Automatic object shape completion from 3D point clouds for object manipulation

Rui Figueiredo, Plinio Moreno and Alexandre Bernardino
Institute for Systems and Robotics (ISR/IST), LARSyS, Instituto Superior Técnico,Universidade de Lisboa
Torre Norte Piso 7, Av. Rovisco Pais 1 1049-001 Lisboa, Portugal
{*ruifigueiredo, plinio, alex*}*@isr.ist.utl.pt*

Keywords: Shape completion, Symmetry, Part-based object representation, Semantic parts

Abstract: 3D object representations should be able to model the shape at different levels, considering both low-level and high-level shape descriptions. In robotics applications, is difficult to compute the shape descriptors in self-occluded point clouds while solving manipulation tasks. In this paper we propose an object completion method that under some assumptions works well for a large set of kitchenware objects, based on Principal Component Analysis (PCA). In addition, object manipulation in robotics must consider not only the shape but the of actions that an agent may perform. Thus, shape-only descriptions are limited because do not consider where the object is located with respect to others and the type of constraints associated to manipulation actions. In this paper, we define a set of semantic parts (i.e. bounding boxes) that consider grasping constraints of kitchenware objects, and how to segment the object into those parts. The semantic parts provide a general representation across object categories, which allows to reduce the grasping hypotheses. Our algorithm is able to find the semantic parts of kitchenware objects in and efficient way

1 INTRODUCTION

Dealing with unknown objects is a strong research topic in the field of robotics. In applications related with object grasping and manipulation (Figueiredo et al., 2012), robots aimed at working in daily environments have to interact with many never-seen-before objects and increasingly complex scenarios.

In this work we adopt a compact object representation, which is based on bounding box sizes and geometrical moments as the main features. More specifically, the proposed representation relies on object dimensions along its main geometrical axes, which can be extracted from 3D point cloud information via Principal Component Analysis (PCA). The proposed representation is low-dimensional, robust to noise and suitable for object categorization and part-based grasp hypotheses generation. However, as any other type of reconstruction based on single views, computing objects bounding boxes is a ill posed problem due to lack of observability of the self-occluded part. Therefore some assumptions about the occluded part must be taken. In this work, we consider that objects present symmetries, so that we are able to reconstruct the unobserved part of the point cloud. This models perfectly simple object shapes like boxes, spheres and cylinders, and is a reasonable assumption for many objects of daily usage when lying on a table. Once the object is completed, global shape characteristics can be extracted and used for object category reasoning, grasp planning and learning.

The remainder of the article is organized in the following manner. In section 2 we overview some related work. Then, in section 3, we describe the proposed methodologies. In section we assess the performance of our approach in a real scenario 4. Finally, in section 5 we draw some conclusions.

2 RELATED WORK

Object grasping and manipulation is one of the most challenging tasks in today's robotics. A fundamental aspect behind the success of a grasping solution, is the choice of the object representation. This should be able to deal with incomplete and noisy perceptual data, and be suitable for real-time applications. Moreover, these should be flexible enough to allow for grasp generalization over multiple object classes.

2.1 Object Representations

Several object representations have been proposed and used in the past to plan and learn grasping and manipulation actions: complete meshes for known objects (de Figueiredo et al., 2015), reconstructed meshes for unknown objects (Aleotti et al., 2012), and object part clusters modeled with superquadrics (Faria et al., 2012), also for unknown objects. Despite these representations allowing for good precision in representing the shape of the objects, they suffer from high-dimensionality and varying description length, thus being hard to define a representation of object categories suitable for generalization. Furthermore, the noise present in the sensor data may negatively influence representations with large number of parameters. Given the current perception technology, the most robust and simple features to represent objects, match their similarity with others, and provide a basis for the definition of categories, must be low-dimensional and rely on gross features, to prevent over-fitting.

2.2 Shape Completion

In recent years shape completion using a single view has been extensively studied, typically in robotics grasping applications. Usually multiple object partial views are acquired from different viewpoints, using 3D range cameras, and the gathered point clouds are then registered and aligned together in a common reference frame. The Iterative Closest Point algorithm (Besl and McKay, 1992) and efficient variants (Rusinkiewicz and Levoy, 2001) are often used to compute the alignment transformations and to build a complete object shape model (Chen and Medioni, 1992). However, when only a single view is available and/or it is not possible to acquire several views due to time constraints or scenario/robot restrictions the shape completion problem becomes harder and some assumptions or pattern analysis must be made. In this direction, a wide range of ideas have been proposed including fitting the visible object surface with primitive shapes such as cylinders, cones, parallelepipeds (Marton et al., 2009; Kuehnle et al., 2008) or with more complex parametric representations like superquadrics (Biegelbauer and Vincze, 2007).

Closely related to our shape completion approach, Thrun and Wegbreit (Thrun and Wegbreit, 2005) proposed a method based on the symmetry assumption. This method considers 5 basic and 3 composite types of symmetries that are organized in an efficient entailment hierarchy. It uses a probabilistic model to evaluate and decide which are the completed shapes, generated by a set of hypothesized symmetries, that best fit the object partial view. More recently Kroemer et al. (Kroemer et al., 2012) proposed an extrusion-based completion approach that is able to deal with shapes that symmetry-based methods cannot handle. The method starts by detecting potential planes of symmetry by combining the Thrun and Wegbreit method with Mitra et al.'s fast voting scheme (Mitra et al., 2006). Given a symmetry plane, an ICP algorithm is used to decide the extrusion transformation to be applied to the object partial point cloud. Despite the fact that these methods were shown to be robust to noise and were able to deal with a wide range of object classes, they are inherently complex in terms of computational effort and thus, not suitable in real-time. Nevertheless, to simplify this problem, one can take advantage of common scenario structures and objects properties that are usually found in daily environments. They mostly involve man-made objects that are typically symmetric and standing on top of planar surfaces. For example, Bohg et al. (Bohg et al., 2011) took advantage of the table-top assumption and the fact that many objects have a plane of reflection symmetry. Starting from the work of Thrun and Wegbreit (Thrun and Wegbreit, 2005) and similar in spirit to Bohg et al. (Bohg et al., 2011), we propose a new computationally efficient shape completion approach which translates a set of environmental assumptions into a set of approximations, allowing us to reconstruct the object point cloud in real-time, given a partial view of the object.

3 Methodologies

The role of our algorithm is to obtain a semantic description of the perceived objects in terms of their pose, symbolic parts and probability distributions over possible object categories. The object segmentation step (Muja and Ciocarlie,) is followed by part detection and object category estimation, which rely on a full object point cloud. When only a partial view of the object is available, we employ a symmetry-based methodology for object shape completion. Next, the extraction of semantical parts is based on the object's dimensions along the main geometrical axes and can be achieved by bounding-box analysis via PCA. The low dimensional and efficient representation obtained guides the division of each object into a set of semantical parts, namely, *top*, *middle*, *bottom*, *handle* and *usable area*. This reduces the search space for robot grasp generation, prediction and planning. The next subsections explain our symmetry-based method for shape-completion and

Figure 1: Objects having rotational symmetry.

the division of the completed point cloud into a set of semantical parts.

3.1 Object Perception with Symmetry Assumptions

As any other type of reconstruction based on single views, computing the bounding-box of the object is an ill posed problem due to lack of observability of the self-occluded part. Thus, as for grasping procedures it is necessary that the robot knows the complete shape of the object of interest, some assumptions about the occluded part must be made. Inspired by the work of Thrun and Wegbreit (Thrun and Wegbreit, 2005) and Bohg et al. (Bohg et al., 2011) and with computational efficiency in mind, we propose a new approach that translates a set of assumptions and rules of thumb observed in many daily environments into a set of heuristics and approximations. They allow us to reconstruct the unobserved part of an object point cloud in real-time, given a partial view.

We consider the following assumptions: a) the objects stand on top of a planar surface (table top assumption); b) the camera is at a higher viewpoint; c) the objects have rotational symmetry; d) their main geometrical axes are either orthogonal or parallel to the supporting plane; e) the axis of symmetry corresponds to one of the main geometrical axes; and f) the direction of the axis of symmetry indicates the object's pose (i.e., upright or sideways).

These constraints model perfectly simple box-like and cylinder-like object shapes, such as kitchen-ware tools, and are reasonable assumptions for many other approximately symmetric objects, such as tools (see Fig. 1). Analogous to (Bohg et al., 2011), we consider only one type of symmetry, however we employ the *line reflection symmetry* (Thrun and Wegbreit, 2005) as it copes better with the object categories that we want to detect.

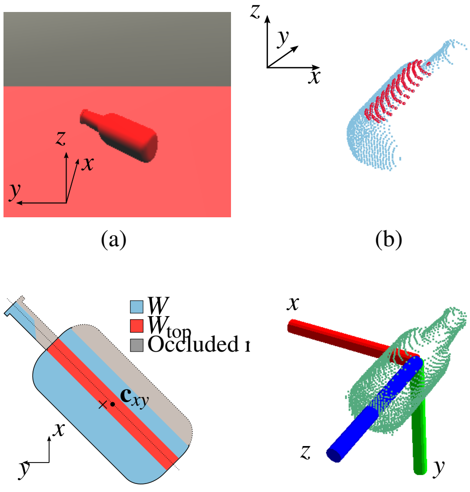
Let $P = \{\mathbf{p}\} \subset \mathbb{R}^3$ be the set of visible object surface points. Our shape completion algorithm finds the object symmetry axis s reflecting all visible surface points across it. This corresponds to rotating P around s by 180°. We determine s by analyzing the box that encloses the set P, considering the principal directions of the box and the dimensions along those directions. The symmetry axis (i.e., principal direction) is orthogonal to the cross product of the bounding box directions whose dimensions are the closest, and passes through the bounding-box centroid. To cope with the supporting plane assumption, we compute the horizontal (i.e., table plane, xy) and vertical (i.e. table normal, z) bounding-box directions and their dimensions separately. The vertical direction of the bounding-box is given by the normal vector to the table plane and its length is given by the furthest point from the supporting plane $d_z = \max_z(P)$. Since the horizontal directions are arbitrarily oriented in the supporting plane, we apply the projection of P onto the table plane and compute the directions and their dimensions in that space. The 2D components of the centroid location on the table plane cannot be correctly estimated from a partial view in most of the cases (as illustrated in Fig. 2). Let $W = \{\mathbf{p}_{z=0}\} \subset \mathbb{R}^2$ be the set containing the projected points. We assume that the top part of the object is visible and holds the symmetry assumptions so that the object's xy-centroid, c_{xy}, is obtained by considering only the top region points $W_{\text{top}} = \{\mathbf{w}_{\text{top}}\} \subset W$, satisfying the condition:

$$\mathbf{w}^i_{\text{top}} = \begin{cases} \mathbf{p}^i_{z=0} & \text{if } p_z > \sigma d_z \\ \emptyset & \text{otherwise} \end{cases} \tag{1}$$

where $\sigma \in [0,1]$ is a parameter tuned according to the camera view-point and the object shape curvature (i.e., σ is higher for cylinder-like shapes and lower for parallelepiped-like ones). The eigenvectors provided by PCA on the set W define the horizontal directions whereas their lengths are given by projecting the points in W onto its eigenvectors and finding the maximum in each direction.

3.2 Part-based object representation

We consider two main types of objects: tools and other objects. A tool has as parts a *handle* and a *usable area*, while the rest of the objects have *top*, *middle*, *bottom* parts and may have *handles*. When the axis of symmetry is parallel to the supporting plane and the lengths of the remaining directions are smaller than a predefined threshold, we consider that the object has a *handle* and a *usable area*. In order to cope with objects such as mugs and pans we detect a handle if a circle is fitted in the projected points W with a large confidence. The points lying outside of the circle are labeled as *handle*. The rest of the points are divided along the axis of symmetry into *top*, *middle* and *bottom*. Fig. 3 illustrates examples of detected



(a) (b)

(c) (d)

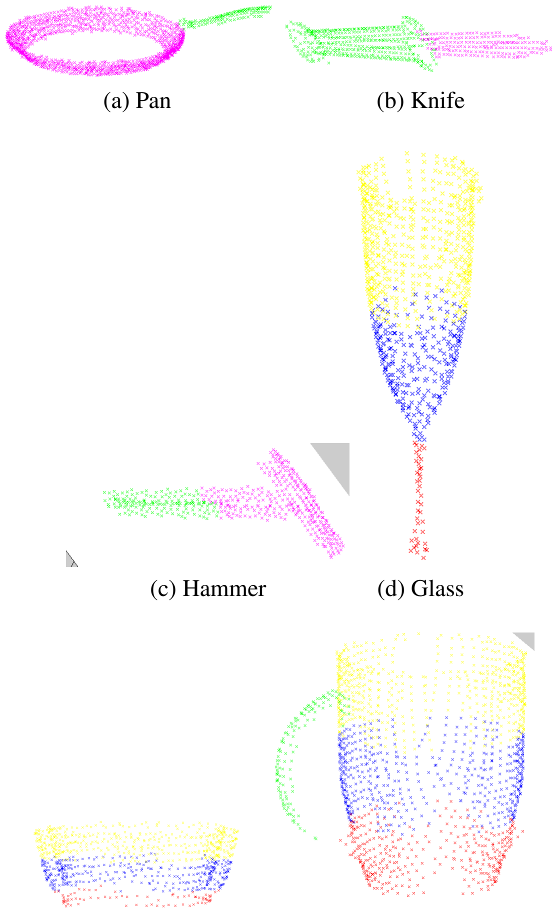
Figure 2: 2D centroid estimation in the presence of self-occlusion. (a) Bottle camera-view. (b) Visible region (blue) and top visible region surface points (red). (c) Bottle planar projection: $\times$ marks the centroid of W (blue), whereas $\bullet$ indicates the centroid of W_{top} (red). (d) After shape completion, an object coordinate frame is defined as having its origin at the bounding box centroid and z-axis aligned with the symmetry axis.

semantic parts for several objects using our completion algorithm.

The bounding boxes of the object parts define the pre-grasp hypotheses, providing two pre-grasp poses for each face of a box, as illustrated in Fig. 4. The final number of pre-grasp hypotheses is pruned in a first stage by the task-dependent logical module and in a second stage by a collision checker and the motion trajectory planner.

4 Experiments

In order to evaluate the proposed approaches, we consider two settings: In the first setting, the experiments are run in a simulated environment (ORCA (Hourdakis et al., 2014)), which provides the sensor (laser range camera Asus Xtion PRO (ASUS,)), objects and interface to physics engine (Newton Game Dynamics library (Jerez and Suero,)) where single objects are placed on top of a table. The object poses considered are *upright* or *sideways* due to the ambiguity between upright and upside-down when using global shape representations. The object semantic parts include: *top*, *middle*, *bottom*, *handle* and *usable area*. The 18 different objects includes instances of categories *pan, cup, glass, bottle, can, hammer, screwdriver, knife*[1]. In the second setting, the experiments are run in an real table-top scenario with 7 objects that belong to the instances of categories *glass*(1), *bottle*(2), *can*(1), *hammer*(1), *screwdriver*(1) and *cup*(1).



(a) Pan (b) Knife

(c) Hammer (d) Glass

(e) Bowl (f) Mug

Figure 3: Semantic parts for several objects after applying the completion algorithm. The colors correspond to parts as follows: yellow - top, blue - middle, red - bottom, green - handle, and magenta - usable area.

Table 1: Accuracy (%) for object part and pose detection.

Dataset	Part detection	Pose detection
Simulation	84.56	100
Real objects	82.14	100

The performance of our algorithm is based on the correct detection of the object parts and their pose. Results are shown in Table 1. We note that the PCA global representation is able to cope well with object pose detection, considering the table-top assumption

[1] Available at `http://www.first-mm.eu/data.html`

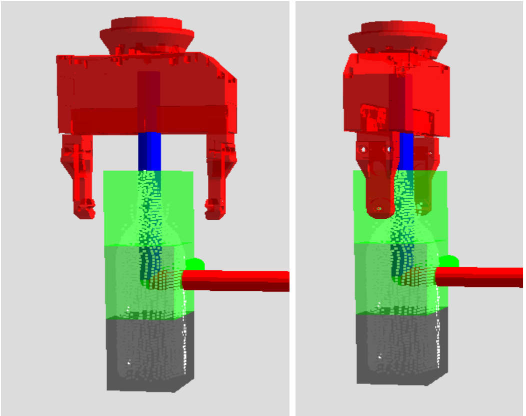


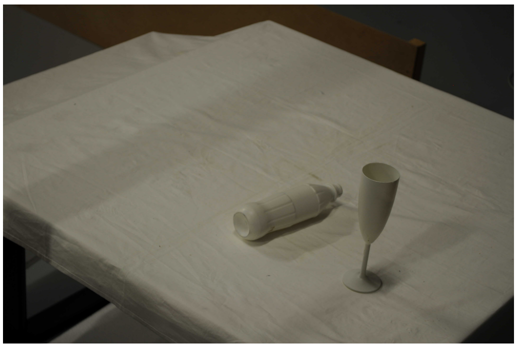
Figure 4: Examples of the pre-grasp gripper poses for a face of the top part of a bottle.

and the object categories assumed. We note that we do not consider the upside-down pose for these tests, as in real-world applications usual poses are upright and sideways. Object part detection suffers from part occlusion for particular object poses, reducing the pipeline performance for object category prediction. In addition to the accuracy, we stress the execution time for the object completion using symmetries and pose detection. The average execution times are 27.5 ms and 15.71 ms on a PC using one core of the Intel Xeon (2.67GHz). These numbers confirm the computational efficiency of our approach, which allows to make fast decisions.

5 CONCLUSIONS

In this work we proposed a novel method for symmetry-based shape completion from single-view 3D point clouds. Furthermore, we introduced bounding box sizes and geometrical moments as features for model-free categorization of every-day objects. We showed that our approach is computationally efficient, robust to noise and, hence, that can be used for pre-grasp reasoning and learning, in real scenarios.

Despite being unsuitable to describe fine details, our representation is robust to noisy perceptual data, convenient for the definition of object categories, which cover a large range of objects. It also provides a basis for reasoning about symmetries and can be used as a fall-back mechanism for grasp planning when other representations fail.



(a) Scenario 1

(b) Scenario 2

(c) Scenario 3

Figure 5: Experimental settings with the real table-top scenario. Each picture shows the objects utilized for each experiment

ACKNOWLEDGEMENTS

This work has been partially supported by the Portuguese Foundation for Science and Technology (FCT) project [UID/EEA/50009/2013]. Rui Figueiredo is funded by FCT PhD grant PD/BD/105779/2014.

REFERENCES

Aleotti, J., Rizzini, D. L., and Caselli, S. (2012). Object categorization and grasping by parts from range scan data. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 4190–4196. IEEE.

ASUS. Asus xtion pro.

Besl, P. J. and McKay, N. D. (1992). A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256.

Biegelbauer, G. and Vincze, M. (2007). Efficient 3d object detection by fitting superquadrics to range image data for robot's object manipulation. In *IEEE International Conference on Robotics and Automation*, pages 1086–1091.

Bohg, J., Johnson-Roberson, M., León, B., Felip, J., Gratal, X., Bergström, N., Kragic, D., and Morales, A. (2011). Mind the gap - robotic grasping under incomplete observation. In *IEEE International Conference on Robotics and Automation*, pages 686–693.

Chen, Y. and Medioni, G. (1992). Object modelling by registration of multiple range images. *Image Vision Comput.*, 10(3):145–155.

de Figueiredo, R. P., Moreno, P., and Bernardino, A. (2015). Efficient pose estimation of rotationally symmetric objects. *Neurocomputing*, 150:126–135.

Faria, D. R., Martins, R., Lobo, J., and Dias, J. (2012). A probabilistic framework to detect suitable grasping regions on objects. *IFAC Proceedings Volumes*, 45(22):247 – 252.

Figueiredo, R., Shukla, A., Aragao, D., Moreno, P., Bernardino, A., Santos-Victor, J., and Billard, A. (2012). Reaching and grasping kitchenware objects. In *System Integration (SII), 2012 IEEE/SICE International Symposium on*, pages 865–870. IEEE.

Hourdakis, E., Chliveros, G., and Trahanias, P. (2014). Orca: A multi-tier, physics based robotics simulator. *Journal of Software Engineering for Robotics*, 5(2):13–24.

Jerez, J. and Suero, A. Newton Game Dynamics. Opensource Physics Engine.

Kroemer, O., Ben Amor, H., Ewerton, M., and Peters, J. (2012). Point cloud completion using symmetries and extrusions. In *Proceedings of the International Conference on Humanoid Robots*.

Kuehnle, J., Xue, Z., Stotz, M., Zoellner, J., Verl, A., and Dillmann, R. (2008). Grasping in depth maps of time-of-flight cameras. In *International Workshop on Robotic and Sensors Environments*, pages 132–137.

Marton, Z. C., Goron, L., Rusu, R. B., and Beetz, M. (2009). Reconstruction and Verification of 3D Object Models for Grasping. In *Proceedings of the 14th International Symposium on Robotics Research*.

Mitra, N. J., Guibas, L., and Pauly, M. (2006). Partial and approximate symmetry detection for 3d geometry. *ACM Transactions on Graphics*, 25(3).

Muja, M. and Ciocarlie, M. Table top segmentation package.

Rusinkiewicz, S. and Levoy, M. (2001). Efficient variants of the icp algorithm. In *International Conference on 3-D Digital Imaging and Modeling*.

Thrun, S. and Wegbreit, B. (2005). Shape from symmetry. In *International Conference on Computer Vision*, pages 1824–1831.