Grasp Generalization Via Predictive Parts

Renaud Detry[1], **Justus Piater**[2]
[1]CVAP, Royal Institute of Technology (KTH), Stockholm, Sweden
detryr@kth.se
[2]Institute of Computer Science, University of Innsbruck, Austria
justus.piater@uibk.ac.at

Objects can be grasped in various ways. Depending on the scenario (object shape, gripper/hand, task objectives, ...), grasps that differ in object-relative pose may greatly differ in their utility. In previous work [1, 2] we introduced the concept of *grasp densities* that represent, for a given scenario, the distribution of successful, object-relative gripper poses

$$p_{X|O=\text{success}}(x), \tag{1}$$

a probability density function over the object-relative gripper poses, represented as a random variable X in $\text{SE}(3) = \text{SO}(3) \times \mathbb{R}^3$, given that the grasp outcome O is success. Using maximum-likelihood reasoning, suitable grasps can be chosen according to this density, even if parts of the density are pruned by influences such as obstructions, kinematic constraints, etc.

Grasp densities are designed to be learned empirically. In principle, a robot attemps to grasp the object a large number of times using a wide variety of object-relative gripper poses. Each successful grasp constitutes a data point drawn from the underlying grasp density (1). In practice, for reasons of efficiency, attempted grasps should be chosen in an informed manner [1]. For resampling and inference, samples are turned into a continuous density by kernel density estimation.

Grasp densities can in principle be used with any kind of 3D object model and any pose estimation algorithm. In our work, we primarily use them in conjunction with our own, learnable, probabilistic object representation [3]. It allows object detection, localization and pose estimation via probabilistic inference, and integrates seamlessly with grasp densities. Figure 1 illustrates a grasp density learned for a toy object.

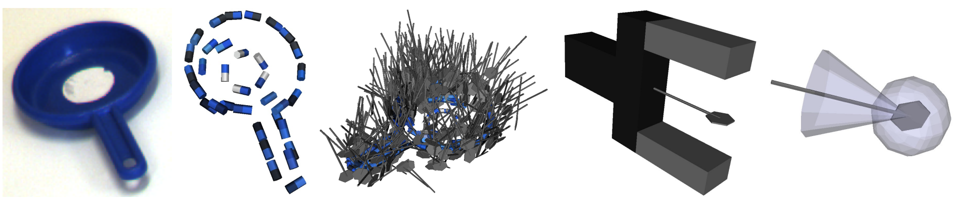

Figure 1: A toy pan, its 3D ECV representation [4], and a learned grasp density, represented as sample grasps drawn from it. Each paddle illustrates an object-relative gripper pose with an associated uncertainty, as shown on the right.

In their original form, grasp densities are associated with an entire object model. There is no explicit notion of contacts, or of which parts of the object a particular grasp is associated with. The present work aims to identify such parts, which can subsequently be used to hypothesize suitable grasps on previously unseen objects that possess similar parts. Given a library $L = \left\{o^{(i)}\right\}_{i\in[1,N]}$ of N known object models $o = (o_v, o_g)$ comprising visual features o_v and a grasp density o_g, we proceed as follows:

1. Randomly segment a set of P object parts $\{p^{(i)}\}_{i\in[1,P]}$ from L, comprising both visual features and a local section of the grasp density (Fig. 2).

2. For each part $p = (p_v, p_g)$, evaluate how well its visual model p_v fits each of the other visual models o_v in L, and how well its grasp density p_g predicts the local sections of the grasp densities o_g associated with those models. This yields a generality measure $m(p, L)$ of part p with respect to the set of known objects L.

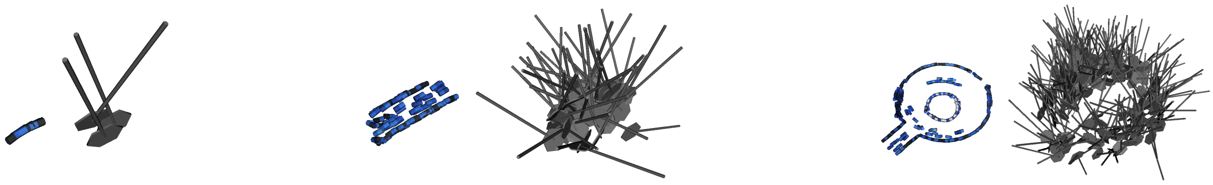

Figure 2: Parts $p = (p_v, p_g)$ of different sizes extracted from the toy pan.

Let $t_x(\cdot)$ denote a rigid transformation by $x \in SE(3)$. The degree of match of a visual part model p_v to an object o_v under an object-relative pose x is given by the cross-correlation

$$d_v(x\,;\,p_v, o_v) = \int p_v(t_x^{-1}(y))o_v(y)\,\mathrm{d}y \tag{2}$$

over the object-centered parameters $y \in SE(3)$; likewise for the grasp models p_g and o_g. A part p now induces a grasp density h on a visual object model $o_v \in L$ as

$$h(x\,;\,p_v, p_g, o_v) = \frac{1}{Z}\int d_v^c(y\,;\,p_v, o_v)p_g(t_x^{-1}(y))\,\mathrm{d}y, \tag{3}$$

where Z is a normalizing factor, and c controls the trade-off between robust prediction and generalization. Intuitively, the density h is computed as the weighted sum of all possible alignments of p_g, where the weights $d_v(y\,;\,p_v, o_v)$ are computed from visual correlation (2). The constant c controls the trade-off between robust prediction and generalization.

The ability of p to predict the grasping properties of o can be measured by the Bhattacharyya distance $f(p, o) = \int \sqrt{h(x\,;\,p_v, p_g, o_v)o_g(x)}\,\mathrm{d}x$ between h and o_g.

Finally, the generality of p with respect to the object library L is computed as

$$m(p, L) = \frac{1}{N-1}\sum_{o\in L_p} f(p, o), \tag{4}$$

where L_p is L minus the object from which p was segmented.

Highly-general parts thus identified can later be used to hypothesize grasps on novel objects by transferring the grasp density of a matching part to the object. Moreover, the set of all matching parts induces a first-guess approximation to a novel object's grasp density, which can subsequently be refined by performing grasps drawn from it.

Experimental results performed both in simulation and on a real robot suggest that parts of intermediate size possess the highest potential to generalize, and demonstrate substantially accelerated empirical acquisition of grasp densities if bootstrapped from previously-learned parts.

References

[1] R. Detry, D. Kraft, A. Buch, N. Krüger, and J. Piater, "Refining Grasp Affordance Models by Experience," in *International Conference on Robotics and Automation*, pp. 2287–2293, 5 2010. Anchorage, AK, USA.

[2] R. Detry, E. Başeski, M. Popović, Y. Touati, N. Krüger, O. Kroemer, J. Peters, and J. Piater, "Learning Object-specific Grasp Affordance Densities," in *International Conference on Development and Learning*, 6 2009. Shanghai, China.

[3] R. Detry, N. Pugeault, and J. Piater, "A Probabilistic Framework for 3D Visual Object Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 1790–1803, 10 2009.

[4] N. Pugeault, F. Wörgötter, and N. Krüger, "Visual Primitives: Local, Condensed, and Semantically Rich Visual Descriptors and their Applications in Robotics," *International Journal of Humanoid Robotics*, vol. 7, no. 3, pp. 379–405, 2010.