subgraphs. (c) The robot reports a room transition on which the human provides feedback. (d) The feedback is used to update the map.

labels can be provided to the robot, see Figure I for an example. This is performed by commanding the robot to stop and telling the robot the name of the room it is in, e.g. "This is the kitchen" (see Figure 1(a)). To handle miscommunication, a powerful grounding-based dialog system is used that can handle complex conversational repair behavior and facilitate a smooth conversation (see [2] for more information). The given label is then added to the next node (image) that is added to the map.

Using the given labels and the structure of the graph the robot can partition the map in different subgraphs. Every node is assigned to that label corresponding to the closest labeled node computed with Dijkstra's shorter path algorithm[9] (see Figure 1(b)). Effectively we are exploiting here the fact that images taken in a convex space, which usually correspond to the notion of rooms, are visually much more similar than images taken while the robot moved through a narrow passage, a door.

III. Interactive mapping

As can be seen in Figure 1(b) the transition from the "living room" to the "dining room" is probably not learned in the way the human had in mind when giving the labels. The human would probably not notice this until it would send the robot to the "Living room" after which the robot would move to the hallway. This can easily be solved by making the robot pro-actively interact with the human.

Every time the robot adds a new image to the map it computes its corresponding label. If this label is different than the label of the previously added node, the robot reports this to the human in the form of a question. In the case of Figure 1(c) the robot asked "We just entered the living room, right?". The human now has the opportunity to provide feedback, possibly reducing the mismatch with its own representation, see Figure 1(d). If later the robot would really enter the "living room" it will again report this to the human confirming that it has correctly learned the transition.

A technical detail is that the robot does not stop driving while reporting room change to the human, so to not interrupt the tour. Thus new nodes are added to the graph while it awaits an answer. The possibly corrected label is put on the node which triggered the robot. This could lead to race conditions if there are a lot of transitions close to each other, e.g. if different locations in the room are also labeled. In the conducted experiments, however, we did not experience such problems.

IV. Results

The new interactive mapping approach was recently implemented on the Biron robot. First test trials were performed in a rented apartment at Bielefeld which was furnished to look like a real home environment. See `http://www.science.uva.nl/~obooij/research/mappingHRI/index.html` which features a video shot during one of the trials illustrating the capabilities of the complete interactive mapping system.

The robot captured panoramic images once every 2 seconds and the tour took around 5 minutes resulting in a total set of 158 images. The complete mapping system, including the image processing, is performed during the tour in real-time on one of the laptops attached to the robot.

In Figures 3(a)-(e) the spatial representation is plotted using hand-corrected odometry data. Note, however, that this odometry data was not used by the mapping algorithm.

In Figure 3(a) the robot drove from the living room at the bottom right of the figure through the hallway to the kitchen on the upper left. By then the only label that was given was in the living room, so it groups every new node with that label. In Figure 3(b) it is provided a new label "Dining room" and as can be seen the graph is split into two groups according to their distance over the graph. The cut between these two groups is located somewhere inside the small hallway. This became apparent to the guide in Figure 3(c) where the robot was guided back to the hallway and asked if it reentered the kitchen. After interacting with the guide the label "Hallway" was added to the map, splitting the graph in three parts, see Figure 3(d). After reentering the living room the robot again asked if this was the "Living room" which was confirmed by the guide resulting in another node being labeled. In Figure 3(e) the final spatial representation is shown as build by the robot.

V. Conclusion

We have shown that using relatively simple human robot interaction techniques we can solve two problems apparent

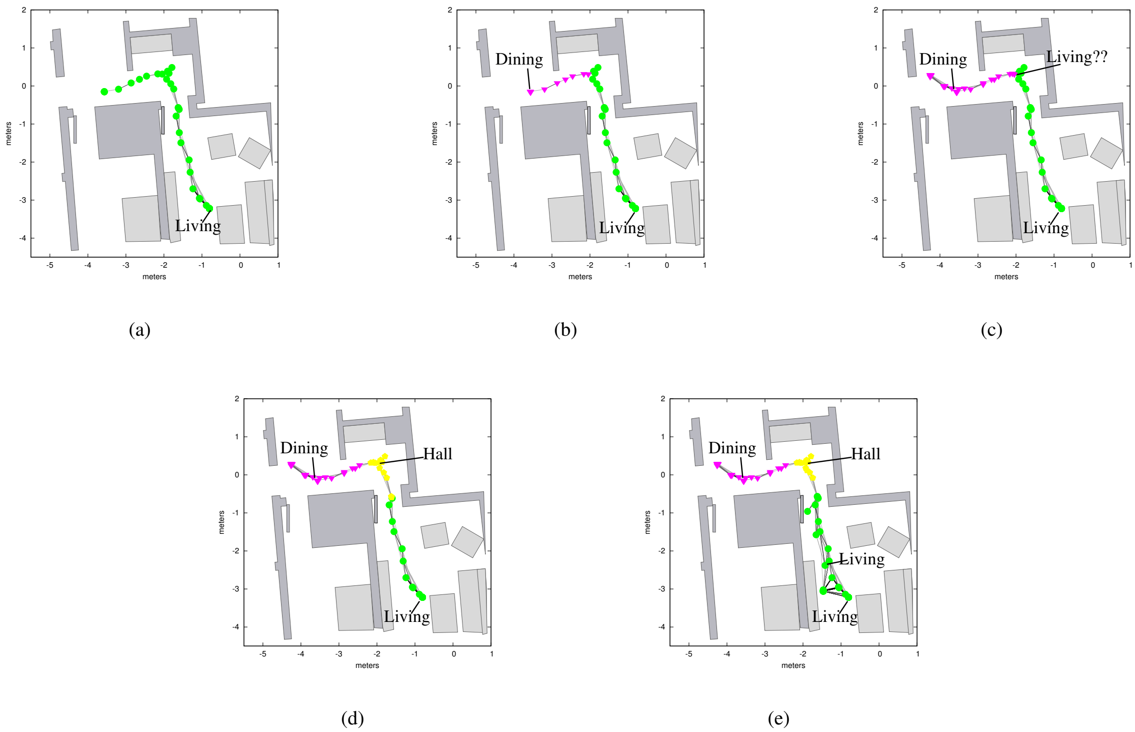


Fig. 3. The spatial representation build by the robot. The different symbols denote nodes (images) of the graph. The lines between the symbols denote links between the lines, with darker colored lines representing links with a smaller distance. Green circles denote nodes belonging to the "Living room", pink squares to the "Dining room" and yellow pentagons to the small "Hallway". Symbols linked with a label represent nodes that were labeled by the guide. In addition part of the ground-truth floor map is plotted on top for reference.

in augmented mapping systems. The robot actively asks the labels of rooms that were not labeled at the first visit and decreases the mismatch between the human representation of room transitions and the robots representation. The complete system can be run in real time on a single laptop and has been shown to work in a real home environment.

Future work is directed to gathering larger evidence for the feasibility of the interactive localization approach. The system scales well to larger environments and is flexible because it uses only a vision sensor.

REFERENCES

[1] G. Taylor A. Diosi and L. Kleeman. Interactive slam using laser and advanced sonar. In *2005 IEEE International Conf. on Robotics and Automation, ICRA 2005*, Barcelona, Spain, April 2005.

[2] Thorsten Spexard, Shuyin Li, Britta Wrede, Jannik Fritsch, Gerhard Sagerer, Olaf Booij, Zoran Zivkovic, Bas Terwijn, and Ben Kröse. Biron, where are you? - enabling a robot to learn new places in a real home environment by integrating spoken dialog and visual localization. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*. IEEE, October 2006.

[3] A. Saffiotti and K. LeBlanc. Active perceptual anchoring of robot behavior in a dynamic environment. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 3796–3802, San Francisco, CA, 2000. Online at http://www.aass.oru.se/˜asaffio/.

[4] E.A. Topp, H. Huettenrauch, H.I. Christensen, and K. Severinson Eklundh. Bringing together human and robotic environment representations - a pilot study. In *In Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006.

[5] Hendrik Zender, Oscar Martinez Mozos, Patric Jensfelt, Geert-Jan M. Kruijff, and Wolfram Burgard. Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems*, 56(6):493–502, June 2008.

[6] O. Martínez-Mozos, R. Triebela, P. Jensfeltb, A. Rottmanna, and W. Burgarda. Supervised semantic labeling of places using information extracted from sensor data. *Robotics and Autonomous Systems*, 55(5):391–402, 2007.

[7] J. Fritsch, M. Kleinehagenbrock, A. Haasch, S. Wrede, and G. Sagerer. A flexible infrastructure for the development of a robot companion with extensible HRI-capabilities. In *Proc. IEEE Int. Conf. on Robotics and Automation*, pages 3419–3425, Barcelona, Spain, April 2005.

[8] O. Booij, Z. Zivkovic, and B. Kröse. Sparse appearance based modeling for robot localization. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 1510–1515. IEEE, October 2006.

[9] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, December 1959.