

Summary: Object classification using tactile sensor information

Peter Kiechle

February 24, 2015

1 Overview

The acquisition of accurate 3D sensor data is a crucial point for robot grasping applications. But all state-of-the-art 3D scanners have to deal with certain limitations such as difficulties with reflective and transparent objects or cluttered scenes. Tactile sensors provide an opportunity to improve the perception of an already touched object which can be used to locally optimize the grasp or perform a regrasp if necessary.

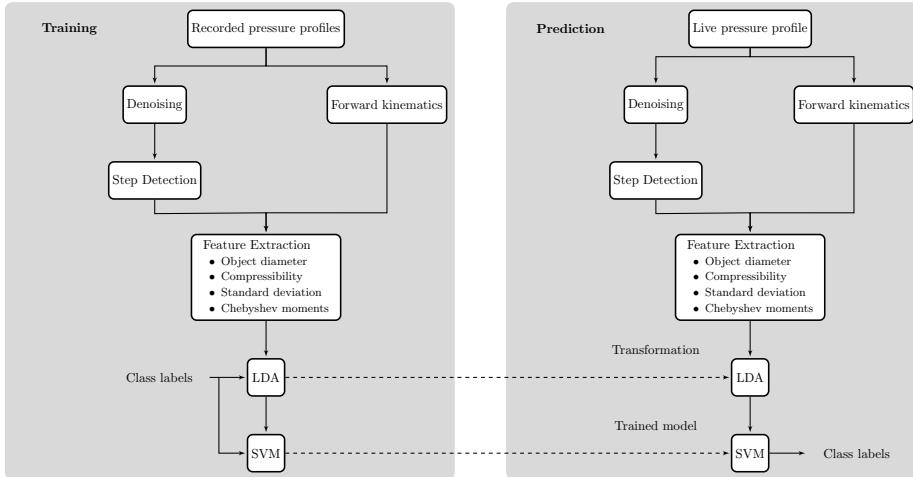


Figure 1: Classification workflow

Figure 1 gives an overview of a working object-grasp classification procedure using SCHUNK's Dexterous Hand 2.0 with tactile sensors by Weiss Robotics. During the training phase various similar grasps of the same object are performed and recorded. The position controller of the SDH-2 delivers joint angle readings at an average data rate of about 110 Hz. The corresponding step size is smaller than 0.02° . The tactile sensor controller transmits data at a rate of roughly 30 frames per second depending on the number of active sensor cells. Due to the internal run-length encoding this value varies significantly over time. By applying a Denavit-Hartenberg-Transformation, the 3D position of each active sensor cell is then computed from the recorded joint angles. Since the tactile

readings are noisy, especially while capturing in high sensitivity mode, a spatial digital filter is applied in order to smooth the signal. A median filter that replaces each value with the median of neighboring entries showed to perform the best in experiments with a variety of denoising algorithms. In the following step, the actual grasps have to be identified and separated from each other. Begin and end frames are therefore determined using a simple step detection algorithm.

The concatenated feature vector consists 54 elements of four different types of features which are further discussed in the following section. Experiments revealed, that the individual features are not equally important. But since they are also not independent from each other, a classical feature selection approach does not work well on the available data. Instead a Linear Discriminant Analysis (LDA), a dimensionality reduction techniques related to Principal Component Analysis (PCA) that maximizes the separation between multiple classes is performed as an intermediate step. A grid-search suggested that the best results are achieved by reducing the 54 dimensions down to 14. A one-vs-one multiclass approach using Support Vector Machines (SVM) is used to execute the final classification task. The trained classifier, that is the support vectors as well as the performed LDA transformation are retained for the prediction phase.

2 Feature Extraction

As mentioned before, the actual feature vector is a concatenation of four different types of features. For the sake of simplicity, all grasps are performed with a pinch grip using two fingers only which restricts the tactile feedback to the sensor matrices at the finger tip.

2.1 Object diameter

The size of a grasped object is intuitively an important property in the context of classification. Training a linear Support Vector Machine before applying a dimensionality reduction shows, that the corresponding size feature has the largest assigned weight to it - a clear indication of its discriminatory power. Based on the hand's dimensions the actual forward kinematics can be computed according to the Denavit-Hartenberg-Transformation (DHT). This results in a 3D point cloud of all active sensor cells. The minimal bounding sphere or enclosing ball has proven to be a versatile metric for the point cloud's extent. The actual implementation relies on an iterative algorithm by Gärtner (1999).

2.2 Compressibility

The SDH-2 is strong enough to actually compress soft objects. It is therefore possible to measure the object's rigidity by comparing the objects diameter at first contact and after the grasping movement has come to rest. It turns out in practice, that the object's initial size cannot be measured accurately enough if the grasp applies torque. In that case, the contact surface between object and both sensor matrices is tilted leading to an overestimation of the real diameter. This asymmetry disappears when all forces reach an equilibrium state. In order to get more robust object size features, the pressure profile's centroids of the end

position frame is used to recalculate the diameter during each step of the grasp. Robust minimal and maximal diameter values are then chosen to estimate the difference.

2.3 Standard deviation of intensity values

Pressure profiles of rigid bodies differ significantly from those of soft objects. The impression on the tactile sensor frame appears smoothed out and lacks sharp edges. The standard deviation of all active sensor cell's intensity values represents an easy way to capture this effect. Note that this feature differs from the second central two-dimensional image moment.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}, \text{ where } \mu = \frac{1}{N} \sum_{i=1}^N x_i$$

2.4 Translation and rotation invariant Chebyshev moments

Since the spatial resolution of the tactile sensor matrices in use is only 6 by 14 taxels, the number of available feature extraction methods is limited. In addition, certain invariance properties are also crucial for the classification success. For example, the extracted features should be invariant to the actual position of the contact area. Rotating the object inside the hand (or vice versa) should not influence the outcome. But similar impressions of different sizes do indicate different objects. To sum it up, the required set of features has to be translation- and rotation invariant but not scale invariant.

Since the original publication of Hu (1962) image moments are a well known concept in computer vision. They are widely used in pattern-recognition and shape description applications as they capture characteristic properties of a segmented object. Originating in mechanics where they describe the mass distribution of physical bodies, moments found their way into statistics to characterize the shape of probability density functions. Features like mean, variance, and skewness for instance are commonly referred to as first-, second central- and third standardized moment. Interpreting an image as a spatial random variable gives rise to the notion of images moments. Flusser et al. (2009) provides extensive information on the topic.

For the two-dimensional tactile sensor matrix $T_s(x, y)$ the *geometric moments* of order $(p + q)$ are defined as:

$$m_{pq} = \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} x^p y^q T_s(x, y)$$

Geometric image moments contain redundant information since the basis set is not orthogonal. Replacing them with Chebyshev polynomials gives rise to Chebyshev moments:

$$T_{pq} = \frac{1}{\rho(p, N)\rho(q, N)} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} t_p(x) t_q(y) f(x, y)$$

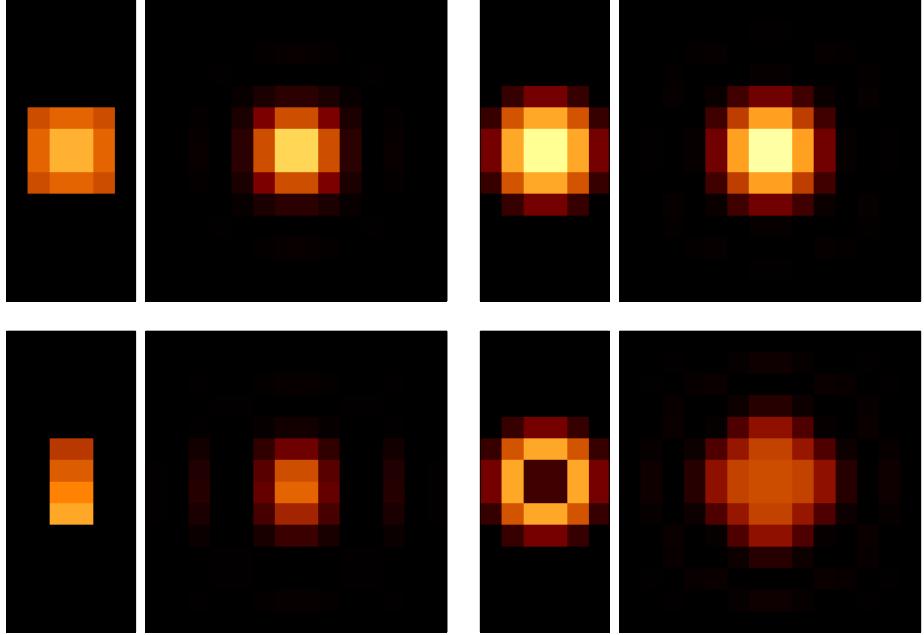


Figure 2: Reconstruction from Chebyshev moments: The left image of each pair represents the original matrix while the right frame is the reconstruction. Maximal moment orders $p, q = 5$

where the Chebyshev polynomials can be defined as:

$$t_n(x) = n! \sum_{k=0}^n (-1)^{n-k} \binom{N-1-k}{n-k} \binom{n+k}{n} \binom{x}{k}$$

The orthogonality property is achieved with:

$$\rho(p, N) = (2n)! \binom{N+n}{2n+1}, \quad n = 0, 1, \dots, N-1$$

Translation and rotation invariance can then be constructed by replacing $t_p(x)$ and $t_p(y)$ with $t_p(x - x_0)$ and $t_p(y - y_0)$ followed by a mapping to the complex number domain. For a fast computation based on a lookup-table approach see Shu et al. (2010).

3 Results and Discussion

Scikit-learn by Pedregosa et al. (2011) offers an easy to use environment to carry out various experiments. In order to estimate the performance of the multiclass SVM approach and to avoid overfitting, 10-fold cross-validation has shown to be the method of choice for most applications. The entire dataset is therefore randomly partitioned into training- and test sets with a ratio of 80/20.

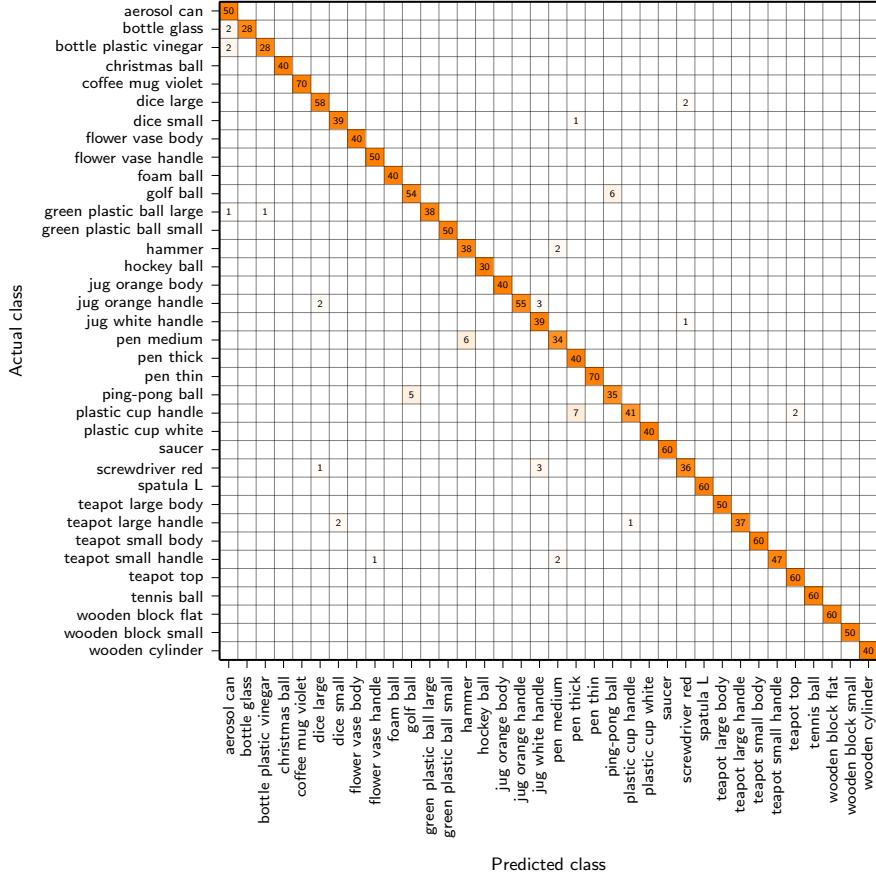


Figure 3: Cumulative confusion matrix of a randomized and stratified 10-fold cross-validation.

In addition, each fold approximately contains the same percentage of samples for each class as the entire set.

The confusion matrix of figure 3 depicts the cumulative result of the performed cross-validation. Symmetric misclassifications occur between the ping-pong- and the golf ball as well as between the medium sized pen and the hammer shaft. The distances to the separating hyperplanes in the one-vs-one classifier scheme do not have to be equal between the same pair of classifiers. A noticeable instance of such a misclassification can be observed in the case of the actual plastic cup handle and the predicted thick pen but not vice versa.

A common metric to compare the performance of classifiers is the average accuracy, i.e. the overall correctness of the model which is calculated as the sum of correct classifications divided by the total number of classifications. Precision also known as reliability describes the classifier's ability not to predict the false class and is computed from the confusion matrix. Recall also known as accuracy intuitively measures if the classifier is able to find all the positive instances of a class. Since there is a trade-off between precision and recall, the F1-score as

a weighted harmonic mean might be more suitable. Table 1 summarizes the results.

A natural extension of the presented concept includes the active exploration of grasped objects. Combining the resulting grasp-object configurations to a single object model can then be seen as an instance of *haptic perception*.

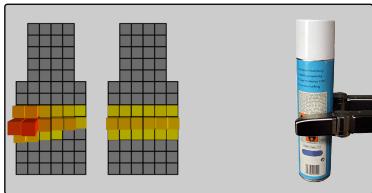
Table 1: Classification performance metrics using a randomized and stratified 10-fold cross-validation

Class	Precision (Reliability)	Recall (Accuracy)	F1 score	Support
<i>aerosol can</i>	0.909	1.000	0.952	50
<i>bottle glass</i>	1.000	0.933	0.966	30
<i>bottle plastic vinegar</i>	0.966	0.933	0.949	30
<i>christmas ball</i>	1.000	1.000	1.000	40
<i>coffee mug violet</i>	1.000	1.000	1.000	70
<i>dice large</i>	0.951	0.967	0.959	60
<i>dice small</i>	0.951	0.975	0.963	40
<i>flower vase body</i>	1.000	1.000	1.000	40
<i>flower vase handle</i>	0.980	1.000	0.990	50
<i>foam ball</i>	1.000	1.000	1.000	40
<i>golf ball</i>	0.915	0.900	0.908	60
<i>green plastic ball large</i>	1.000	0.950	0.974	40
<i>green plastic ball small</i>	1.000	1.000	1.000	50
<i>hammer</i>	0.864	0.950	0.905	40
<i>hockey ball</i>	1.000	1.000	1.000	30
<i>jug orange body</i>	1.000	1.000	1.000	40
<i>jug orange handle</i>	1.000	0.917	0.957	60
<i>jug white handle</i>	0.867	0.975	0.918	40
<i>pen medium</i>	0.895	0.850	0.872	40
<i>pen thick</i>	0.833	1.000	0.909	40
<i>pen thin</i>	1.000	1.000	1.000	70
<i>ping-pong ball</i>	0.854	0.875	0.864	40
<i>plastic cup handle</i>	0.976	0.820	0.891	50
<i>plastic cup white</i>	1.000	1.000	1.000	40
<i>saucer</i>	1.000	1.000	1.000	60
<i>screwdriver red</i>	0.923	0.900	0.911	40
<i>spatula</i>	1.000	1.000	1.000	60
<i>teapot large body</i>	1.000	1.000	1.000	50
<i>teapot large handle</i>	1.000	0.925	0.961	40
<i>teapot small body</i>	1.000	1.000	1.000	60
<i>teapot small handle</i>	1.000	0.940	0.969	50
<i>teapot top</i>	0.968	1.000	0.984	60
<i>tennis ball</i>	1.000	1.000	1.000	60
<i>wooden block flat</i>	1.000	1.000	1.000	60
<i>wooden block small</i>	1.000	1.000	1.000	50
<i>wooden cylinder</i>	1.000	1.000	1.000	40
average total	0.971	0.969	0.969	1720

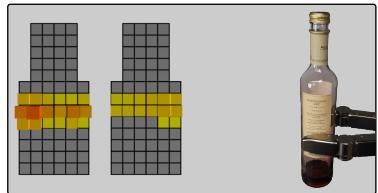
References

- Jan Flusser, Barbara Zitova, and Tomas Suk. *Moments and Moment Invariants in Pattern Recognition*. Wiley Publishing, 2009. ISBN 0470699876, 9780470699874.
- Bernd Gärtner. Fast and robust smallest enclosing balls. In *Proceedings of the 7th Annual European Symposium on Algorithms*, ESA '99, pages 325–338, London, UK, UK, 1999. Springer-Verlag. ISBN 3-540-66251-0. URL <http://dl.acm.org/citation.cfm?id=647909.740295>.
- Ming-Kuei Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2):179–187, February 1962. ISSN 0096-1000. doi: 10.1109/TIT.1962.1057692.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830, 2011.
- Huazhong Shu, Hui Zhang, Beijing Chen, Pascal Haigron, and Limin Luo. Fast computation of tchebichef moments for binary and grayscale images. *IEEE Transactions on Image Processing*, 19(12):3171–3180, 2010. URL <http://dblp.uni-trier.de/db/journals/tip/tip19.htmlShuZCHL10>.

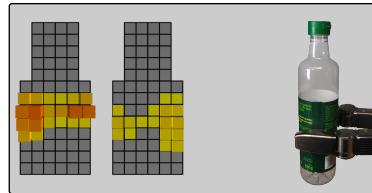
A Appendix



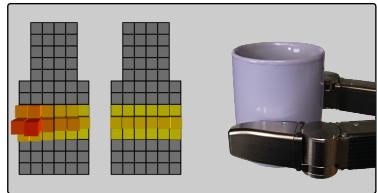
(4.1) Aerosol can



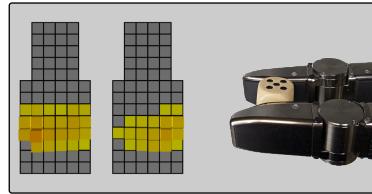
(4.2) Bottle glass



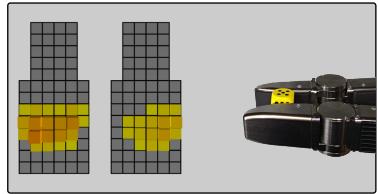
(4.3) Bottle plastic vinegar



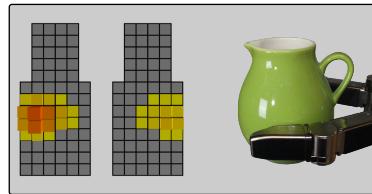
(4.4) Coffee mug violet



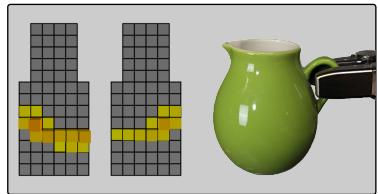
(4.5) Dice large



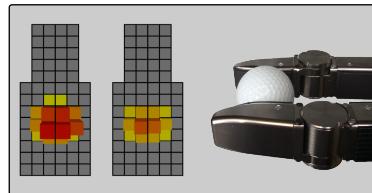
(4.6) Dice small



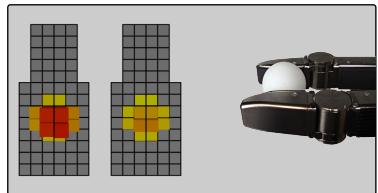
(4.7) Flower vase body



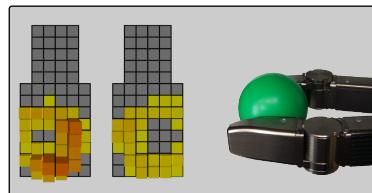
(4.8) Flower vase handle



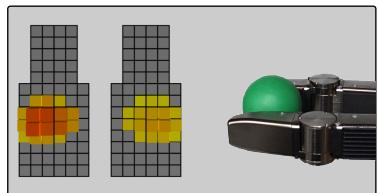
(4.9) Golf ball



(4.10) Ping-pong ball

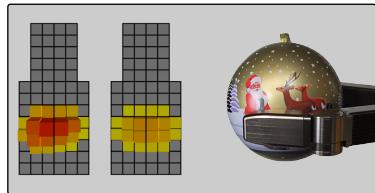


(4.11) Green plastic ball large

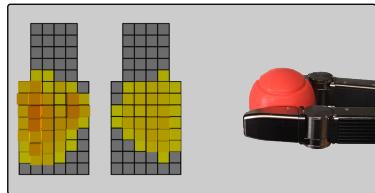


(4.12) Green plastic ball small

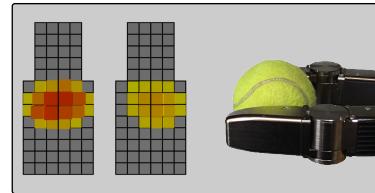
Figure 4: Learned grasp-object configurations



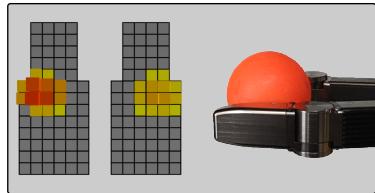
(4.13) Christmas ball



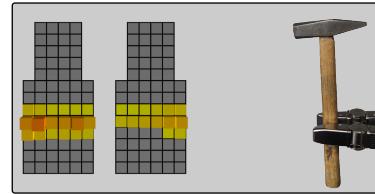
(4.14) Foam ball



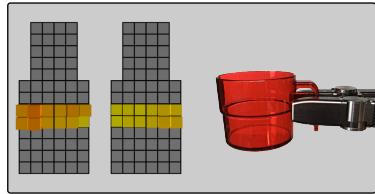
(4.15) Tennis ball



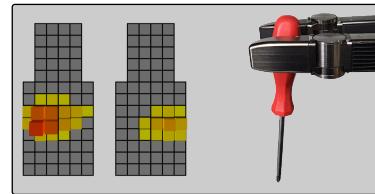
(4.16) Hockey ball



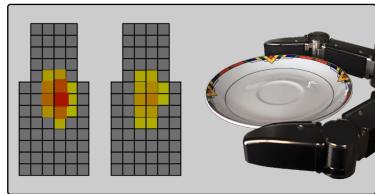
(4.17) Hammer



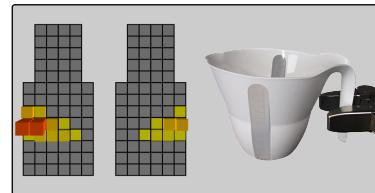
(4.18) Plastic cup handle



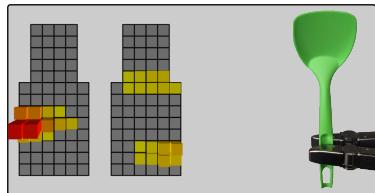
(4.19) Screwdriver



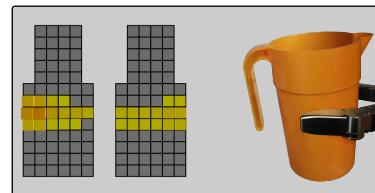
(4.20) Saucer



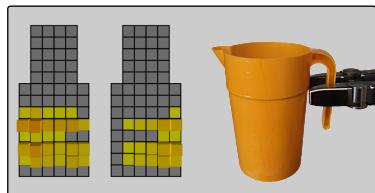
(4.21) Jug white handle



(4.22) Spatula (R)

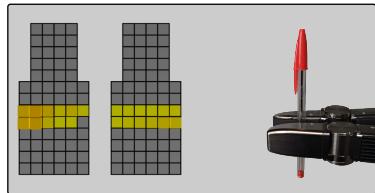


(4.23) Jug orange body

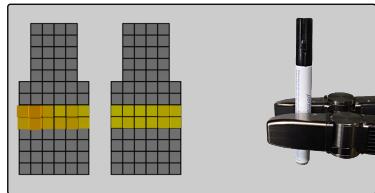


(4.24) Jug orange handle

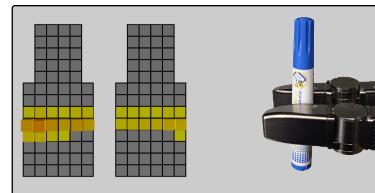
Figure 4: Learned grasp-object configurations (continued)



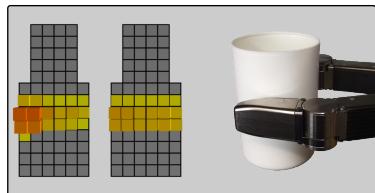
(4.25) Pen thin



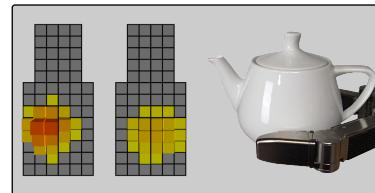
(4.26) Pen medium



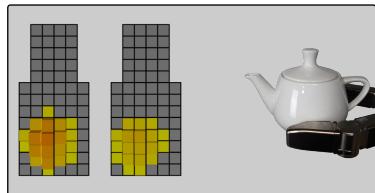
(4.27) Pen thick



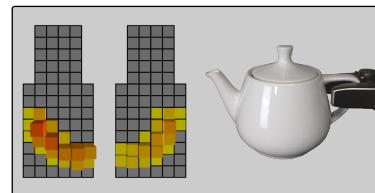
(4.28) Plastic cup white



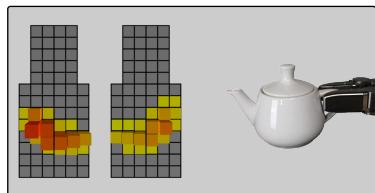
(4.29) Teapot large body



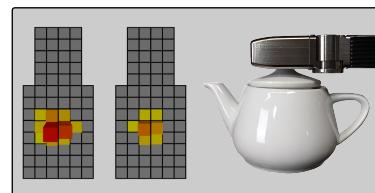
(4.30) Teapot small body



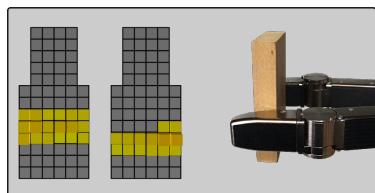
(4.31) Teapot large handle



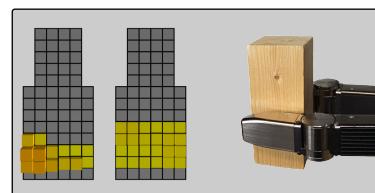
(4.32) Teapot small handle



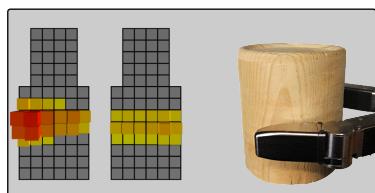
(4.33) Teapot top



(4.34) Wooden block flat



(4.35) Wooden block small



(4.36) Wooden cylinder

Figure 4: Learned grasp-object configurations (continued)