Interactive Object Modelling for a Humanoid Service Robot

Regine Becher[1], Peter Steinhaus[1], and Rüdiger Dillmann[1]

IAIM, University of Karlsruhe (TH)
Haid-und-Neu-Str. 7
76131 Karlsruhe, Germany
{becher,steinhau,dillmann}@ira.uka.de

Abstract. In this paper, we discuss our concept for object modelling for a humanoid service robot in a human-centered environment. We motivate the use of a training center with appropriate sensor systems for object learning. After discussing the types of information that are relevant for object modelling and where to get this information from, we introduce our concepts of flexibility, extendibility and interactive modelling by explaining the object representation we propose and the multimodal ways of interaction included in our system. The paper ends with some conclusions and a perspective on future works.

1 Introduction

For a humanoid service robot to be able to work in a human-suited environment, having detailed models of objects and of its environment is essential. As the environment of such a robot is necessarily dynamical, pre-built models only are not very useful. In our project, we are thus developing an object and environment representation which is dynamical and extendible. In particular, our goal is to enable a service robot in a household to correct and complete a representation of its specific environment and the objects it contains with the aid of a human user. It is possible for the user to interactively take influence on the robots models, e.g. give a special name to preferred or often-used objects or introduce a feature of an object which is especially important to him.

In this paper, we are discussing our concept of interactive object modelling. First, we introduce our system concept, in particular the use of a specific training area for object modelling, and discuss the need for interactivity in robot object modelling (section 2). Section 3 gives an overview of interactive object modelling for robots in the current research. In section 4, we describe our training area in more detail, especially the sensor systems we are using. Section 5 discusses the type of information that a robot needs for object modelling in a human-centered environment and how this information can be acquired in detail. Sections 6 and 7, respectively, explain how interaction is realized in our concept, and how object information is represented flexibly and extendibly. Finally, section 8 concludes this paper and gives some insights into our future work.

2 System Concept

In our system concept, there is a clear division between the modelling of objects (and their features and functionalities) and the modelling of the robot's environment. We have chosen a setting which is prototypical for service robots: service tasks in a household, in particular in the kitchen. Thus, a model of the kitchen environment has to be built up by the robot – naturally, this should take place in the kitchen itself. In contrary, object modelling is done mainly in a so-called training area. The training area is equipped with other sensor systems than the robot, i.e. the object models can be built up much more precise than it would be possible in the kitchen itself. Note that, although the object modelling takes place in the training center, the objects must be recognized etc in the experimental kitchen setup in everyday robot use. The advantage of modelling objects in a dedicated training area is that the objects can be modelled under ideal conditions more precisely than by the robot.

The system concept that we are proposing includes an interactive approach for object modelling. As we will show later, this is a very essential part of the system. Some sorts of knowledge about objects can not be gained automatically, but only with the help of the user (e.g. knowledge about preferences of the user). For a service robot in a household, it is this knowledge which makes it useful at all.

In addition, our object model is extendible and flexible such that new objects as well as new information about objects can be included into the model easily, even by an unexperienced user. Again, having a predefined object model as some other systems do would not allow for the flexibility which is necessary in a human household. After all, there are slight differences from kitchen to kitchen, and having your personal robot fetch your favorite cup means that it has to have an object model where you can include the cup (i.e. which is flexible and extendible) and that it allows for interactive object modelling (otherwise you could not tell the robot that you prefer this specific cup).

For these two reasons, we are developing a flexible, extendible object representation which allows for interactive modelling by the user. Before introducing our concept in more detail, the next section discusses the state of the art in interactive environment modelling.

3 State of the Art

A lot of research projects all over the world are working on humanoid robots, but the main focus of research is on mechatronics (e.g. a human-like walk). The issue of object and environment modelling is not that much considered. In our system, we argue for a dynamic, flexible object model which is built up and extended interactively. Both aspects (dynamic modelling and interactive modelling) are not observed as much in existing systems as they should.

Dynamic modelling is mostly done for navigation and path planning (e.g. [Asoh et. al. (2001)], [Graefe/Bischoff (1997)]) and also to build up geometri-

cal world models ([Baader/Hirzinger (1995)]). In contrast, features or functionalities of objects are usually not modelled dynamically. We argue that for a humanoid service robot in a household environment, representing features and functionalities of objects dynamically is essential. Also, xml for object representation is so far mostly used for web applications or multi-agent systems ([Chella et. al. (2002)]).

Interaction between robot and user takes place in various systems, but is usually not used for object modelling. It is even stated that learning can take place either by programming the robot or by independent learning without any interaction with the human ([Graefe/Bischoff (1997)]). In contrast to this opinion, we argue for interactive object modelling and thus interactive learning of a humanoid service robot and explain why it is important for successfully using service robots in human-centered environments like a household. Human-robot interaction for purposes other than object modelling if often restricted to one modality, mainly speech, in many cases there is no true dialogue between human and robot, or pointing gestures are only used as assisting information in case of underspecified speech input (e.g. [Asoh et. al. (2001)], [Knoll et. al. (1997)], [Lopes/Teixeira (2000)], [Perzanowski et. al. (2001)]).

4 Sensor Systems in the Training Area

As introduced above, we are using a training area which is equipped with different sensor systems which allow for a precise modelling of the objects themselves as well as of their features and their functionalities. In addition, multi-modal human-machine interaction is used to supply missing information to the system and to allow the user to correct, delete and add information where necessary.

Our current sensor systems for interactive object modelling are a Sony firewire high-resolution colour camera with 1280x1024 pixels, 7.5 frames per second and a 24 bit colour depth used in combination with the Matrox Imaging Library as software tool. A flexible and mobile Videre Design Mega-D stereo camera system consisting of two high-resolution colour cameras is used with the Small Visions System (SVS) library. Furthermore, a Minolta laser scanner system with structured light with a precision of up to 0.2mm is used. A tactile data glove (CyberGlove by Immersion) with 22 strain gauges to measure joint angles and a magnetic field tracker (Flock of Birds by Ascension) are used to track the position and state of the user hand. Finally, scales are included to determine the weight of objects, and microphones are used for speech recognition.

5 Information about Objects

For service tasks in a household, objects have to be detected, localized, and manipulated (i.e. approach, grasp and disapproach tasks under specific manipulation constraints). Additionally, information is needed about objects such as their form, size, or whether they are combined objects.

Starting from these considerations, we have identified relevant categories of information about objects. These categories include: relevant types of grasps, appropriate grasp forces, contact points and typical approach and disapproach trajectories to handle and manipulate objects; object geometry including e.g. geometry representations, size, stable positions, preferred positions, weight and texture; manipulation constraints like maximum speed, maximum acceleration, maximum contact forces or the restriction to preferred orientations; and algorithms for detecting and localizing objects and for detecting their state.

These categories of information about objects have in common that they can only partially be computed automatically from sensor data. In each case, there is a considerable amount of information which must be supplied interactively by the user. A third possible way of gathering information is to infer new knowledge from already known object features. In this case, the system can build up new hypotheses and present them to the user for acknowledgement or adaption.

In this section, we only describe information which can be obtained from sensor data directly without interaction with the user. More detailed information about user-system interaction and the information gathered in this way is given in the next section, a description of our object representation in section 7.

The sensor systems which are described in the previous section are used to obtain object models as follows:

Colour camera From the colour camera, we gain information about object texture.

Stereo camera system The stereo camera system allows us to gain 3D information about objects, and thus to compute information about an object's geometry, like surface models or a bounding box.

Structured light We use structured light to get information about the object geometry as from the stereo camera system, but with a higher precision.

Scales For getting information about an object's weight, we are planning to integrate scales into our training area whose result will be processed automatically.

As can be seen from this list, most information that is necessary or very desirable for a service robot in a human-centered environment can hardly be won from automatic processing of sensor data only. In contrary, even the information which on a first glance can be gathered automatically depends on interaction with the user in some cases. An example would be a scan of an object (by camera and/or structured light) with occlusions in the scan. In this case, the system could point the user interactively to turn the object into a different position for a further scan. We are thus concluding that human-machine interaction is a very important part of a sensible object modelling process.

6 Multi-modal Human-Robot Interaction

To enable the user to take influence on the robot system's object modelling, we are using methods of multi-modal interaction: speech, gestures, and graphical

user interfaces. Whereas speech and gestures are very natural and intuitive means of communication for the human, graphical user interfaces have the advantage to allow a very precise display of information in space and of complex information. This is both true for input the user is giving the robot as well as for feedback he is getting from the robot. We are thus using speech recognition, gesture recognition and graphical interfaces for human input to the robot, and speech synthesis and graphical interfaces as output of the robot (gestures as output are not available in the training center, and the information they could convey can be given much more precisely and at least equally fast by a graphical interface, and they are thus not necessary for our system).

The sensor systems of the training area are used interactively in the following way:

Colour camera and structured light In the case of occlusions during a scan, either the system can point the user to the problem and ask him to present the object in a different position, or the user can detect the problem himself, turn the object and let the system take an additional scan of the object.

Stereo camera system In addition to gathering information about the object geometry etc, the stereo camera system can be used for gesture recognition and tracking of the user hand ([Rogalla et. al. (2002)].

Tactile data glove The data glove with tactile sensors allows the user to interactively present potential grasps, grasp points and grasp forces.

Magnetic field tracker Using the magnetic field tracker to track the user hand, approach and disapproach trajectories and the position of the hand can be obtained.

Microphones The microphones are essential for speech recognition.

Therefore, modes of user input in our system are gestures, grasps, grasp positions and forces, trajectories, speech and graphical user interfaces. The system is also able to give output to the user, this takes place in form of speech output, acoustic feedback and the graphical user interface. To have true interaction, some form of dialogue between the system and the user is necessary. Dialogues are realized as spoken dialogues (speech recognition and synthesis) and via the graphical user interface. Additionally, gestures can be used on the side of the human to present dialogue information.

A typical example of interactive object modelling could be the following: The user presents a new object to the system which is scanned. Object features are computed automatically as far as possible, but as only one scan was taken, parts of the object were occluded to the system. The user is pointed (speech output or graphically) to turn the object into another position for a second scan. The user confirms the turning of the object, and the system takes another scan. In a next step, the system generates hypotheses about stable object positions which are judged by the user. The user can also give information about object features which seem to be important to him and the system point the user to information (especially about features of objects) which is still missing.

At the end of this learning process, the system has built up a representation of the object which is confirmed by the user and contains all information the

user found to be interesting. Note that it is easily possible to add or change information about objects afterwards, e.g. during application in the experimental kitchen setup. The next section explains in more detail how objects are represented in our system.

7 Object Representation

As we have motivated above, the object model of a service robot in a human-centered environment should be flexible and extendible. In addition, it is not only information about objects which has to be available to the robot system, but also information about the context and the environment (cf. section 8). Therefore, a representation should be chosen which can express very different kinds of information in order to make processing as easy as possible.

From these considerations, we are proposing an attribute-value structure for object modelling which is represented as xml-structure. The advantages of this representation are its flexibility, its extendibility and the fact that all types of information that are needed for a service robot to have complete models of its environment, objects, tasks etc can be represented in this structure. Additionally, basic data types are defined for data structures which are very common, e.g. trajectories, triangle meshs or grasps. These basic data types allow for faster processing of information, and they also correspond to human information processing categories and are thus easily understandable for the human user during the interactive modelling process.

Potential grasps for an object are represented using the Cutkosky hierarchy. The information about a grasp contains its Cutkosky type and parameters typical for this type of grasp. Additionally, the grasp forces as absolute value and direction, the contact forces relative to an object coordinate system, and the approach and disapproach trajectories belonging to a grasp are stored with it. All of this information is gathered interactively as the grasps and movements are demonstrated by the user.

For representing object geometry in a broad sense, we are trying to acquire a surface model of the object from sensor data. The surface model is built up from polygonal surface patches (another basic data type), and the texture for each surface patch. Additionally, the system determines a bounding box for each object (represented as box relative to the object coordinate system, and additionally as length, depth and height of the box), and estimates the center of gravity of the object. The object's weight is gained from the scales in the training center. Stable positions and preferred positions are also determined interactively: stable positions can either be demonstrated by the user, or the system can build up hypotheses about stable positions which the user has to confirm or discard. Preferred positions (e.g. a cup standing upright) are a subset of the stable positions which the user has to choose from them.

To ease detection, localization and handling of objects in the experimental kitchen setup and later on in everyday use, the robot system should know for each object how such operations can best be performed, i.e. which sensor systems

delivers the best results and with which parameters. It is the user who judges about the possible functionalities of the system, e.g. with which sensor system the fill level of a container can be determined most reliably. In such cases, the system presents the potential processing results of different sensor systems, and the user ranks them. The result are cognitive operators for functionalities of objects. The cognitive operators supply information about the sensor system which has to be used, its functions and parameters.

Finally, information about manipulation constraints has to be given to the system. This information is gained from user demonstrations of handling the object. The maximum speed and acceleration for each axis of the coordinate system are represented as three-dimensional vectors; the maximum contact forces (e.g. to stack objects) are given for each stable object position; and finally, the user can restrict the system to preferred positions of the object.

As explained above, all of this information is stored in the object model of the robot as attribute-value pairs in an xml representation.

8 Conclusions and Future Work

In this paper, we have introduced our concept for object modelling for a humanoid service robot in a human-centered environment. We have motivated the use of a training center with appropriate sensor systems for object learning. After discussing the types of information that are relevant for object modelling and where to get this information from, we introduced our concepts of flexibility, extendibility and interactive modelling by explaining the object representation we propose and the multimodal ways of interaction included in our system.

Current and future work in this area includes environment modelling in the experimental kitchen setup, task modelling and context modelling.

9 Acknowledgements

The work of the Collaborative Research Center 588 "Humanoid robots – learning and cooperating multimodal robots" is supported by the Deutsche Forschungsgemeinschaft (DFG).

References

[Asoh et. al. (2001)] H. Asoh, Y. Motomura, F. Asano, I. Hara, S. Hayamizu, K. Itou, T. Kurita, T. Matsui, N. Vlassis, R. Bunschofen, B. Kröse: Jijo-2: An Office Robot that Communicates and Learns. IEEE Intelligent Systems, 2001.

[Baader/Hirzinger (1995)] A. Baader, G. Hirzinger: World Modeling for a Sensor-in-Hand Robot Arm. Intelligent Robots and Systems 95. 'Human Robot Interaction and Cooperative Robots', Proceedings. 1995 IEEE/RSJ International Conference on , Volume: 2 , 5-9 Aug 1995.

[Chella et. al. (2002)] A. Chella, M. Cossentino, R. Pirrone, A. Ruisi: Modeling Ontologies for Robotic Environments. SEKE02, Ischia, Italy, 2002.

[Graefe/Bischoff (1997)] V. Graefe, R. Bischoff: A Human Interface for an Intelligent Mobile Robot. IEEE International Workshop on Robot and Human Communication 1997.

[Knoll et. al. (1997)] A. Knoll, B. Hildebrandt, J. Zhang: Instructing Cooperating Assembly Robots through Situated Dialogues in Natural Language. Proceedings of the 1997 IEEE International Conference on Robotics and Automation. 1997.

[Lopes/Teixeira (2000)] L. Seabra Lopes, A. Teixeira: Human-Robot Interaction through Spoken Language Dialogue. Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems.

[Perzanowski et. al. (2001)] D. Perzanowski, A. Schultz, W. Adams, E. Marsh, M. Bugajska: Building a Multimodal Human-Robot Interface. IEEE Intelligent Systems, (1) 2001.

[Rogalla et. al. (2002)] O. Rogalla, M. Ehrenmann, R. Zöllner, R. Becher, R. Dillmann: Using Gesture and Speech Control for Command a Robot Assistant, ROMAN 2002, Sep.02, Berlin, Germany