Discovering Morphology from Action Observation

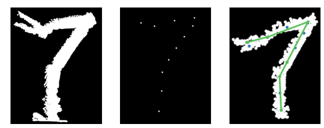
Neil R. Bell, Tim Oates, Cynthia Matuszek
bell5|oates|cmat@umbc.edu
Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County, Baltimore, Maryland 21250

Task assignment in heterogeneous groups often relies on the existence of communication or shared models of capability, either between members or on the part of a centralized planner. However, in practice, when two robots are brought together, the effort required to construct shared action and structure models is significant; in *ad hoc* teams, hand-crafting shared planning and communication solutions on a per-team basis becomes impractical.

One approach is to enable robots to learn about one another solely from observation. We describe work on deriving a model of a robot's capabilities by estimating its structure from observations of its actions, adapting confidence as new observations occur. We continuously model physical structure based on sensor observations; this morphology can then be used to plan an optimal strategy to achieve a cooperative goal.

We consider a manipulator represented as an unknown number of non-branching rigid links connected by joints with an unknown range of motion. Observations are taken using a low-cost depth sensor. For any given frame, the point cloud is reduced to a set of centroids from which structural hypotheses can be exhaustively generated and ranked by frequency of occurrence. The initial probability space is uniform, that is, the observed system may have any number of links.

Given a single observation, the point cloud is skeletonized using K-Means and possible decompositions of the skeleton into hypothetical structures are generated (see Figure 1). These hypotheses are pruned to eliminate those that represent extremely unlikely robot shapes. This process generates many hypotheses from a single observation, merging them into a statistical model that represents the overall likelihoods of each, which converges as additional observations are obtained. With a flexible enough decomposition algorithm, our statistical modeling should be robust in handling diverse structures.



(a) (b) (c)

Fig. 1. Application on a JACO arm using a Kinect 2. (a) Original point cloud; (b) centroids after K-Means decomposition; (c) hypothesized morphology after one frame.

Our work is most similar to, and is informed by, a rich body of work on articulated pose estimation [3] and kinematic modeling [2]. However, we are not aware of other work on estimating the capabilities of an unknown robot by estimating morphology from observation. Our approach is unsupervised; structural hypotheses are exhaustively generated and ranked by frequency of occurrence. In future, we plan to explore initializing our learning model using a supervised approach in which we use a simulator to generate a large number of arms with varying morphologies and learn a mapping from observables (e.g., simulated point clouds) to capabilities using an auto-encoder for which the input is a depth images concatenated with a vector of link lengths padded with zeroes [1] or RNNs.

[1] A. Kumar and T. Oates. Connecting deep neural networks with symbolic knowledge. *30th Int'l Joint Conference on Neural Networks*, 2017.
[2] L. Sciavicco and B. Siciliano. *Modelling and control of robot manipulators*. Springer Science & Business Media, 2012.
[3] Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In *Computer Vision and Pattern Recognition*, 2011.