BIRON, where are you?

Enabling a robot to learn new places in a real home environment by integrating spoken dialog and visual localization

Thorsten Spexard, Shuyin Li,
Britta Wrede, Jannik Fritsch and Gerhard Sagerer
Applied Computer Science, Faculty of Technology
Bielefeld University
33594 Bielefeld, Germany
Email: tspexard@techfak.uni-bielefeld.de

Olaf Booij, Zoran Zivkovic,
Bas Terwijn and Ben Kröse
Informatics Institute, Faculty of Science
University of Amsterdam
1098 SM Amsterdam, The Netherlands
Email: obooij@science.uva.nl

Abstract— **An ambitious goal in modern robotic science is to build mobile robots that are able to interact as companions in real world environments. Especially for caretaking of elderly people a system robustly working at private homes is essential, requiring a very natural and human oriented way of communication. Since home environments are usually very individual a first task for a newly acquired robot is to get familiar with its new environment. This paper gives a short overview on how we integrated a vision based localization using the advantages of a very modular architecture and extending a spoken dialog system for online labeling and interaction about different locations. We present results from the integrated system working in a real, fully furnished home environment where it was able to learn the names of different rooms. This system enables us to perform real user studies in future without the need to fall back to Wizard-of-Oz experiments. Ongoing work aims at enabling the robot to take initiative by asking for unknown locations. A future extension is the ability to generalize over features of known rooms to make predictions when encountering unknown rooms.**

I. Introduction

Based on the observation that robots ‘living’ in home environments need to be ‘socially aware’, there has been a trend in robotics research to develop features facilitating social behavior of robot companions. However, the next step that needs to be taken now is to actually move robots out of the lab into real home environments.

It has been shown [1] that for an improved acceptance of assistive robotic products, the functionality needs to move beyond task-based interactions, and systems need to be attractive, affordable, and (especially in care applications) be non-stigmatizing. From the user interaction perspective, perceived social intelligence is more important than artificial intelligence.

In the context of the European project ‘The Cognitive Robot Companion’ (COGNIRON, see [2]) we work on a robot ’companion’ that is able to learn new skills and grow its capacities in constant interaction with humans. Learning conceptual representations of space and objects is an important basis in order to enable a grounded interaction between the user and the robot. In this paper we describe our results in building and testing a system that can learn representations of space through verbal interaction with humans.

An important aspect for such a system is its architecture. One of the goals of an architecture is to have a description of the system from a functional view: how should the dialog system be integrated to support learning. The other goal of an architecture is a system description in modules which can be implemented separately. Since COGNIRON is an ’Integrated Project’, the development of software modules for different tasks is carried out by multiple developers. Therefore, software tools and standardization have to be used to support the developers. With the right approach, integrating the different software components is a matter of ”configuration” rather than programming.

Furthermore, the goal of our research activities is not only developing new algorithms and pieces of software capable of performing a certain task, but also applying them to the real world or scenarios as close as possible to reality. In order to test our system we define a scenario, the so-called Robot Home Tour. In this scenario a person without any knowledge of robotics has just bought a robot companion and gives it a tour through its private home so to familiarize it with its new habitat. The human points to and names locations and objects which she believes are necessary for the robot to remember. The robot should have strong human robot interaction capabilities in order to understand and interact with the human guide as well as robust mapping and localization methods to build a representation of the totally new environment. For such a system it is important to be evaluated not only in a lab but in a real home environment in order to perform real-world user studies with an autonomously working system.

In this paper we briefly describe related work (section II), the methods for map-building and dialog, and their integration into the system architecture (section III). Finally we present results from tests in a real home environment performed within the home tour scenario (section IV).

II. Related Work

For a long time roboticists and researchers in Artificial Intelligence have been developing mobile robots that can interact with humans. At first the robots were specifically designed to operate in office environments and interact with humans with a background in robotics. The Jijo-2 robot [3] for example was developed to carry out fetch and carry tasks and guide people to locations in an office environment. The robot was equipped with a dialog system designed to communicate in the Japanese language. The navigation system relied heavily on the speech input from the human and did not use any other sensors except odometry. Also a localization system was implemented using an omnidirectional camera, but the degree of integration with the other systems is unclear.

In recent years robots are put in less controlled environments with a lot of people who are unfamiliar with robots. For example, the robot Robox [4] was developed to give tours to visitors at crowded places and was tested in an exhibition involving Robotics. Another project was the museum tour-guide robot Rhino [5], which was deployed in the museum of Bonn. Both robots could use speech to communicate with the visitors, although the dialog was limited with little to no interaction. Also both robots were provided with an accurate metric map, so mapping the environment was not considered an issue.

In [6] a study is described that aimed at developing a domestic robot which is able to naturally communicate with humans in their homes. The resulting robot Lino did have a dialog system as well as a navigation and localization system, however these functionalities were not integrated. Popular robot systems developed for the entertainment industry such as AIBO and PARO, are also meant for living in close relation with humans (see e.g. [7]). None of these however have both speech and localization systems built in.

Although the objective of various studies is to develop a robot that can aid people in their own homes, very few experiments are conducted in a real household.

III. Implementation

Since acting in a real household implies many different abilities for a robot, a modular software design was chosen, consisting of individual modules running separately. In this section we describe the overall architecture as well as the Localization and Dialog modules, before explaining how they are integrated.

A. Architecture

A three layer architecture [8] was chosen consisting of a deliberative, an intermediate, and a reactive layer (s. Fig. 1). The dialog system for complex user interaction is located in the top deliberative layer and in the bottom layer reactive modules capable of adapting to sudden changes in the environment are placed. Since neither the deliberative layer dominates the reactive layer nor the reactive layer dominates the deliberative one, a module called Execution Supervisor (ESV) was developed [9] located in the intermediate layer, where the Localization and its knowledge base are also positioned. The ESV coordinates the different tasks of the individual modules by reconfiguring the parameters of each module. For example, the Player Interface for controlling the hardware is configured to receive movement commands from different modules.

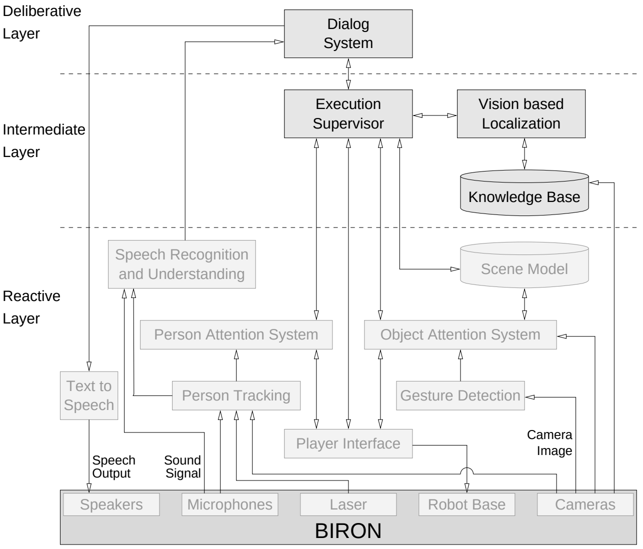


Fig. 1. Integration of Localization into the HRI architecture of BIRON.

The ESV can be described as a finite state machine. The different HRI abilities are represented as states and a message sent from a module to the ESV can result in a transition from state A to state B. For each transition the modules in the different layers are reconfigured. Additionally to the new configuration, data like movement speed is exchanged between the modules. All data exchange via the ESV is based on the XML based Communication Framework (XCF, see [10]) using four predefined XML structures:

Event: data sent from modules to ESV
Status: data sent from ESV to deliberative modules
Order: data sent from ESV to non deliberative modules
Reject: data is not accepted by ESV since it is too old or the robot is not in the appropriate state

All data exchange between the ESV and each module are automatically established after reading a configuration file. This file also contains the definition of the finite state machine and the transitions that can be performed. This makes the system easily extendable for new HRI capabilities, by simply changing the configuration file for adding new components like the Localization without changing one line of source code.

A former implementation already contained modules for multiple person tracking with attention control [11] and an object attention system [12] based on deictic gestures for learning new objects and storing them in a scene model. Here we detail how the human-robot interaction is extended for teaching the robot location names and retrieving location names. To achieve this functionality, the dialog is extended and a vision-based localization module is added to the overall architecture as depicted in Fig. 1. This enables the robot to learn names of new locations during the interaction with a human to generate a human-augmented spatial representation of the home environment.

B. Map-building and localization

Map building and localization are crucial functionalities for mobile robots. The most commonly used approach is the landmark-based SLAM method, which uses a recursive filter, such as the Extended Kalman Filter to build a metric map of the environment. In order to make a precise and consistent map, data from different sensors is integrated, such as odometry, vision and laser [5]. Another approach of modeling the environment is to construct an appearance based map. In such a model the sensor readings taken at different locations are not merged but kept separate as characterization of the locations. In the following we will describe how we build such a map and how we use it for localization. For a more complete description see [13].

In the exploration phase the robot collects a series of sensor-readings, which in our case are panoramic images obtained from an omnidirectional camera which is mounted on the top of the robot [14], see figure 3 for the mounting and figure 6 for some example images. From every sensor reading a feature vector is extracted, consisting of a set of scale invariant local image features (SIFT) per image [15] which is stored in the Knowledge Base. If the human guide stops the robot and provides the robot with the name of the location (see Section IV-B), the next set of features is augmented with that label.

In an off-line map building phase, all the feature vectors are compared pairwise. Various distance metrics can be chosen to calculate an affinity between two images. In this study we will compare them by finding corresponding SIFT-features and imposing the epipolar constraint on the corresponding image-locations, see [16], [13], [17]. The minimum number of corresponding image points needed for computing the relative position between two images, given that the robot moves on a plane, is 4. Thus if 4 or more features are constrained by the epipolar constraint, we state that the two images match.

The information obtained by matching all the feature vectors pairwise is then put into a graph. The nodes of the graph represent the sensor readings and the links between the nodes indicate if two sensor readings are in some way close to each other, given the matching criteria that was used. This 'appearance based' graph contains, in a natural way, the information about how the space in an indoor environment is separated by the walls and other barriers. Images from a convex space, like a room, will have many connections between them and just a few connections to some images that are from neighboring space, such as a corridor. To find the links that connect different highly connected subgraphs we apply a normalized graph cut approximation algorithm originating from graph theory [18]. In this way we obtain clusters of the sensor-readings, that correspond with the convex spaces that make up the environment. For the used clustering algorithm the number of clusters must be known or selected using some graph theoretic criteria [19]. We will assume that every convex space is labeled exactly once, so we set the number of clusters to the number of labels that were given.

Now we can use the labels given by the tour guide during the exploration to label these clusters. In the ideal case the labeled sensor-readings all belong to a separate cluster. Every cluster then gets labeled according to the name of its single labeled sensor-reading. However, due to the mismatch in human conceived spaces and the spaces that result from the automatic clustering, it may occur that some clusters bear no labeled node, while others may have multiple labeled nodes. We currently resolve this in the following manner. Clusters with multiple labels are again clustered into two group. This is repeated as long as there are clusters with multiple labels. Clusters with no labeled node, will not be labeled at all. In the future we hope to develop methods to solve these ambiguities by interaction with the human guide.

The robot can use the clustered appearance based graph, augmented with human given labels, to localize itself in the environment. Because the sensor readings are retained in this representation localization is straight-forward: the robot takes a new sensor reading and matches it with the sensor readings in the representation. The label that is associated with the cluster to which the sensor readings belong corresponds to the current location. If the matching nodes come from different clusters, for example if the robot is on the border of two or more rooms, we find the node with the largest amount of matching neighbors, according to the graph, and use the label associated with the cluster of that node. If the current image did not match any in the graph or the found cluster has no label then the robot can not localize itself.

C. Dialog

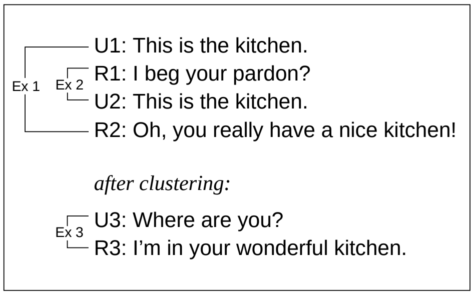
To provide labels for and retrieve position data from the Localization a spoken dialog system is used. In general a dialog system is responsible for carrying out interactions with the user including transferring user commands to the robot control system and reporting task execution results to the user. During the conversation a dialog system should be able to regulate the initiative distribution, handle miscommunication, draw inferences between interlocutors' contributions and organize and maintain the discourse. To enable these abilities we implemented a powerful grounding-based dialog model for BIRON.

Clark [20] proposed the notion of grounding: during a conversation the interlocutors need to coordinate their mental states based on their mutual understanding about the current intentions, goals and tasks. He termed this process as 'grounding'. Furthermore, a speaker can only be sure that her account (presentation) was fully understood if her interlocutor provides some evidence of understanding (acceptance), i.e., if the 'common ground' is available. Only then will the speaker be willing to proceed to another account. Our dialog system is based on this idea. We represent interlocutors' contributions as exchanges, i.e., pairs of contributions. They achieve the state 'grounded' only if the acceptance of the presentation is available which depends on the communication success (e.g., if the speech input is clearly understood) and the robot task execution status. These exchanges are organized in a stack which represents the ungrounded discourse up to the current state. The grounding status of the whole stack is dependent on the status of the individual exchanges and the relations between them. We introduced 4 types of such relations (*default*, *support*, *correct* and *delete*) and they can also have local effects on their previous exchanges. According to the execution results of the robot control system the dialog system formulates contributions for the robot. Each contribution of both the user and the robot is categorized in terms of its roles, i.e., if it initiates an exchange of a certain relation to the previous exchange or if it is the acceptance of an existing one. According to this role, either a new exchange is pushed onto the stack or an old one (or a group of old ones) is popped because it reaches the status 'grounded'. All the popped exchanges are collected into a vector which records the complete dialog history. We thus model the grounding process using an augmented push-down automaton which exhibits local flexibility in contrast to conventional approaches ([21], [22]). The implemented system enables a mixed-initiative dialog style and a well-organized discourse maintaining mechanism. It can also handle complex conversational repair behavior and facilitate a smooth conversation.

D. Integration

To start a conversation and processing a command by the Dialog the user needs to be tracked by the system and speech processing has to be activated based on a person looking at the robot while speaking. This avoids the robot reacting to sound sources like a radio or a TV. Speech understanding will interprete the request, sending a semantic representation of the user utterance to the Dialog. If the utterance is identified as being Localization related, e.g. because it contains the name of a location and a deictic reference, it will be forwarded to the Localization by the ESV. This way the Localization is activated to learn new location names. When the Dialog processes a query for a room the Localization will provide the name of the location if known. The result will be handed to the Dialog which generates a verbal response. In future, the Localization could inform the Dialog in a proactive interaction via the ESV that a recent location is unknown, causing the Dialog to ask the user for the name of the location. In order to allow these different types of interactions, both modules send and receive information to and from the Execution Supervisor using the data structures as described above.

In the following we explain how the dialog system works with the example of a localization-related task (as shown in Fig. 2). The user starts by naming the location (U1) but is not understood possibly due to speech recognition problems. This presentation creates a new exchange Ex 1. Without the need to consult the rest of the robot system, the dialog system immediately starts a clarification question (R1), i.e., it creates a new exchange Ex 2 with the grounding relation 'support' to Ex 1. When the user answers the robot's question (U2) Ex 2 is popped from the stack as it is now grounded. Since Ex 2 has a *support* relation to the previous not-understood exchange Ex 1, Ex 1 is updated with the newly collected information that the user names the location with 'kitchen'.



U1: This is the kitchen.
R1: I beg your pardon?
U2: This is the kitchen.
R2: Oh, you really have a nice kitchen!

after clustering:

U3: Where are you?
R3: I'm in your wonderful kitchen.

Fig. 2. Dialog example for localization (U: User, R: Robot)

The dialog system then tries to provide acceptance for Ex 1 by sending the command `SetLocName` with the parameter 'kitchen' to the Localization. Once the dialog system receives a positive result about the successful operation `SetLocName:kitchen` the status of Ex 1 is changed to *grounded* and Ex 1 is popped from the stack with an acceptance being issued to the user (R2). The stack, i.e., the currently ungrounded discourse, is now empty.

After the offline-clustering process of the Localization, the robot is able to answer questions like "Where are you?" (U3) which creates a new exchange Ex 3. To provide the acceptance for this exchange, the dialog system sends the command `GetLocName` to the Localization which then successfully delivers the name of the location 'kitchen'. Thus, the dialog system can ground the current Ex 3 and pop it from the stack while informing the user about the current location (R3).

IV. Experiment - BIRON @ Home

After integrating and testing the new components successfully at the laboratory, the overall system including Localization and Dialog were used in a real, less structured home environment. To perform the home tour scenario as described in Section I the mobile robot BIRON was used as a demonstrator.

A. Hardware

BIRON is based on the Pioneer PeopleBot from ActiveMedia. The platform is equipped with several sensors to obtain information of the environment and the surrounding humans. A pan-tilt color camera is mounted on top of the robot for acquiring images of objects and the upper body part of humans interacting with the robot. Behind the pan-tilt unit the omnidirectional camera for localization is positioned (see Fig. 3). Two farfield microphones are located at the front of the upper platform, right below

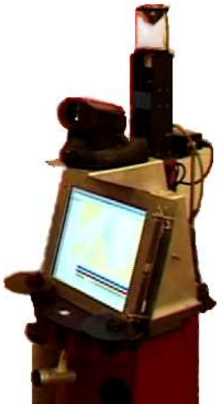


Fig. 3. Camera setup.

a touch screen display, to localize sound sources. Below the microphones, an iSight firewire camera for detecting deictic gestures can be found. A SICK laser range finder is mounted at the front on the base platform. All software components are running on a network of distributed computers. The onboard PC in the robot's base is used for controlling the drive and the on-board sensors as well as for sound localization. An additional PC inside the robot's upper extension is used for person tracking and person attention.

The two on-board PCs are linked by Fast-Ethernet to a router with wireless LAN. Three additional laptops are linked to the on-board PCs via the router. One laptop is used for gesture, object, and face recognition. The second one is used for localization and the third laptop performs speech processing and is linked via wireless connection.

B. Performing the Home Tour

Equipped with this hardware, BIRON was taken to a real and fully furnished home environment, much more unstructured than common laboratory surroundings, including plants, pictures, wooden desks and cupboards. The robot was guided from the hallway to the office, returning to the hall way a second time to proceed to the living room (see Fig. 4). Since the clustering for learning new locations (see Section III-B) had to be done separately, the home tour was divided into two subtasks: first, mapping new locations and second asking the robot about its position after the offline clustering.

Beginning with the mapping, the home tour started at the office, where the user asked the robot about its position. Since the location was not introduced to the robot the Localization returned *unknown* triggering the Dialog to generate the speech output:“Well, I don't know where I am.” The user subsequently told the robot the name of its recent position which causes the Dialog to send the label of this room to the Localization and to generate a confirmation answer. The user, continuously tracked by the person tracking and attention system [11] proceeded the tour alternately asking the robot to follow her and telling it the names of new rooms using the Dialog (see Fig. 5). During the whole tour the Localization was periodically taking pictures of the surrounding every 2 seconds using the omnidirectional camera. After stopping the tour at the living room and clustering the data, a second tour starting at the living room was performed, now the user asked the robot about its position at different rooms. For both, naming the rooms and asking for locations, a position close to the middle of that space was chosen for a better view of the omnidirectional camera, but no fixed positions were used. Videos of the different interactions in the home environment are available at [23].

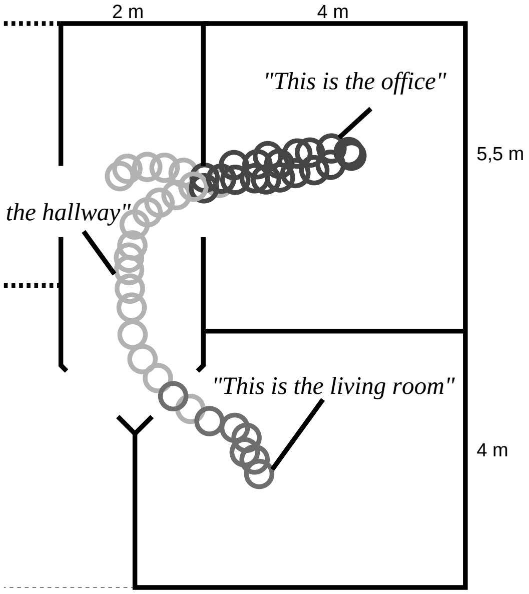


Fig. 4. Floor plan of the flat the home tour was performed at, containing the robot positions and results of the Localization.

Fig. 5. Interacting with BIRON at a real home environment.

C. Results

The robot successfully completed the home tour, while acquiring a map of the environment. The interaction with the robot, naming the places and asking the robot its location, went smoothly due to the Dialog system. We will now give some more detailed results for the mapping and localization system.

During the home tour 253 omnidirectional images were shot. In figure 4 the approximate positions from where the images were taken are sketched. The images of the two rooms are feature rich, while the hallway is quite plain, as can be seen in some panoramic images in figure 6. In some images the view is blocked by persons walking through the environment in addition to the tour guide standing in front of the robot. As described in the Section III-B, the images were matched pairwise to construct a graph. Because we do not make use of odometry it is hard to visualize this graph. In the upper part of figure 7 the graph is visualized by a ‘connectivity matrix’, which depicts for every pair of images if they do or do not match. The big black squares that are visible indicate

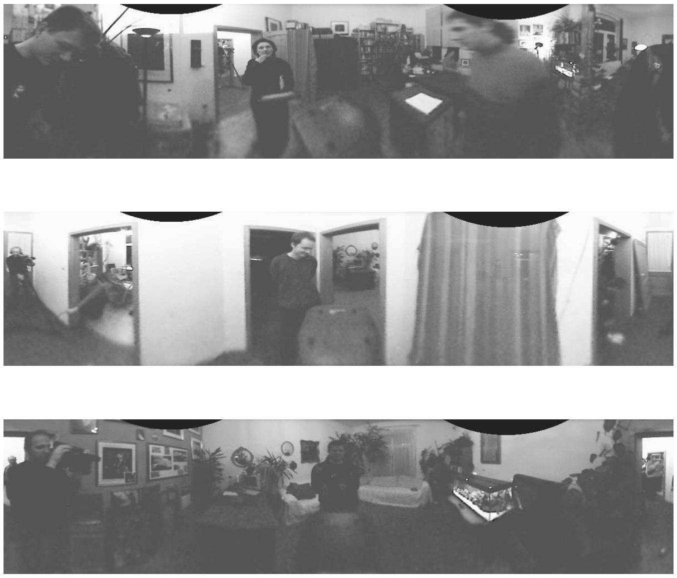


Fig. 6. Panoramic images as taken by the omnidirectional camera of the hallway, the office and the living room (from top to bottom).

large subsets of the images, which all match with each other, because they look alike. The biggest square corresponds to the matching images taken from the office and the square at the lower right to the living room. The first thirty images were taken from the hallway as well as the images 155 to 170, when the robot reentered it after leaving the office (see also figure 4). This can be seen in the connectivity matrix, where the images match. Some of the images falsely matched, because of persons blocking the view, or sensor noise, as is visible in the connectivity matrix by the white areas with the sparse dots.

Despite the false matches, the graph was successfully clustered into three separate subgraphs which corresponded to the three spaces. Because every space was given a label, each cluster contained exactly one node that was labeled and could thus be labeled accordingly. The lower part of figure 7 shows to which labels the images were clustered. This is also visualized in Fig. 4, where the image positions of different clusters have a different color. The algorithm was sometimes indecisive at the borders between two spaces, e.g., when the robot was entering the office, some images in the hallway near the doorpost were clustered in the office and vise versa. This however is reasonable given the fact that a large portion of the room is already visible when standing near to it.

To test the localization method it had to localize 300 images taken from a test run given the labeled appearance based map. The test images were annotated by hand, which was fairly easy because the spaces were clearly separated by the doorposts. An example of a test image that matches with a training image is given in Fig. 8. Of all test images 10 were misclassified, from which 3 were close to a border, and 1 image was not localized at all. All the misclassified images were shot from the hallway, so the errors are probably due to the lack of features.

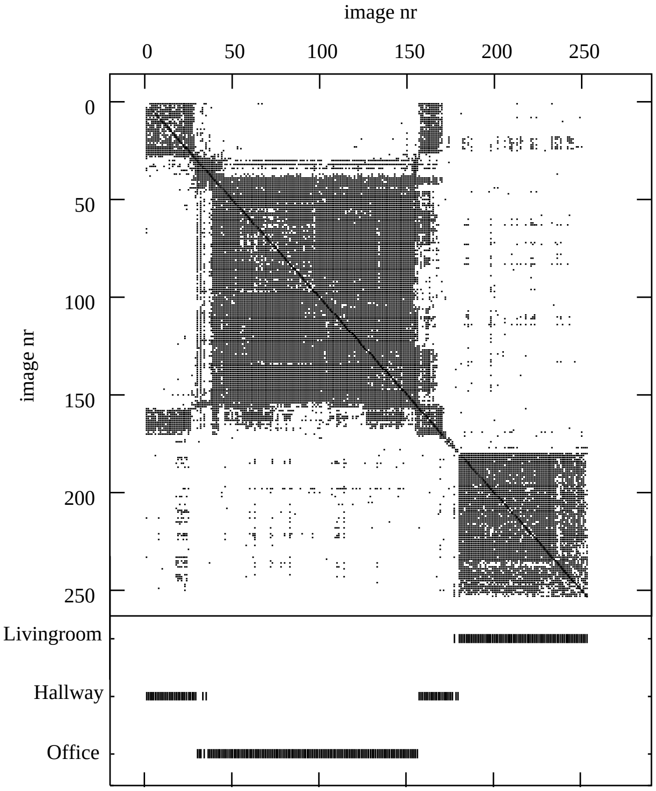


Fig. 7. The connectivity graph and the clustering results shown for the images taken during the home tour.

V. Conclusion

In this paper we presented the results of integrating Localization and Dialog on a mobile robot platform enabling the robot to interact with a human via speech in an unknown home environment. We described an approach for estimating the recent position using only images of an omnidirectional camera. A spoken dialog system for human oriented communication is used for setting labels to newly learned locations and generating a verbal output according to the position data delivered by the Localization. The integrated system was tested in a real home environment successfully performing both learning new locations and identifying them after an clustering phase.

By creating a robot that can operate in real home environment, we have a testbed that serves as a basis for more complex experiments and user studies referring to different scientific issues. Taking advantage of the integrated system, it is now possible to use the different kinds of knowledge within the system to improve the interaction quality. For example, as a next step we will enable the robot to take the initiative, by asking for the names of unknown places or stating which room it enters by itself. This proactive behavior increases its human-robot interaction capabilities and thus the social acceptance.

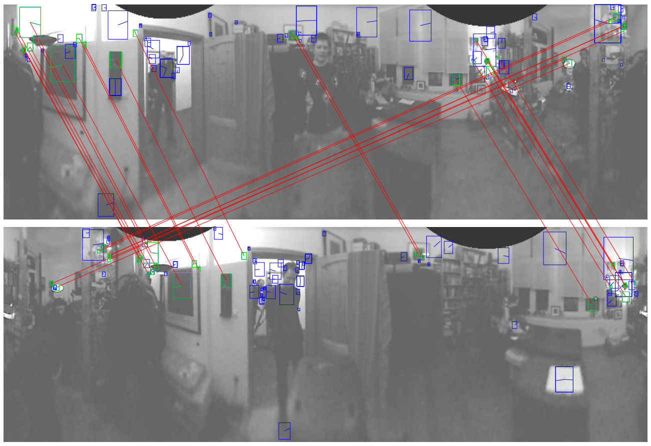


Fig. 8. Matching a test image with a training image. The lines indicate matching image features.

Acknowledgment

This work has been supported by the European Union within the ‘Cognitive Robot Companion’ (COGNIRON) project (FP6-IST-002020) [2].

The authors want to thank the inhabitants of the flat where the home tour was performed.

References

[1] J. Forlizzi, C. DiSalvo, and F. Gemperle, “Assistive robotics and an ecology of elders living independently in their homes,” *Journal of HCI Special Issue on Human-Robot Interaction*, vol. 19, no. 1/2, pp. 25–59, January 2004.

[2] COGNIRON, “The cognitive robot companion,” http://www.cogniron.org.

[3] H. Asoh, N. Vlassis, Y. Motomura, F. Asano, I. Hara, S. Hayamizu, K. Itou, T. Kurita, T. Matsui, R. Bunschoten, and B. Kröse, “Jijo-2: An office robot that communicates and learns,” *IEEE Intelligent Systems*, vol. 16, no. 5, pp. 46–55, Sep/Oct 2001.

[4] R. Siegwart and et al., “Robox at expo.02: A large scale installation of personal robots,” *Special issue on Socially Interactive Robots, Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 203–222, 2003.

[5] W. Burgard, A. Cremers, D. Fox, D. Hhnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, “The interactive museum tour-guide robot,” in *Proc. of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, 1998.

[6] B. Kröse, J. Porta, K. Crucq, A. van Breemen, M. Nuttin, and E. Demeester, “Lino, the user-interface robot,” in *Proceedings of the First European Symposium on Ambience Intelligence (EUSAI)*, E. Aarts, R. Collier, E. van Loenen, and B. Ruyter, Eds. Eindhoven, The Netherlands: Springer, Nov. 2003, pp. 264–274, iSBN 3-540-20418-0.

[7] H. H. Lund and J. Nielsen, “An edutainment robotics survey,” in *In Proceedings of the Third International Symposium on Human and Artificial Intelligence Systems: The Dynamic Systems Approach for Embodiment and Sociality*, Fukui, Dec 2002.

[8] 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*, Barcelona, Spain, April 2005, pp. 3419–3425.

[9] M. Kleinehagenbrock, J. Fritsch, and G. Sagerer, “Supporting advanced interaction capabilities on a mobile robot with a flexible control system,” in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, vol. 3, Sendai, Japan, September/October 2004, pp. 3649–3655.

[10] S. Wrede, J. Fritsch, C. Bauckhage, and G. Sagerer, “An XML Based Framework for Cognitive Vision Architectures,” in *Proc. Int. Conf. on Pattern Recognition*, no. 1, 2004, pp. 757–760.

[11] J. Fritsch, M. Kleinehagenbrock, S. Lang, G. A. Fink, and G. Sagerer, “Audiovisual person tracking with a mobile robot,” in *Proc. Int. Conf. on Intelligent Autonomous Systems*, F. Groen, N. Amato, A. Bonarini, E. Yoshida, and B. Kröse, Eds. Amsterdam: IOS Press, March 2004, pp. 898–906.

[12] A. Haasch, N. Hofemann, J. Fritsch, and G. Sagerer, “A multi-modal object attention system for a mobile robot,” in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*. Edmonton, Alberta, Canada: IEEE, August 2005, pp. 1499–1504.

[13] Z. Zivkovic, B. Bakker, and B. Kröse, “Hierarchical map building using visual landmarks and geometric constraints,” in *Intl. Conf. on Intelligent Robotics and Systems*. Edmundton, Canada: IEEE/JRS, August 2005.

[14] Z. Zivkovic and O. Booij, “How did we built our hyperbolic mirror omnidirectional camera - practical issues and basic geometry,” University of Amsterdam, Tech. Rep. IAS-UVA-05-04, 2005.

[15] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[16] J. Kosecká, F. Li, and X. Yang, “Global localization and relative positioning based on scale-invariant keypoints.” *Robotics and Autonomous Systems*, vol. 52, no. 1, pp. 27–38, 2005.

[17] F. Schaffalitzky and A. Zisserman, “Multi-view matching for unordered image sets, or “How do I organize my holiday snaps?”,” in *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, vol. 1. Springer-Verlag, 2002, pp. 414–431.

[18] J.Shi and J.Malik, “Normalized cuts and image segmentation,” *IEEE Transactions on Pattern Anlysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–904, 2000.

[19] Z. Zivkovic, B. Bakker, and B. Kröse, “Hierarchical map building and planning based on graph partitioning,” in *Proc. IEEE Int. Conf. on Robotics and Automation*, 2006, to appear.

[20] H. H. Clark, Ed., *Arenas of Language Use*. University of Chicago Press, 1992.

[21] D. Traum, “A computational theory of grounding in natural language conversation,” Ph.D. dissertation, University of Rochester, 1994.

[22] J. E. Cahn and S. E. Brennan, “A psychological model of grounding and repair in dialog,” in *Proc. Fall 1999 AAAI Symposium on Psychological Models of Communication in Collaborative Systems*, 1999.

[23] Applied Computer Science - Bielefeld University, “Home tour videos,” http://www.techfak.uni-bielefeld.de/ags/ai/projects/BIRON/.