

Graz University of Technology (Austria)

I am in charge of the Vision Based Measurement Group at Graz University of Technology. The research group is focused on two main areas: Object Category Recognition & Detection; and Structure & Motion Analysis. We have excellent first results on “active object categorization” with NAO, where NAO is handed an object, inspects it visually, and classifies the object category. To explore object detection, manipulation, and recognition further, it is desirable to develop autonomous grasping capabilities on NAO.

While NAO has quite versatile arms, hands and fingers, autonomous grasping of objects with NAO poses a number of hard problems. First, there is (still, although the longer arms of V3 have certainly improved this issue) a rather limited area that can be confidently reached by NAO’s arms. Second, there is the pincer-shape of the three fingers, and only one degree of freedom (open/close) in grasping.

Third, NAO’s monocular vision and rather limited computational power call for a slim, reconstruction-free approach. Therefore, we restrict ourselves to a study case, where NAOs “world” is unsaturated, in terms of colour, i.e. white or grey walls, table surfaces, floor. Furthermore, the objects to be grasped are limited to highly saturated, coloured (red, green, blue) foam cubes. We have addressed the autonomous grasping problem in two distinct steps:

### 1. Grasping from a small table in front of NAO.

NAO is covering behind the table, and only NAO’s arms are actuated. The area of reach is limited to the part of the table that is in the field of view of the bottom camera. This task addresses mainly issues of camera calibration, coordinate transformations, and visual servoing.



Grasping from a table



NAOs “room”



Grasping from the floor

## 2. Autonomous grasping from the floor.

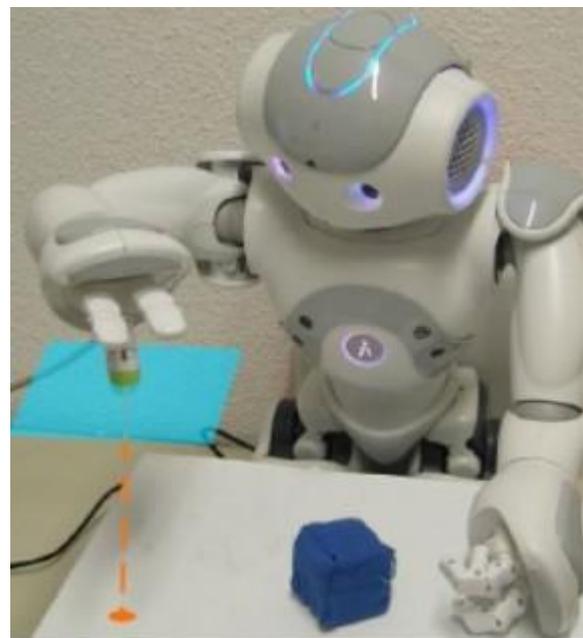
We have built a small “room” for NAO, with white walls and a homogeneous dark floor. NAO is placed in this room, and a foam cube is thrown in. NAO detects the cube, walks towards it, bends down, and grasps it. While this task is certainly more complex than the grasping from a table, NAO can control its approach towards the cube, ending up in a well-suited pose for grasping. Thus, we even achieve higher success rates for grasping from the floor than for grasping from the table.

**“ To explore object detection, manipulation, and recognition further, it is desirable to develop autonomous grasping capabilities on NAO. ”**

In both cases, visual servoing is performed as follows.

We calculate the orientation of the floor plane – it coincides with the normal of NAO’s feet. Next, we calculate a plane parallel to the floor plane, in which we position the hand above the cube. In the case of the table, this plane is approx. 10cm above the table plane. When picking up from the floor, the plane is 10cm above the floor plane. We calibrate NAO’s bottom camera and project the position of NAO’s thumb on the table/floor surface. We call this, the “virtual thumb” concept. Finally, by visual servo control, we try to align the position of the virtual thumb with the colour blob (image of the coloured cube) in the image.

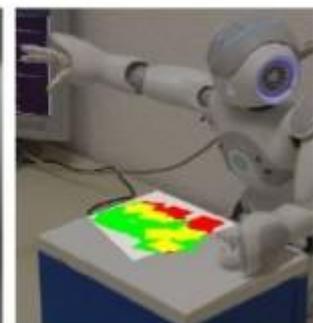
For the “pick up from the floor” task, NAO first visually inspects the room. He takes a number of views by rotating his head. A cube is detected as a highly saturated colour blob in the image. If no cube can be found, NAO rotates around his horizontal axis and takes more views. Once a cube has been found, NAO approaches the cube, walking towards it. This is achieved by colour blob tracking and head motions, trying to keep the cube in NAO’s field of view while walking.



The motion plane, where the hand is positioned, and the projection of the “virtual thumb”.



NAO's view, thumb, table surface normal, the “virtual thumb” (orange dot).



74% successful grasps

Once NAO is close to the cube, he stops walking, and starts a fine-positioning move, before bending down to start the grasping.

In repeated experiments (50 independent tries each, grasping from the table and from the floor), we reach a success rate of 74% successful grasps from the table, and 88% successful grasps from the floor.

NAO is ideally suited for the research performed in our group. Work with humanoids is attractive to students, and we offer an excellent working environment in our image-based measurement lab. The NAO hardware platform is sufficiently stable, and it can be rather easily accessed and programmed. At the same time, compared to industrial robots, NAO is inherently imprecise. This calls for qualitative, goal-oriented solutions. Furthermore, such a robotic system will force the student to use sensorial input (in our case mostly visual input), and to find ways to communicate processing results. While in Computer Vision and Image Understanding research, the output is often in the form of processed/augmented images, NAO is well suited to demonstrate success of computer vision algorithms by actions, motion, or by voice output.

I am working on NAO with groups of my students (up to 6 students per semester). Inside our vision-based measurement lab, we have set up a “living room” for NAO. This “room” is about 2×2 meters, with white walls of 60cm height, so that NAO’s visual input will be constrained to the interior of its room, while it is easy to watch from outside. We have equipped the room with various colour-coded furniture that can be used as chairs and tables of variable height, and also (turned upside-down) as colour-coded baskets.



Thomas Höll, and the experimental setup for grasping from the floor

Current projects include active object categorization of toy objects that are handed to NAO, the autonomous vision-based grasping described here, and we have also built a flexible maze that can be put into the room in various configurations. We wish to make NAO find its way through this maze.

Nao is mainly programmed using C++ and OpenCV. Behaviours are naturally and easily implemented in Choregraphe. These are first, yet quite encouraging, results on autonomous grasping. The major current limitation is the colour and size of our foam cubes. They must be homogeneous (untextured), highly saturated (uniformly and intensely coloured), soft objects of a size that is convenient to grasp for NAO's hand.

**“ While in Computer Vision and Image Understanding research, the output is often in the form of processed/ augmented images, NAO is well suited to demonstrate success of computer vision algorithms by actions, motion, or by voice output.”**

However, with completely unchanged parameters, NAO also accidentally succeeded in picking up a green Duplo brick from the floor.

Improvements in autonomous grasping capabilities will be the key to further progress in this area. This would probably require more sophisticated algorithms and higher computational power to calculate grasping points for more complex shaped and textured objects.

For further information about Axel Pinz' research group and his research activities, please visit: [www.emt.tugraz.at](http://www.emt.tugraz.at)

Published papers:

T. Höll and A. Pinz. Vision-based grasping of objects from a table using the humanoid robot Nao. In Proc. ARW'2011, Austrian Robotics Workshop