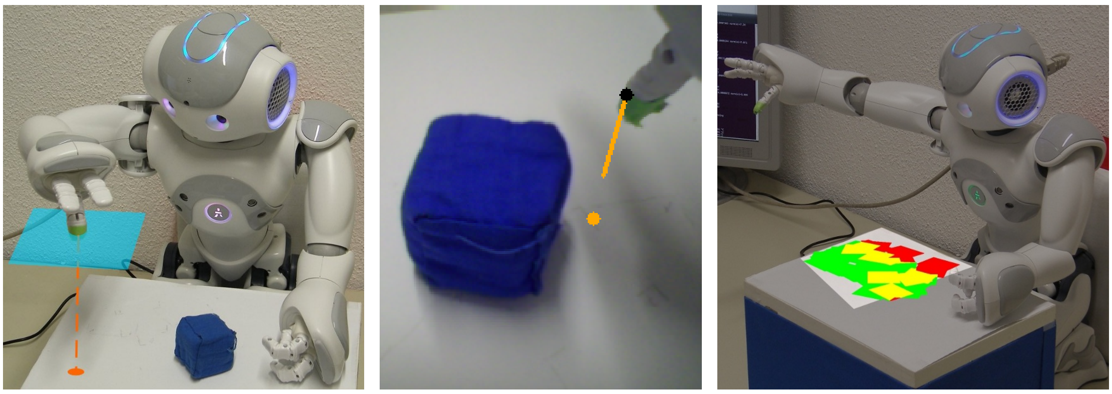
Vision-based grasping of objects from a table using the humanoid robot Nao

Thomas Höll, Axel Pinz
Inst. of Electrical Measurement and Measurement Signal Processing
Graz Univ. of Technology, Austria
tomh@sbox.tugraz.at, axel.pinz@tugraz.at

Introduction: There is a large amount of work in robotics that addresses hand-eye coordination ([1], [2]) and vision-based calculation of objects and of grasping pose, but autonomous grasping on Nao poses particular problems: We face rather limited on-board processing power, inherent inaccuracies in repositioning due to the building quality of the robot, reduced stiffness of joints in case of continuous operation, a quite limited grasping ability of Nao's hands with their pincer-shaped three fingers, as well as severe self-occlusion when the hand approaches the object in its final grasping position (the robot's camera will mostly see the back of the hand with self-occlusion of the fingers and severe occlusion of the object). Therefore, we propose a solution to autonomous vision-based grasping of objects from a table that consists of:

- computationally slim algorithms,
- estimation of the spatial orientation of the table surface,
- projection of the thumb onto this surface ($\rightarrow$ *virtual thumb*),
- control of the arm motion according to the position of this *virtual thumb*.



(a) The *virtual thumb* (b) Nao's view (c) 74% successful grasps

Figure 1: Our solution to the grasping of a colored foam cube by Nao. Sitting behind a "table", the table normal vector is estimated relative to Nao's feet position. (a) shows a cyan plane that visualizes, in which plane the positioning of Nao's arm is done. The *virtual thumb* (orange dot) is calculated as the projection of Nao's thumb (green) onto the table plane. (b) shows an image of Nao's bottom camera augmented by this *virtual point* (orange dot) and the normal vector of the table plane (orange line). (c) shows the result of our grasping experiment. Green areas show successful grasps, red ones failure cases. Yellow indicates occasional failure. The highlighted area marks the field of view of Nao's bottom camera.

The major algorithmic components of our approach are outlined in the paragraphs below.

Parallel planes recovered from joint sensor readings: We require an estimation of the table plane orientation in 3D space. To achieve this, we assume that the table plane is parallel to the ground plane, which can be measured by the 3D orientation of Nao's feet. The Nao framework delivers transformation matrices relative to a given reference coordinate system [3]. Define $\mathbf{T}_{\mathcal{T}}^{\mathcal{R}}$ as the homogeneous transformation of the torso reference coordinate system to the right foot. According to our assumption, the table plane has an offset in z-direction (i.e. the table's height) relative to the right foot coordinate system, given by $\mathbf{T}_{\mathcal{T}}^{\mathcal{R}}$. In our experimental verification we measure an average deviation angle between the normal vectors of the ground plane $\vec{n}_{gr}$ and the calculated table plane $\vec{n}_{tb}$ by actuating Nao's arm in a plane that should be parallel to the table plane (cyan plane in fig. 1(a)) and measuring the variation in the actual distance between hand and table surface. The deviation angle is then calculated as $\angle(\vec{n}_{gr}, \vec{n}_{tb}) = 4.8° \pm 1.6°$.

Projecting the thumb onto the surface of the table: With the help of transformation matrices, provided by Nao's framework, and the knowledge of the spatial orientation of the table plane $\vec{n}_{tb}$, it is now possible to project the finger tip of Nao's thumb onto the table plane. A first series of experiments showed us that the position of the thumb is essential for a successful grasp. Based on this observation, we design a control mechanism to move the *virtual thumb* to a desired goal position relative to the object. To calculate the desired goal position, a blob detection algorithm [4] is used, which provides a bounding box around the foam cube. The desired goal position is calculated as the center of the bottom line of this bounding box, adding an offset that depends on the arm pose. Please note, that all these calculations are performed in 2D image coordinates (Nao's view of the scene, see fig. 1(b)). To control the arm motion, a *visual servoing* method is used.

Visual servo control: With the help of *visual servo control*, we want to control the hand of Nao such that the grasping of the foam cube is accomplished. Denote $\vec{s} \in \mathbb{R}^2$ as the point feature of the *virtual thumb* and $\vec{s^*} \in \mathbb{R}^2$ as the desired feature, at which a grasping is possible (in image coordinates of Nao's view). $\vec{e}(t) = \vec{s} - \vec{s^*}$ is the error function that should be minimized. In [1] the control law $\vec{v}_h = -\lambda\, \mathbf{L}_e^+ \vec{e}$ is proposed, where $\vec{v}_h$ is the hand velocity, λ is a gain factor and $\mathbf{L}_e^+$ is the so called "interaction matrix". According to [2], there are several ways to estimate $\mathbf{L}_e^+$. In our approach, we use a constant interaction matrix, estimated by a few exploratory movements, observing the changes of $\vec{s}$. According to [5], this is an eligible assumption for small work spaces.

Summary and Conclusion: Hand-eye coordination and autonomous grasping are well-researched fields. However, we face a number of practical limitations and problems due to Nao's building quality, kinematics and pincer-shaped hands. The monocular vision system forces us to search for a solution that works based on measurements in 2D image space. To our knowledge, this is the first implementation of autonomous grasping on Nao. In our experimental validation (see fig. 1(c)), we manually placed a blue foam cube at various postions on the table. In 50 experiments, we achieved 37 successful grasps (74%). Most of the failures occured in the red area close to Nao's body and are due to the restricted freedom of motion of Nao's hand.

References

[1] F. Chaumette and S. Hutchinson, “Visual servo control. Part I: Basic approaches,” *Robotics Automation Magazine, IEEE*, vol. 13, no. 4, pp. 82 –90, 2006.

[2] F. Chaumette and S. Hutchinson, “Visual servo control. Part II: Advanced approaches [tutorial],” *Robotics Automation Magazine, IEEE*, vol. 14, no. 1, pp. 109 –118, 2007.

[3] Aldebaran Robotics, *NAO user guide*, 2010.

[4] T. Schlager, “Vision-based control on the Nao academic robot.” Bachelor thesis, Inst. of El. Measurement and Measurement Signal Processing, Graz Univ. of Technology, 2010.

[5] J. Su, Y. Zhang, and Z. Luo, “Online estimation of image Jacobian matrix for uncalibrated dynamic hand-eye coordination,” *Int. J. Syst., Control Commun.*, vol. 1, pp. 31–52, July 2008.