# Template-Based Learning of Grasp Selection

Alexander Herzog, Peter Pastor, Mrinal Kalakrishnan, Ludovic Righetti, Tamim Asfour, Stefan Schaal

Autonomous robotic grasping is one of the pre-requisites for personal robots to become useful when assisting humans in households. Seamlessly easy for humans, it still remains a very challenging task for robots. An essential aspect of robotic grasping is to automatically choose an appropriate grasp configuration given an object as perceived by the sensors of the robot. The high variety in the size and geometry of objects to be grasped (for example, household objects, see Fig. 4) makes it very hard to develop an algorithm that provides promising grasp hypotheses. Model free approaches have been proposed that directly operate on point clouds provided by 3D sensors (for example a stereo camera system, or the Microsoft Kinect). Hsiao et. al. developed an algorithm that searches among feasible top and side grasps to maximize the amount of object mass between the finger tips of the robot gripper [1]. Klingbeil et. al. developed an algorithm that searches for a good grasp configuration by maximizing the contact area between the robot's gripper and the perceived point cloud [2]. Both these approaches generate a ranked list of grasp hypotheses suitable for execution on the robot. Usually, the ranking of these grasp hypotheses is fixed and does not adapt over time. Thus, whenever the robot is required to grasp a particular object for which the best grasp hypothesis fails, the robot will need to retry choosing different grasp candidates from the list every single time. Such algorithms lack the ability to adapt and improve the ranking of grasp hypotheses based on previous successful grasp executions. Furthermore, these algorithms do not allow to input an appropriate grasp hypothesis if none of the generated grasp configurations leads to a successful grasp. In this paper, we propose a novel model free grasp selection algorithm that overcomes these limitiations. Our algorithm has the following favorable characteristics:

- Appropriate grasp configurations can be taught through kinesthetic teaching.
- Our algorithm is able to autonomously improve the ranking of generated grasp candidates over time based on feedback from previous grasp executions.
- Our proposed local shape descriptor, the template, encodes regions on the object that are suitable for grasping such that it generalizes accross different objects.
- Finally, our proposed method is computationally efficient.

Our approach is based on the simple assumption that similar objects can be grasped with similar grasp configurations. For example, a pen can be grasped from the table with a strategy similar to that used to grasp a screwdriver of the same size. To measure such similarities between objects

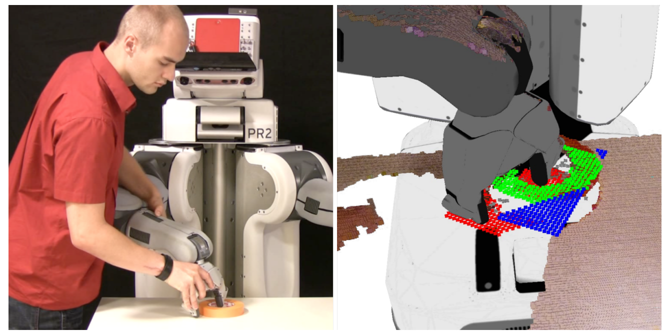

Fig. 1. User demonstrates an appropriate grasp candidate to the PR2 robot (left). A visualization of the corresponding template extracted from the perceived point cloud along with the corresponding gripper pose is shown on the right.

we propose a local shape descriptor that we refer to as a *template*. Recently, templates have been successfully used to encode local regions of terrains enabling a quadruped robot to choose good footholds [3]. In [3], templates have been used to encode terrain heightmaps. In contrast, in our work, we use templates to encode object heightmaps that are sampled from various heightmap axes (purple arrows in Fig. 3). The algorithm is initialized by teaching the robot a set of grasp configurations and storing the extracted templates along with the associated gripper pose in the template library (see Fig. 1). Grasp hypotheses for a new object are generated by sampling candidate templates from the perceived object

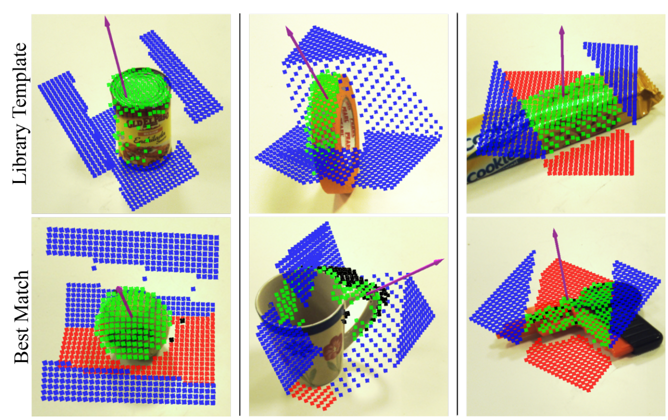


Fig. 3. A *template* consists of a raster of height values. Additionally, each tile also contains information about whether it is one of (1) *object surface*: points on the object (green), (2) *background*: points that do not belong to the object, e.g. table (red), (3) *empty regions*: points that are outside the bounding box of the gripper (blue), or (4) *self-occluded regions*: points that may or may not be part of the object and are not directly visible from the current view angle (black). The top row shows templates contained in the library that have been learned from demonstration; the bottom row shows the best corresponding match for new objects.

point cloud and matching them against the templates stored in the library. The grasp configuration associated with the best match is used as a grasp hypothesis for the new object. Candidate templates are sampled from the convex hull of the object point cloud, one template for each face, at different rotations around the heightmap axis. A template consists of $n \times n$ tiles, each of which contains a height value and one out of four types: *object surface*, *background*, *empty regions*, *self-occluded* (see Fig. 3). A similarity measure between two templates is computed using a weighted $\ell_1$ distance between the height values of the templates. The weights are selected according to the types of the tiles. We use only two different weights for the $4 \times 4 = 16$ possible combinations. A higher weight is assigned to tile combinations which include a surface type. This gives more importance to differences in heights of tiles that contain an object. A normalizing factor is used to multiply the weighted distance to make templates with different number of surface tiles comparable.

The proposed algorithm is also able to learn from failed trials. After each failed trial, we store the template that led to the failure in the library. At each new grasp attempt we compute the similarity between the new candidate templates and the failed ones in order to improve the ranking.

We evaluated the performance of the proposed grasp selection algorithm on the Willow Garage PR2 robot. The experimental setup consisted of the robot positioned in front of a table. A calibrated Microsoft Kinect sensor mounted on the PR2 robot was used to obtain dense point clouds from the object. Initially, a user demonstrated a total of 15 feasible grasps on 7 different household objects (see Fig. 1). The test set consists of 38 objects shown in Fig. 4. We placed each object in various positions and orientations on the table, one at a time. The best ranked feasible grasp hypothesis provided by our algorithm was chosen for execution on the robot. A particular grasp was considered a success if the robot was able to lift the object off the table, indicating a stable grasp. Our algorithm achieved a success rate of 87%, i.e. 83 out of 95 grasp hypotheses lead to a stable grasp. A subset of the achieved grasps are shown in Fig. 2 as well as in the video supplement. The results indicate that our template representation is able to generalize from only 15 demonstrations to a large variety of objects (see Fig. 4). The reason for the few unsuccessful grasps

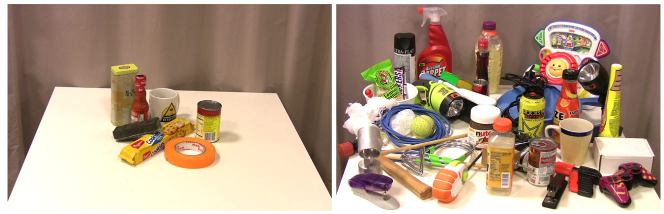

Fig. 4. Seven objects used in the training set (left) and 38 objects used as test set to evaluate the proposed algorithm (right).

is mainly due to slightly mis-oriented grasp poses. For example when grasping the whiteboard marker, premature contact of the gripper caused the marker to roll away. However, our algorithm was able to use feedback from previous unsuccessful trials to autonomously improve the grasp selection process. As shown in the video supplement, the robot was able to autonomously improve the grasp selection leading to only successful grasp attempts. All processes were running onboard the PR2 robot and the algorithm took between 7 and 30 seconds to generate new grasp hypotheses. Therefore the algorithm is fast enough to be used in a realtime planning pipeline.

In this article we presented a template-based grasp selection algorithm which uses demonstrated grasp configurations and generalizes them to novel objects. We showed that the proposed template representation is able to capture object features enabling our algorithm to achieve a good success rate on a challenging set of differently shaped objects. Additionally, the proposed method was able to improve the grasp selection process autonomously through trial-and-error. The algorithm has been implemented efficiently and results have been presented on the PR2 robot. Future work includes applying the proposed method to more complex robot hands, e.g. the Barrett three finger hand.

## References


[1] K. Hsiao, S. Chitta, M. Ciocarlie, and G. E. Jones, “Contact-reactive grasping of objects with partial shape information,” in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems*, 2010.

[2] E. Klingbeil, D. Drao, B. Carpenter, V. Ganapathi, O. Khatib, and A. Y. Ng, “Grasping with application to an autonomous checkout robot,” in *Proc. of the IEEE Intl Conf. on Robotics and Automation*, 2011.

[3] M. Kalakrishnan, J. Buchli, P. Pastor, and S. Schaal, “Learning locomotion over rough terrain using terrain templates,” in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems*, 2009, pp. 167–172.


Fig. 2. Subset of achieved grasps on the test set.