# Visuomotor Learning

We work on probabilistic representations and learning methods that allow a robotic agent to infer useful behaviors from perceptual observations. We develop a probabilistic model of 3D visual and haptic object properties, along with means of learning the model autonomously from exploration: A robotic agent physically experiences the correlation between successful grasps and local visual appearance by "playing" with an object; with time, it becomes increasingly efficient at inferring grasp parameters from visual evidence. Our visuomotor object model relies on (1) a grasp model representing the grasp success likelihood of relative hand-object configurations, and (2) a 3D model of visual object structure, which aligns the grasp model to arbitrary object poses (3D positions and orientations). These models are discussed below.

3D Pose Estimation and Recognition

Our visual model is a generative object representation which allows for 6D pose estimation in stereo views of cluttered scenes (Detry 2009). In this work, we model the geometry and appearance of object edges. Our model encodes the spatial distribution of short segments of object edges with with one or more density functions defined on the space of three–degree-of-freedom (3DOF) positions and 2DOF orientations. Evidence for these short segments is extracted from stereo imagery with a sparse-stereo edge reconstruction method (Biologically Motivated Multi-modal Processing of Visual Primitives, 2004, Krüger et.al); evidence is turned into edge densities through kernel methods.
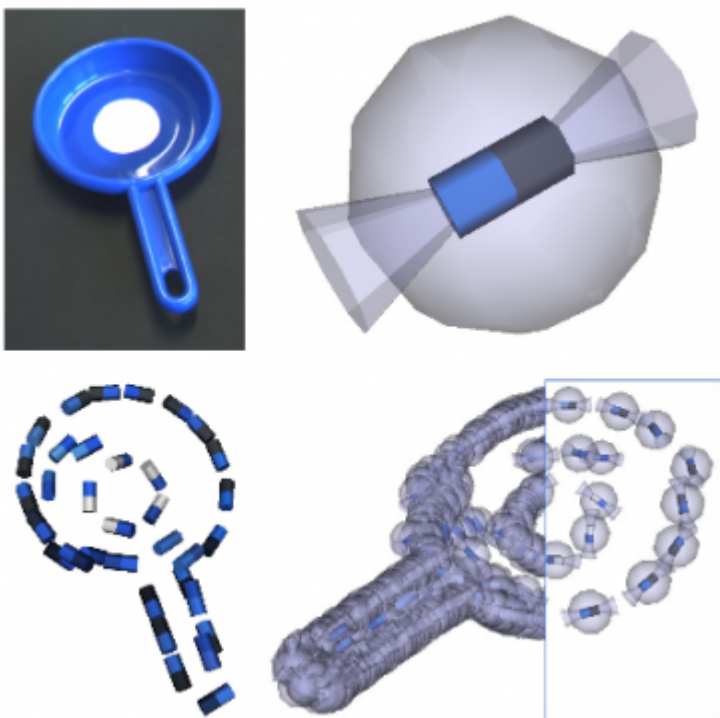
Object appearance – mainly color – is modeled by searching for clusters in the object appearance domain, then defining a different edge distribution for each color. The object model effectively consists in a Markov network containing one latent variable that represents the object pose, and a set of latent variables for the edge densities described above. Edge densities readily allow one to compute the 6DOF pose distribution of an object within an arbitrary scene by propagating scene evidence through the model. Pose inference is implemented with generic probability and machine

learning techniques – belief propaga

Pose estimation. The right image shows the maximum-likelihood pose for the toy pan, which is extracted from the largest mode of the pose distribution over the scene shown on the left.

Interactive, Visuomotor Learning of Grasp Models

We are also working on the modeling and learning object grasp affordances, i.e. relative object-gripper poses that yield stable grasps. These affordances are represented probabilistically with grasp densities (Detry 2010, which correspond to continuous density functions defined on the space of 6D gripper poses – 3D position and orientation.

Projection of a 6DOF grasp density on a 2D image. Grasp success likelihood is proportional to the intensity of the green mask.

Grasp densities are linked to visual stimuli through registration with a visual model of the object they characterize, which allows the robot to grasp objects lying in arbitrary poses: to grasp an object, the object's model is visually aligned to the correct pose; the aligned grasp density is then combined to reaching constraints to select the maximum-likelihood achievable grasp. Grasp densities are learned and refined through exploration: grasps sampled randomly from a density are performed, and an importance-sampling algorithm learns a refined density from the outcomes of these experiences. Initial grasp densities are computed from the visual model of the object.

Combining the visual model described above to the grasp-densities framework yields a largely autonomous visuomotor learning platform. In a recent experiment (Detry 2010), this platform was used to learn and refine grasp densities. The experiment demonstrated that the platforms allows a

robot to become increasingly efficient at inferring grasp parameters from visual evidence. The experiment also yielded conclusive results in practical scenarios where the robot needs to repeatedly grasp an object lying in an arbitrary pose, where each pose imposes a specific reaching constraint, and thus forces the robot to make use of the entire grasp density to select the most promising achievable grasp. This work led to publications in the fields of robotics (Detry 2010) and developmental learning (Detry 2009).

Acknowledgments

From:
https://iis.uibk.ac.at/ - **IIS**

Permanent link:
https://iis.uibk.ac.at/research/visuomotor-learning?rev=1329330962

Last update: **2018/09/03 14:57**