Assistance Robotics: A Survival Guide for Real World Scenarios

Hans-Joachim Böhme, Sven Hellbach, Frank Bahrmann, Marc Donner, Johannes Fonfara, Marian Himstedt, Mathias Klingner, Peter Poschmann, Mathias Rudolph, and Richard Schmidt*

University of Applied Sciences Dresden, Artificial Intelligence Lab
boehme@htw-dresden.de
http://www.htw-dresden.de/

1 Introduction

The research goal of the Artificial Intelligence Lab at the University of Applied Sciences Dresden is the development of intelligent and interactive mobile service and assistance systems. For the application of such systems, different scenarios are imaginable, like guidance and information systems within public institutions (museums, airports, train stations, etc.), assistance or supporting systems in markets (do-it-yourself store, shopping malls,etc.), systems to support the elderly in home environments, and systems for industrial tasks (mobile measuring and surveillance systems).

Fig. 1. One of our robots.

For the current research work the robot, which is depicted in Fig. 1, is used as a guidance system in a museum in the city of Dresden (Technische Sammlungen Dresden). The research platform is a commercially available SCITOS G5 equipped with two laser range finders, an omnidirectional camera, a depth camera, a ring of sonar sensors, and a head with two eyes that has no further sensor capabilities, but is used for interaction purposes. The robot is meant to serve as companion or tutor, which guides through the exhibition and provides background information to the exhibits. To start the dialog with the visitors of the exhibition, the robot possesses a broad spectrum of multimodal abilities or channels, like language, gestures, a touch screen and multimedia presentations. Furthermore, the system is meant to be able to adapt the way those different channels are combined during the dialog to the specific needs and preferences of its interaction partner.

In addition to the research and development of a dialog system, our robot needs to be able to navigate the exhibition autonomously. However, our intention is not to provide help by augmenting the environment with additional sensors. Instead, the robot is equipped with a number of sensors as already mentioned. Only these sensors are used to generate an inner map of the robots surroundings.

* This work was supported in part by ESF grands number 100071902 and 100076162, as well as by SMWK grand 4-7531.60-02-5120-11/5

S. Wölfl (Ed.): Poster and Demo Track of the 35th German Conference on Artificial Intelligence (KI-2012), pp. 8-11, 2012.

In particular in an environment like the museum, which contains irreplaceable exhibits, a reliable, collision-free navigation is essential. A further difficulty for the navigation task are visitors, which in fact are dynamical obstacles that need to be avoided. The obstacle avoidance should happen in a socially acceptable way, so that the visitors are not frustrated or even frightened.

2 Ongoing Research Work

For our application, a lot of subsystems have to work in conjunction, in particular navigation (mapping, localization, obstacle avoidance) and interaction (people detection and tracking, speech recognition). The remainder of this paper gives a short overview of the researched and applied system components.

Navigation and Obstacle Avoidance: For being able to navigate autonomously within the exhibition, we use a navigation system based on Monte Carlo Localization using a laser range finder that is supported by a 3D obstacle segmentation system. Our navigation strategy is based on Fox et al. [1], who introduced the Dynamic Window Approach (DWA) for collision avoidance that became very popular over the last decade. Further details on our navigation system can be found in [2,3,4].

3D obstacle segmentation: The museum offers a challenging environment for the application of an autonomous mobile robot platform. Using only a laser range finder for navigation would not allow to detect each and every obstacle. That is why we complement the laser detection with data from a depth camera.

Our obstacle detection approach consists of two steps. The parameters of the ground plane are estimated, followed by a segmentation of the depth values into those belonging to the floor plane and those that represent obstacles.

The fact that our depth camera is mounted on top of the robot leads to undesired pitch movements. To compensate this effect, we correct the plane parameters in each incoming frame. A more elaborated description of this process can be found in [2,5].

Autonomous awareness behavior: To support the human-robot interaction, the robot should show an observable awareness behavior, e.g. move the head towards its interaction partners and give them the feeling that the robot is listening. Hence, they are encouraged to address the robot in a natural way. In order to do so, the robot has to perceive the persons within its vicinity in the first place, for which we need a people tracking system.

We use a tracking-by-detection approach, where a number of sensor cues detects people and passes those detections to the Kalman-filter-based tracker. Details of the people tracking approach can be found in [6].

To give surrounding people the impression that the robot is aware of them, the robot turns its head and it appears as it would look at them. In order to fortify this impression of being looked at, the robot changes the person he looks at every ten seconds. The person to look at is chosen at random, but not each person has the same probability of being chosen. The probability is computed from its distance to the robot, the necessary head movement and the time since the person was looked at before. Therefore, the robot prefers nearby persons, little head movement, and persons it has not looked at for a while or at all.

Dialog system and Wizard-of-Oz: We believe that having a spoken dialog with museum visitors is an important aspect of a tour guide robot. It attracts people and allows to demonstrate an intelligent system. One reason why today's guide robots still lack complex dialog capabilities, is that speech recognition is a major unresolved problem. Developing such a dialog system under real world conditions is a challenging task. An unfinished system would leave the visitors unsatisfied or even frustrated. Hence, we have decided to create the illusion for the visitor, that the robot is already able to talk to them in a natural way. This is achieved by replacing the speech recognition system by a human operator using a Wizard of Oz method. With this idea, high level parts of the dialog system can then be evaluated, and other subsystems can be tested under real world conditions without annoying visitors too much by a highly experimental or even malfunctioning system.

For Wizard of Oz inspired experiments a hidden operator controls the robot's dialog system. In order to allow the operator to still react to the visitors, the images from the omnidirectional camera as well as an audio stream are transferred to the operator's laptop. To select the available phrases, different parts of the dialog were defined. For all of those parts, answers or reactions to possible questions or situations were assembled beforehand. For this task, a GUI has been developed that allows the operator to select pre-defined phrases or text fragments. It is not possible to enter arbitrary text.

Furthermore, the operator can select the robot's target position and initiate the driving mode, which is completely autonomously. While the robot navigates, his head is facing in the direction of movement. This allows persons walking towards the robot to guess the robot's intention to pass. Such a behavior is one of the many subtle steps towards a socially acceptable navigation. During the dialogs, the head of the robot is facing the dialog partners. From our first experiments with a real audience we derived first corner stones for an automated dialog system. In [2] these ideas and aspects are gathered and discussed.

Automated Prosody generation: As the dialogs themselves are held in natural human language, real time speech recognition and generation are crucial. So far, external software solutions (Loquento S.p.A.) are employed for those tasks. An evaluation of Text-To-Speech (TTS) software showed a general lack of the ability to generate speech with an entertaining and natural sounding prosody, which is of particular interest for long, tentatively boring information. Therefore a focus of our research is the automatic labeling of text with prosodic information, applying relevance learning algorithms [7].

Speaker Localization: To capture the acoustic environment, a special microphone array, consisting of four directional microphones, is used. A blind source separation approach utilizing the special microphone array characteristics [8] currently gets implemented into our framework. Furthermore, the characteristics of the microphone array allows the localization of the sound source. This possibility leads to a potential further cue for our people tracker.

SOM based upper body pose estimation: For our work on motion classification we apply an enhancement of the approach presented in [9]. It relies on the data of a depth camera from which a Self-Organizing Map (SOM) is extracted to model the human upper body. Crucial in that context is the correct assignment of the SOM neurons to a specific region of the tracked person's upper body. Various sources of error and noise lead to the situation that sometimes neurons migrate from one part of the upper body to another.

Without a verification of the SOM a future subsequent classification of the pose will produce incorrect results and may lead to wrong interpretations of the actual situation. Therefore we extended the approach in [9] by reshaping the trained SOM to a skeleton model to estimate the anatomical correctness of the pose. Having generated the skeleton model, incorrect Self-Organizing Maps will be rejected if the subsequent verification fails. The further goal is to eliminate this problem by integrating adaptive metrics [7]. A deeper insight in our approach can be gained in [10].

3 Summary

In this paper summary of the current field of work of the AI Lab at the UAS Dresden was presented. As it could be seen, for the moment, the focus of our research efforts is set to a museum's tour guide robot. However, our future plans are to concentrate on the idea of assisted living for the elderly, as well.

References

1. Fox, D., Burgard, W., Thrun, S.: The dynamic window approach to collision avoidance. IEEE Robot. Automat. Mag. **4**(1) (March 1997) 23 –33
2. Poschmann, P., Donner, M., Bahrmann, F., Rudolph, M., Fonfara, J., Hellbach, S., Böhme, H.J.: Wizard of Oz revisited: Researching on a tour guide robot while being faced with the public. In: RO-MAN. (2012) In press.
3. Himstedt, M., Hellbach, S., Böhme, H.J.: Feature extraction from Occupancy Grid Maps using Non-negative Matrix Factorization. In: DAGM NC2. (2012) In press.
4. Bahrmann, F., Hellbach, S., Böhme, H.J.: Please tell me where I am: A fundament for a semantic labeling approach. In: Poster and Demo Track KI-2012. (2012)
5. Donner, M., Poschmann, P., Klingner, M., Bahrmann, F., Hellbach, S., Böhme, H.J.: Obstacle detection for robust local navigation through dynamic real-world environments. In: IROS. (2012) in press.
6. Poschmann, P., Hellbach, S., Böhme, H.J.: Multi-modal people tracking for an awareness behavior of an interactive tour-guide robot. In: ICIRA. (2012) in press.
7. Hammer, B., Villmann, T.: Generalized relevance learning vector quantization. Neural Networks **15** (2002) 1059 – 1068
8. Gunel, B., Hachabiboglu, H., Kondoz, A.: Acoustic source separation of convolutive mixtures based on intensity vector statistics. Trans. on ASL **16**(4) (2008) 748 –756
9. Haker, M., Böhme, M., Martinetz, T., Barth, E.: Self-organizing maps for pose estimation with a time-of-flight camera. In: DAGM. (2009) 142–153
10. Klingner, M., Hellbach, S., Kästner, M., Villmann, T., Böhme, H.J.: Modeling Human Movements with Self-Organizing Maps using Adaptive Metrics. In: DAGM NC2, Graz (AT) (2012) In press.