Scene Understanding using Part-Based Object Affordances

Jana Pavlasek[1], Karthik Desingh[2], Odest Chadwicke Jenkins[1,2]

I. Introduction

The ability to effectively use tools would enable robots to perform assembly, repairs and maintenance of equipment in remote locations. To effectively perform tasks involving tool use in these situations, a robot needs to identify, localize, grasp and manipulate the object that is most suitable for a given task in an unstructured, cluttered environment. For tasks such as these, it is useful to represent the scene as a collection of all possible actions the robot can take. In other words, the scene can be represented as a collection of *object affordances* [1], rather than in terms of object classes. Knowledge of the object affordances as well as the pre- and post-conditions of the actions facilitates high level task planning along with low-level motion planning towards a given goal.

Tool affordances are inherently attached to parts of the tool, for example, a screwdriver handle affords grasping and the tip affords driving screws. Therefore, localizing the parts of the object individually facilitates scene understanding in terms of object affordances. A parts-based representation of the objects is robust to occlusions due to clutter or self-occlusions because the object model can be used to infer the location of occluded parts. The parts-based method also allows us to obtain an efficient pose estimate of the object by constraining the belief to hypotheses which respect the object part model. Accurate pose estimation of the objects allows the robot to calculate grasps needed to carry out the task.

II. Methodology & Previous Work

Deformable part models have been used extensively in the literature to represent objects in terms of their parts [2]. This is typically done over 2D data where models are represented using geometric constraints and geometric object features. To represent the relation between object parts, we propose to instead use semantic relationships which describe the object model, for example, the head of a hammer is attached to the end of the handle. This method generalizes over object instances without requiring a precise description of the object geometry for each instance. In addition, we propose to demonstrate this in 3D, with RGB-D data. The semantic representation of the objects is more interpretable to human users. Semantic relationship descriptions are also compatible with human understandable task goals.

We model the object part poses using a Markov Random Field (MRF) and perform efficient inference on the graph using Pull Message Passing for Nonparametric Belief Propagation (PMPNBP) [3]. PMPNBP is a Bayesian inference method which maintains a belief over object pose hypotheses as a set of particles. In this work, the messages are constructed using potentials describing correspondence of the current observation to a given hypothesis and the agreement of the hypothesis to the parts-based model described by semantic relations. Previous work has shown the efficiency of using PMPNBP to estimate the pose of articulated objects represented as an MRF in terms of their parts. Additional potentials that encode semantic relationships provided as prior knowledge are used to improve robustness and consistency in the scene estimation, as in [4]. The formulation of the parts-based estimation can be extended to represent inter-object semantic relationships given additional prior knowledge of the domain for more efficient full scene estimation (ie the hammer is on the table and next to the tool chest).

III. Conclusion

In this work, we aim to demonstrate a method for understanding scene as a collection of available affordances which encodes semantic information seamlessly and is efficient and robust to occlusions. This will enable a robot to make use of the object part affordances towards completion of a given task, such as tool use for equipment assembly.

References

[1] J. J. Gibson, "The theory of affordances," in *Perceiving, acting and knowing: toward an ecological psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers, 1977, pp. 67–82.

[2] P. F. Felzenszwalb, R. B. Girshick, D. Mcallester, and D. Ramanan, "Object detection with discriminatively trained part based models," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, pp. 1627–1645, 2009.

[3] K. Desingh, S. Lu, A. Opipari, and O. C. Jenkins, "Efficient nonparametric belief propagation for pose estimation and manipulation of articulated objects," *Science Robotics*, vol. 4, no. 30, 2019.

[4] Z. Zeng, Y. Zhou, O. C. Jenkins, and K. Desingh, "Semantic Mapping with Simultaneous Object Detection and Localization," in *IEEE International Conference on Intelligent Robots and Systems*, 2018, pp. 911–918. [Online]. Available: https://arxiv.org/pdf/1810.11525.pdf

[1]Robotics Institute, University of Michigan, Ann Arbor, USA

[2]Department of Electrical Engineering & Computer Science, University of Michigan, Ann Arbor, USA

{pavlasek,kdesingh,ocj}@umich.edu