Exemplar-based Prediction of Object Properties from Local Shape Similarity

Jeannette Bohg*, Daniel Kappler* and Stefan Schaal*†
*Autonomous Motion Department
Max Planck Institute for Intelligent Systems, Tübingen, Germany
Email: <first.lastname>@tue.mpg.de
†Computational Learning and Motor Control lab
University of Southern California, Los Angeles, CA, USA
Email: sschaal@usc.edu

I. Introduction

Autonomous grasping of any kind of object in arbitrarily complex environments is still unattainable for today's robots. There have been great advances in robust grasping and even manipulation of *known* objects in environments of moderate complexity, e.g. by Righetti et al. [18], Hudson et al. [10], Kazemi et al. [12]. However, the higher the uncertainty about crucial aspects of a manipulation task, the harder it becomes for the robot to successfully plan and execute its actions.

For example, there are theoretically well-founded metrics to evaluate the performance of a grasp given complete information about the object, the hand and their relative poses [8] that are implemented in all the major simulators such as GraspIt! [16] or OpenRave [6]. But how to let a robot grasp an object of uncertain global shape is an active area of research. There is little agreement in the community on how to best represent partial object information and infer a grasp given this. This is however a very common problem in real-world scenarios. Especially in cluttered scenes, large parts of an object may be occluded and segmentation of the visible parts from its surroundings becomes more difficult. A comprehensive overview of the different approaches towards this problem is given by Bohg et al. [3].

There are a few methods that try to estimate global object shape from partial information, e.g. [2, 1, 7]. They are often motivated to provide the basis for grasp planning methods that assume knowledge of a full object model.

Other methods do not attempt to predict global object shape but rather to predict graspability directly from the partial and local information. Commonly, these methods employ supervised learning techniques on annotated grasp experience databases to predict where and how to grasp an object in a scene [19, 17, 5, 15, 9, 13, 14]. Local methods have several advantages over method representing global object shape. They allow to generalize learned models across different objects that may have a very different global shape but are locally similar. Because they only rely on local information, they are also less sensitive to segmentation errors or occlusions. Furthermore, no prior semantic knowledge on e.g. object identity or category is necessary. All these factors reduce the complexity of information extraction from raw sensory data. However, global object shape has a large influence on whether a grasp will succeed or not. This is naturally not captured by local information. If two objects share two similar parts but have otherwise vastly different global shapes, different grasps may be required. In this paper, we propose a method that predicts (i) graspability and (ii) global object shape given only local information. Thereby, we inherit the advantages of local methods and still yield a prediction of global object shape. This can form the input to subsequent grasp planners or controllers.

II. Approach

An overview of the proposed system is shown in Fig. 1. In detail, we propose a novel method that enables a robot to infer a grasp pre-shape for an object of unknown identity, category or shape given only noisy and partial information that is obtained from an RGB-D camera. In line with some of the aforementioned related work [19, 15, 9, 13, 14], we formulate this as a classification problem that takes in a local shape representation and outputs the probability of a grasp applied at this location to be successful. We learn the function that maps our local shape feature to grasp stability based on a recently proposed large-scale synthetic database Kappler et al. [11]. In total, it contains around 500k data points of annotated local shape representation, referred to as *templates* throughout the remainder of this abstract. This is by far the largest dataset available in the community with stability metrics that are verified through crowdsourcing.

As a classifier, we use a Random Forest.

For our second aim of predicting global object shape from local information we exploit the ability of the Random Forest to cluster the dataset into locally similar templates. Given a classified query template, obtained from the trained Random Forest, we can extract all the exemplars of the training data set that ended up at the same leaf nodes. Since we have access to complete information about these exemplars, we can use it to make predictions about some latent properties of the target object. These can be object categories but also global object shape. Here, we will focus on the latter as this constitutes important information for grasp planners and robot controllers.

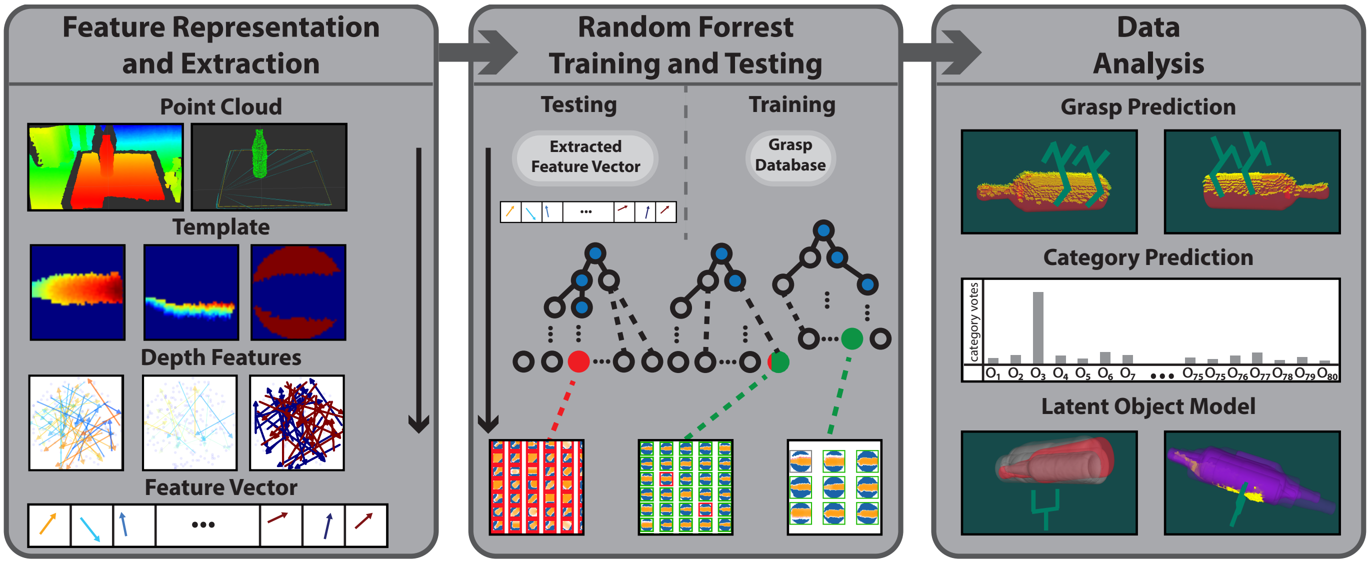

Fig. 1. Overview of the proposed system from extracting the point cloud, suggesting a grasp to predicting global shape. **(Left)** A point cloud of the whole scene is recorded with an RGB-D camera. The point cloud needs to be sampled for grasp candidates. We use a simple table-plane based segmentation to restrict the sampling to the remaining clusters on top of the plane. Each grasp candidate is represented by a local shape representation: *template*. Here we show the three channels this template consists of: surface, occlusion and free space. We extract a feature from each channel of this template. It consists of pairs of probes as also used in [20, 4]. and is stacked into one feature vector. This feature vector is then used for training and at test time. **(Middle)** The feature vector serves as an input to a Random Forest Classifier which has been trained offline on a database. By averaging over the response of each tree in the forest, the input feature vector is classified as either stable or not. Additionally, each leaf node at which a query data point ends up is associated with a subset of the training data. This is shown in the 'mosaics' of the bottom row in which each square represents a template of a training data point. A green frame indicates a positive and a red frame a negative example. **(Right)** This locally similar data can then not only be used for predicting the set of stable grasps per point cloud, but also for predicting other global object properties such as object category and global object shape. Regarding the latter, the left image shows a shape distribution (gray) given synthetic data from the database as input. Since ground truth object models (red) are available for all test data points in the database, we can quantitatively evaluate how well this distribution predicts global object shape. On the right, you see an example for the prediction of global object shape for real point clouds.

Specifically, we model the global shape of the target object as a non-parametric distribution that is populated with the mesh models of the retrieved training data points. As these point ended up in the same leaf nodes of the random forest, their feature description must be similar to the one of the query data point. Therefore, they are locally similar to each other. We are interested in analysing if they are also globally similar.

III. Results

In experiments, we will show that our trained model for predicting grasp stability achieves the same performance as the current state-of-the-art on this data set which uses a Convolutional Neural Net (CNN). In terms of our second aim, we can quantitatively show that global shape of unknown objects can be coherently predicted from locally similar training data points. We also show that this works particularly well when only considering positively-labeled exemplars, i.e. exemplars that yielded a successful grasp. This suggests that they have a particularly high predictive power for global object shape. This makes sense, as global object shape has a high influence on grasp success. Furthermore, we show qualitative examples of grasp retrieval and object shape prediction on real data.

References

[1] M. Bjorkman, Y. Bekiroglu, V. Hogman, and D. Kragic. Enhancing visual perception of shape through tactile glances. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ Int. Conf. on*, pages 3180–3186, Nov 2013.

[2] Jeannette Bohg, Matthew Johnson-Roberson, Beatriz León, Javier Felip, Xavi Gratal, Niklas Bergström, Danica Kragic, and Antonio Morales. Mind the Gap - Robotic Grasping under Incomplete Observation. In *IEEE Int. Conf. on Robotics and Automation*, May 2011.

[3] Jeannette Bohg, Antonio Morales, Tamim Asfour, and Danica Kragic. Data-driven grasp synthesis: A survey. *IEEE Transactions on Robotics*, 2014.

[4] Jeannette Bohg, Javier Romero, Alexander Herzog, and Stefan Schaal. Robot arm pose estimation through pixel-wise part classification. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, Hongkong, China, 2014.

[5] Renaud Detry, Carl Henrik Ek, Marianna Madry, and Danica Kragic. Learning a dictionary of prototypical grasp-predicting parts from grasping experience. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2013.

[6] Rosen Diankov. *Automated Construction of Robotic Manipulation Programs*. PhD thesis, CMU, Robotics Institute, Aug 2010.

[7] Stanimir Dragiev, Marc Toussaint, and Michael Gienger. Uncertainty aware grasping and tactile exploration. In *IEEE Int. Conf. on Robotics and Automation*, 2013.

[8] C. Ferrari and J. Canny. Planning optimal grasps. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 1992.

[9] A. Herzog, P. Pastor, M. Kalakrishnan, L. Righetti, J. Bohg, T. Asfour, and S. Schaal. Learning of grasp selection based on shape-templates. *Autonomous Robots*, 2013.

[10] Nicolas Hudson, Thomas Howard, Jeremy Ma, Abhinandan Jain, Max Bajracharya, Steven Myint, Calvin Kuo, Larry Matthies, Paul Backes, Paul Hebert, Thomas J. Fuchs, and Joel W. Burdick. End-to-end dexterous manipulation with deliberate interactive estimation. In *IEEE Int. Conf. on Robotics and Automation, ICRA 2012, 14-18 May, 2012, St. Paul, Minnesota, USA*, pages 2371–2378, 2012.

[11] Daniel Kappler, Jeannette Bohg, and Stefan Schaal. Leveraging big data for grasp planning. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2015.

[12] Moslem Kazemi, Jean-Sebastien Valois, J. Andrew Bagnell, and Nancy S. Pollard. Robust object grasping using force compliant motion primitives. In *Robotics: Science and Systems VIII, University of Sydney, Sydney, NSW, Australia, July 9-13, 2012*, 2012.

[13] O. Kroemer, E. Ugur, E. Oztop, and J. Peters. A kernel-based approach to direct action perception. In *Int. Conf. on Robotics and Automation (ICRA)*, 2012.

[14] Quoc V. Le, David Kamm, A. F. Kara, and Andrew Y. Ng. Learning to grasp objects with multiple contact points. In *ICRA*, pages 5062–5069, 2010.

[15] Ian Lenz, Honglak Lee, and Ashutosh Saxena. Deep learning for detecting robotic grasps. *The Int. Jour. of Robotics Research (IJRR)*, 2014.

[16] A. T. Miller and P. K. Allen. Graspit! a versatile simulator for robotic grasping. *Robotics & Automation Magazine, IEEE*, 2004.

[17] Mila Popović, Gert Kootstra, Jimmy Alison Jørgensen, Danica Kragic, and Norbert Krüger. Grasping unknown objects using an early cognitive vision system for general scene understanding. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, San Francisco, USA, 2011.

[18] Ludovic Righetti, Mrinal Kalakrishnan, Peter Pastor, Jonathan Binney, Jonathan Kelly, Randolph Voorhies, Gaurav S. Sukhatme, and Stefan Schaal. An autonomous manipulation system based on force control and optimization. *Auton. Robots*, 36(1-2):11–30, 2014.

[19] Ashutosh Saxena, Justin Driemeyer, and Andrew Y. Ng. Robotic grasping of novel objects using vision. *The Int. Jour. of Robotics Research (IJRR)*, 27(2):157–173, February 2008.

[20] Jamie Shotton, Ross Girshick, Andrew Fitzgibbon, Toby Sharp, Mat Cook, Mark Finocchio, Richard Moore, Pushmeet Kohli, Antonio Criminisi, Alex Kipman, and Andrew Blake. Efficient human pose estimation from single depth images. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 99:1, 2012.