


Our approach to robot behavior might be called semi-autonomy: the robot must dispose of a repertoire of skills that are carried out autonomously, but the actual control of overall behavior must be left to a human operator. Although classical techniques are readily available for this control, we have chosen to implement a gesture interface which is suited for use by technically untrained people, a decision dictated by the quest to provide a prototype for a service robot.

2 Architecture

All individual robot components are controlled by autonomous agents that communicate with each other and with other agents that implement specific skills, basic operations that can be composed in various ways to implement specific tasks. Central control is maintained by a separate agent that defines task timelines and contains a user interface. On the implementation level, each agent basically is one executable under Unix or QNX, respectively (see figure 1 for an overview of all agents). A central skill is fixation, i.e., directing the axes of both eyes towards a point of interest. It is calibrated by an autonomous procedure [6, 7] and serves as the principal transformation for the extraction of 3D information. The three-dimensional position of the fixated point can be estimated by simple triangulation from the pan, tilt and vergence angles of the camera head.
Fig. 2. The user points at an object with a gesture (t = 2.6s, top), which is recognized and localized by graph matching (t = 82.2s, bottom).
Fig. 3. The object pointed at is recognized and localized ($t = 142.9s$).

3 Example behavior

Here, the skills are only listed as part of a standard example behavior, but they can be recombined in a variety of ways. Figs. 2 through 6 are taken from a video, which can be viewed from our website (http://www.neuroinformatik.ruhr-uni-bochum.de/inl/VDM/research/robotics/contents.html). That behavior is as follows.

*Initial situation:* GripSee stands before a table with various known objects and waits for a human hand to transmit instructions (Fig. 2, top).

*Hand tracking and fixation:* The hand is tracked to follow the pointing movement. After the hand movement stops, the hand is fixated.

*Hand posture analysis* is executed on the distortion-reduced small central region of interest, and the type of posture is determined (Fig. 2, bottom, and [9]).

*Object Fixation:* It is expected that the hand points at an object, thus the fixation point is lowered about 10 cm. Now, a simple localization algorithm detects the object, which is then fixated. After three or four iterations, the object is fixated well, and a new region of interest is defined.
Fig. 4. The arm is moved towards the object ($t = 164.8\text{s. top}$), and it is picked up using the specified grip($t = 189.6\text{s. bottom}$).
Fig. 5. The user depicts a point for putting down the object using another gesture ($t = 233.1s$, top), which is localised ($t = 266.5s$, bottom) by graph matching.
Object recognition determines the type of object, its exact image positions and sizes and its orientation (Fig. 3 and [4,3]).

Fixation: The recognition module yields a more precise object position than the preceding fixation. That precision is crucial for reliable grasping, thus the object is fixated once more. Neural networks are used for fixation and triangulation in order to cope with the nonlinearities of the camera head.

Grip planning: A grip suitable for the object and the hand posture is selected from a database and transformed into the object’s position.

Trajectory planning and grip execution: A trajectory is planned. arm and gripper are moved to the object, and finally the object is grasped. Another trajectory transfers it to a default position conveniently located in front of the cameras for possible further inspection (Fig. 5).

Hand tracking, hand fixation and position analysis: GripSee again watches the field of view for the operator’s hand to reappear, this time pointing with a predefined gesture to the three-dimensional position of the place to put the object down (Fig. 5 and [9]).
Trajectory planning, execution, and release: A trajectory is planned, the arm moves the object to the desired location and the gripper releases it. In the absence of further clues, the 3D-precision of that release is an immediate measure for the quality of hand localization (Fig. 6).

4 Concluding remarks

Additional functionality, which is currently being worked on, includes the construction and incorporation of tactile sensors and the semi-autonomous learning of new objects. More detailed information can be found on our website and in the articles [1, 2].

This work is funded by grants from the DFG (Graduiertenkolleg KOGNET) and the German Federal Minister for Education and Research (01 IN 504 EQ). We thank Michael Pötzsch, Michael Rinne, Bernd Fritzke, Herbert Janßen, Percy Dahm, Rainer Menzner, Thomas Bergener, and Carsten Bruckhoff for invaluable help to get various aspects to the point of actual functioning and to Carsten Brauer for filming the video.

References