Semantic-MultiKinFu: Multiple Semantic-TSDF 3D grid Maps for Warehouse Robots

Dinh-Cuong Hoang*, Todor Stoyanov*, and Achim J. Lilienthal*

Abstract— **Grasping systems that build upon meticulously planned hand postures rely on precise knowledge of object geometry, mass and frictional properties — assumptions which are often violated in practice. In this work, we propose an alternative solution to the problem of grasp acquisition in simple autonomous pick and place scenarios, by utilizing the concept of grasp envelopes: sets of constraints on gripper postures. We propose a fast method for extracting grasp envelopes for objects that fit within a known shape category, placed in an unknown environment. Our approach is based on grasp envelope primitives, which encode knowledge of human grasping strategies. We use environment models, reconstructed from noisy sensor observations, to refine the grasp envelope primitives and extract bounded envelopes of collision-free gripper postures. Also, we evaluate the envelope extraction procedure both in a stand alone fashion, as well as an integrated component of an autonomous picking system.**

I. Introduction

Despite significant advances in gripper hardware design and in robot planning and control algorithms, autonomous grasp acquisition under uncontrolled conditions remains a challenging research problem. On one hand, this is due to the necessity to solve a high-dimensional grasp- and motion planning problem for the full gripper-manipulator chain.

The classical approach to reduce this complexity is to decouple the grasp synthesis problem by planning separately for the gripper and the manipulator. To this end, state of the art grasping systems [1], [2], [3], [4] often rely on sets of pre-computed grasps. At runtime, these grasps are evaluated by a motion planner and executed in an open-loop fashion, or in combination with perceptual feedback ($e.g.$, visual servoing).

II. Related Work

As a comprehensive overview of data-driven grasp planning is beyond the scope of this work, we direct the interested reader to a recent survey by Bohg et al. [6] and only overview relevant approaches in constraint-based grasp representations.

In [7], TSRs are defined manually, while the main focus of our work is on extracting grasp envelopes in an automatic manner. Our grasp envelopes can be seen as a generalization of TSRs with application to the grasping task, and can be readily integrated in the motion planning framework proposed by Berenson et al.

The second set of approaches closely related to our work focus on the task of locating grasping affordances in various perceptual sensor data: for example, RGB images [8] or point clouds [9]. While there is a vast body of literature on extracting candidate grasps from various sensing modalities, most works concentrate on finding a single grasping configuration.

III. Grasp Envelope Extraction

A. Grasp Envelopes

Let $\boldsymbol{p} = [x, y, z, o_x, o_y, o_z, o_w]^T$ be the pose of the gripper frame relative to a fixed frame connected to the robot kinematic model, represented by a 3D translation and a quaternion-encoded orientation. In addition, let $\boldsymbol{q} \in \mathbb{R}^n$ be the configuration vector describing the joint angles of all n gripper joints. We represent the posture of the gripper as a vector $\boldsymbol{x} = \left[\boldsymbol{p}^T, \boldsymbol{q}^T\right]^T$. We then define a grasp envelope $\mathcal{G}$ as a set of gripper postures satisfying constraints imposed on the gripper pose and joint configuration:

$$\mathcal{G} = \left\{\boldsymbol{x} \in \mathbb{R}^{n+7} \,|\, c_i(\boldsymbol{x}) \leq 0, \, i = 1, \ldots, m\right\} \tag{1}$$

In this work, we restrict our analysis to linear inequality constraints, in order to facilitate an efficient implementation. In addition, we concentrate our analysis on an underactuated gripping device with a single actuated degree of freedom, thus $q \in \mathbb{R}$. For simplicity, we only impose box constraints on the gripper opening angle, $i.e.$, $q_{min} \leq q \leq q_{max}$.

The definition of a grasp envelope $\mathcal{G}$ in (1) allows us to encode various constraints on the robot end effector pose. In this work we define two prototypical sets of grasp envelope constraints, which we term *grasp envelope primitives* — a cylindrical and a spherical primitive. At runtime, the algorithms in Sec. III-C and III-D are used to refine the bounds on these primitives and obtain valid collision-free grasp envelopes for a particular object placement in the workspace.

An illustration of a cylindrical grasp envelope primitive is shown in Fig. 3(a), while a grasp envelope obtained by imposing additional bounds[1] on the primitive is shown in Fig. 3(b).

Our treatment of the spherical shell grasp primitive is done in an analogous manner, by constraining the gripper pose in between two concentric spheres while aligning the gripper's approach axis with the vector pointing towards the sphere center and maintaining the gripper's lateral axis (y in Fig. 3) approximately parallel to a horizontal plane.

*Centre for Applied Autonomous Sensor Systems (AASS); Orebro University.

[1] In previous work [10] we refer to this as a *truncated* grasp envelope.

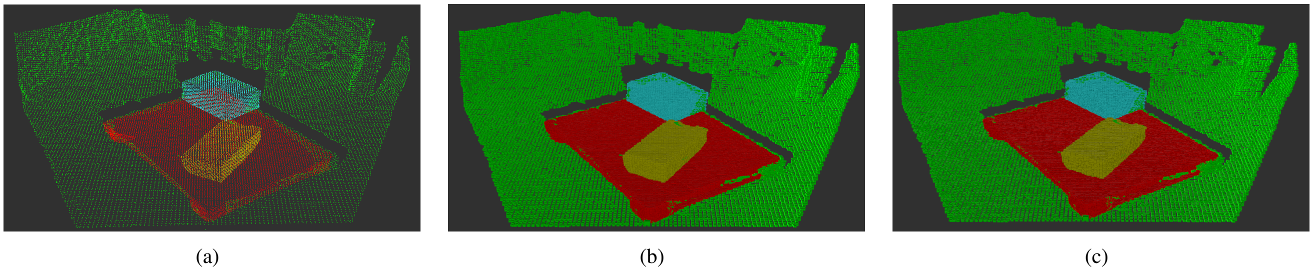

(a)
(b)
(c)

Fig. 1. Two-dimensional illustrations of the three main steps of our approach: (a) approach to pre-computing a collision map for the initial grasp envelope primitive; (b) three types of collisions which can occur and which are handled differently by our approach; (c) process used for finding the maximum-volume cluster of valid grasp postures. The maximum-volume cluster in (c) is shown in the shaded rectangle labeled **b**.

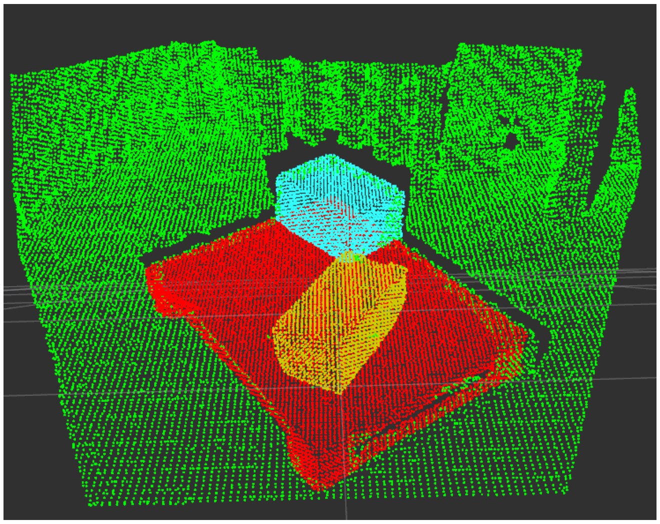


Fig. 2. The grasp envelope extraction procedure proposed in this work was used in an autonomous picking system, composed of the underactuated Velvet Fingers gripper [5] mounted on a KUKA LWR4+ robot arm (background). Experiments were performed to evaluate grasp acquisition success rates for the five test objects in the foreground.

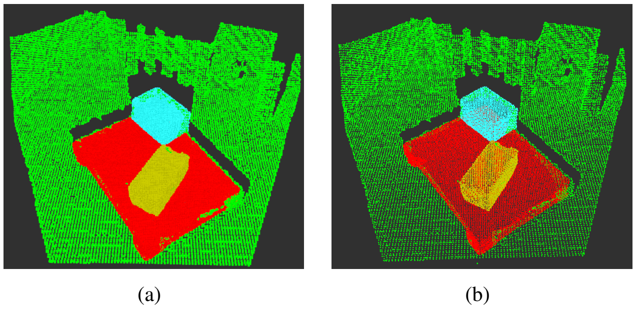

(a)
(b)

Fig. 3. An illustration of a cylindrical grasp envelope primitive in (a) and a corresponding grasp envelope in (b). The constraints are satisfied for any end effector pose which brings the gripper reference frame inside the shaded cylindrical shell sector, while maintaining orientation of the x and z axis within their respective shaded cone regions.

B. Pre-computing a Collision Map

The first step of our approach is to pre-compute a 3D collision map for each of the two grasp envelope primitives. This step is performed offline and greatly speeds up the subsequent online primitive refinement. The procedure, illustrated by the 2D projection in Fig. 4(a), begins by sampling gripper poses from the initial envelope primitive. We obtain the samples with a regular grid in the parametric space of each primitive and store them in a regular sample grid $\mathcal{S}$. For the cylindrical shell primitive we sample along the distance to the center d, the orientation α (w.r.t. the cylinder coordinate frame), and the height above the horizontal plane h. Similarly, the spherical shell primitive is parametrized by the radius r, and the polar and azimuth angles θ, ϕ.

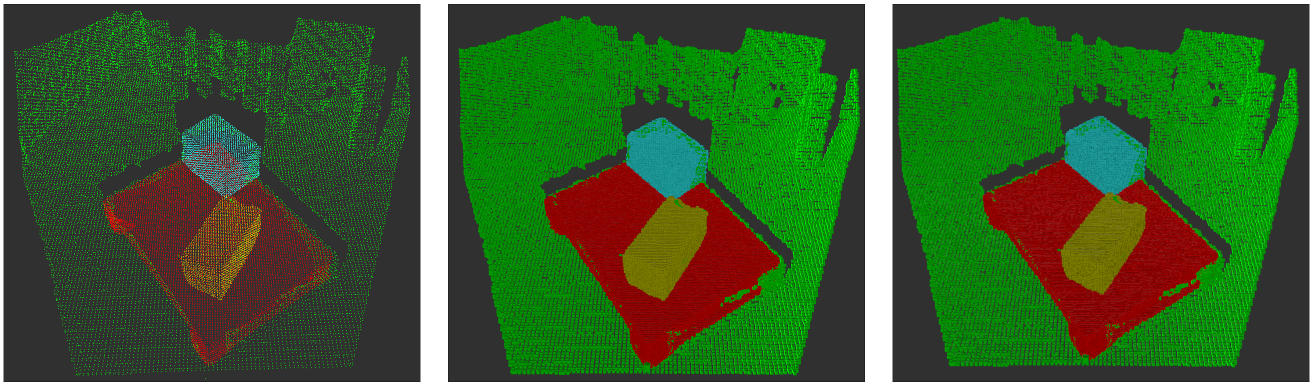
Next, for each sampled gripper pose we generate the gripper footprint in a collision map. As shown in Fig. 4(a), for this step we use an idealized gripper model, consisting of a bounding box around the palm and a semi-cylinder covering all possible placements of the fingers under different opening angles. Each cell of the collision map covered by the ideal gripper footprint is then marked as occupied by the currently evaluated sample from $\mathcal{S}$. By iterating this procedure over all samples in $\mathcal{S}$ we obtain a collision map which encodes in every cell the list of gripper postures that would potentially occupy it.

C. Finding Valid Configurations in Reconstructed Scenes

Once we obtain the intersection between the two maps, we iterate through all affected cells in the collision map and check the affected gripper pose samples from $\mathcal{S}$. As illustrated in Fig. 4(b), we distinguish between three types of collisions: 1) between the gripper palm and the map; 2) between the gripper fingers and the target object; and 3) between the fingers and other objects in the environment. Case 1) entails that the considered gripper pose would result in a collision, and thus we mark the corresponding sample in $\mathcal{S}$ as invalid. Cases 2) and 3) would result in a collision only for specific configurations of the gripper opening angle q, thus we use them to impose tighter bounds on q_{min} and q_{max} associated with the respective sample in $\mathcal{S}$. At the end of this procedure we check the bounds on q_{min}, q_{max} for all remaining valid samples in $\mathcal{S}$ and invalidate those that impose infeasible constraints ($i.e.$, $q_{min} \geq q_{max}$) and those that would not result in a grasp of the target object ($q_{min} = 0$ when no contact with the object was detected).

D. Fitting a Maximum Volume Envelope

Following the procedures outlined in the previous subsections, we obtain a grid of discrete samples $\mathcal{S}$ representing a set of gripper postures. Some of the postures in $\mathcal{S}$ have been



(a) (b) (c)

Fig. 4. Two-dimensional illustrations of the three main steps of our approach: (a) approach to pre-computing a collision map for the initial grasp envelope primitive; (b) three types of collisions which can occur and which are handled differently by our approach; (c) process used for finding the maximum-volume cluster of valid grasp postures. The maximum-volume cluster in (c) is shown in the shaded rectangle labeled **b**.

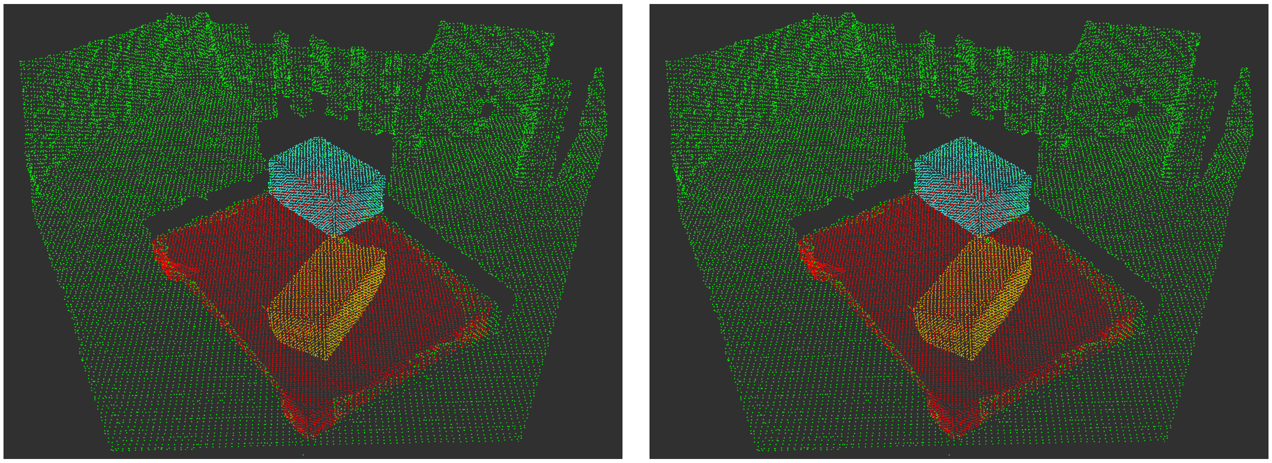
invalidated by the collision checks in Sec. III-C, while the remaining ones are valid under certain bounds on q. In order to obtain a refined grasp envelope following the definition in (1), we need to find a set of inequality constraints in Cartesian space containing only valid samples from $\mathcal{S}$. Thanks to our regular grid sampling strategy from Sec. III-B, it is straightforward to obtain a set of Cartesian-space constraints for any axis-aligned bounding box in $\mathcal{S}$. Thus, we can obtain a refined grasp envelope by looking for the largest axis-aligned inscribed box in $\mathcal{S}$: $i.e.$, the maximum volume box containing only valid samples.

Finding maximum volume/area inscribed shapes is in general a hard problem. Our problem instance is however additionally constrained to a regular sample grid and axis aligned shapes, and can therefore be solved efficiently. If we treat invalid samples in $\mathcal{S}$ as occupied space, and conversely valid samples as free space, we can use a variant of the distance transform to speed up our search. In general, we are looking for a k dimensional axis-aligned bounding box, with k being the dimensionality of the sampling grid $\mathcal{S}$. For simplicity, in this section we discuss a 2D version of our approach (illustrated in Fig. 4(c)), which is readily extendable to kD. The main idea here is to look for the best position of the lower-right corner of the maximum bounding rectangle (the shaded rectangle labeled **b** in Fig. 4(c)). We first construct a 2D distance field which for every free cell encodes the distance to the closest occupied cell in the negative direction ($i.e.$, towards the upper-left corner), along each dimension. This provides us with an absolute upper bound on the volume of a box which uses a particular cell as a lower-right corner: $e.g.$, the example labeled **a** in Fig. 4(c) would result in a maximal volume of $8 \times 5 = 40$ samples. Using this as a criterion to prune the search space, Algorithm 1 can be used to find the maximal-area box. Lines 8-15 in the algorithm describe the search procedure along the x direction of the distance grid, as shown also in case **b** of Fig. 4(c). In essence, we keep track of the maximum area rectangle verified so far (m_i, m_j), and update it for every slice along x in lines 10-11. We next check if it is possible to obtain a larger area rectangle (in the best case) if we continue searching. If so, we refine the boundary d_j along y (lines 12-13), otherwise we stop the search (line 15). Last, we check if the rectangle obtained in this manner is larger than the current best candidate and move to the next possible bottom-right corner (lines 18-19). For the envelope primitives discussed in this work, $k = 3$ and we obtain a straight-forward extension of the presented algorithm to 3D by searching in addition along the third dimension. At a moderate computational cost, this algorithm can be modified to extract the top N largest volume regions. Finally, since the zero position for sampling dimensions associated with varying orientation is arbitrary, in these cases we allow regions to span across the first-to-last element border and modify the distance field computation accordingly.

IV. Evaluation

A. Envelope extraction evaluation

As a first step, we evaluated the proposed grasp envelope fitting approach on a set of five test objects placed in ten different scenes. Here, the purpose was to evaluate the computational effort of the envelope extraction procedure, as well as the quality of the obtained envelopes. Under the assumption that all grasps which respect the envelope constraints are successful, envelope quality Q is defined as the number of valid gripper pose samples from S that fall within the grasp envelope. This measure directly relates to the enclosed volume of the envelope constraints. Thus, envelopes with larger quality Q allow for more gripper pre-grasp posture redundancy and easier grasp acquisition. Two plush toys (a teddy bear and a pig), a water bottle, a large cup and a cardboard box (shown in Fig. 6(a)), all graspable by the considered gripping device, were chosen as target objects. We placed the objects in different poses in two different types of scenes, models of which are shown in Fig. 5. For the first set of five scenes, the objects were placed on shelves in an office environment, resulting in a moderately cluttered setup. Conversely, in the second set, objects were placed on a table top and were spaced further apart from



(a) (b)

Fig. 5. Illustration of the test environments used in this paper: (a) shows the TSDF models of the five scenes from the *shelves* data set, reconstructed at 5 mm resolution; (b) shows the models of the *table* data set at 10 mm resolution.

Algorithm 1: BBSEARCH2D: find largest rectangle

```
Input: Distance transform $D$
$V_{max} \leftarrow 0$
for $\forall$ cells in $D$ do
    Let $i, j \leftarrow$ index of current cell, $D^x_{i,j}, D^y_{i,j} \leftarrow$ value of $D$ at $i, j$ along $x, y$ dimensions
    if $(i - D_i)(j - D_j) > V_{max}$ then
        $m_i, m_j \leftarrow 0$
        $d_i, d_j \leftarrow D^x_{i,j}, D^y_{i,j}$
        //search along $x$
        for $k \leftarrow i$ to $i - D_i$ do
            $d_{min} \leftarrow min(d_j, D^y_{k,j})$
            if $(k - i)d_{min} > m_i m_j$ then
                $m_i, m_j \leftarrow k - i, d_{min}$
            if $d_i d_{min} > m_i m_j$ then
                $d_j \leftarrow d_{min}$
            else
                break
        //search along $y$ equivalent
        ...
        if $(i - m_i)(j - m_j) > V_{max}$ then
            $V_{max} \leftarrow (i - m_i)(j - m_j)$, store bounds
return $V_{max}$, bounds
```

each other. These two sets of environments (referred to as the *shelves* and *table* data sets) were scanned using a handheld Asus Xtion Pro[2] RGB-D camera. The sensor pose was tracked using the SDF Tracker algorithm [11]. The shelves data set was reconstructed at grid resolutions of 5 mm and 10 mm, while the table data set was reconstructed only at a resolution of 10 mm, as the lower number of geometrical features in the latter caused bad tracker performance at higher grid resolutions. Object poses and bounding sphere/cylinder size were manually determined for each test case.

We pre-computed two sets of spherical and cylindrical grasp envelopes, using collision grid resolutions of 5 mm and 10 mm respectively. For the cylindrical envelopes, we sampled poses at 15 radial distances with $0.1 \text{ m} \leq d \leq 0.3$ m, 100 orientations with $0 \leq \alpha \leq 2\pi$ rad and 7 vertical slices over a span of 0.2 m. The spherical primitive was sampled at 15 distances r and 100 orientations, using the same metric bounds, as well as 7 different inclinations. In both cases, this resulted in a 3D sampling grid $\mathcal{S}$ of 10500 gripper poses. A visual example of the results obtained for the extracted grasp envelope of one of the objects in the shelves data set is shown in Fig. 6(b).

In addition to visual inspection, we also measured for each grasp envelope the computation time spent in extracting it and the quality metric Q. Fig. 6(c) shows a boxplot of the obtained envelope qualities per object and data set, with each box centered on the mean and spanning the area between the 25th and 75th percentile. We can draw several conclusions from the results shown in Fig. 6(c). First, it is evident that the less cluttered table data set results in substantially larger grasp envelopes, validating that the proposed method is able to find grasp envelopes with higher quality for objects that are easier to grasp. Second, by comparing results on the two reconstructions of the shelves data set, we note that the grasp envelopes extracted from higher resolution models are slightly larger. This effect is due to the more precise collision checks performed. However, in our tests the performance gain by using a higher resolution does not seem significant, most likely owing to the fact that the amount of clutter in the scenes is not extremely high. As this performance gain comes at a price of an order of magnitude slower computational time, we conclude that very precise environment models should only be utilized when necessary. Finally, we note that the performance of the proposed method is worst for the cup object. Upon further investigation, this artifact can be explained by a particularity of the TSDF representation used to model the environment: it is particularly susceptible

[2] http://www.asus.com/Multimedia/Xtion_PRO_LIVE/

to modeling errors on thin objects observed from multiple viewpoints. As both the outer and inner walls of the cup are often visible in our data sets, the fidelity of the models is often not optimal, resulting in lower certainty on the extracted grasp envelopes. Regarding computational performance, our approach extracts grasp envelopes in roughly 200 ms at a resolution of 5 mm and $20 - 40$ ms at a 10 mm resolution (using a single core of an Intel Xeon CPU E5-1620 v3 at 3.50GHz). Most of the time is spent on computing the valid samples as per Sec. III-C, with only a small fraction of resources expended on extracting the maximum volume envelope.

B. Grasp acquisition success evaluation

In order to evaluate the usefulness of the extracted envelopes for the purpose of autonomous grasp acquisition, we also integrated the proposed envelope extraction algorithm into the inequality Stack-of-Tasks (SoT) [12] manipulator control framework implementation we presented in [10]. In our experimental setup, we mounted the Velvet Fingers gripper (augmented with an Asus Xtion Pro depth camera) on a KUKA LWR 4+ robot. We roughly (by hand) placed one of the target objects shown in Fig. 2 at a known picking location, while the other four objects were placed pseudo-randomly in the workspace to simulate clutter (sample scene configuration shown in Fig. 7(a)). The five objects used in these experiments were: a plush teddy bear (Teddy, 82 g), a stack of duplo blocks (Duplo, 60 g), a water bottle (Bottle, 103 g), a coffee mug (Cup, 134 g), and a toy ball (Ball, 54 g). The two grasp envelope primitives generated in the previous sub-section (at a model resolution of 10 mm) were used also in this trial: the cylindrical primitive was associated to the Bottle, Cup and Duplo objects, while the spherical one was used for the Teddy and Ball objects.

In each experimental run we first controlled the manipulator to move the gripper-mounted camera to three predefined scene observation poses. Simultaneously we built a TSDF model from the depth images, using camera poses obtained through the robot's forward kinematic model. We then used our grasp envelope extraction procedure to obtain constraints on the gripper posture for the target object, which were subsequently used to form control tasks used in the SoT framework (see [10] for a more in-depth description). At this point, we made a slight modification to the envelope extraction procedure in order to obtain envelopes that were more likely to be reachable for the employed robot arm. To this end, we imposed additional constraints on the gripper orientation prior to extracting the grasping envelopes, requiring the final end effector orientation to be within a cone of width $\frac{\pi}{2}$ rad, centered at the initial end effector orientation. A sample grasp envelope extracted in this manner from a training scene is shown in Fig. 7(b). Once the manipulator motion control satisfied the grasp envelope constraints, we executed the grasp acquisition routine described in [3], in order to obtain an enveloping grasp of the target object. A trial was judged successful if the target object could be lifted and extracted from the scene. We performed twenty trials for each target object, with varying placement poses of the surrounding objects in the scene.

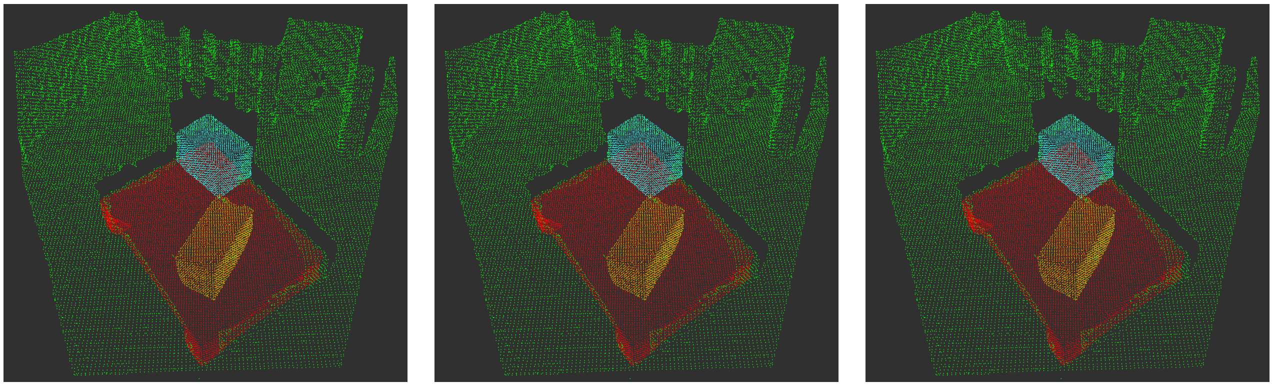
The results of this set of experiments are summarized in Table I. For each of the test objects we report respectively: the number of experiments performed; the number of successful grasps; the respective success rate; the grasp envelope quality Q; the trajectory length $l = \int_0^{t^m} \|\dot{\boldsymbol{q}}(t)\|_1 dt$ as the sum of angular distances traveled by all joints of the manipulator; the times t^p, t^e, t^m, and t^g for the pre-positioning, envelope extraction, grasp pose approach motion and grasp acquisition phases respectively; as well as the total time for each run $\sum t$. In eight of the trials the target object was occluded from the sensor view point and our approach did not find a suitable grasping envelope. These trials are excluded from the statistics in Table I.

The obtained statistics are relatively uniform across the different trials and different objects, with two minor exceptions. First, the success rates for the Teddy are notably lower than for the rest of the target objects. This discrepancy is unlikely to be linked to the size of the extracted envelopes, which are almost on par with the other object associated to a spherical envelope primitive (the Ball). We attribute the lower success rate chiefly to the lower friction coefficient of the Teddy, which leads to frequent slippage against the belts of the Velvet Fingers gripper and consequently a lower likelihood of attaining a stable grasp. The second discrepancy is in the substantially longer trajectory execution times (t^m) for the Duplo object. The reason for this is that the underlying task dynamics parameters in the SoT framework were adjusted after the tests on the Duplo object, in order to speed up testing. We also note the reasons for failure in grasping for the remaining objects: for the Duplo object, two failures were due to the object toppling over and one failure was due to a collision with an obstacle during approach; for the Bottle, both failures were due to the object toppling; finally, for the Ball all three failures were due to the object rolling out of the gripper upon contact. Finally, the grasp envelope extraction procedure produced slightly smaller volumes in comparison to the offline tests from Sec. IV-A, due to the additional constraint on end effector orientation. The reported envelope extraction run-times t^e were consistent, allowing for some overhead for message passing between different nodes.

V. Discussion

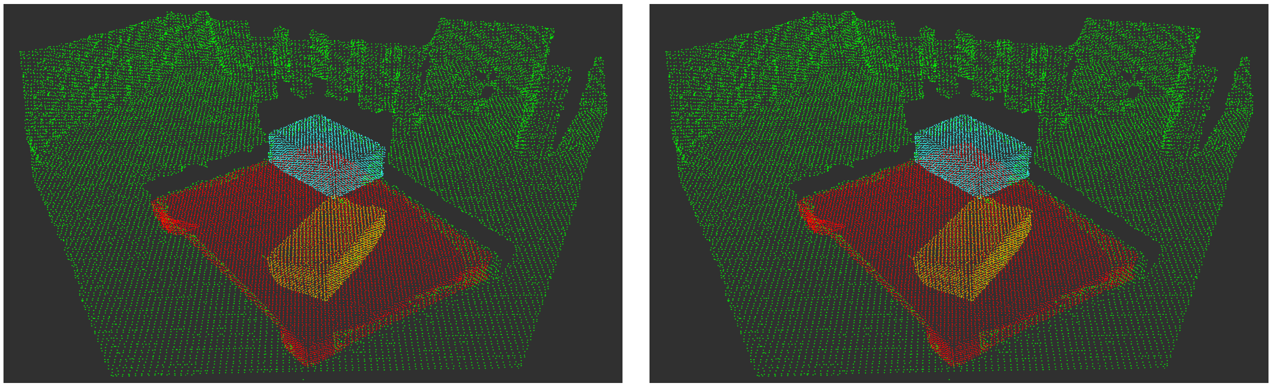
In this article we propose a method for extracting grasp envelopes — constraints on gripper pose and joint configurations — from online reconstructed workspace models. We utilize knowledge of basic human grasping strategies to define two grasp envelope primitives favoring grasps along the target object PCA directions.

A limitation of the current work is the lack of rigorous treatment of kinematic feasibility of the extracted envelopes. Therefore, we plan to extend our approach using information of the likelihood of achieving grasp poses, *e.g.* by using capability maps [13].



(a) (b) (c)

Fig. 6. (a) shows the five test objects used in the evaluation. The two plush toys were evaluated with spherical envelope primitives fit to the head, while the remaining objects were associated with cylindrical primitives. (b) illustrates typical constraint envelopes extracted for objects in cluttered environments; (c) summarizes results of the proposed grasp envelope extraction method, showing the distribution of envelope quality Q over different objects and scenes.



(a) (b)

Fig. 7. Setup used for the experiments in Sec. IV-B. A sample scene used in the evaluation is shown in (a). A reconstructed test scene model and a corresponding grasp envelope (visualized as a set of selected grasping configurations) are shown in (b).

TABLE I

GRASP ACQUISITION EVALUATION

Object	# of exp.	Success	Rate [%]	Q	$l\,[rad]$	$t^p\,[s]$	$t^e\,[s]$	$t^m\,[s]$	$t^g\,[s]$	$\sum t\,[s]$
Duplo	20	17	85.0	161.4 ± 109.4	4.3 ± 0.9	14.9 ± 3.0	0.06 ± 0.11	25.9 ± 3.8	14.7 ± 3.9	55.6 ± 5.3
Cup	16	16	100.0	51.2 ± 27.5	4.7 ± 1.3	13.7 ± 3.5	0.06 ± 0.05	10.8 ± 0.2	6.9 ± 2.6	31.5 ± 3.7
Bottle	17	15	88.2	67.1 ± 68.3	4.9 ± 2.3	14.0 ± 3.3	0.03 ± 0.05	10.8 ± 0.4	8.5 ± 3.9	33.3 ± 6.1
Ball	20	17	85.0	32.7 ± 16.6	5.6 ± 0.4	13.9 ± 3.4	0.06 ± 0.05	10.8 ± 1.4	13.9 ± 4.3	38.7 ± 4.9
Teddy	19	12	63.2	28.9 ± 13.8	5.6 ± 0.5	15.5 ± 2.8	0.06 ± 0.07	10.6 ± 0.9	11.5 ± 4.6	37.8 ± 4.3
Total	92	77	83.7	69.46 ± 78.06	5.0 ± 1.3	14.4 ± 3.2	0.06 ± 0.07	14.1 ± 6.6	11.4 ± 4.9	39.9 ± 10.0

REFERENCES

[1] V. Tincani, M. G. Catalano, E. Farnioli, M. Garabini, G. Grioli, G. Fantoni, and A. Bicchi, "Velvet fingers: A dexterous gripper with active surfaces," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 1257–1263.

[2] D. Berenson, R. Diankov, K. Nishiwaki, S. Kagami, and J. Kuffner, "Grasp planning in complex scenes," in *Proc. IEEE/RAS International Conference on Humanoid Robots*, 2007, pp. 42–48.

[3] S. Srinivasa, D. Ferguson, C. Helfrich, D. Berenson, A. Collet, R. Diankov, G. Gallagher, G. Hollinger, J. Kuffner, and M. VandeWeghe, "Herb: A home exploring robotic butler," *Autonomous Robots*, vol. 28, no. 1, pp. 5–20, 2010.

[4] R. Krug, T. Stoyanov, M. Bonilla, V. Tincani, N. Vaskevicius, G. Fantoni, A. Birk, A. J. Lilienthal, and A. Bicchi, "Velvet fingers: Grasp planning and execution for an underactuated gripper with active surfaces," in *Proc. of the IEEE International Conference on Robotics and Automation*, 2014, pp. 3669–3675.

[5] T. Stoyanov, N. Vaskevicius, C. Müller, *et al.*, "No more heavy lifting: Robotic solutions to the container unloading problem," *IEEE Robotics and Automation Magazine*, 2016, in press.

[6] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis — a survey," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, 2014.

[7] D. Berenson, S. Srinivasa, and J. Kuffner, "Task space regions: A framework for pose-constrained manipulation planning," *The International Journal of Robotics Research*, vol. 30, no. 12, pp. 1435–1460, 2011.

[8] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," *The International Journal of Robotics Research*, vol. 27, no. 2, pp. 157–173, 2008.

[9] A. t. Pas and R. Platt, “Localizing grasp affordances in 3-d points clouds using taubin quadric fitting,” *arXiv preprint arXiv:1311.3192*, 2013.

[10] R. Krug, T. Stoyanov, V. Tincani, H. Andreasson, R. Mosberger, G. Fantoni, and A. J. Lilienthal, “The next step in robot commissioning: Autonomous picking & palletizing,” *IEEE Robotics and Automation Letters*, 2016, in press.

[11] D. R. Canelhas, T. Stoyanov, and A. J. Lilienthal, “Sdf tracker: A parallel algorithm for on-line pose estimation and scene reconstruction from depth images.” in *IROS*. IEEE, 2013, pp. 3671–3676.

[12] O. Kanoun, F. Lamiraux, and P.-B. Wieber, “Kinematic control of redundant manipulators: Generalizing the task-priority framework to inequality task,” *IEEE Transactions on Robotics*, vol. 27, no. 4, pp. 785–792, 2011.

[13] F. Zacharias, C. Borst, and G. Hirzinger, “Capturing robot workspace structure: representing robot capabilities,” in *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*. Ieee, 2007, pp. 3229–3236.