Generalizing Grasps Across Partly Similar Objects

Renaud Detry   Carl Henrik Ek   Marianna Madry   Justus Piater   Danica Kragic

***Abstract*— The paper starts by reviewing the challenges associated to grasp planning, and previous work on robot grasping. Our review emphasizes the importance of agents that generalize grasping strategies across objects, and that are able to transfer these strategies to novel objects. In the rest of the paper, we then devise a novel approach to the grasp transfer problem, where generalization is achieved by *learning*, from a set of grasp examples, a dictionary of object parts by which objects are often grasped. We detail the application of dimensionality reduction and unsupervised clustering algorithms to the end of identifying the size and shape of parts that often predict the application of a grasp. The learned dictionary allows our agent to grasp novel objects which share a part with previously seen objects, by matching the learned parts to the current view of the new object, and selecting the grasp associated to the best-fitting part. We present and discuss a proof-of-concept experiment in which a dictionary is learned from a set of synthetic grasp examples. While prior work in this area focused primarily on shape analysis (parts identified, e.g., through visual clustering, or salient structure analysis), the key aspect of this work is the emergence of parts from *both* object shape *and* grasp examples. As a result, parts intrinsically encode the intention of executing a grasp.**

I. Introduction: Challenges in Grasp Planning

This paper studies the planning of grasping actions, or, in other words, the problem of exploiting perceptual data to select a wrist position and finger configuration to which a hand can be transported in order to grasp an object. The wrist position (or *grasping point*) corresponds to the region of the object towards which the hand will move. The finger configuration (or *hand preshape*) corresponds to the angles to which finger joints are set prior to coming in contact with the object.

Grasp planning is a complex problem. A grasp must bind a hand to an object, and prevent the object from subsequently slipping or escaping. Configurations which lead to a collision between the hand and the object or other obstacles must be avoided, and task-related constrains must be verified (certain tasks restrain the number of possible grasps, as a knife should be held specifically by its handle when the task is to cut something). Perceptual data, usually provided by vision, are noisy and often limited to a single viewpoint. For dexterous grasping, the space of action parameters (hand positions and configurations) quickly becomes high-dimensional (a human hand has twenty-five degrees of freedom – six for the wrist position and orientation, and nineteen for the finger joint angles). Yet, despite the complexity of the problem, the frequent recurrence of grasping in everyday tasks imposes an ability to plan grasps quickly.

In robotics, grasp planning traditionally relies on contact-force analysis [3], [34]. Force analysis bases planning on a reconstruction of the geometry and physical properties of the objects that surround the agent. Provided that such a reconstruction is available, the agent searches the space of hand configurations for the configuration that best verifies grasping constraints (binding configuration, no collisions, task compatibility). In practice, the applicability of force analysis is limited by the difficulty of obtaining accurate models of object geometry, mass, and friction characteristics. Also, as the space of hand configurations is high dimensional, the optimization procedure underlying force analysis is computationally expensive. These shortcomings motivated the community to rethink the planning problem, leading for instance Borst et al. [5] to demonstrate that finding the globally optimal grasp is often not strictly worth the computational effort, as for many tasks an average grasp (in the force-analysis sense) is acceptable. The bigger leap however came with a class of methods that parted drastically from the traditional planning philosophy. Instead of searching for a grasp that optimally satisfies the various (vision-dependent) grasping constraints, these methods extract, from the agent's experience, a function that directly maps visual perceptions to grasp parameters, with the advantage of *implicitly* capturing the object's physical properties, and avoiding a costly search through the high-dimensional space of hand configurations [7], [21], [25], [30], [36].

Numerous behavioral studies tend to support the existence of similar processes in the human grasping system. It has been shown for instance that humans often grasp objects by preshaping their hand during its transportation towards the object [18], then compliantly refining the grip upon contact [19]. Concurrently, neurophysiological studies suggested that, in monkeys, the cortex encodes a set of prototype grasps, which are selectively triggered by visual stimuli [26]. It thus seems plausible, as proposed, for instance, by Johansson et al. [19], that the human grasping system relies on a set of prototypical motor programs that are selected and parametrized by visual input, therefore acting as a direct mapping from vision to action. Humans arguably possess the most sophisticated grasping system known today, being able to plan complicated grasps in just a few hundreds of

R. Detry, C. H. Ek, M. Madry, and D. Kragic are with the Centre for Autonomous Systems and the Computer Vision and Active Perception Lab, CSC, KTH Royal Institute of Technology, Stockholm, Sweden. Justus Piater is with the Intelligent and Interactive Systems group, Institute of Computer Science, University of Innsbruck, Austria. Email: {detryr,chek,madry,danik}@csc.kth.se, justus.piater@uibk.ac.at

This work was supported by the Swedish Foundation for Strategic Research, the Belgian National Fund for Scientific Research (FNRS), and the EU projects CogX (FP7-IP-027657) and TOMSY (IST-FP7-Collaborative Project-270436).

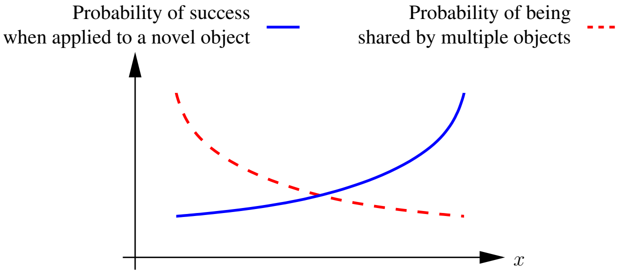


Fig. 1: Robustness-transferability trade-off in feature-based grasp planning. The x axis corresponds to the amount of information encoded by a part. Highly informative parts allow for a robust grasp application. However, these are less likely to be shared across objects.

milliseconds [17]. We believe that the possibility that such an efficient system be based on a direct vision-action mapping is a strong argument for researching vision-action mappings for robotics.

To learn a vision-to-grasp mapping for one specific object, an agent usually collects a set of grasp examples, and lets machine-learning algorithms construct a grasp predictor from these. Such a model allows the agent to quickly produce grasping plans for the object on which it trained. However, collecting grasp examples is an expensive, time-consuming process. A major focus in grasp learning is to develop methods that produce useful manipulation models from as few data as possible. A natural means of limiting the need for examples is to try and adapt memories of previous objects to the planning of a grasp onto a novel object. Many objects share similarities in shape, and similarities in grasp affordances, and both are often correlated. When a novel object appears, instead of starting to learn from scratch, an agent may instead attempt to apply to it the strategies it has acquired for partly similar objects. To this end, means of linking grasps to certain object *features* have been researched, in the hope of transferring grasps across objects that share the same features. The challenge of this task is to decide which visual cues should be captured by the features. Intuitively, a feature should capture no more no less than the specific cues that predict the applicability of a grasp. If a feature misses important cues, it risks predicting faulty grasps. If a feature includes cues that are not directly related to grasping, its transferability to other objects will be impeded. Designing a feature for grasp generalization thus involves a robustness-transferability trade-off, as illustrated in Fig. 1.

A number of methods for vision-based grasping learn a mapping from image features, such as local gradients or SIFT, to grasp parameters [24], [25], [30]. One advantage of these methods is their conceptual elegance:

1) Extract features from images of a set of objects.
2) Label these features as good or bad grasping point, either with the help of a teacher [30] or through autonomous exploration [24].
3) Learn a grasp classifier.

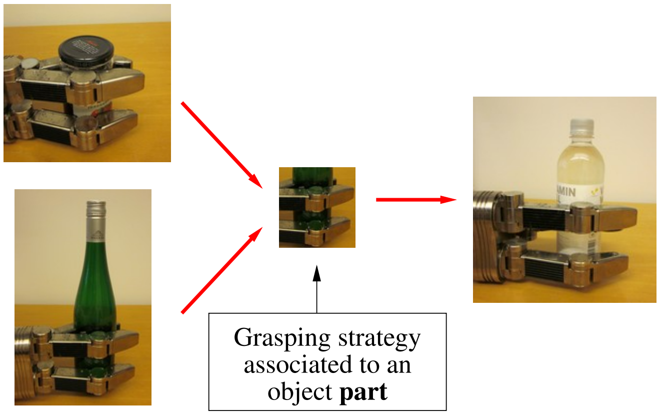


Fig. 2: Learning part-grasp associations. The agent will identify, within its visuomotor experience, recurrent associations of object *parts* and successfully executed grasps. These grasps will then be applicable to novel objects that share the same part.

4) Transfer grasps by classifying features obtained from images of novel objects.

Unfortunately, these methods also come with their shortcomings. From a practical viewpoint, the geometric information provided by a local feature detector is generally poor. As grasping is an intrinsically 3D interaction, it largely relies on 3D object properties, such as shape, which are only partly captured by 2D image features. It is thus difficult to link, for example, a 3D gripper orientation to an image feature.

Across the range of visual cues that have been used for designing grasp planners, 3D shape has lead to particularly good results. By contrast to methods based on image features, methods that link grasp parameters to a shape model [1], [9], [11], [14], [23] benefit from an increased geometric robustness, which makes it easier to preshape the hand to approximate object shapes, and accurately position and orient the wrist and fingers with respect to the object. Mapping grasps to 3D cues is supported by behavioral and neurophysiological studies. Behavioral studies have demonstrated the reliance of human grasping on 3D shape [16], while neurophysiologists have observed a mapping from 3D shape to action prototypes in monkeys [27].

II. Learning Shape Prototypes for Generalizing Grasps

In the rest of the paper, we present an adaptive grasp planner that learns a mapping from object shape to grasp parameters.

A. From Part to Grasp

Linking grasp parameters to the shape of the whole body of an object limits the applicability of the model to that particular object. In order to transfer grasps across objects, we instead explore the linking of grasp parameters to object parts. In order to allow the agent to generalize its acquired knowledge to novel objects, we propose to provide it with means of identifying, within its visuomotor experience, recurrent associations of object *parts* and successfully executed grasps. For instance, the agent may have successfully transported objects such as bottles, cans, and jars, which have different sizes, but which can be seized by applying the same power grasp to their side. We propose to provide the agent with means of understanding, from a set of such examples, that any object that presents a cylindrical part can be grasped sideways with a wide-palm grasp (Fig. 2).

B. *Previous Work on Part-based Grasping*

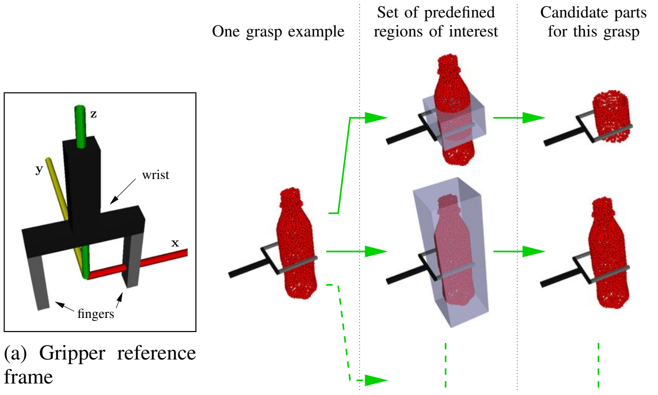
Part-grasp associations have been previously suggested and studied by several research groups [2], [23], [36]. In the earlier work, the definition of parts was often either hard-coded [23], or driven by shape analysis [1], [2], [36]. There is however an increasing interest for defining parts based on grasping experience [10], [12], [15], [22], [37]. For instance, Herzog et al. [15] and Zhang et al. [37] presented two exciting data-driven approaches where a part describes an object's shape in a fixed-size region around a grasping point. These approaches are further discussed below.

C. *Method*

Our work aims at learning, from a set of grasp examples, a dictionary of prototypical parts by which objects are often grasped. A key property that we wish to allow our agent to extract from experience is the spatial extent of grasp-predicting parts. For instance, in the case presented in Fig. 2, we wish our agent to learn that the relevant part is a 10cm-high cylinder. The the tap of the jar or the conic upper part of the bottle should be ignored, as they are not shared by the two objects.

Training data are provided to the agent in the form of a set of grasps demonstrated onto objects known to the agent. (The agent has previously acquired 3D point clouds that model the shape of the objects.) A grasp is parametrized by the 6D pose of the wrist (3D position and 3D orientation), and by the 6D pose of the object. Our method works as follows: First, the agent generates, from the grasp examples, a large number of part candidates of varying sizes (Section III). Most of the candidates will not generalize well. However, it is our hope that for every set of objects that share a graspable part, each object will yield one candidate that approximately captures that part. The candidates that recur across objects are identified by clustering part candidates (Section IV). Dense clusters will contain parts by which objects are often grasped, which are thus promising for grasping novel objects.

The central parts of all clusters will form the dictionary used by the agent to grasp novel objects. An important aspect of our work appears at this point. As the dictionary of parts is only formed from cluster centers, it is allowed to be orders of magnitude smaller than the set of grasp examples initially provided to the agent. In the data-driven approaches discussed above [15], [37], each grasp example yields a part. By contrast, in our work, a grasp example only "votes" for the potential inclusion of a part into the dictionary, which provides us with a means of controlling the size of the dictionary in order to keep the computational cost of planning a grasp onto a novel object reasonably low.


One grasp example
Set of predefined regions of interest
Candidate parts for this grasp
z
y
wrist
x
fingers

(a) Gripper reference frame

(b) Generating part candidates

Fig. 3: Generating part candidates. The black and grey renderings on each image represent the pose of the gripper set for a sideways grasp on the soda bottle. Parts of varying sizes are generated by defining several box-shaped regions of interest centered on the gripper.

Also, in our work, parts emerge from both object shape and grasp examples. A key result is our ability to optimize the robustness-transferability trade-off discussed above. Not only the shape, but also the spatial extent (or size) of the parts that form the dictionary depend on the available grasp data. Our approach involves an explicit search for recurrent patterns within the agent's visuomotor experience, which leads to the identification of parts that directly predict grasp applicability.

III. Generating Part Candidates

Part candidates are generated by extracting object surface segments of varying size in the vicinity of grasps demonstrated by a teacher. Parts are thus represented, as the object from which they are extracted, by point clouds. This process is illustrated for a soda bottle in Fig. 3. Surface segments are extracted using a set of predefined regions of interest (ROI). These regions are centered on the gripper, as the applicability of a grasp is largely conditioned by the shape of the surface in the direct vicinity of the grasping point. ROI sizes should *a priori* vary in all directions. However, the preshape of the gripper at the time of the grasp can limit the number of regions that are interesting to look at. For instance, in the case shown in Fig. 3, it is reasonable to limit the ROI width along the x axis of the gripper to the distance that separates both fingers, as the object will usually not be larger that this gap. With more sophisticated hands, grasp preshapes can further constrain the definition of ROIs.

IV. Extracting Dense Clusters of Parts

Graspable parts that generalize are discovered by clustering part candidates. Dense groups of similarly-shaped candidates correspond to shapes onto which grasps can be applied in order to seize several different objects. These shapes are thus likely to predict grasp applicability for novel objects.

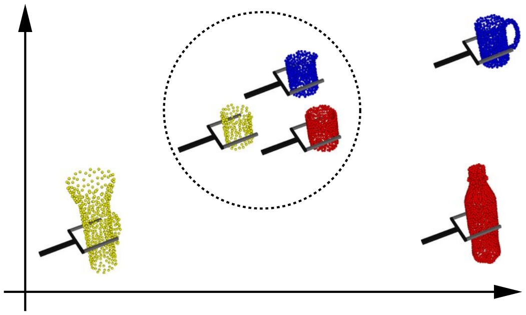


Fig. 4: Finding parts that allow for transferring grasps to a novel object. The three outer "parts" (which correspond to entire objects), will not generalize well. By contrast, the three center parts, which represent a piece of the flashlight, cup, and soda bottle, are very similar to each other. As there exist a shape similarity across these three parts extracted from different objects of the training database, it seems reasonable to assume that the grasps related to these parts are potentially applicable to novel objects.

In Fig. 4, none of three outer parts would be applicable to other objects. The three middle parts, by contrast, encode a shape-grasp relation that would be applicable to an object that has a cylindrical part of a similar diameter.

Clustering part candidates requires the definition of a measure of shape (dis)similarity. This measure is defined in the next section. Section IV-B details the clustering algorithm.

A. *Measuring Part Dissimilarity*

This section defines a measure part dissimilarity. We note that, as we ultimately aim at using parts for predicting gripper poses, we must measure the (dis)similarity of *gripper-relative* shapes. In other words, a cylindrical part grasped from the side should *not* be similar to the same cylindrical part grasped from the bottom.

In this work, a part is represented by a point cloud defined in a reference frame that corresponds to the 6D pose of the grasp associated to that part. Let $P = \{x_i\}_{i\in[0,n]}$ and $Q = \{y_i\}_{i\in[0,m]}$ denote the point-cloud representations of two parts, with all x_i's and y_i's belonging to $\mathbb{R}^3$. Let us then denote by d^* an asymmetric measure of dissimilarity of P and Q, with

$$d^*(P,Q) = \sum_{i=0}^{n} \min_{j\in[0,m]} f(x_i, y_j), \tag{1}$$

where

$$f(x,y) = \begin{cases} \frac{\|x-y\|}{T} & \text{if } \|x-y\| \le T, \\ 1 & \text{if } \|x-y\| > T. \end{cases} \tag{2}$$

The dissimilarity d^* is often used as error function for point-cloud alignment. In our experiments, the threshold T is set to two centimeters.

We define the dissimilarity of two parts P and Q as

$$d(P,Q) = d^*(P,Q) + d^*(Q,P). \tag{3}$$

The dissimilarity d is symmetric in its arguments. It amounts to the sum of the Euclidean distances between the points of P and their nearest neighbor in Q, and the points of Q and their nearest neighbor in P.

B. *Clustering Parts*

The dissimilarity measure defined in the previous section provides us with a qualitative tool for reasoning on the recurrence of shape-gripper associations across grasp examples. As expressed in the conceptual illustration of Fig. 4, we wish to find a geometric configuration with dense clusters of parts induced by our similarity measure. Dense clusters will correspond to parts that frequently occur within our database. These parts are therefore likely to be useful for grasping novel objects.

The measure described in IV-A provides a global dissimilarity measure between each item in the database from which we can generate a distance matrix

$$D_{ij} = d(P_i, P_j) \tag{4}$$

for all the entries in the database. In order to interpret the data we wish to find a geometrical configuration of the datapoints where the Euclidean distance corresponds to the dissimilarity measure we defined. One possibility is to directly apply classical multi-dimensional scaling [8] to the distance matrix. However, in this paper we are interested in finding a geometrical configuration which suits interpreting the data in terms of clusters. In order to do so we introduce additional flexibility by first interpreting the distance matrix in terms of an inner-product of Gram matrix. Distance matrices and Gram matrices can be interchanged [29] as data inducing representations. Dependent on applications there are benefits associated with each view-point. Here the use of a Gram matrix allows us to view the matrix as a covariance matrix; this approach is well known as the "kernel-trick" [4]. To that end, we use a squared exponential function to apply a non-linear transform of the space that the dissimilarity measure induces,

$$k(P,Q) = e^{-\frac{d(P,Q)^2}{\sigma}}. \tag{5}$$

The squared exponential function induces a geometrical space well-suited for clustering as it will push points that are close together closer and move points far apart even further apart. The parameter σ controls the strength of this transformation.

Discovering part clusters could be achieved directly on the distances defined above (4). However, in order to facilitate the illustration of our method in the experiments presented below, we first recover a low-dimensional approximation of the data, then cluster the data in this low-dimensional space. We recover a d dimensional approximation of the data by solving the following minimization problem,

$$\hat{\mathbf{C}} = \text{argmin}_{\mathbf{C}} ||\mathbf{K} - \mathbf{C}||_{\text{F}}^2, \tag{6}$$

where $\mathbf{K}$ is the Gram matrix whose elements are defined by $k(P_i, P_j)$ for all the entries in the database, and the rank of $\mathbf{C}$ is constrained to be at most d. The solution can be found in close form through an eigenvalue problem and is well-known as kernel principal component analysis [31].

Having resolved a geometrical representation of the data, we wish to partition the space in such a manner that we can discover atomic classes of grasps independent of object type. We proceed through a two-stage process. First, we want to group each point in the database into a small number of classes. Secondly, we wish to explain each class by a single representative grasp. Underpinning our approach is the notion that the dissimilarity measure contains this desired structure. This assumption implies that the grouping can be cast as a clustering problem. Clustering is a well-studied problem within computer science and datamining. It has been used extensively to create compact representations of data using mixture models [35] or for application scenarios where a significant amount of prior information about the partitioning is available [6].

The dissimilarity measure $d(\cdot,\cdot)$ is defined between each point in the database. This allows us to construct a graph $G \in \{\mathcal{V},\mathcal{E}\}$ where each grasp is represented by a node $v_i \in \mathcal{V}$ with edges $e_{ij} \in \mathcal{E}$ connecting associated nodes. We wish to find a partitioning that respects the dissimilarity measure $d(\cdot,\cdot)$. To that end, we construct a fully connected graph. The edge weights are $e_{ij} = \mathbf{C}_{ij}$, *i.e.*, inversely proportional to the dissimilarity between the grasps according to our measure. In order to partition the space, it now remains to cut the graph into disjoint regions each representing a cluster.

In this paper we employ the normalized cuts [33] approach to partition the graph. The $\text{cut}(\mathcal{A},\mathcal{B})$ of a graph G into two sets of disjoint nodes $\mathcal{A}$ and $\mathcal{B}$ is defined as,

$$\text{cut}(\mathcal{A},\mathcal{B}) = \sum_{i\in\mathcal{A},j\in\mathcal{B}} e_{ij}. \tag{7}$$

The normalized cuts algorithm finds the partitioning of the graph that minimizes the following objective function,

$$\text{cut}_{\text{normalized}}(\mathcal{A},\mathcal{B}) = \frac{\text{cut}(\mathcal{A},\mathcal{B})}{\text{assoc}(\mathcal{A},\mathcal{V})} + \frac{\text{cut}(\mathcal{A},\mathcal{B})}{\text{assoc}(\mathcal{B},\mathcal{V})}, \tag{8}$$

$$\text{assoc}(\mathcal{A},\mathcal{V}) = \sum_{i\in\mathcal{A},j\in\mathcal{V}} e_{ij}. \tag{9}$$

The denominator grows with increasing node sets which works to penalize creating very small clusters.

V. Proof of Concept

We now present a proof-of-concept experiment which illustrates the method suggested above. The experiment is realized on synthetic data consisting of seven two-finger grasps demonstrated on four objects (see Fig. 5a and Fig. 5b).

Three sets of regions of interest were defined for the three grasp types present in the database. Three ROIs were defined for "cylindrical" grasps, which correspond to the grasps number 1, 2 and 3 of Fig. 5b. Four ROIs were defined for the parallel grasps (4, 5, 6), and six ROIs for the pinch grasp (7). We note that, in the case of the synthetic data studied in this paper, considering cylindrical, parallel and pinch grasps is purely anecdotal. However, in a real-case scenario, the hand preshape used for a given grasp would allow us to limit the number of parts that need to be considered as candidates. For instance, with a cylindrical grasp (Fig. 6), generating ROIs that differ in size in a direction perpendicular to the palm of the hand is more important than considering variations along directions parallel to the palm. With a parallel grasp (for instance, Fig. 2), ROIs of various lengths in a direction parallel to the palm are necessary. These observations motivated the definition of different sets of ROIs for the different types of grasps shown in Fig. 5. The part candidates generated with these ROIs are shown in Fig. 5c.

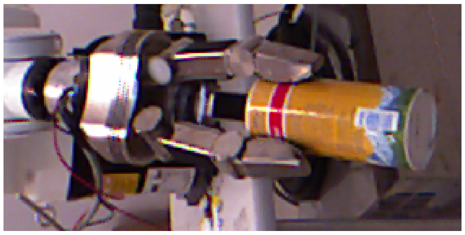


Fig. 6: Cylindrical grasp preshape. The finger-surface normals at the contact points are 120° apart.

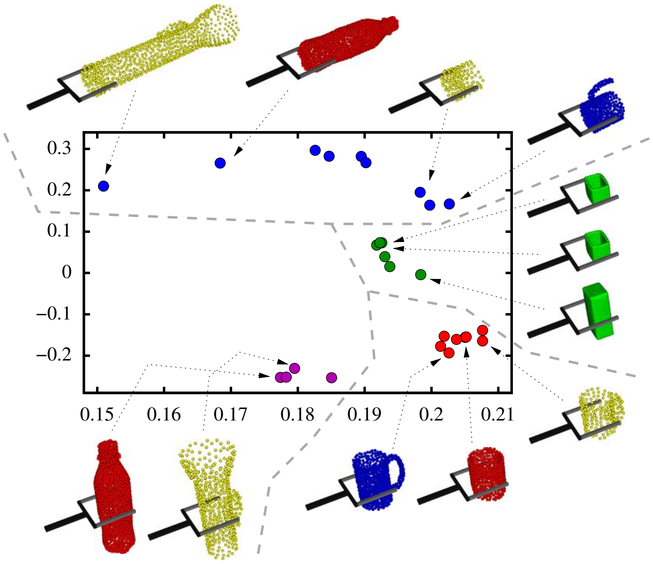
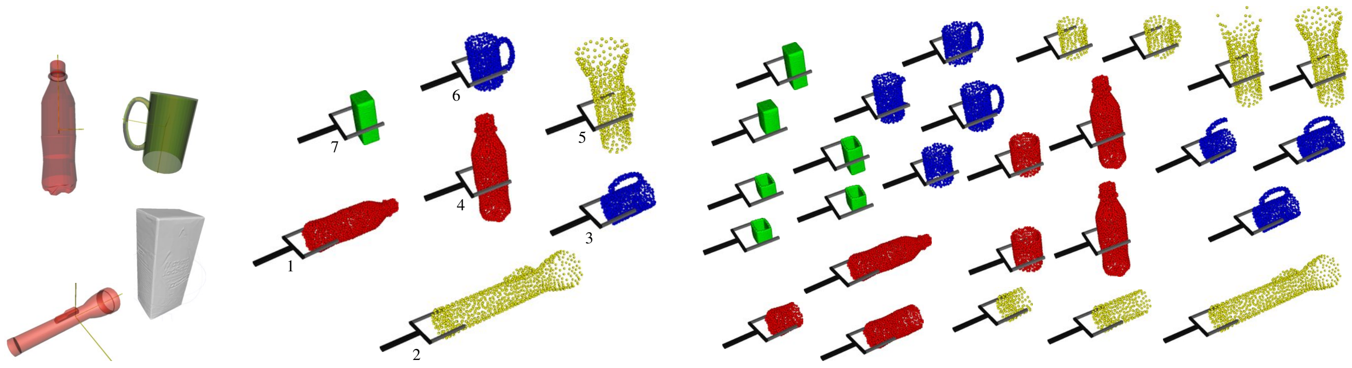


Fig. 7: Two-dimensional approximation of candidates' geometric configuration, computed from the dissimilarity measure of Section IV-A. Dot colors indicate the data cluster to which a datapoint (part candidate) belongs (see text for details). The colors of the dots within the plot and the colors of the parts surrounding the plot are unrelated. We note that the vertical and horizontal axes are not equally scaled.

As explained in Section IV-B, kernel PCA provides us with low-dimensional approximations of our data. A two-dimensional approximation is show in Fig. 7. This plot shows that the dissimilarity measure of Section IV-A properly separates candidate parts in groups of similarly-shaped parts. These groups can be correctly identified by the clustering algorithm of Section IV-B, as reported by the colors associ-



(a) Object set (b) Demonstrated grasps (c) Part candidates

Fig. 5: Experimental data. Three of the objects are cylinders of different sizes, and one is a box. Seven grasps are synthetically demonstrated to the agent. for the cylinders, both sideways and top-down grasps are demonstrated. Fig. (c) shows the candidate parts computed from the grasps of Fig. (b). Part colors indicate which object a part is segmented from.

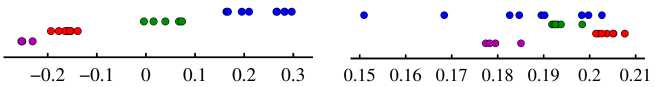


Fig. 8: Projection of the data (candidate parts) onto the first (left) and second (right) principal components of the data. Colors indicate the data cluster to which a datapoint belongs (see text for details). The elevation of the datapoints above horizontal axes is meant to help identifying clusters.

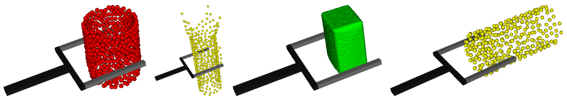


Fig. 9: Prototype parts. These parts correspond to the centers of the clusters of Fig. 7.

ated to the datapoints. In this paper, the number of clusters was determined by inspection. However, BIC-like criterions that compute an optimal number of clusters could be used instead [32]. We note that the two axes of this plot are not equally scaled. The data shows a larger variance along the vertical axis than along the horizontal axis. Fig. 8 shows the projection of the data onto its first and second principal components (which correspond to the vertical and horizontal axes of Fig. 7, respectively). Fig. 8 indicates that the first component contains enough information to identify most of the clusters computed from the dissimilarity measure. The second component leads to a clear separation of the purple and red clusters.

Despite the modest number of data, computing the central point of each cluster allows us to identify a set of prototypical graspable parts. These parts are shown in Fig. 9. We emphasize that despite its reliance on complete object shape models for learning prototypical parts, the method presented above is applicable to predicting grasps onto novel objects perceived through a single 3D snapshot. Fig. 10 illustrate the application of the first and last prototypes of Fig. 9 to a novel object. The right side of Fig. 10 shows the point-cloud representation of the scene (captured by a depth sensor), and the two grasps suggested by the prototypes. The parts are aligned to the object using the pose estimation method of Detry et al. [13].

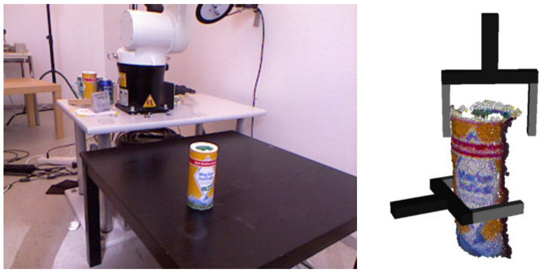


Fig. 10: Grasping a novel object using a dictionary of parts. The rightmost image shows the grasps suggested by the first and last prototypes of Fig. 9, respectively approaching the object from the side and from the top.

VI. Discussion

The dissimilarity measure of Section IV-A provides a direct channel for injecting expert knowledge into to the method presented above. By choosing suitable dissimilarities, one can let a variety of desirable visuomotor strategies emerge from data clustering. For instance, one may argue that similarly-shaped parts may predict similar grasps despite a scale difference. Basing a similarity measure on a mix of local shape features (Spin images [20], or FPFH [28]) and global shape features (for instance, the first few moments of a point cloud) has the potential of robustly representing shape while being invariant, to some extent, to scale. Such a measure would allow an agent to understand that cylinders of different radii can be grasped in similar ways. Simultaneously, the distance matrix of Eq. 4 would be much simpler to compute from a set of compact shape features than from the original point-cloud representations. Using shape features would effectively move some of the computational effort out of the distance-matrix computation (quadratic in the number of candidate parts), into a process linear in the number of candidate parts.

Grasp preshapes were discussed in the previous section, albeit remaining of anecdotal use. In a real-world scenario involving a dexterous hand, preshape is an essential grasping property. In such a scenario, a dissimilarity measure would benefit from the availability of preshape parameters, as it would provide an additional cue for separating unrelated parts.

VII. Conclusion

We reviewed the challenges associated to robotic grasping and the importance of devising means of transferring grasping strategies across objects. We then depicted a method that allows an agent to identify, within its visuomotor experience, graspable parts that generalize across objects. Part candidates are first generated by extracting object surface segments in the vicinity of grasps demonstrated by a human. Candidates are then clustered by means of nonlinear dimensionality reduction and unsupervised learning algorithms. The central elements of the resulting clusters are selected to form a dictionary of prototypical parts that can then be used for grasping novel objects. As the dictionary of parts is only formed from cluster centers, it is allowed to be orders of magnitude smaller than the set of grasp examples initially provided to the agent. A grasp example only “votes” for the potential inclusion of a part into the dictionary, which provides us with a means of controlling the size of the dictionary in order to keep the computational cost of planning a grasp onto a novel object reasonably low. Finally, not only the shape, but also the spatial extent (or size) of the parts that form the dictionary depend on the available grasp data. Prototypical parts are selected based on their recurrence across experienced grasps, which leads to the identification of parts that strongly predict grasp applicability.

References

[1] J. Aleotti and S. Caselli. Part-based robot grasp planning from human demonstration. In *IEEE International Conference on Robotics and Automation*, 2011.

[2] C. Bard and J. Troccaz. Automatic preshaping for a dextrous hand from a simple description of objects. In *IEEE International Workshop on Intelligent Robots and Systems*, pages 865–872. IEEE, 1990.

[3] A. Bicchi and V. Kumar. Robotic grasping and contact: a review. In *IEEE International Conference on Robotics and Automation*, 2000.

[4] C. M. Bishop. Pattern recognition and machine learning, 2006.

[5] C. Borst, M. Fischer, and G. Hirzinger. Grasping the dice by dicing the grasp. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 4, pages 3692–3697, 2003.

[6] Y. Boykov and M.-P. Jolly. Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images. In *International Conference on Computer Vision*, 2005.

[7] J. Coelho, J. Piater, and R. Grupen. Developing haptic and visual perceptual categories for reaching and grasping with a humanoid robot. In *Robotics and Autonomous Systems*, volume 37, pages 7–8, 2000.

[8] M. Cox and T. Cox. Multidimensional scaling. *Handbook of data visualization*, Jan. 2008.

[9] C. de Granville, J. Southerland, and A. H. Fagg. Learning grasp affordances through human demonstration. In *IEEE International Conference on Development and Learning*, 2006.

[10] R. Detry. *Learning of Multi-Dimensional, Multi-Modal Features for Robotic Grasping*. PhD thesis, University of Liège, 2010. Supervisor: Justus Piater.

[11] R. Detry, D. Kraft, O. Kroemer, L. Bodenhagen, J. Peters, N. Krüger, and J. Piater. Learning grasp affordance densities. *Paladyn. Journal of Behavioral Robotics*, 2(1):1–17, 2011.

[12] R. Detry and J. Piater. Grasp generalization via predictive parts. In *Austrian Robotics Workshop*, 2011.

[13] R. Detry, N. Pugeault, and J. Piater. A probabilistic framework for 3D visual object representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(10):1790–1803, 2009.

[14] C. Goldfeder, M. Ciocarlie, H. Dang, and P. Allen. The Columbia grasp database. In *IEEE International Conference on Robotics and Automation*, 2009.

[15] A. Herzog, P. Pastor, M. Kalakrishnan, L. Righetti, T. Asfour, and S. Schaal. Template-based learning of grasp selection. In *The PR2 Workshop (Workshop at IROS’11)*, 2011.

[16] Y. Hu, R. Eagleson, and M. A. Goodale. Human visual servoing for reaching and grasping: The role of 3-d geometric features. In *IEEE International Conference on Robotics and Automation*, 1999.

[17] L. S. Jakobson and M. A. Goodale. Factors affecting higher-order movement planning: a kinematic analysis of human prehension. *Experimental Brain Research*, 86(1):199–208, 1991.

[18] M. Jeannerod. The timing of natural prehension movements. *Journal of Motor Behavior*, 1984.

[19] R. S. Johansson. Sensory input and control of grip. In *Novartis Foundation Symposium*, pages 45–58, 1998.

[20] A. E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21:433–449, 1999.

[21] I. Kamon, T. Flash, and S. Edelman. Learning to grasp using visual information. In *IEEE International Conference on Robotics and Automation*, volume 3, pages 2470–2476, 1996.

[22] J. Kim. M. eng. Master’s thesis, Massachusetts Institute of Technology, 2007.

[23] A. T. Miller, S. Knoop, H. Christensen, and P. K. Allen. Automatic grasp planning using shape primitives. In *IEEE International Conference on Robotics and Automation*, volume 2, pages 1824–1829, 2003.

[24] L. Montesano and M. Lopes. Learning grasping affordances from local visual descriptors. In *IEEE International Conference on Development and Learning*, 2009.

[25] A. Morales, E. Chinellato, A. H. Fagg, and A. P. del Pobil. Using experience for assessing grasp reliability. *International Journal of Humanoid Robotics*, 1(4):671–691, 2004.

[26] G. Rizzolatti, R. Camarda, L. Fogassi, M. Gentilucci, G. Luppino, and M. Matelli. Functional organization of inferior area 6 in the macaque monkey. *Experimental Brain Research*, 71(3):491–507, 1988.

[27] G. Rizzolatti and G. Luppino. The cortical motor system. *Neuron*, 31(6):889–901, 2001.

[28] R. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (FPFH) for 3D registration. In *IEEE International Conference on Robotics and Automation*, 2009.

[29] M. Saric, C. H. Ek, and D. Kragic. Dimensionality Reduction via Euclidean Distance Embeddings. Technical report, KTH, Royal Institute of Technology, Stockholm, 2011.

[30] A. Saxena, J. Driemeyer, and A. Y. Ng. Robotic Grasping of Novel Objects using Vision. *International Journal of Robotics Research*, 27(2):157, 2008.

[31] B. Schölkopf and A. Smola. Kernel principal component analysis. *Artificial Neural Networks—ICANN’97*, 1997.

[32] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[33] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[34] K. B. Shimoga. Robot grasp synthesis algorithms: A survey. *The International Journal of Robotics Research*, 15(3):230, 1996.

[35] D. Song, C. H. Ek, K. Huebner, and D. Kragic. Multivariate Discretization for Bayesian Network Structure Learning in Robot Grasping. In *International Conference on Robotics and Automation*, pages 1–8. Royal Institute of Technology, 2011.

[36] J. D. Sweeney and R. Grupen. A model of shared grasp affordances from demonstration. In *International Conference on Humanoid Robots*, 2007.

[37] L. E. Zhang, M. Ciocarlie, and K. Hsiao. Grasp evaluation with graspable feature matching. In *RSS Workshop on Mobile Manipulation: Learning to Manipulate*, 2011.