# 3D-Compositional Model associated with Grasp Templates and Haptic Data

**Alexander Rietzler** ALEXANDER.RIETZLER@UIBK.AC.AT
*Institute of Computer Science*
*University of Innsbruck*
*Innsbruck, Austria*

## 1. Problem Statement

The problem dealt with here is to associate graspability and haptic characteristics to parts produced by a 3D compositional models in order to perform robust grasping of even unknown objects exhibiting known parts. The work is strongly tied to requirements of Task 2.1 and 2.2 of the PaCMan project Workpackage 2.

## 2. Experimental Setup

As a vision sensor, the depth sensor Microsoft Kinect is used. The 3D pointcloud data captured by the sensor is represented as a continuous surface distribution in SE(3) (Detry and Piater, 2010).

In order to perform grasp experiments and to gather tactile data our setup is equipped with a Kuka Lightweight Robot and a Schunk Dextrous Hand 2 with 6 tactile arrays at its 6 finger links.

First a set of successful grasps are shown on a training set, capturing object-relative wrist pose of the gripper, tactile signature and kinesthetic signature (gripper joints), see Figure 1.

(a) Training set (b) Test set

Figure 1: IKEA objects used for training and test set.

After recognizing meaningful parts of the training objects with the 3D-compositional model (Rezapour-Lakani et al., 2014) the grasps including haptic characteristics are associated to the part the grasp is performed on.

To evaluate the system, a grasp experiment is performed measuring grasp success rates.

## 3. Method

### Part and Grasp Representation

The method chosen builds on part recognition software developed by Rezapour-Lakani et al. (2014). Given a novel scene, the software segments unseen objects containing similar parts into their known parts. Each part belongs to a previously learned part-cluster marked with a cluster id. To restrict grasp template alignment to the associated part, its oriented bounding box (OBB) is computed and passed to a grasp inference algorithm.

To each part-cluster a set of grasps is attached. In order to match grasps on an unseen but familiar part, the shape-relevant information of graspability is captured by a 3D template $G(x)$, where each point $x$ is represented by a 6D Kernel resulting in a continuous shape distribution. The template is generated by cropping the shape distribution with a cube-shaped region of interest (ROI) close to the gripper contact points during a trained grasp. Grasps are further represented by a pre-grasp pose and a retreat-pose, and by three different grasp types – pinch, spherical and parallel. Figure 2 shows a set of selected templates.

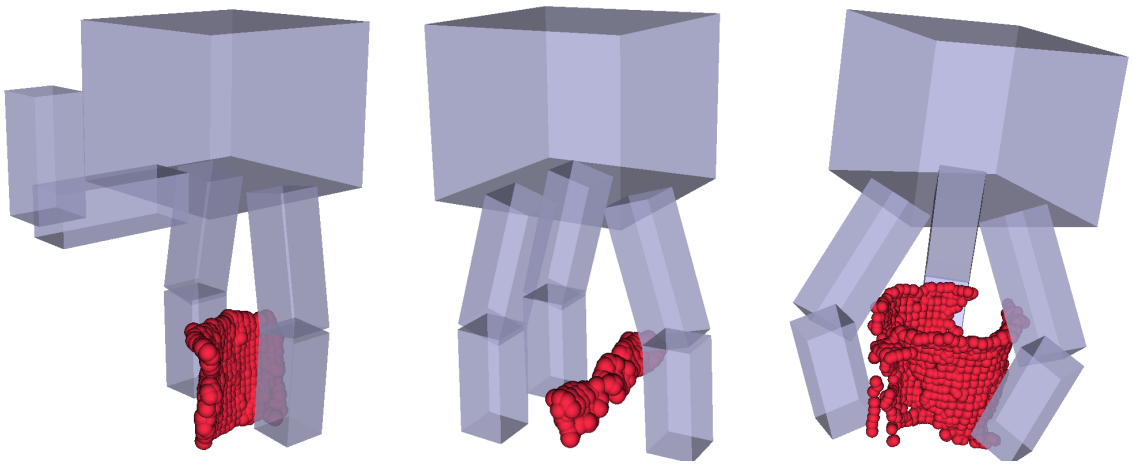


Figure 2: Three selected templates for visualization. From left to right: body-pinch (container), handle-parallel (pot), body-spherical (mug).

Tactile data is gathered during a training grasp. Each of the six phalanges of the gripper is equipped with a tactile array of 6×14 taxels for proximal and 6×13 for distal phalanges, and a sensor value ranging from 0 to 5000, which is proportional to taxel pressure, here treated as continuous values. For our training grasps only the distal arrays are active, which yields $3 \times (13 \times 6)$ continuous values.

### Grasp Inference

The grasp template alignment procedure is based on a pose estimation algorithm,[1] where the 6D pose-dependent cross-correlation between the shape distribution of the scene $Q(x)$ and the template $G(x)$ is computed as:

$$S(G,Q) = \int G(x)Q(x)\,\mathrm{d}x \tag{1}$$

This integral is computed by Monte Carlo approximation. In addition, to force the alignment close to the selected part only the OBB of the grasp-template ($G$) has to intersect with the OBB of the selected part ($P$):

$$C(G,P) = \begin{cases} 1 & OBB(G) \cap OBB(P) \neq \{\} \\ 0 & OBB(G) \cap OBB(P) = \{\} \end{cases} \tag{2}$$

As the output of our grasp inference algorithm computes a single best grasp per template, we check for collisions with the environment as well as for reachability of the gripper wrist pose hypothesis (r,t) via inverse kinematics. The reachability term reads as

$$R_{r,t} = \begin{cases} 1 & (r,t) \text{ reachable} \\ 0 & (r,t) \text{ not reachable} \end{cases} \tag{3}$$

Combing equations (1–3), our optimization over wrist poses yields

$$S^*_{r^*,t^*} = \mathrm{argmax}_{r,t}\{S(T_{r,t}(G),Q) \cdot C(T_{r,t}(G),P) \cdot R_{r,t})\} \tag{4}$$

where $T_{r,t}(\cdot)$ denotes a rigid body transformation.

### Success Classification

To measure the additional value of using haptic data for grasping, we calculate success prediction performance based on haptics. We represent grasp success as categories $\{1 = \text{success}, -1 = \text{failure}\}$. Grasps that are unstable in the sense that the object could be lifted but moved loosely in hand after pickup are marked as a failure.

As raw data, we measure:

- Tactile: 3 fingers × 13×6 sensor matrix: 234 sensor values
- Kinesthetic: 7 joints: 7 sensor values
- Total: 241 sensor values

To classify the performance, we use the following features as motivated in Section 4.

- Centroid of the active sensor matrix: $S = \left(\frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}}\right)$, where $M_{ij} = \sum_x \sum_y x^i y^j I(x,y)$ are the raw image moments

1. `http://nuklei.sourceforge.net`

- Average pressure per active taxel: $\frac{M_{00}}{N}$
- Sum of absolute difference of measured gripper joint positions and target joint positions $\sum_i |q_{mes}(i) - q_{learned}(i)|$
- Total number of features: 10

To capture the assumed nonlinearities in classification we choose Gaussian Process Classification with a diagonal square-exponential covariance function, which is proven to be robust with respect to a large number of feature dimensions (large compared to our dataset size).

As our grasp inference algorithm computes an alignment score $S^*$, we compare grasp success prediction performance based on haptic features and $S^*$.

## 4. Experiment and Results

For our experiment we chose 3 training objects from the PaCMan IKEA dataset, where we performed 11 training grasps in total to generate the grasp templates, see Figure 1(a). We associated each of the training grasps with the corresponding part-clusters, which were recognized on the training objects.

The two meaningful part-clusters that were generated contained mostly body-parts and handle-parts respectively, associated with 5 and 4 templates each. For the reason of fitting a template in a stable manner, we chose to perform our grasp experiments only with templates from the part-cluster containing body-parts.

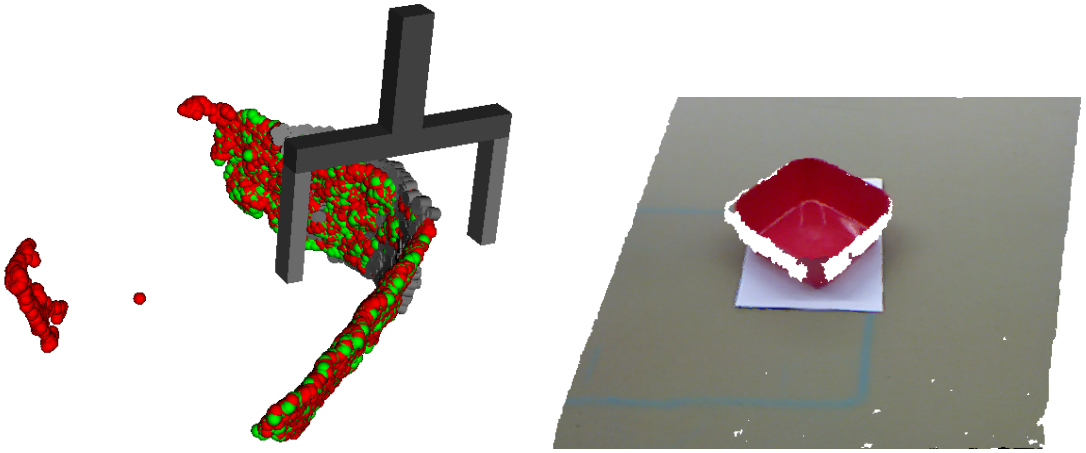
Our experiment was performed in scenes containing a single object. For each object we used the 3D compositional model to compute the parts, fitted all the 5 templates on the parts, and executed their grasp. Figure 3 shows a scene with a selected part and corresponding aligned grasp template with highest alignment score.

In total, we planned 50 grasps on 12 parts of 9 different objects. To measure our grasp success rate we make a distinction between test and training objects. Not all grasps could be executed due to planning failures. For training objects we executed 19 grasps, where 15 of them were successful. This gives a success rate of 79%. For novel objects we executed 21 grasps, where 11 of them were successful, which a success rate of 52%. Note that in a real application scenario success rates would be higher since we would choose the grasp with highest alignment score for an object, whereas here we execute all grasps associated with a given part. The goal of this experiment is to obtain a balanced training set for success classification. Still the experiment shows that, especially for highly similar objects like mug-blue and mug-brown, the grasp transfer works very well. A summary of the experiment can be found in Table 1.

### Success Prediction

We use Gaussian Process classification to predict grasp success rates as mentioned in Section 3. Our dataset is slightly imbalanced with 26 positive compared to 14 negative samples.

To evaluate the performance we use leave-one-out cross-validation and compute accuracy, precision, recall and F1 measure. In addition we compute the same performance



(a) Pointcloud (b) RGB image

Figure 3: Instantiated scene containing the red box. Points in green show the selected part the grasp template is aligned to. Points in gray show the aligned grasp template. The gripper indicates wrist pose and grasp type. Here a pinch grasp has the highest alignment score.

| Executed Grasps | Grasps/Part | Parts | Objects | Success Train | Success Novel |
|---|---|---|---|---|---|
| 40/50 | 5 | 12 | 9 | 15/19 | 11/21 |

Table 1: Overview of the grasp experiment. Training ojects with success rates: mug-blue(5/7), container-1 (7/8), pot(3/4). Test objects with success rates: mug-brown (3/4), box-red(3/4), funnel(2/3), frying-pan(0/2), watering-can(2/5), plate(1/3).

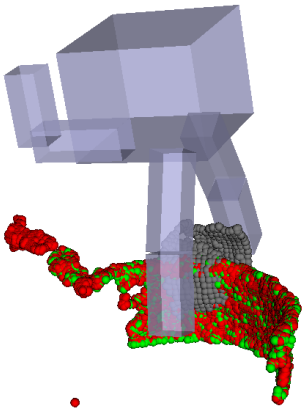
measures with respect to only classifying the data with respect to the alignment score $S^*$. Our results are shown in Table 2.

| Features | Accuracy | Precision | Recall | F1 | TP | TN | FP | FN |
|---|---|---|---|---|---|---|---|---|
| Haptic&Alignment Score | 0.78 | 0.79 | 0.88 | 0.84 | 23 | 8 | 6 | 3 |
| **Haptics** | **0.85** | **0.83** | **0.96** | **0.89** | 25 | 9 | 5 | 1 |
| Alignment Score | 0.70 | 0.73 | 0.85 | 0.79 | 22 | 6 | 8 | 4 |

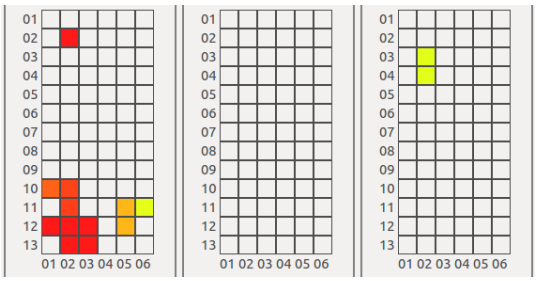
Table 2: Success classification performance for haptics features and or alignment score. TP, TN, FP, FN are the number of true positives, true negatives, false positives and false negatives.

First of all, our performance measures show that we can quite well predict grasp success from haptic features. Interestingly using only haptic features instead of all gives us the best result. Of importance is especially the precision value, as this expresses the confidence that a grasp will work after measuring its haptic features at contact. In our case that means

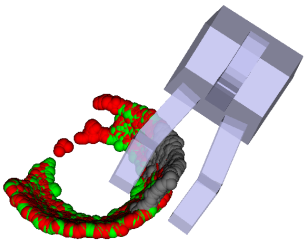
that no matter what we do, based on our precision value using haptic features, 83% of successfully predicted grasps will subsequently succeed when lifting up the object. From our intuition as humans, this rate could be higher. One reason for that can be that the when a pinch-grasp doesn't close around the object at all, touch sensors may be activated due to self-collision. This is why we created the kinesthetic feature (see Sec. 3) that should detect when the fingers are fully closed without object. Still if the the grasp closes around a thin rim, the signatures can be very similar for successful and failed grasps.



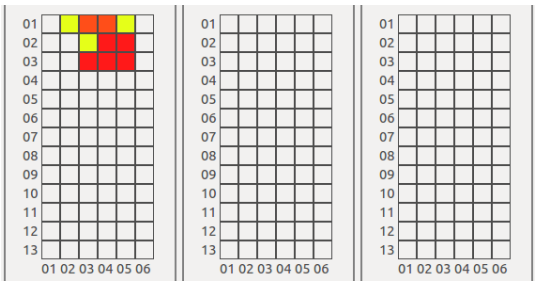
(a) Grasp Pot




(b) Tactile Reading Pot



(c) Grasp Plate




(d) Tactile Reading Plate

Figure 4: (a,b) Grasp and tactile reading of a successful grasp on the pot despite a-non matching template ($S^*$=0.18). Also the grasp is correctly predicted by the classifier. One can see a stable tactile reading only on one of the pads due to convexity of the object. (c,d) Grasp and tactile reading of an unsuccessful grasp on the plate. The grasp is classified wrongly, probably due to the strong tactile reading. Also the template matches well but is too far from the real contact point. The tactile reading is very close to the edge of the sensor pad and the object is heavier than most other objects in the test set.

Also for some successful grasps the tactile sensor doesn't show any value because the fingers touch the object at an oblique angle. The Schunk Hand is not compliant, and the sensors are not sensitive enough to detect such contacts.

The fact that classification performance of haptics is significantly better than prediction based on alignment score indicates the importance of haptics for grasping in unstructured environments with novel objects. Also there exist cases where the alignment score and the template alignment are bad, but still the resulting grasps are stable, see Figure 4(a). This increases the number of possible positive grasps in a scenario where grasps are rejected based on an alignment score threshold.

## 5. Discussion

We performed a grasping experiment where we successfully showed that our method produces successful grasps on partly similar but unseen objects. Most importantly it also shows that using haptic features are useful to predict grasp success. This allows overall grasp success rate to be increased by triggering regrasps based on tactile readings.

Another important point of the method is that due to the pre-selection of parts with our 3D-compositional model, we are able to reduce computation time compared to a scenario where we try to fit templates to the whole object or cluttered scenes, which is especially important in the PacMan project.

The methods also has its weaknesses. The templates need a certain size to be matched in a stable way. In the future we plan to make templates more local and adjust their sizes based on the part in use.

## References

Renaud Detry and Justus Piater. Continuous surface-point distributions for 3D object pose estimation and recognition. In Ron Kimmel, Reinhard Klette, and Akihiro Sugimoto, editors, *Asian Conference on Computer Vision*, volume 6494 of *LNCS*, pages 572–585, Heidelberg, 2010. Springer. doi: 10.1007/978-3-642-19318-7_45. URL `https://iis.uibk.ac.at/public/papers/Detry-2010-ACCV.pdf`.

Safoura Rezapour-Lakani, Mirela Popa, Antonio J. Rodríguez-Sánchez, and Justus Piater. Scale-Invariant, Unsupervised Part Decomposition of 3D Objects. In *Parts and Attributes*, 9 2014. URL `https://iis.uibk.ac.at/public/papers/Rezapour-2014-PA.pdf`. Workshop at ECCV.