Presented at the Workshop on Approaches to Sensorimotor Learning on Humanoid Robots,
IEEE International Conference on Robotics and Automation, 2009.
For final version, see http://www.montefiore.ulg.ac.be/~piater/papers/Detry-2009-SLHR.pdf

Autonomous Learning of Object-specific Grasp Affordance Densities

Renaud Detry Emre Başeski Norbert Krüger Mila Popović Younes Touati Justus Piater

In this paper, we address the issue of learning and representing object grasp affordances. Our first aim is to organize and memorize, independently of grasp information sources, the whole knowledge that an agent has about the grasping of an object, in order to facilitate reasoning on grasping solutions and their likelihood of success.

By *grasp affordance*, we refer to the the different ways to place a hand or a gripper near an object so that closing the gripper will produce a stable grip. The grasps we consider are parametrized by a 6D gripper pose and a grasp (preshape) type. The gripper pose is composed of a 3D position and a 3D orientation, defined within an object-relative reference frame. We represent the affordance of an object for a certain grasp type through a continuous probability density function defined on the 6D object-relative gripper pose space $SE(3)$, similar to the approach of de Granville et al. [2]. The computational encoding is *nonparametric*: A density is simply represented by the samples we see from it. The samples supporting a density are called *particles*; the probabilistic density in a region of space is given by the local density of the particles in that region. The underlying continuous density is accessed by assigning a kernel function to each particle – a technique generally known as *kernel density estimation* [6]. The kernel functions essentially capture Gaussian-like shapes on the 6D pose space $SE(3)$ (see Fig. 1). The expressiveness of a single kernel is rather limited: location and orientation components are both isotropic, and within a kernel, location and orientation are modeled independently. Nonparametric methods account for the simplicity of individual kernels by employing a large number of them: a grasp density will typically be supported by a thousand particles. An object is linked to a separate grasp density for each type of grasp it affords, e.g. one density for pinch grasp affordance and another density for or power grasps.

The second contribution of this paper is a framework that allows an agent to learn initial affordances from various grasp cues, and enrich its grasping knowledge through experience.

Affordances are initially constructed from human imitation, or from model-based methods [1]. The grasp data produced by these *grasp sources* is used to build continuous *grasp hypothesis densities*. Given the nonparametric representation, building a density from a set of grasps is straightforward – grasps can directly be used as particles representing the density. These densities are attached to a (pre-existing) 3D visual object model [3], [5], which will allow a robotic agent to execute samples from a grasp hypothesis density under arbitrary object poses, by using the visual model to estimate the 3D pose of the object. The visual model has the form of a hierarchy of increasingly expressive object parts called *features*; the single top feature of a hierarchy represents the whole object. A hierarchy is implemented in a Markov tree, features corresponding to hidden nodes. A grasp affordance is attached to this model simply as a new *grasp feature* linked to the top feature of the network. Probabilistically speaking, this effectively stores an expression of the joint distribution $\mathbf{p}(X_o, X_g)$, where X_o is the pose object, and X_g is the grasp affordance.

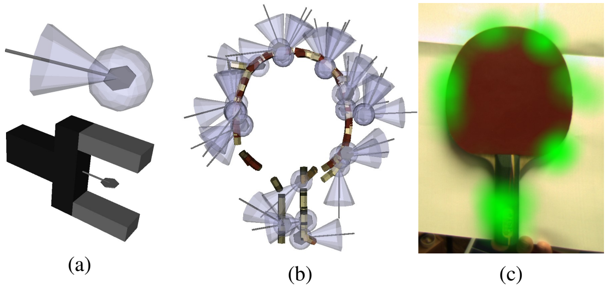
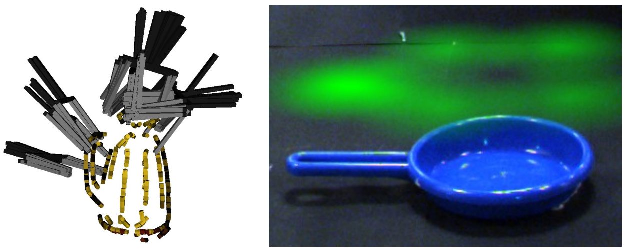


Fig. 1. Grasp density representation. The top image of Fig. (a) illustrates a particle from a nonparametric grasp density, and its associated kernel widths: the translucent sphere shows one position standard deviation, the cone shows the variance in orientation. The bottom image illustrates how the schematic rendering used in the top image relates to a physical gripper. Fig. (b) shows a 3D rendering of the kernels supporting a grasp density for a table-tennis paddle (for clarity, only 30 kernels are rendered). Fig. (c) indicates with a green mask of varying opacity the values of the location component of the same grasp density along the plane of the paddle (orientations were ignored to produce this last illustration).

Visual inference of the hierarchical model is performed using a belief propagation algorithm (BP) [4], [7], [3]. BP derives a posterior pose density for the top feature of the hierarchy, thereby producing a probabilistic estimate of the object pose. When an object model has been visually aligned to an object instance, the grasp affordance of the object *instance* is computed through the same BP inference. Intuitively, this corresponds to transforming the grasp density to align it to the current object pose, yet explicitly taking the uncertainty on object pose into account to produce a posterior grasp density that acknowledges visual noise.

Fig. 2a shows *samples* from a hypothesis density learned from imitation of human grasps at a tea jug, along with the visual 3D jug model. Fig. 2b shows another imitation-based hypothesis density, projected on a 2D image in the same way as Fig. 1.

R. Detry and J. Piater are with the University of Liège, Belgium. (Renaud.Detry@ULg.ac.be, Justus.Piater@ULg.ac.be).
E. Başeski, M. Popović, Y. Touati, and N. Krüger are with the University of Southern Denmark. (mila@mmmi.sdu.dk, emre@mmmi.sdu.dk, norbert@mmmi.sdu.dk).



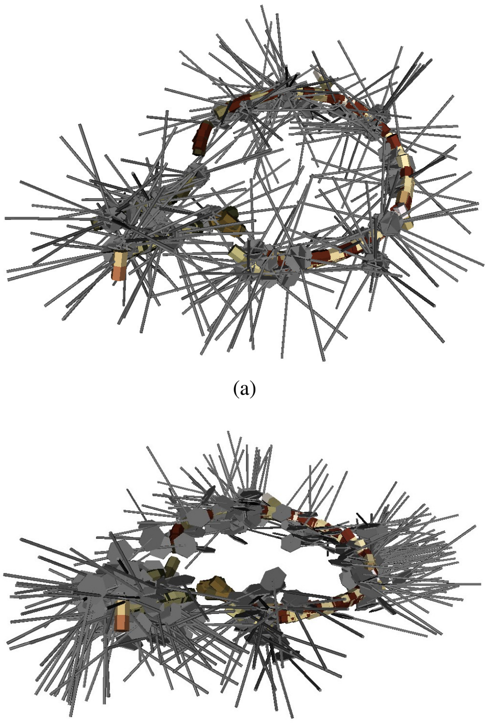
(a) (b)

Fig. 2. Hypothesis densities learned from imitation. See text for details.

As the name suggests, hypothesis densities do not pretend to reflect the true properties of an object. Their main defect is that they may strongly suggest grasps that might not be applicable at all, for instance because of embodiment discrepancies between the demonstrator and the robot in imitation-grounded hypotheses. A second, more subtle issue is that the grasp data used to learn hypothesis densities will generally be afflicted with a source-dependent spatial bias. A very good example can be made from grasps computed from a visual model [1], which will be more numerous around parts of the object that have a denser visual resolution, incidentally biasing the corresponding region of the hypothesis density to a higher value. The next paragraphs explain how grasping experience can be used to compute new densities (*empirical* densities) that better reflect object properties.

Empirical densities are leaned from the execution of *samples* from a hypothesis density, allowing the agent to familiarize itself with the object by discarding wrong hypotheses and refining good ones. Familiarization thus essentially consists in autonomously learning an empirical density from the outcomes of sample executions. A simple way to proceed is to build an empirical density directly from a set of successful grasp samples. However, this approach would inevitably propagate the spatial bias mentioned above to the new densities. Instead, we use *importance sampling* to properly weight successful grasps, allowing us to draw samples from the physical grasp affordance of an object. The weight associated to a grasp sample x is computed as $\mathbf{a}(x) \,/\, \mathbf{q}(x)$, where $\mathbf{a}(x)$ is 1 if the execution of x has succeeded, 0 else, and $\mathbf{q}(x)$ corresponds to the value of the continuous hypothesis density at x. A set of these weighted samples directly forms a grasp empirical density. Each empirical density is associated to the object model in the same way as hypothesis densities, through a new feature in the hierarchical network.

Fig. 3a shows the particles supporting a hypothesis density computed from visual cues [1]. Fig. 3b shows an empirical density learned from the hypothesis density of Fig. 3a. The feedback on grasp execution needed to build the empirical density of Fig. 3b was provided by a human teacher, through visualization of the grasps and the object in a 3D rendering software. The main evolution from Fig. 3a to Fig. 3b is the removal of a large number of grasps for which the gripper wrist collides with the object. Grasps also tend to approach the paddle along paths contained in the paddle plane, preventing fingers from colliding with the object during hand servoing.



(a)

(b)

Fig. 3. Particles from grasp densities. Fig. (a) corresponds to a hypothesis density learned from visual cues. Fig. (b) is an empirical density learned from Fig. (a). See text for details.

A unified representation of grasp affordances can potentially lead to many applications. An interesting example is their use within a grasp planner, which would combine a grasp density with hardware physical capabilities (robot reachability) and external constraints (obstacles) in order to select the grasp that has the largest chance of success within the subset of achievable grasps.

Acknowledgments

This work was supported by the Belgian National Fund for Scientific Research (FNRS) and the EU Cognitive Systems project PACO-PLUS (IST-FP6-IP-027657).

References

[1] Daniel Aarno, Johan Sommerfeld, Danica Kragic, Nicolas Pugeault, Sinan Kalkan, Florentin Wörgötter, Dirk Kraft, and Norbert Krüger. Early reactive grasping with second order 3D feature relations. In *ICRA*, 2007.

[2] Charles de Granville, Joshua Southerland, and Andrew H. Fagg. Learning grasp affordances through human demonstration. In *ICDL*, 2006.

[3] Renaud Detry, Nicolas Pugeault, and Justus H. Piater. Probabilistic pose recovery using learned hierarchical object models. In *ICVW*, 2008.

[4] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[5] Nicolas Pugeault. *Early Cognitive Vision: Feedback Mechanisms for the Disambiguation of Early Visual Representation*. Vdm Verlag Dr. Müller, 2008.

[6] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, 1986.

[7] Erik B. Sudderth, Alexander T. Ihler, William T. Freeman, and Alan S. Willsky. Nonparametric belief propagation. In *CVPR*, 2003.