Model-free Monocular Visual Servoing

Benjamin Panreck
Beuth University of Applied Sciences Berlin, Germany
bpanreck@beuth-hochschule.de

Simon Untergasser

Manfred Hild

***Abstract*—In this document, we present an unusual approach to pathfinding for grasping objects. Since for robots in human-machine interaction different requirements apply than for robots used in industry, there is also a need for other approaches. We present a sensorimotor approach based on gaze heuristics, which is based exclusively on easy-to-learn implicit environmental models and which uses the advantages of robots with few degrees of freedom.**

I. Challenge

Robots in human-machine interaction can behave differently than robots in industrial applications because the requirements are fundamentally different. Speed, accuracy, reliability and reproducibility are required from industrial systems, but they are less important in the interaction with humans. Here, intuitively comprehensible actions are required. Failures and lower speed are explicitly allowed and can be dealt with in human environments. We use the humanoid robot Myon with an anthropometric morphology comparable to a seven year old child. The robot was originally developed for research in the field of artificial language evolution with minimum complexity in morphology [1]. Each arm has only four degrees of freedom (DOF). In comparison with the human arm, which has seven DOF plus two assisting DOF in the shoulder, this is a massive reduction of flexibility in the actuation space. The advantages of limiting the mechanical DOF are the reduction of cost and complexity.

Furthermore Myon uses only one monocular camera for visual input without depth information. This is rather unusual, because manipulating robots normally use depth information to plan their grasping [2]. Our tests have shown that monocular distance estimates using head motions are subject to extreme errors due to noise and distortion of the optical system and are not suitable for predictive grasp planing or typical approaches from the field of visual servoing.

With this restricted platform, our challenge is to grasp objects of unknown size and distance using only hand-eye-coordination.

II. Solution

To address this challenge, we use a behavioral heuristic based on local sensorimotor loops and an online learned implicit model of the environment in the form of quadrics. This makes it adaptive to changes in morphology of the robot or of the environment.

Our approach is, to control the arm without an explicit model of itself, the environment or of the object to grasp, but by using tightly coupled sensorimotor loops (SML). Basically, we use six independent SML to control the head- and arm-movement. The first two SML are used to couple two DOF of the head (yaw and pitch) directly to the visual input by bringing the coordinates (x and y) of the object to be grasped to the center of the camera image. The third loop is a coupling of the visual orientation of the lower arm and the roll-joint of the head to bring the lower arm in a vertical orientation to the camera image. This allows us to couple the two DOF of the shoulder (roll and pitch) directly to the visual input with the goal to move the robot's hand coordinates (x and y) to the center of the image. This means that the hand and the object to be grasped meet in the center of the image. In behavioral psychology this behavior is called gaze-heuristics [3] and is a known behavior of humans to cause (or avoid) collisions with an object. The robot's wrist is controlled automatically in a fixed orientation to the ground by using the acceleration sensors in the lower arm. When additionally the elbow is actuated, the hand either gets bigger (closer to the camera), or smaller (further away from the camera). When we use the elbow to bring the hand further away from the camera, the hand will collide with the object. To avoid this collision, we need a measure for the distance without knowledge about the object size. This problem can be tackled by using an implicit model of the table where the object is placed. One possibility are quadrics. The table where the object of interest is placed, is a surface in $\mathbb{R}^3$. Using quadrics we are able to model such surfaces in the three dimensional joint space (shoulder pitch, shoulder roll, elbow) of the robot. Such a model can be learned (using RLS or similar) by moving the robot's hand to different positions on such a surface while keeping the elbow joint relaxed. As usual for heuristic behaviors, the gaze-heuristic used here is a problem-specific solution.

The final grasping movement performed by the hand of the robot needs to be considered separately. For this additional problem, there are promising approaches which make use of the environmental constraints [4].

A major problem that still needs to be solved are parallax errors on the optical axis of camera, hand and object.

References

[1] M. Hild et al., "Myon: Concepts and Design of a Modular Humanoid Robot Which Can Be Reassembled During Runtime," in *Proceedings of the 14th CLAWAR*, Paris, France, September 2011.

[2] R. Li and H. Qiao, "A survey of methods and strategies for high-precision robotic grasping and assembly tasks—some new trends," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 6, pp. 2718–2732, 2019.

[3] G. Gigerenzer and P. Todd et al., *Simple Heuristics That Make Us Smart*, 1st ed. Oxford University Press, 1999.

[4] C. Eppner, *Robot grasping by exploiting compliance and environmental constraints.* Technische Universitaet Berlin (Germany), 2019.