

Robust cylinder detection and pose estimation using 3D point cloud information

Rui Figueiredo, Plinio Moreno and Alexandre Bernardino
Institute for Systems and Robotics, Instituto Superior Técnico, Lisboa, Portugal
Email: {ruifigueiredo, plinio, alex}@isr.tecnico.ulisboa.pt

Abstract—This work deals with the problem of detecting cylindrical shapes, commonly found in household and industrial environments, using 3D point cloud information from consumer RGB-D cameras. Existing approaches are fragile in the presence of clutter, in particular flat surfaces, leading to errors during the orientation estimation process that compromise the whole method. We address the aforementioned problem with a novel soft voting scheme that incorporates curvature information in the orientation voting phase. For each potential cylinder point, the principal curvature direction is combined with the normal vector to disambiguate candidate orientations. A set of experiments with synthetically generated data are used to assess the robustness of our method with different levels of clutter and noise. The results demonstrate that incorporating the principal curvature direction within the orientation voting process allows for large improvements on cylinders' parameters estimation. Qualitative results with point clouds acquired from consumer RGB-D cameras, confirm the advantages of the proposed approach.

I. INTRODUCTION

Due to recent technological advances in the field of 3D sensing, range sensors have become financially affordable to the average consumer, boosting the proliferation of robotics applications requiring accurate 3D object recognition and pose estimation capabilities. Mainly in tasks involving interaction with the surrounding environment, such as manipulation, robots require accurate object pose information in order to achieve successful grasps. Efficiency is another important requirement in robots with power limitations [1], where fast and accurate perception is required, e.g. for the manipulation of kitchenware objects [2]. Therefore, it is of the utmost importance to build efficient perceptual systems that are not only robust to sensory noise, but also to occlusion and clutter.

In this work, we focus on cylindrical shaped objects which are commonly found in everyday environments, ranging from domestic (e.g. cups, cans, bottles) to industrial contexts (e.g. pipes, pillars, scaffolds). One of the most successful cylinder fitting approaches in the state-of-the-art [3] is based on a computationally efficient 2-step Generalized Hough Transform (GHT). We extend the former method with a set of improvements that allow coping with large levels of clutter, in particular flat surfaces which often introduce problematic biases during the orientation estimation.

Our main contribution is threefold: first, we propose a novel randomized sampling scheme for the creation of orientation Hough accumulators. Our sampling method allows the incorporation of prior structure knowledge during the creation



(a) RGB-D kinect data

Fig. 1: A snapshot of a RGB-D point cloud and overlaid cylinders (green) detected with our methodology. Figure best seen in color.

of the orientation Hough accumulator, which improves accuracy with the same computational resources. Second, we introduce a novel soft-voting scheme, which considers surface curvature information, in order to cope with flat surfaces which previously voted for erroneous and arbitrary tangential orientations. Third, we perform a systematic and thorough quantitative assessment of the influence of noise and clutter on detection and pose estimation error of cylinder fitting methods, comparing our proposed method with that of [3]. Our ROS [4] C++ implementation runs in real-time allowing an easy and direct integration in general robotics systems, e.g. in grasping and manipulation pipelines.

The remainder of this paper is structured as follows. In section II we overview previous related work available in the literature. In section III we describe in detail the various steps involved in the proposed cylinder detection methodology. In section IV we quantitatively evaluate the benefits of the proposed contribution. Finally, in section V we draw some conclusions and future work ideas.

II. RELATED WORK

Object recognition and pose estimation with 3D depth data is an important subject in computer vision with many applications in robotics. There are two main approaches to the problem that depend on the availability of 3D object models: 3D model based and learning based. If one has a description of the 3D shape of the object, either given by a parametric surface representation or by a CAD mesh representation, the 3D model-based methods are typically used. If such representations are not available, the dominant approaches rely in machine learning techniques that “learn a model” given a set samples of the object images acquired by the robot sensors. In this paper we focus on cylinder parametric representations, thus we will use 3D model-based techniques. One of the most successful approaches for model-based 3D object recognition using point clouds are based on [5], [6] and create a global descriptor for a given object shape model, using point pair features. The CAD model of the object is used to create a large database of features. At run-time, the matching process is done locally using an efficient voting scheme similar to the Generalized Hough Transform [7]. Each point pair detected in the environment casts a vote for a certain object and 3D pose.

However in unstructured environments, existing CAD based methods tend to suffer from clutter and occlusion. In semi-structured environments (e.g. industrial pipelines), strategies based on the detection and estimation of parametric geometric primitives are generally more robust and flexible. For the extraction of simple geometric shape primitives like planes, cylinders, cones and spheres, the two most common paradigms are the Hough transform [7] and Random Sample Consensus (RANSAC) [8], which are robust to outliers and noisy data.

RANSAC-based approaches are typically preferred over the former since they are more general and do not require the definition of complex transformations from 3D input to parametric spaces. In the RANSAC paradigm, the data is used directly to compute best-fit models. Despite their proven applicability for the extraction of geometric primitives in noisy 3D data [9] [10], in particular in tabletop object segmentation, RANSAC-based approaches have high memory requirements. Being a non-deterministic iterative algorithm, computational time is greatly dependent on the allowed iterations to produce reasonable results, hence becoming impractical for scenarios with large levels of clutter [11]. In other words, the large number of random selections in large-scale point clouds may compromise the method applicability in applications with real-time constraints. Furthermore, their lack of flexibility hinders the incorporation of model-specific heuristic knowledge, that enables the creation of more effective and efficient specialized methodologies.

The problem of detecting and estimating the pose of cylinder structures using 3D range data and Hough transform is naturally formulated on 5-dimensional parametric spaces, but this results in prohibitive computational complexity due to the curse of dimensionality (the size of the Hough accumulator is exponential in the number of dimensions). A more efficient

approach [3] uses a 2D Hough transform to estimate orientation followed by a 3D Hough transform to simultaneously detect radius and position. Though reducing the exponential complexity factor, this approach still lacks speed and robustness in dense point cloud data. In [12] the authors proposed a coarse-to-fine voting procedure that speeds-up the former method by several orders of magnitude. Another interesting idea is the incorporation of environment structural constraints (e.g. cylinders are standing vertically or horizontally on the floor) to reduce the search space [11] to a small subset of possible orientations.

Despite the improvements on computational complexity of the previous approaches, the lack of robustness to clutter of the aforementioned approaches still sets the main setback to their usage in real applications. To address this problem, the focus and the main contributions of our work are: a novel randomized sampling scheme for the creation of orientation Hough accumulators which allows the incorporation of environment structural priors to improve orientation estimation accuracy with the same resources; voting scheme that significantly improves the robustness of Hough methods in cylinder detection and pose estimation.

III. METHODOLOGY

Our approach is based on the former work of Rabbani et al. [3] that splits the cylinder detection and pose estimation problem in two independent Hough transform stages. In the first stage, 3D point normals cast votes for possible cylinder orientations, in a 2D orientation accumulator. In the second stage, the point cloud is rotated according to the determined orientation and each point votes for a position and radius of the cylinder in a 3D Hough accumulator. In that work the unit sphere of orientations is uniformly and deterministically sampled at a predefined number of points [13], to generate a discrete Hough accumulator space, in which voting is subsequently performed. A larger number of cells on the unit sphere improves the accuracy of the orientation estimate, at the cost of increased computational effort. In the present work, we propose several improvements to the orientation voting stage of [3].

In this section we describe in detail our methodology for improved orientation estimation during cylinder detection. First, we introduce a novel randomized sampling scheme which enables the creation of non-uniform, problem-specific orientation Hough accumulators. Then we present a novel and more efficient Hough voting scheme that relies on simple inner products. As opposed to [3] we avoid the computational burden of explicitly voting in spherical coordinates, which requires the computation of rotation matrices and, consequently, of inefficient trigonometric functions. Furthermore, our voting scheme is richer than the one of [3] since it allows incorporating curvature information. When compared with the work of [3], the proposed methodology is able to cope with higher levels of clutter, including flat surfaces such as ground planes, hence avoiding the need of prior plane detection and removal.

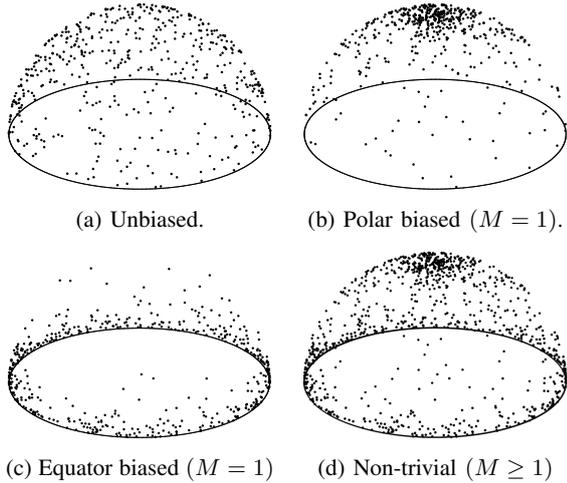


Fig. 2: Different sampled unitary spheres, where each point defines an orientation cell in the Hough accumulator.

1) *Randomized Orientation Hough Accumulator*: The proposed orientation Hough accumulator space is composed of a set of cells \mathcal{D} lying on a unit sphere. The center of each cell corresponds to a unique absolute orientation. The accumulator is analogous to a Voronoi diagram defined on a spherical 2-manifold \mathbb{S}^2 in 3D space, as depicted in Fig. 2, and is represented by set of N_d 3D Cartesian sample points with unit norm, centered in the reference frame origin (center of the sphere)

$$\mathcal{D} = \{\mathbf{d}^i \in \mathbb{R}^3, i, \dots, N_d : \|\mathbf{d}^i\| = 1\} \quad (1)$$

which are i.i.d. and randomly generated from a three dimensional Gaussian Mixture Model (GMM) distribution

$$\mathbf{d}^i = \frac{\mathbf{v}^i}{\|\mathbf{v}^i\|} \text{ where } \mathbf{v}^i \sim p(\boldsymbol{\theta}) = \sum_{m=1}^M \phi^m \mathcal{N}(\boldsymbol{\mu}_d^m, \boldsymbol{\Sigma}_d^m) \quad (2)$$

where M is the number of mixture components and where each $\mathbf{d}^i \in \mathcal{D}$ represents an orientation, allowing for efficient voting with observed surface normals, using inner products (equation 3).

The statistics of the GMM components are chosen according to task at hand (e.g. find vertically aligned cylinders) or prior knowledge on how likely specific orientations are (e.g. cylinders are unlikely to be in diagonal orientations). On one hand, in order to produce uniform and unbiased accumulator structures, the surface should be sampled from a rotationally symmetric distribution, i.e., from a single Gaussian with zero mean and variance equal in all dimensions [14] (Fig. 2a). On the other hand, non-uniform, task-dependent sampling biasing can be achieved by manipulating the Gaussian Mixture Model parameters (Fig. 2b and 2c).

Hypothetical accumulator spaces that may be suitable for different priors are depicted in Fig. 2. In the absence of prior information or task definition, one should sample from a single component Gaussian, with zero mean and standard deviation

equal in all dimensions (Fig. 2a). If for instance the task is to find cylinders that are vertically aligned with the reference frame, one should privilege orientations at the pole (Fig. 2b) rather than the equator (Fig. 2c). In the latter case, varying the Gaussian mean is not sufficient. One could sample from a single-component zero mean GMM with larger variance in the horizontal directions. Finally, more complex detection tasks or prior knowledge can benefit from GMMs with many component (Fig. 2d).

Our randomized sampling scheme offers several advantages over the one of [3], namely:

- it is easier to implement than its deterministic counterpart [13] and allows for the fast creation of biased orientation voting spaces.
- the non-deterministic nature of the representation offers a convenient mechanism for encoding task-related biases or probabilistic prior knowledge about possible orientations, depending on the environment (e.g. cups are typically oriented vertically on tables). Biasing the orientation Hough accumulator space leads to more efficient, flexible and adaptable resource allocation and to more accurate orientation estimation, for the same memory resources.

2) *Fast Robust Orientation Voting Scheme*: At run-time, the input of our algorithm is a scene input point cloud which comprises a finite set of 3D Cartesian points $\mathcal{P} \subset \mathbb{R}^3$, where $\mathcal{P} = \{\mathbf{p}^s, s = 1, \dots, N_s\}$.

First, we estimate the surface normals at each scene point $\mathbf{p}^s \in \mathcal{P}$ using the Principal Component Analysis (PCA) [15] of the covariance matrix created from its k -nearest neighbors. Let $\mathcal{N} = \{\mathbf{n}^s, s = 1, \dots, N_s\}$ denote the set of surface normals. Then, we proceed with the computation of the principal curvatures as follows. For each scene point \mathbf{p}^s , we compute a projection matrix for the tangent plane given by the associated normal \mathbf{n}^s . After, we project all normals from the k -neighborhood onto the tangent plane. Finally, we compute the centroid and covariance matrix in the projected space. We finally employ eigenvalue decomposition of this covariance matrix to obtain the principal curvature direction $\mathbf{c}_{\max}^s \in \mathbb{R}^3$ and the corresponding eigenvalue $k_{\max} \in \mathbb{R}$ (see Fig. 3). Let $\mathcal{C} = \{\mathbf{c}_{\max}^s, s = 1, \dots, N_s\}$ denote the set of principal curvature directions and $\mathcal{K} = \{k_{\max}^s, s = 1, \dots, N_s\}$ the set of the corresponding eigenvalues.

The orientation voting procedure goes as follows: For each direction cell \mathbf{d}^s in the orientation Hough accumulator A , we compute the inner product with all the scene surface normals $\mathbf{n}^s \in \mathcal{N}$ and their associated principal curvature directions $\mathbf{c}_{\max}^s \in \mathcal{C}$ to cast continuous votes in the accumulator according to the function

$$A(d) = \sum_{s=1}^{N_s} k_{\max}^s |(1 - \mathbf{d}^s \mathbf{c}_{\max}^s)| |(1 - \mathbf{d}^s \mathbf{n}^s)| \quad (3)$$

This voting function gives more weight to directions that are simultaneously, orthogonal to the the normal and the principal curvature directions. Moreover, the eigenvalue k_{\max}^s functions as a curvature high-pass filter, that suppresses low curvature

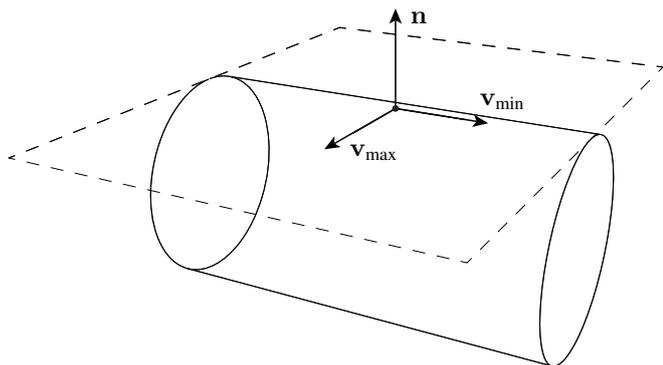


Fig. 3: Normal and principal curvatures' directions for a cylinder surface point.

candidates, since points belonging to flat surfaces have very low k_{max}^s .

After determining the cylinder orientation we proceed with the estimation of the cylinder position and radius, as detailed in [3].

IV. RESULTS

Several experiments were conducted in order to quantitatively evaluate the quality of the cylinder parameters recovered by the method in Rabbani et al. [3] and by our proposed method, when dealing with increasing levels of clutter and noise. In all experiments, we generated 200 synthetic scenes, each containing a single instance of a cylinder. By using synthetically generated scenes, we were able to compare the algorithm pose results with a known ground truth. The selected parameters for both methodologies were the following: The radius was fixed to $r = 0.3m$ and the height was uniformly sampled from the interval $[0.05, 2.0] m$. The number of cylinder surface points was fixed and set to 900 and the number of orientation sample points in the Hough accumulator space was set to $N_d = 450$.

A. Robustness to clutter

In order to assess the performance gains of the proposed strategies in the presence of flat surfaces (further designated as clutter)

$$\text{clutter} = 1 - \frac{\text{cylinder surface points}}{\text{total scene points}} \quad (4)$$

we added synthetically generated planar extremities to cylinders, that simulate realistic cylindrical shapes such as containers with lids and cans. Surface points on cylinder tops are problematic for orientation estimation since they vote for orthogonal directions, and in this experiment were considered as planar clutter. The surfaces were generated with increasing point density levels, to each previously generated cylinders' bottom and top extremities (see Fig. 4). Moreover, each cylinder was set at a random pose. The quantitative results illustrated in Fig. 7 demonstrate the advantage of considering both the surface curvature and the surface normal in the

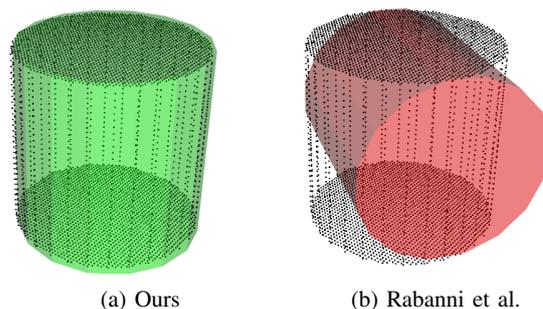


Fig. 4: Our method against Rabbani et al. when dealing with flat surfaces.

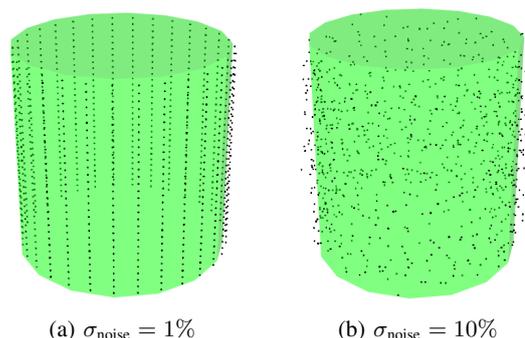


Fig. 5: Estimated cylinder parameters with our method, from a point cloud corrupted with different levels of additive Gaussian noise.

orientation voting step. When dealing with flat surfaces that belong to real-life cylinders, our method estimates better the cylinder orientation, as show by the absolute orientation errors in Fig. 7b . As expected, these improvements have a direct and positive impact in the quality of the position and radius estimations, as show by the absolute radius and position errors plots in Fig. 7c. As opposed to [3], our method is able to cope with large amounts of clutter while keeping the performance at the levels of uncluttered scenes.

B. Robustness to noise

In pursuance of quantifying the behavior of the Rabbani et al. algorithm [3] and our proposed extensions in the presence of noisy visual sensors, each of the 200 generated scenes was corrupted by different levels of additive Gaussian noise, with standard deviation proportional to the cylinder radius (see Fig. 5). This time, in order to validate the advantages of our randomized sampling scheme for the creation of the orientation Hough accumulator, the orientation of the cylinders was fixed and aligned with the frame of reference z -axis. We considered and compared two different sampling strategies (see Table I):

- an unbiased distribution reflecting the absence of prior knowledge about the cylinder orientation.
- a biased distribution that favours vertical orientations.

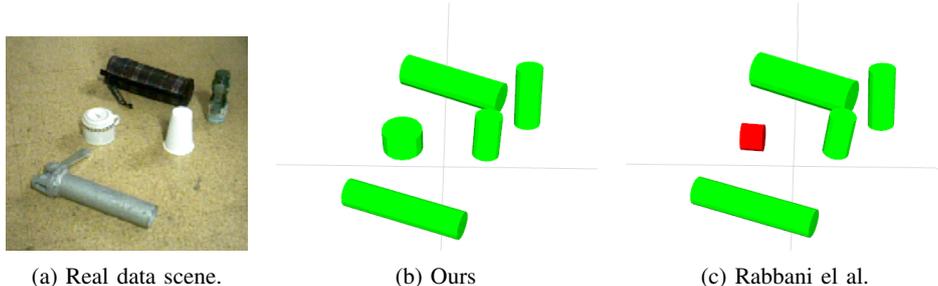


Fig. 6: Qualitative estimations from data acquired with an Asus Xtion 3D camera. Good and bad detections in green and red, respectively.

Bias	μ_p			Σ_p		
	x	y	z	xx	yy	zz
unbiased	0	0	0	0.5	0.5	0.5
biased	0	0	1.0	0.05	0.05	0.5

TABLE I: Orientation Hough accumulator biasing parameters.

Figure 8 depicts the cylinder parameters estimation errors for both methodologies in the presence of noise. The results show that both methodologies have similar robustness to noise, hence, demonstrating the benefit of our approach when considering the superior performance of our method in cluttered scenes. Additionally, biasing the orientation accumulator in the face of prior structural knowledge improves significantly estimation accuracy. Overall, our extensions results in dramatic improvements on robustness to clutter, without sacrificing robustness to noise. Furthermore, a simple qualitative assessment of our method with data acquired from a RGB-D camera demonstrates its applicability to real-scenarios, as exhibited in Fig. 1 and Fig. 6, and its superior robustness to planar clutter.

V. CONCLUSIONS

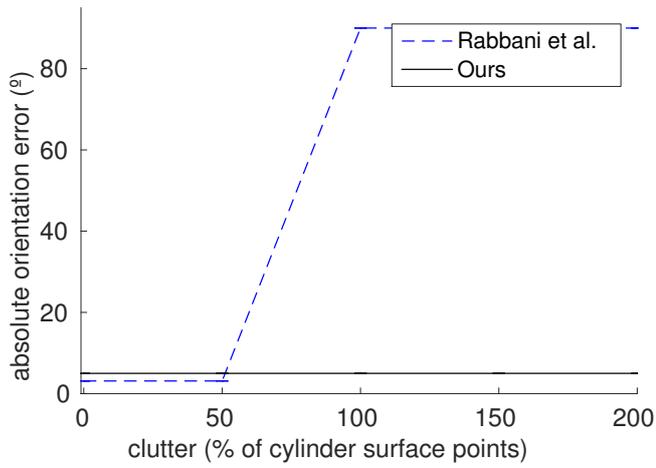
In this paper, we have proposed a robust soft-voting scheme based on the Generalized Hough Transform for the detection and pose estimation of arbitrary cylindrical structures from 3D point clouds. The proposed method incorporates curvature information in the voting scheme that improves the rejection of outliers, mainly those arising from planar surfaces that pollute the orientation voting space and introduce erroneous biases in cylinder orientation estimation. The results show significant improvements on the detection rates and pose estimates with respect to previous schemes. A systematic quantitative analysis of robustness to clutter and noise validates our approach and sets a benchmark for future research. For future work, we note that robustness to noise could be further enhanced by sequentially integrating cylinder detections through temporal Bayesian filtering [16].

ACKNOWLEDGEMENTS

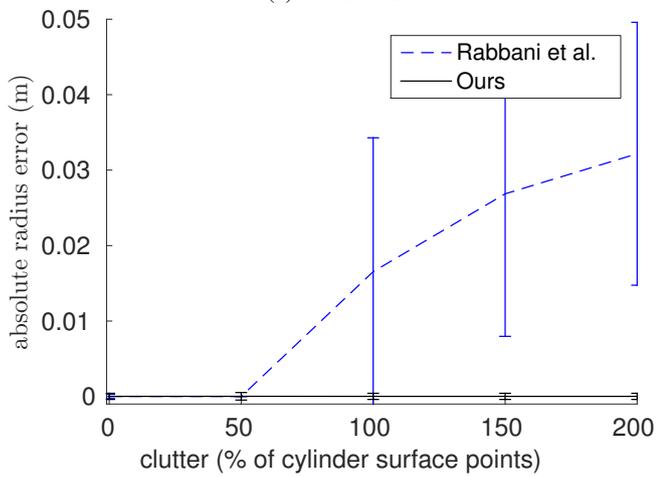
This work has been partially supported by the Portuguese Foundation for Science and Technology (FCT) project [UID/EEA/50009/2013]. Rui Figueiredo is funded by FCT PhD grant PD/BD/105779/2014.

REFERENCES

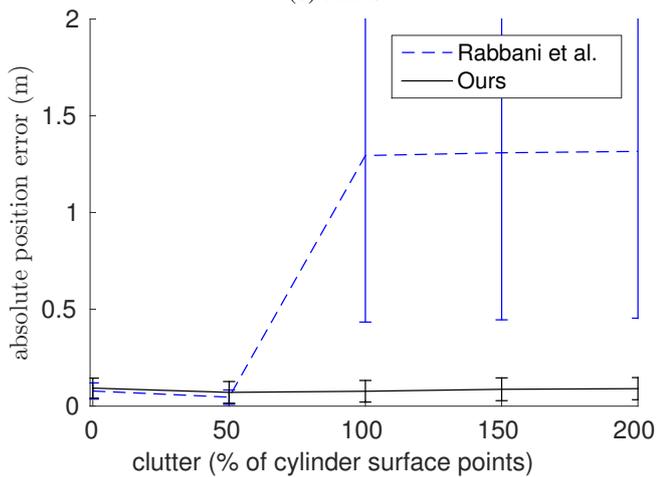
- [1] P. Moreno, R. Nunes, R. Figueiredo, R. Ferreira, A. Bernardino, J. Santos-Victor, R. Beira, L. Vargas, D. Aragão, and M. Aragão, "Vizzy: A humanoid on wheels for assistive robotics," in *Robot 2015: Second Iberian Robotics Conference*. Springer International Publishing, 2016, pp. 17–28.
- [2] R. Figueiredo, A. Shukla, D. Arago, P. Moreno, A. Bernardino, J. Santos-Victor, and A. Billard, "Reaching and grasping kitchenware objects," in *IEEE/SICE International Symposium on System Integration (SII)*, 2012, pp. 865–870.
- [3] T. Rabbani and F. Van Den Heuvel, "Efficient hough transform for automatic detection of cylinders in point clouds," *ISPRS WG III/3, III/4*, vol. 3, pp. 60–65, 2005.
- [4] M. Quigley, J. Faust, T. Foote, and J. Leibs, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, no. 3.2., 2009.
- [5] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3d object recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Ieee, 2010.
- [6] R. P. de Figueiredo, P. Moreno, and A. Bernardino, "Efficient pose estimation of rotationally symmetric objects," *Neurocomputing*, vol. 150, pp. 126–135, 2015.
- [7] P. Hough, "Method and Means for Recognizing Complex Patterns," U.S. Patent 3.069.654, Dec. 1962.
- [8] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
- [9] R. Schnabel, R. Wahl, and R. Klein, "Efficient ransac for point-cloud shape detection," in *Computer graphics forum*, vol. 26, no. 2. Wiley Online Library, 2007, pp. 214–226.
- [10] L. C. Goron, Z.-C. Marton, G. Lazea, and M. Beetz, "Robustly segmenting cylindrical and box-like objects in cluttered scenes using depth cameras," in *Robotics: Proceedings of ROBOTIK 2012; 7th German Conference on*. VDE, 2012, pp. 1–6.
- [11] Y.-J. Liu, J.-B. Zhang, J.-C. Hou, J.-C. Ren, and W.-Q. Tang, "Cylinder detection in large-scale point cloud of pipeline plant," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 10, pp. 1700–1707, 2013.
- [12] Y.-T. Su and J. Bethel, "Detection and robust estimation of cylinder features in point clouds," in *ASPRS Conference*, 2010.
- [13] E. Lutton, H. Maitre, and J. Lopez-Krahe, "Contribution to the determination of vanishing points using hough transform," *IEEE transactions on pattern analysis and machine intelligence*, vol. 16, no. 4, pp. 430–438, 1994.
- [14] M. E. Muller, "A note on a method for generating points uniformly on n-dimensional spheres," *Commun. ACM*, vol. 2, no. 4, pp. 19–20, Apr. 1959.
- [15] K. P. F.R.S., "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine Series 6*, vol. 2, no. 11, pp. 559–572, 1901.
- [16] R. P. de Figueiredo, P. Moreno, A. Bernardino, and J. Santos-Victor, "Multi-object detection and pose estimation in 3d point clouds: A fast grid-based bayesian filter," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2013, pp. 4250–4255.



(a) orientation

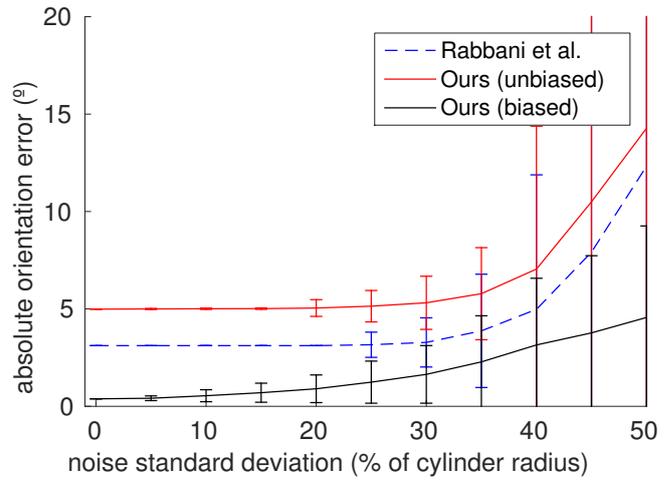


(b) radius

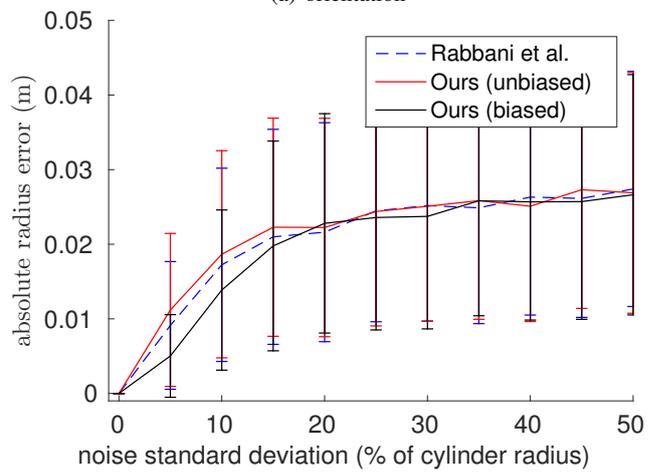


(c) position

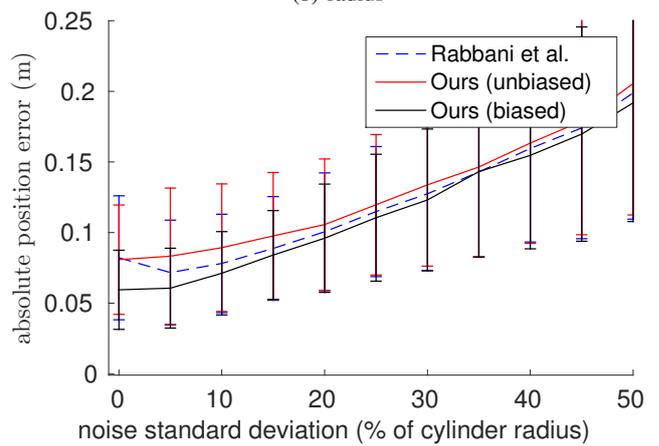
Fig. 7: Robustness of our method against the method of Rabbani et al. et al. to different levels of flat surface clutter.



(a) orientation



(b) radius



(c) position

Fig. 8: Robustness of our method against the method of Rabbani et al. to different levels of noise.