Robotics and Autonomous Systems 117 (2019) 17–28

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

A robust and efficient framework for fast cylinder detection

Rui Figueiredo [a,*], Atabak Dehban [a,b], Plinio Moreno [a], Alexandre Bernardino [a], José Santos-Victor [a], Helder Araújo [c]

[a] Institute for Systems and Robotics, Instituto Superior Tecnico, Universidade de Lisboa, Lisbon, Portugal
[b] Champalimaud Centre for the Unknown, Lisbon, Portugal
[c] Institute for Systems and Robotics, Universidade de Coimbra, Coimbra, Portugal

ARTICLE INFO

Article history:
Available online 10 April 2019

ABSTRACT

In this work, a complete solution is provided for detecting and identifying cylindrical shapes, which are commonly found in household and industrial environments, using consumer-grade RGB-D cameras. Most standard approaches to detect and identify cylinders are not robust to outliers (e.g. points on other objects in the scene), which limits their applicability in realistic scenes. In addition, these methods fail to benefit from environmental constraints, e.g. the fact that cylinders often lie or stand on flat surfaces. To tackle the aforementioned limitations, we introduce three main novelties: (i) a point cloud soft voting scheme with curvature information that reduces the influence of outliers and noise, (ii) a selective sampling of the orientation space that favors orientations known *a priori*, and (iii) a deep-learning based classifier to filter out objects with non-cylindrical appearance in the 2D images, thus further improving robustness to outliers.

A set of experiments with synthetically generated data are used to assess the robustness of our fitting method to different levels of outliers and noise. The results demonstrate that incorporating the principal curvature direction within the orientation voting process allows for large improvements on cylinders parameters estimation. Furthermore, we demonstrate that combining the 2D deep-learning cylinder classifier with the 3D orientation voting scheme allows for large speed-up and accuracy improvements on cylinder identification. The qualitative and quantitative results with real data acquired from a consumer RGB-D camera, confirm the advantages of the proposed framework.

1. 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. More specifically, in the tasks that involve interaction with the surrounding environment, e.g. manipulation, an artificial agent would require to accurately recognize objects and estimate their pose. These tasks include successful manipulation and grasping, obstacle avoidance and self localization with respect to known landmarks, to name a few.

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 outliers.

A key aspect behind the success of a grasping solution resides in the choice of the object representation, which can deal with incomplete and noisy perceptual data and is flexible enough to cope with inter and intra-class variability, allowing the generalization to never-seen objects. Furthermore, in order to cope with transmission bandwidth and computational processing capacity limitations, efficient and fast perception is an essential requirement for real-time performance.

In this work, we propose a novel computationally efficient attentional framework for the task of simultaneously detecting, recognizing and identifying particular object shapes. We focus on cylindrical shaped objects which are commonly found in domestic (e.g. cups, bottles) and industrial environments (e.g. pipes, pillars, scaffolds), and identifying them plays an important role in many robotic grasping applications [2,3].

The proposed framework relies on the tabletop assumption, i.e., objects are placed on flat surfaces, which is another widely adopted scenario in robotics [4,5] (Fig. 1). In order to deal with

* Corresponding author.
E-mail addresses: ruifigueiredo@isr.tecnico.ulisboa.pt (R. Figueiredo), adehban@isr.tecnico.ulisboa.pt (A. Dehban), plinio@isr.tecnico.ulisboa.pt (P. Moreno), alex@isr.tecnico.ulisboa.pt (A. Bernardino), jasv@isr.tecnico.ulisboa.pt (J. Santos-Victor), helder@isr.uc.pt (H. Araújo).

https://doi.org/10.1016/j.robot.2019.04.002
0921-8890/

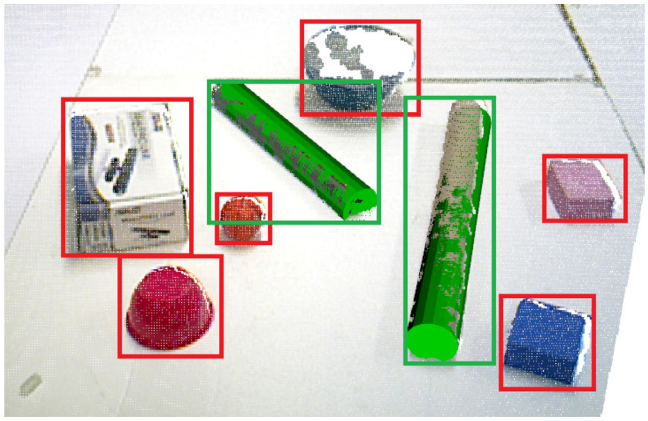


Fig. 1. A snapshot of a RGB-D point cloud and overlaid cylindrical (green) and non-cylindrical (red) shapes detected with our methodology. Figure best seen in color . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cluttered environments which are often populated with multiple non-cylindrical shapes i.e. distractors, we take advantage of the recent advances in deep learning architectures to introduce an efficient recognition module that learns to filter out irrelevant object candidates. More specifically, we incorporate a pre-attentive shape-based selection mechanism, that avoids the need of time-consuming, top-down cylinder parameter identification at an early stage, on irrelevant salient candidate objects. Furthermore, the most successful cylinder fitting approaches in the 3D shape fitting literature are based on a computationally efficient 2-step Generalized Hough Transform (GHT) [6]. We extend this method with a set of improvements that allow coping with large levels of outliers, mainly residing on bases of cylinders, which often introduce problematic biases during the orientation estimation. The cylinder fitting approach described in this paper was originally proposed in [7], but the reviewed literature and experimental evaluation here is significantly expanded.

Our main contribution is threefold: first, and unlike previous approaches that are only based on 3D depth information, we combine a state-of-the-art [6,7] cylinder fitting approach which is based on a robust and computationally efficient 2-step GHT with a 2D image-based top-down Deep Convolutional Neural Network proposal rejection mechanism to increase the quality and speed of estimations. Since gathering a large dataset, required for deep learning based recognition techniques is laborious and time consuming, we provide a semi-automatic data gathering procedure, using 3D information, which greatly facilitates acquiring and labeling relatively large amounts of data. Second, we propose a novel randomized sampling scheme for the creation of orientation Hough accumulators. Our sampling method allows the incorporation of prior structure knowledge which improves accuracy with the same computational resources. And finally, as our third contribution, we introduce a novel soft-voting scheme, which considers surface curvature information, in order to cope with points that exist on flat surfaces which vote for erroneous and arbitrary tangential orientations.

We perform a systematic and thorough quantitative assessment of the influence of noise and outliers on detection and pose estimation error of cylinder fitting methods, comparing our proposed method with that of [6]. Our ROS [8] and Caffe [9] C++ implementation can identify multiple cylinders under a second, allowing an easy and straightforward integration in general robotics systems, e.g. in grasping and manipulation pipelines. The code[1] and datasets[2] of our experiments are publicly available online.

The remainder of this paper is structured as follows. In Section 2 we overview previous related work available in the literature. In Section 3 we describe in detail the various steps involved in the proposed cylinder detection and identification methodology, as well as the datasets used for training and evaluating the pipeline. In Section 4 we quantitatively evaluate the benefits of the proposed contributions. Finally, in Section 5 we draw our conclusions and propose promising future work ideas.

2. Related work

As described in the previous section, successful identification of objects in an environment requires not only the development of robust and efficient object detection architectures, but also the definition of flexible shape representations that should facilitate generalization to never-seen-objects, via the integration of different visual sensing modalities. Therefore, we organize the present section in two distinct parts. First, an overview of the state-of-the-art methods in visual attention, with an emphasis on shape-based models of selective attention is presented. Afterwards, we analyze various object identification paradigms proposed in the literature, suitable for applications that require identification and localization of parametric shapes.

2.1. Shape-based selective attention

Visual attention plays a central role in biological and artificial systems to control perceptual resources [10,11]. The classic artificial visual attention systems use salient features of the image, benefiting from the information provided via hand-crafted filters. Recently, deep neural networks have been developed for recognizing thousands of objects and autonomously generate visual characteristics that are optimized by training with large datasets. Besides their application in object recognition, these features have been very successful in other visual problems such as object segmentation [12], tracking [13] and visual attention [14].

Evidence from neurophysiology studies [15] suggests that people consider shape as an important feature dimension among other low-level visual features (e.g. texture and color). In [16] the authors found that subjects looking for a particular shape (e.g. flowers or pillows) are more accurate in reporting other features of that object (e.g. color) meaning that people have attention mechanisms for shape features. Furthermore, infants rely more on shape than on color when learning new objects, which in turn allows them to generalize to other objects with similar visual features while interacting with them [17]. This fact motivates the need of developing more sophisticated, shape-biased and bottom-up attentional architectures [18].

2.2. Object identification in robotics

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 this 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 often used for simultaneous object recognition and 3D pose estimation [19]. If such representations are not available, the dominant approaches rely on machine learning techniques that "learn a

[1] [code] https://github.com/ruipimentelfigueiredo/shape_detection_fitting.
[2] [dataset] http://soma.isr.tecnico.ulisboa.pt/vislab_data/facyl/facyl.zip.

model'' given a set of image samples of the object, acquired by the robot sensors [20]. Despite being flexible and capable of generalizing to novel objects in detection and classification tasks, these methods are often unsuitable for estimating some shape properties, such as 3D pose or size of the object. In this work we leverage the accuracy and generalization capabilities of state-of-the-art deep learning techniques in recognition tasks, with robust 3D model-based fitting approaches to develop a multi-modal, fast, and robust cylinder identification pipeline.

One of the most successful approaches for model-based 3D object recognition using point clouds are based on [21,22] where a global descriptor for a given object shape model is created, 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 and robust voting scheme similar to the Generalized Hough Transform [23]. 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 outliers and occlusion. In semi-structured environments (e.g. industrial pipelines), strategies based on the detection and estimation of parametric shapes are generally more robust and flexible [24–26]. For the extraction of simple geometric shape primitives like planes, cylinders, cones and spheres, the two most common paradigms are the Hough transform [23] and Random Sample Consensus (RANSAC) [27], 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 [28,29], in particular in tabletop object segmentation, RANSAC-based techniques 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 outliers [30]. 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 (2 orientations, 2 locations plus the radius), 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). The most efficient parametric shape fitting methods are based on Hough transforms that estimate cylinder parameters, i.e. orientation, position and radius, in two sequential voting steps [6,7]. More specifically, they rely on a 2D Hough transform to estimate orientation, i.e. the direction of the cylinder axis, followed by a 3D Hough transform to simultaneously detect radius and position. Though reducing the exponential complexity factor, this approach still lacks speed in dense point cloud data. In [31] and [32] 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 [30] to a small subset of possible orientations.

Despite the improvements on computational complexity of the previous approaches, their lack of robustness to outliers still sets the main drawback to their usage in real applications. Palánz et al. [33] introduces a method that finds the cylinder that fits better in a point cloud, modeled as a mixture of two Gaussians. One Gaussian models the data samples belonging to the cylinder and the other Gaussian models the outliers. The random variable of the model is the fitting error, which is lower for the inliers and larger for the cylinder outliers. The error considered in their work is the sum of the perpendicular distance from the point to the estimated cylinder, and its parameters are estimated using the Expectation Maximization algorithm for the mixture of Gaussians. Although they show a large robustness to outliers, the method is computationally demanding and not parallelizable. Tran et al. [24] propose an algorithmic approach that starts from individual cylinder detection, followed by a mean shift clustering in the cylinder space parameters. The individual cylinder detection algorithm finds promising cylinder hypotheses based on weighted point cloud normal estimation and an inlier point selection. The normals are utilized to find the cylinder axis orientation by selecting the eigenvector corresponding to the smallest eigenvalue of the covariance matrix C of normal vectors of inliers. The inliers are selected by projecting the cylinder points to a plane normal to the cylinder axis orientation and fitting the projected points to a circle. This approach is robust to outliers and finds multiple cylinders, but is computationally more expensive than [6], which is the baseline of our approach. Nurunnabi et al. [25] propose an algorithmic approach that relies on Robust Principal Component analysis (RPCA) to find the cylinder orientation and Robust Least Trimmed Squares (RTLS) regression to remove outliers from the RPCA cylinder parameter estimation. The RTLS removes outliers that do not fit the projected circle from the cylinder points. This approach is limited to find just one cylinder in the point cloud.

In this paper we propose a novel fitting approach that leverages an efficient implementation of the Hough-based method of [6] with the increased robustness of using statistical models to encode domain-specific knowledge. More specifically, 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 computational resources; a voting scheme that significantly improves the robustness of Hough methods in cylinder detection and pose estimation.

Still, all the aforementioned fitting approaches are incapable of filtering, at an early stage, different object shapes that act as irrelevant visual distractors. The time consuming process of fitting shapes to distractors, marks another limitation of fitting approaches, which hinders their applicability in real world scenarios. Kostavelis et al. [34] have incorporated Graph-Based Visual Saliency algorithm (GBVS) as a pre-processing step in training a biologically inspired Hierarchical Temporal Memory (HTM) network. According to these results, the introduction of a bottom-up attention mechanism significantly improves the efficiency and performance of down-stream tasks, however, it is not clear how much their approach can generalize to the detection of occluded objects. Similarly, we incorporate a mediating shape-based pre-attention bottom-up mechanism to reduce the space of possible cylindrical shapes to a small subset of prominent objects in the field of view, in a bottom-up manner. The 2D image patches, coming from 3D segmentation are first classified using a Deep Convolutional Neural Network (DCNN), which is robust to occlusion. Object classes of interest (i.e. cylinder), are further considered for parameter identification, which results in faster and more accurate estimates.

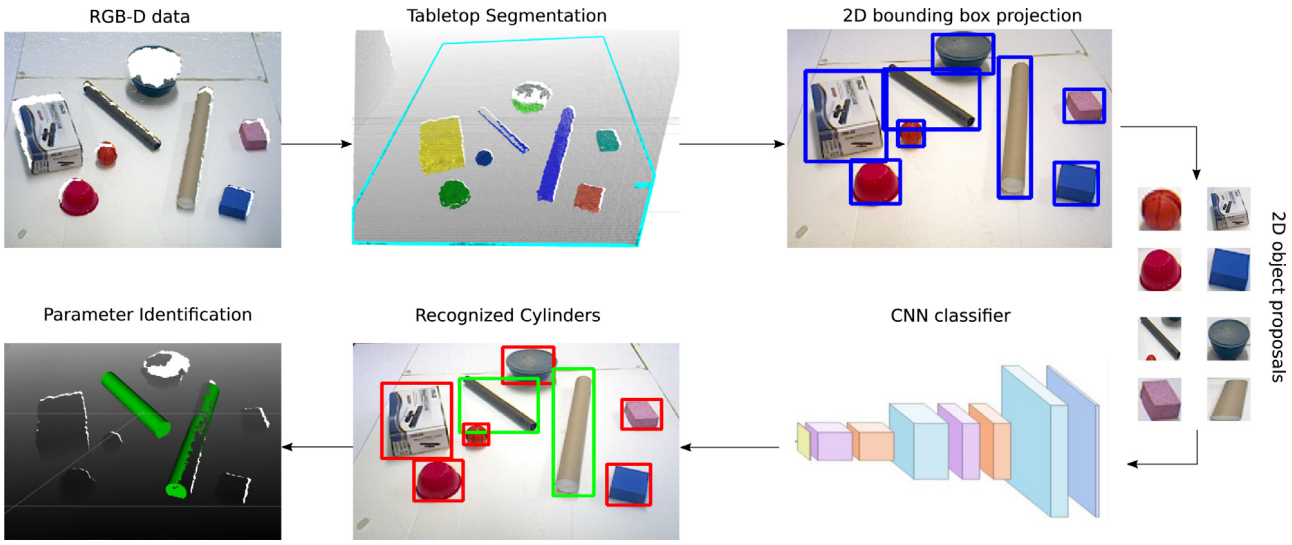


Fig. 2. General overview of our shape-based attention framework . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Methodology

In this section we describe our framework for efficient detection and identification of cylindrical shapes using multiple visual sensing modalities: color and depth. The proposed architecture, depicted in Fig. 2, is an integration of different cognitive blocks which are responsible for object segmentation and shape recognition, fitting and localization. In the remainder of this section we describe in detail the multiple components of our pipeline.

3.1. System overview

We start by detecting tabletop objects using 3D point cloud information, since points above tables are considered to belong to potentially graspable objects. Therefore, the first component of our cylinder detection and identification pipeline is a bottom-up segmentation module that is triggered by salient objects laying on flat surfaces [35]. First, we use a RANSAC-based fitting approach, which efficiently operates on organized point cloud data [36], in order to detect planes on the scene and segment objects above these planes. We rely on Euclidean clustering [36] to identify individual objects. Afterwards, these objects are projected on the 2D camera plane to extract bounding boxed 2D focused images from a stream of monocular images, which are used to recognize cylindrical shapes via a deep artificial neuronal network classifier. The proposed Convolutional Neural Network (CNN) is trained offline via transfer learning, and acts as a shape-based mediating pre-attentive selective mechanism that filters out non-cylindrical shapes. Finally, the parameters of the identified cylindrical shapes are estimated in 3D Cartesian space, using an efficient and robust top-down depth-based Hough transform.

3.2. Transfer learning for early shape-based attention

In order to reject region proposals and avoid parametric identification of non-cylindrical objects, we propose to use deep neural networks. Inspired by recent advances of deep learning in achieving state of the art performance in recognition tasks, we use a deep CNN as a binary classifier to decide if a particular object is a cylinder or not.

However, using a deep neural network for the task at hand can pose several challenges. Firstly, most deep neural network architectures are notoriously data-hungry, usually trained on millions of labeled images. Secondly, designing a neural network architecture for a new task is time consuming and involves a large amount of trial and errors. And last, storing and using them on most embedded systems is impractical due to the substantial size and the computations they require.

3.2.1. Data acquisition and training

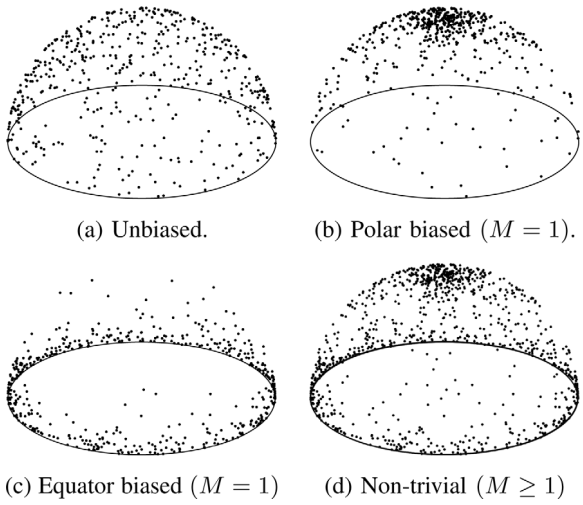
To solve the first problem, we propose a fast and convenient procedure for semi-automatic gathering of labeled data, which does away with the need of manual labeling. The procedure relies on the 3D tabletop segmentation method and the 3D bounding box projection to 2D approach described in the previous subsection. For the creation of positive samples, we first place many different cylindrical shaped objects on tabletops and acquire data, from multiple views, using an hand-held RGB-D camera. Then for the creation of the negative examples dataset, we repeat the same procedure with all the non-cylindrical objects, commonly found in the testing environment.

3.2.2. Cylindrical-shapes recognition

For the second problem, i.e. architecture design, we propose to use transfer learning [37]. More specifically, we have used a network previously trained on imagenet dataset [38] and fine-tuned it as a cylinder classifier. This way, the architecture of the network is pre-defined and it is only necessary to change the last layer such that instead of predicting probability classes of 1000 objects, it only outputs the probability that an input image is a cylinder or not. Moreover, it is generally assumed that if a network performs well on a recognition task, it means it has learned *informative* features which are useful for different tasks. As a result, it is possible to train the network on significantly smaller datasets and only slightly change the previously learned features.

3.2.3. Performance speed-ups

In order to have a small network which performs reasonably fast even in the absence of powerful GPUs, we used a neural network called SqueezeNet [39]. This network achieves AlexNet accuracy score on imagenet while being 50 times smaller. Taking advantage of this reduction in parameters of the network, it is possible to have a fast and reliable classifier which is more suited towards real-time applications.



(a) Unbiased. (b) Polar biased $(M = 1)$. (c) Equator biased $(M = 1)$ (d) Non-trivial $(M \geq 1)$

Fig. 3. Different sampled unitary spheres, where each point on the unit sphere represents the center of a candidate Voronoi cell orientation.

3.3. Cylinder parametric fitting

Our approach is based on the former work of Rabbani et al. [6] 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 [40], 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 [6].

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 [6], 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 [6] since it allows incorporating curvature information. When compared with the work of [6], the proposed methodology is able to cope with higher levels of outliers, including flat surfaces such as ground planes, hence avoiding the need of prior plane detection and removal.

3.3.1. 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. 3, and is represented by a set of N_d 3D Cartesian sample points with unit norm, centered in the reference frame origin (center of the sphere)

$$\mathcal{D} = \{\mathbf{d}^i \in \mathbb{R}^3, i, \ldots, N_d : \|\mathbf{d}^i\| = 1\} \tag{1}$$

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

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

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

The parameters 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 relative 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 [41] (Fig. 3a). On the other hand, non-uniform, task-dependent sampling biasing can be achieved by manipulating the GMM parameters (see Fig. 3).

Hypothetical accumulator spaces that may be suitable for different priors are depicted in Fig. 3. 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. 3a). If for instance the task is to find cylinders that are vertically aligned with the reference frame (e.g. table reference plane), one should privilege orientations at the pole (Fig. 3b) rather than the equator (Fig. 3c). 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, prior knowledge or more complex detection tasks (e.g. locating diagonal pipes or machine handles) can benefit from GMMs with many components (Fig. 3d).

Our randomized sampling scheme offers several advantages over the one of [6], namely:

- it is easier to implement than its deterministic counterpart [40] 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 and computational resources.

3.3.2. Fast robust orientation voting scheme

At run-time 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 $P = \{\mathbf{p}^s, s = 1, \ldots, N_s\}$.

First, we estimate the surface normals at each scene point $\mathbf{p}^s \in \mathcal{P}$ using the Principal Component Analysis (PCA) [42] of the covariance matrix created from its k-nearest neighbors. Let $\mathcal{N} = \{\mathbf{n}^s, s = 1, \ldots, N_s\}$ 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 directions $\mathbf{c}^s_{\text{max}} \in \mathbb{R}^3$ and $\mathbf{c}^s_{\text{min}} \in \mathbb{R}^3$ and the corresponding eigenvalues $k_{\text{max}} \in \mathbb{R}$ and $k_{\text{min}} \in \mathbb{R}$ (see Fig. 4). Let $\mathcal{C}_{\text{max}} = \{\mathbf{c}^s_{\text{max}}, s = 1, \ldots, N_s\}$, $\mathcal{C}_{\text{min}} = \{\mathbf{c}^s_{\text{min}}, s = 1, \ldots, N_s\}$

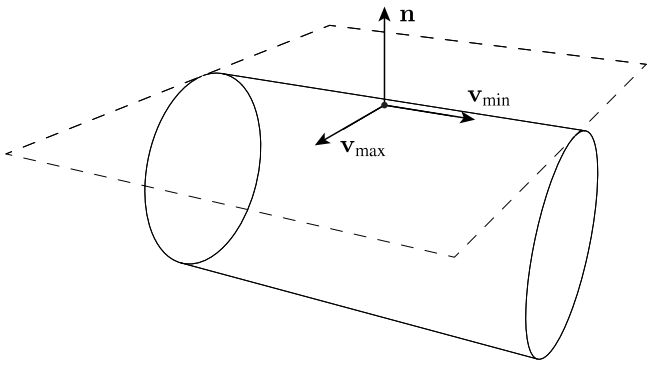


Fig. 4. Normal ($\mathbf{n}$) and principal curvatures' directions ($\mathbf{c}_{\max}$ and $\mathbf{c}_{\min}$) for a cylinder surface point.

denote the sets of principal curvature directions and $\mathcal{K}_{\max} = \{k^s_{\max}, s = 1, \ldots, N_s\}$, $\mathcal{K}_{\min} = \{k^s_{\min}, s = 1, \ldots, N_s\}$ the sets of the corresponding eigenvalues.

The orientation voting procedure goes as follows: For each direction cell $\mathbf{d}^i$ 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 larger principal curvature directions $\mathbf{c}^s_{\max} \in \mathcal{C}$ to cast continuous votes in the accumulator according to the function

$$A(i) = \sum_{s=1}^{N_s} \frac{k^s_{\max} - k^s_{\min}}{k^s_{\max} + k^s_{\min}} \left|\left(1 - \mathbf{d}^i \mathbf{c}^s_{\max}\right)\right| \left|\left(1 - \mathbf{d}^i \mathbf{n}^s\right)\right| \tag{3}$$

This soft voting function gives more weight to directions that are simultaneously orthogonal to the normal and the principal curvature directions. The term $\frac{k^s_{\max} - k^s_{\min}}{k^s_{\max} + k^s_{\min}}$ benefits surface points with large and low curvature along directions $v_{\max}$ and $v_{\min}$, respectively.

After determining the cylinder orientation we proceed with the estimation of the cylinder position and radius, as detailed in [6]. First, we align the estimated cylinder axis with the camera z-axis. Then, we project the inlier points on the camera xy plane and use a Circular Hough Transform (CHT) [43] to estimate the cylinder position and radius.

3.3.3. Goodness-of-fitting criterion

Finally, the goodness of the fitting of a cylinder is evaluated using the following conditional confidence measure:

$$p(\text{cylinder}|\text{object}) = \frac{N_{model}}{N_{\text{cluster}}} \tag{4}$$

where N_{model} represents the number of points that fit the estimated cylinder parametric model (i.e. inliers) and N_{cluster} the total number of 3D points belonging to the object. Estimations below a user-defined quality threshold are discarded and considered as non-cylindrical shapes. We have used this criterion as a *baseline* for cylinder detection.

3.4. Datasets description

In this subsection we introduce the details of the datasets created for assessing the proposed pipeline, as well as their generation and gathering procedures.

3.4.1. Simulation environment

To be able to quantitatively measure the robustness of the proposed cylinder fitting approach, when dealing with variable levels of outliers, noise and occlusion, we created a simulation environment, to synthetically generate cylindrical point clouds with user-specified characteristics, namely:

- cylinder parameters: radius, orientation, position and height;
- outlier levels: the percentage of points belonging to cylinder bases, compared with points belonging to the cylinder surface (Fig. 7);
- noise levels: the standard deviation of additive Gaussian noise (Fig. 8);
- occlusion levels: partial cut length along the axial direction of the complete cylinder surface (Fig. 9)

By using synthetically generated scenes, one is able to assess the robustness of 3D cylinder fitting algorithms, in the face of noise, outliers, and occlusion, with known ground truth.

3.4.2. Real data

In order to assess the proposed CNN classifier impact on the fitting pipeline, we created multiple tabletop scenarios, containing cylindrical and other object shapes, that were recorded from various view points, with a hand-held Asus Xtion RGB-D sensing device. This dataset was partitioned in the following two sets:

3.4.2.1. Classifier dataset. this set was collected with the purpose of training, validating and testing the performance of the classifier. Each scene contained either cylinders or non-cylindrical shapes, which facilitates automatic generation of labeled datasets (see 5).

3.4.2.2. Run-time benchmark dataset. the goal of gathering this set is to benchmark the whole framework performance improvements in the presence of salient visual distractors, which differentiates from the previous set, as each scene contains both cylindrical and distracting shapes (see Fig. 6).

Table 1 contains the statistics of the two sets.

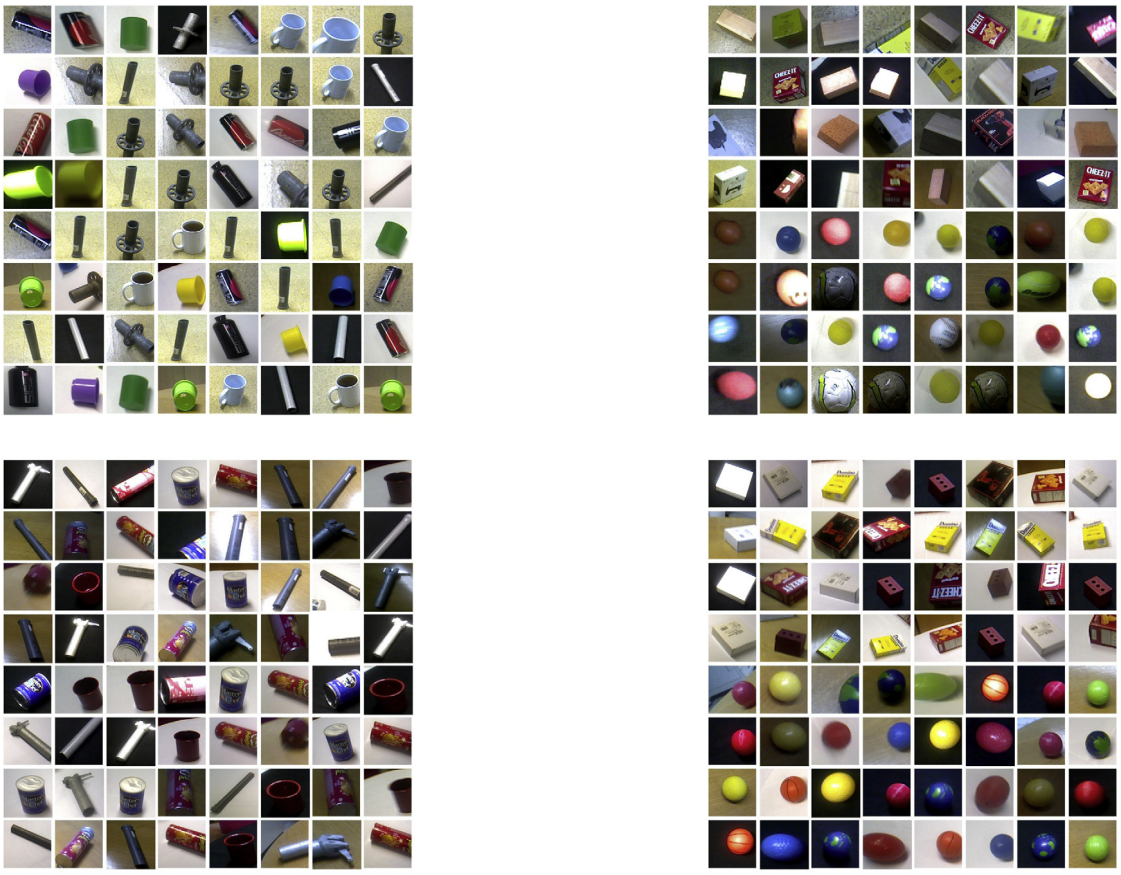
4. Experiments and results

In this section we describe the experiments carried to evaluate the components of our fast cylinder identification framework. First, we evaluate individually the proposed classification and fitting approaches, and then we report the performance of the whole methodology, with an emphasis on the computational benefits introduced by the proposed cylinder classifier.

4.1. Cylinder fitting performance

Several experiments were conducted in order to quantitatively evaluate the quality of the cylinder parameters recovered by our method and the one of Rabbani et al. [6] , when dealing with increasing levels of outliers, noise and occlusion. The fitting performance comparison was assessed using the simulation environment outlined in 3.4.1. By using synthetically generated scenes, we were able to compare the algorithm pose results with a known ground truth.

In all fitting experiments, we generated 1000 scenes, each containing a single instance of a cylinder. The selected cylinder parameters were the following: The radius was fixed to $r = 0.3$ m 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 $|\mathcal{P}| = 900$ and the number of orientation sample points in the Hough accumulator space was set to $N_d = 450$. To validate the advantages of our randomized sampling scheme for the creation of the orientation Hough accumulator, in all generated scenes the orientation of the cylinder was fixed and aligned with the z-axis of the frame of reference. We considered and compared the following different sampling distributions for creating the orientation Hough accumulator space (see Table 2):



(a) cylindrical samples (b) non-cylindrical samples

Fig. 5. Object image crop examples from the created classifier training (top row) and testing (bottom row) datasets.

Fig. 6. Scene samples from the collected 200 frame RGB-D benchmark dataset.

Table 1
Real dataset statistics.

	Training/validation			Test			Run-time benchmark		
	Cylinders	Distractors	**Total**	Cylinders	Distractors	**Total**	Cylinder	Distractor	**Total**
#scenes	1387	669	2056	694	679	1373	200		200
#objects	5657	3725	9382	2300	2219	4519	480	900	1380

- an unbiased distribution reflecting the absence of prior knowledge about the cylinder orientation.
- a mildly and a strongly biased distribution that favors vertical orientations.

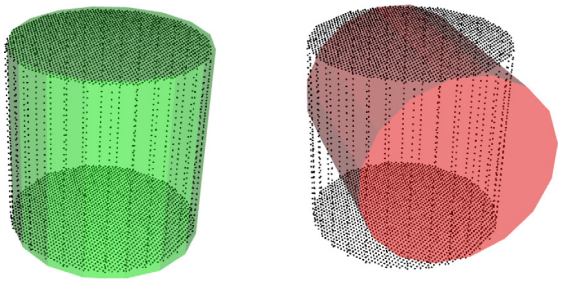
Finally, for each scene we generated 30 Hough accumulators to reduce estimation error bias and variance.

4.1.1. Robustness to outliers

In order to assess the performance gains of the proposed strategies in the presence of flat surfaces (i.e. outliers)

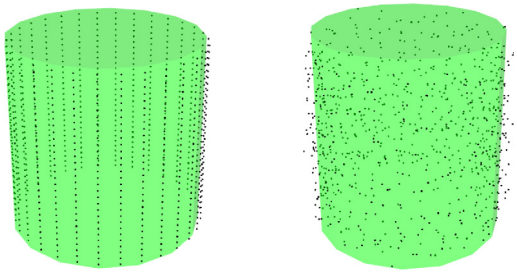
$$\text{outliers} = \frac{\text{total scene points}}{|\mathcal{P}|} - 1 \tag{5}$$

we added synthetically generated planar extremities to cylinders, that simulate realistic cylindrical shapes such as containers/cans with lids. 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 (i.e. statistical outliers). The surfaces were generated with a total of 10, increasing point density levels, to each previously generated cylinders' bottom and top extremities (see Fig. 7). The quantitative results illustrated in Fig. 10 (center column) demonstrate the advantage of considering both the surface curvature and the surface normal in the orientation voting step. When dealing with



(a) Ours (b) Rabanni et al.

Fig. 7. Our method against Rabbani et al. when dealing with flat surfaces.



(a) $\sigma_{\text{noise}} = 1\%$ (b) $\sigma_{\text{noise}} = 10\%$

Fig. 8. Estimated cylinder parameters with our method, from a synthetically generated point cloud with increasing levels of noise.

flat surfaces that belong to cylinders, our method estimates better the cylinder orientation, as shown by the absolute orientation errors in Figs. 10a and 10b.

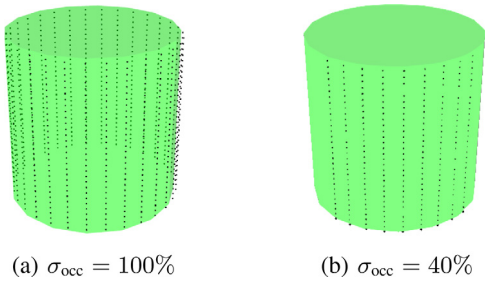
According to our implementation, the original method of Rabbani et al. can deal with cases where up to 50% of the points are outliers, without failing. When the number of outliers exceed 150% of the relative number of candidate points belonging to the cylinder surface, the method exhibits an orientation error of 90 degrees, since points belonging to flat surfaces (i.e. outliers) vote for orthogonal directions to the ground truth cylinder orientation. Our method is able to cope with up to 200% of planar outliers, with minimal impact in orientation estimation. The linear transition in between can be justified by the fact that the error increases linearly with the number of outliers voting for orthogonal, wrong orientations. This is an artifact of the soft-voting scheme, resulting in consistent response to small and large amount of outliers. In between, the response exhibits a linear decrease in the pose estimation accuracy.

As expected, these improvements have a direct and positive impact in the quality of the position and radius estimations, depicted through the absolute radius and position errors plots in Figs. 10e and 10d.

4.1.2. Robustness to noise

In pursuance of quantifying the behavior of the Rabbani et al. algorithm [6] and our proposed extensions in the presence of noisy visual sensors, each of the 1000 generated scenes was corrupted by 10 different levels of additive Gaussian noise, with standard deviation proportional to the cylinder radius (see Fig. 8).

Fig. 10 (left column) 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 significantly improves the



(a) $\sigma_{\text{occ}} = 100\%$ (b) $\sigma_{\text{occ}} = 40\%$

Fig. 9. Estimated cylinder parameters with our method, for different levels of occlusion.

Table 2
Orientation Hough accumulator biasing parameters used for the creation of the orientation Hough accumulators in the experiments with synthetic data.

Bias	$\boldsymbol{\mu}_p$			$\boldsymbol{\Sigma}_p$		
	x	y	z	xx	yy	zz
Unbiased	0	0	0	0.5	0.5	0.5
Mildly top-biased	0	0	1.0	0.5	0.5	0.5
Strongly top-biased	0	0	1.0	0.05	0.05	0.05

estimation accuracy. Overall, our extensions result in dramatic improvements regarding 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 exemplified in Figs. 1 and 12, and its superior robustness to outliers.

4.1.3. Robustness to occlusion

To evaluate the robustness of our methodology to occlusion we simulated cylinder partial views by cutting the original cylinder along the axial directions by different amounts (see Fig. 9). The amount of occlusion is given by the ratio of points in the original and occluded cylinders $|\mathcal{P}|$, according to:

$$\text{occlusion} = 1 - \frac{|\mathcal{P}'|}{|\mathcal{P}|} \tag{6}$$

where $\mathcal{P}'$ is the set of points of the occluded cylinder. Fig. 10 (right column) demonstrates that the performance of our soft voting scheme sightly improves on the method of [6]. Including the sampling bias in the direction of the cylinder orientation, the improvement becomes significant for large levels of missing data.

4.2. Shape-based attention

To train and evaluate the performance of the CNN classifier we have used the dataset outlined in Section 3.4.2.1. As explained in the previous section, we fine-tune the final layer of SqueezeNet with our which contains 9382 train samples (out of which, we used 10% for validation) and 4519 test samples of unseen objects. 5 shows a few samples that were used to train and test the network. The original training dataset contained less than 10 000 samples and, in order to gain more robustness to different orientations, they were mirrored in vertical and horizontal directions, effectively quadrupling the amount of available data. The learning rate for fine-tuning the network was empirically selected as 0.0005 and we kept other parameters as their proposed values by [39]. Fig. 11a shows the performance of the classifier at various points during training.

Our experiments with the neural network classifier demonstrates generalization to unseen cylindrical and non-cylindrical objects. In order to quantitatively evaluate the performance of the 2D image-based deep neural network classifier, it is compared

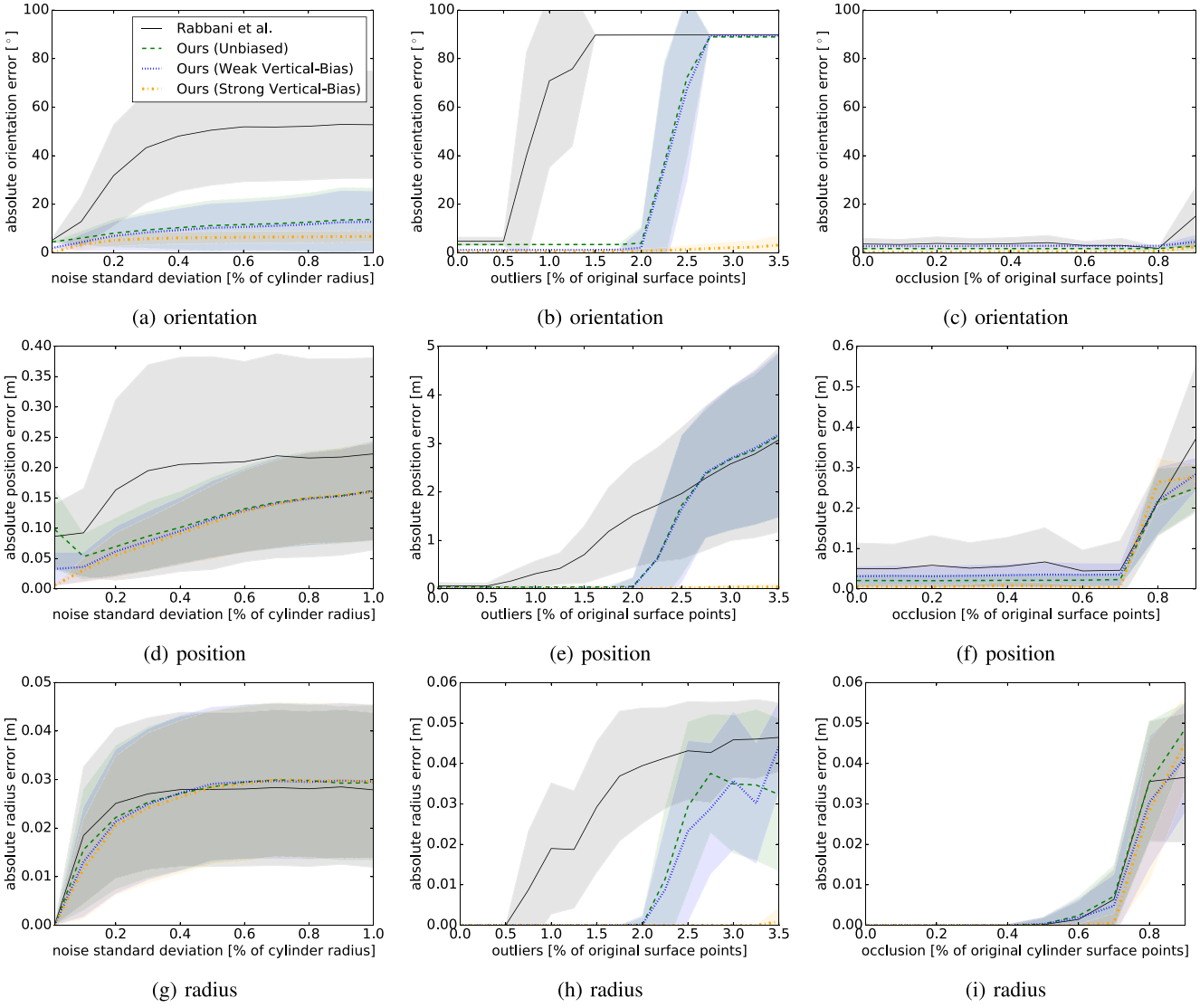


Fig. 10. Robustness of our method against the method of Rabbani et al. Left: different levels of noise. Center: different levels of flat surface outliers. Right: different levels of occlusion.

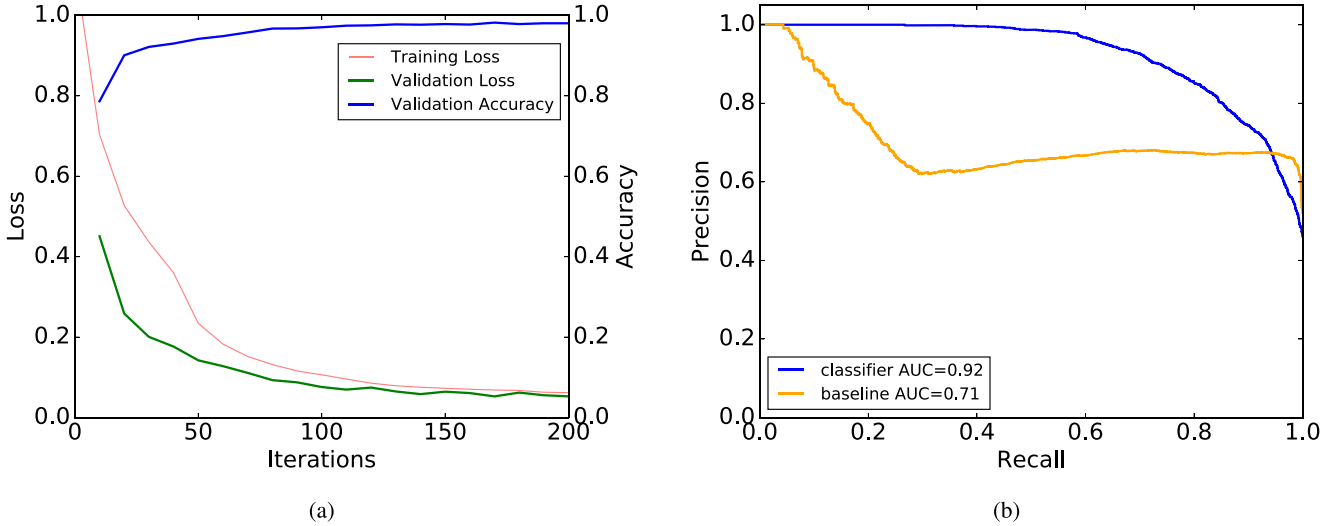


Fig. 11. Evaluation of the performance of the binary classifier. (a) Loss and accuracy evolution of the classifier on training and validation data. (b) Precision–Recall curves of the cylinder class for baseline and SqueezeNet classifier on the test data. AUC: Area Under the Curve.

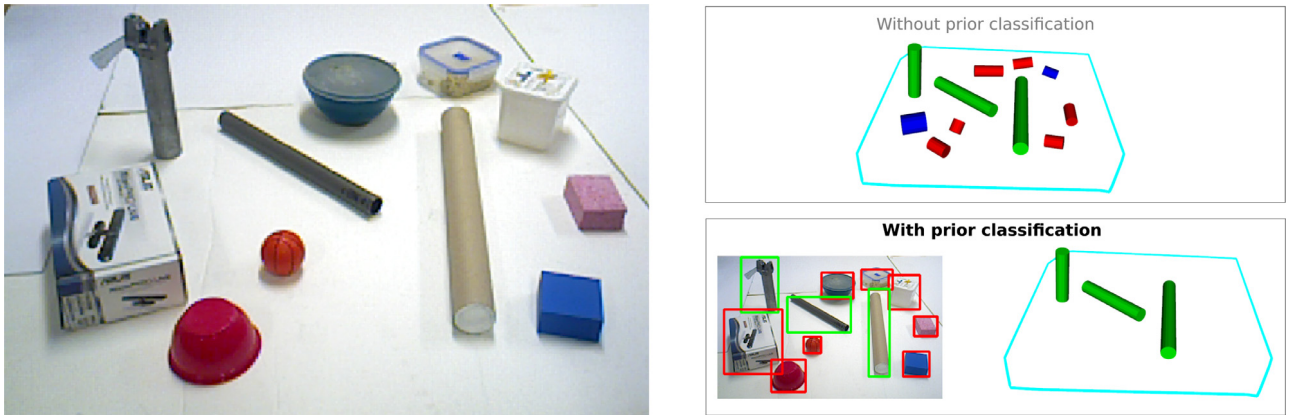


Fig. 12. Qualitative assessment of our framework with data acquired with an Asus Xtion 3D camera. Cylinder identification for an example scene. . Detection: Good and bad classifications in green and red, respectively. Parameter identification: green represents correct parameter estimation; blue represents correct non-cylindrical shape objects identified by the baseline quality of fitting criterion; red represents wrong estimations without the classifier . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with a baseline indicator of the fit quality criteria defined in Section 3.3.3. Fig. 11b compares the precision–recall curves of the two classifiers on the test set, demonstrating the superior performance of the proposed classifier.

4.3. Overall framework assessment

The complete framework was evaluated using the dataset described in Section 3.4.2.2, where each scene contains on average 7 objects (see Table 3a). Fig. 12 depicts an example of the cylinder parameters estimation quality for the proposed methodology in the presence of noisy 3D point cloud data. The use of prior classification, results not only in temporal gains, but also on early filtering of non-cylindrical distractors, hence improving the reliability of the 3D cylinder fitting approach (see Table 3). Overall, improvements on detection speed and robustness to visual distractors can be achieved by incorporating the shape-based pre-attention mechanism, which results in improvements on detection speed and robustness to visual distractors without sacrificing robustness to noise. Furthermore, the evaluation of our method with data acquired from a consumer RGB-D camera demonstrates our method applicability to real-scenarios and its advantages in scenes populated with salient visual distractors. In order to better ground the time complexity of this pipeline, we have also experimented with an off-the-shelf state-of-the-art object detector (Faster-RCNN) [44], which similar to SqueezeNet was also fine-tuned to detect cylinders in RGB images. This detector uses ResNet101 as the classifier. Using the detector, one can achieve a constant run-time with respect to the number of objects in a scene, however, according to Table 3, only the segmentation and classification provided by Faster-RCNN takes more time than our complete pipeline, even with an average of 7 visible objects. Furthermore, unlike off-the-shelf object detectors, 3D tabletop segmentation allows the definition of a table coordinate frame and, hence, the incorporation of prior knowledge in the fitting process.

Table 3
Quantitative analysis of the time performance of the proposed pipeline in a set of 200 RGB-D frames acquired with an Asus Xtion camera.

(a) Detected objects per scene

	Cylinders	Distractors	**Total**
Ground truth	2.4 ± 0.68	4.5 ± 1.50	6.9 ± 1.8
No classifier	4.00 ± 0.77	4.00 ± 0.77	8.00 ± 0.00
With classifier	1.90 ± 0.70	6.10 ± 0.70	8.00 ± 0.00

(a) Processing times (ms)

	Segmentation	Classification	Identification	**Total**
No classifier	37.38 ± 9.29	–	97.89 ± 44.46	135.27 ± 48.27
With classifier	37.38 ± 9.23	14.52 ± 5.21	28.71 ± 23.29	$\mathbf{80.61 \pm 28.53}$
F-RCNN	142.12 ± 6.61		–	–

5. Conclusions

In this paper, we have proposed a complete, robust and, efficient cylinder detection and parameter identification framework. Unlike previous approaches that are only based on 3D depth information, our methodology incorporates RGB information by means of a novel shape-based pre-attentive top-down attentional mechanism that filters out visual distractors at an early stage. Furthermore, we have developed 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 demonstrate significant detection accuracy and time speed-ups as well as major improvements on the detection rates and pose estimations with respect to previous schemes. A systematic quantitative analysis of robustness to outliers 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 sequential Bayesian filtering [45]. In addition, the current classifier is trained with a limited number of cylinders, however, it is expected to improve the generalization to unseen cylinders if the training set contains multiple cylindrical objects of various shapes and colors. On the other hand, even finetuning such methods commonly require training with large amounts of data which is time consuming and sometimes unfeasible. We have circumvented this issue by devising an automated data collection and annotation scheme, however, recent advances in using simulated data for finetuning is another promising approach to overcome this challenge [46,47].

In this paper we have focused on cylindrical shapes but the proposed core ideas can be easily extended to other shape types, depending on training data availability. Combining a generic multi-label classifier with the proposed randomized Hough accumulator and the soft voting scheme, paves the way to extend

the current cylinder identification pipeline to various shapes (e.g. cuboids, ellipsoids, cones). As a final remark, we emphasize that the computational complexity of the proposed solution scales linearly with the number of objects in the scene, which may become problematic in environments with many distractors. However, all components of the pipeline are parallelizable and, depending on the application requirements, one can benefit from an increase in the available hardware resources to further improve run-time performance. Finally, complex objects such as cylindrical containers require more elaborate representations such as semantic or relational. In the case of cylindrical containers one can consider that containers have two object primitives: planes and cylinders. Future work should consider these type of representations through the use of Probabilistic Graphical Models [48] to further improve the pipeline performance.

Acknowledgments

This work has been partially supported by the Portuguese Foundation for Science and Technology (FCT) project [UID/EEA/50009/2019]. Rui Figueiredo and Atabak Dehban are funded by FCT, Portugal PhD grant PD/BD/105779/2014 and PD/BD/105776/2014, respectively. Helder Araújo would like to thank FCT grant [UID/EEA/0048/2013]. Finally, we gratefully acknowledge the support of NVIDIA Corporation, United States with the donation of the Titan Xp GPU used for this research.

References

[1] P. Moreno, R. Nunes, R. Figueiredo, R. Ferreira, A. Bernardino, J. Santos-Victor, R. Beira, L. Vargas, D. Aragão, 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. Aragao, P. Moreno, A. Bernardino, J. Santos-Victor, A. Billard, Reaching and grasping kitchenware objects, in: IEEE/SICE International Symposium on System Integration (SII), 2012, pp. 865–870.

[3] A.T. Miller, S. Knoop, H.I. Christensen, P.K. Allen, Automatic grasp planning using shape primitives, in: Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on, Vol. 2, IEEE, 2003, pp. 1824–1829.

[4] A. Dehban, L. Jamone, A.R. Kampff, J. Santos-Victor, A deep probabilistic framework for heterogeneous self-supervised learning of affordances, in: IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids), 2017, IEEE, 2017, pp. 476–483.

[5] T. Mar, V. Tikhanoff, L. Natale, What can i do with this tool? self-supervised learning of tool affordances from their 3d geometry., IEEE Trans. Cognit. Deve. Syst. (2017).

[6] T. Rabbani, F. Van Den Heuvel, Efficient hough transform for automatic detection of cylinders in point clouds, ISPRS WG III/3, III/4 3 (2005) 60–65.

[7] R. Figueiredo, P. Moreno, A. Bernardino, Robust cylinder detection and pose estimation using 3d point cloud information, in: IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), 2017.

[8] M. Quigley, J. Faust, T. Foote, J. Leibs, Ros: an open-source robot operating system, in: ICRA Workshop on Open Source Software, Vol. 3, no. 3.2., 2009.

[9] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: convolutional architecture for fast feature embedding, in: Proceedings of the 22nd ACM International Conference on Multimedia, ACM, 2014, pp. 675–678.

[10] D. Amso, G. Scerif, The attentive brain: insights from developmental cognitive neuroscience, Nat. Rev. Neurosci. 16 (10) (2015) 606–619.

[11] R. Parasuraman, S. Yantis, The Attentive Brain, Mit Press Cambridge, MA, 1998.

[12] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: Computer Vision (ICCV), 2017 IEEE International Conference on, IEEE, 2017, pp. 2980–2988.

[13] D. Held, S. Thrun, S. Savarese, Learning to track at 100 fps with deep regression networks, in: European Conference on Computer Vision, Springer, 2016, pp. 749–765.

[14] S. Zagoruyko, N. Komodakis, Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer, in: International Conference on Learning Representations, 2017.

[15] V. Gal, L.R. Kozák, I. Kóbor, E.M. Bankó, J.T. Serences, Z. Vidnyánszky, Learning to filter out visual distractors, Eur. J. Neurosci. 29 (8) (2009) 1723–1731.

[16] B.B. Stojanoski, M. Niemeier, Late electrophysiological modulations of feature-based attention to object shapes, Psychophysiology 51 (3) (2014) 298–308.

[17] R.W. Fleming, Visual perception of materials and their properties, Vision Res. 94 (2014) 62–75.

[18] S. Tek, G. Jaffery, L. Swensen, D. Fein, L.R. Naigles, The shape bias is affected by differing similarity among objects, Cognit. Dev. 27 (1) (2012) 28–38.

[19] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, 3d object recognition in cluttered scenes with local surface features: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 36 (11) (2014) 2270–2287.

[20] C.R. Qi, H. Su, K. Mo, L.J. Guibas, Pointnet: deep learning on point sets for 3d classification and segmentation, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, 2017, pp. 77–85, [Online]. Available: https://doi.org/10.1109/CVPR.2017.16.

[21] B. Drost, M. Ulrich, N. Navab, S. Ilic, Model globally, match locally: efficient and robust 3d object recognition, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Ieee, 2010.

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

[23] P. Hough, Method and Means for Recognizing Complex Patterns, U.S. Patent 3.069.654, 1962.

[24] T.-T. Tran, V.-T. Cao, D. Laurendeau, Extraction of cylinders and estimation of their parameters from point clouds, Comput. Graphics 46 (2015) 345–357.

[25] A. Nurunnabi, Y. Sadahiro, R. Lindenbergh, Robust cylinder fitting in three-dimensional point cloud data, ISPRS - Int. Archives Photogrammetry, Remote Sens. Spatial Inform. Sci. XLII-1/W1 (2017) 63–70, [Online]. Available: https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-1-W1/63/2017/.

[26] K. Huebner, S. Ruthotto, D. Kragic, Minimum volume bounding box decomposition for shape approximation in robot grasping, in: Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on, IEEE, 2008, pp. 1628–1633.

[27] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.

[28] R. Schnabel, R. Wahl, R. Klein, Efficient ransac for point-cloud shape detection, in: Comput Graphics Forum, Vol. 26, no. 2, Wiley Online Library, 2007, pp. 214–226.

[29] L.C. Goron, Z.-C. Marton, G. Lazea, 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.

[30] Y.-J. Liu, J.-B. Zhang, J.-C. Hou, J.-C. Ren, W.-Q. Tang, Cylinder detection in large-scale point cloud of pipeline plant, IEEE Trans. Visualizat. Comput. Graphics 19 (10) (2013) 1700–1707.

[31] Y.-T. Su, J. Bethel, Detection and robust estimation of cylinder features in point clouds, in: ASPRS Conference, 2010.

[32] A.K. Patil, P. Holi, S.K. Lee, Y.H. Chai, An adaptive approach for the reconstruction and modeling of as-built 3d pipelines from point clouds, Autom. Constr. 75 (2017) 65–78.

[33] B. Paláncz, J. Awange, A. Somogyi, N. Rehány, T. Lovas, B. Molnár, Y. Fukuda, A robust cylindrical fitting to point cloud data, Aust. J. Earth Sci. 63 (5) (2016) 665–673.

[34] I. Kostavelis, L. Nalpantidis, A. Gasteratos, Object recognition using saliency maps and htm learning, in: Imaging Systems and Techniques (IST), 2012 IEEE International Conference on, IEEE, 2012, pp. 528–532.

[35] M. Muja, M. Ciocarlie, Table Top Segmentation Package.

[36] R.B. Rusu, N. Blodow, Z.C. Marton, M. Beetz, Close-range scene segmentation and reconstruction of 3d point cloud maps for mobile manipulation in domestic environments, in: Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on, IEEE, 2009, pp. 1–6.

[37] K. Weiss, T.M. Khoshgoftaar, D. Wang, A survey of transfer learning, J. Big Data 3 (1) (2016) 9.

[38] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. (IJCV) 115 (3) (2015) 211–252.

[39] F.N. Iandola, M.W. Moskewicz, K. Ashraf, S. Han, W.J. Dally, K. Keutzer, SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size, 2016, arXiv:1602.07360.

[40] E. Lutton, H. Maitre, J. Lopez-Krahe, Contribution to the determination of vanishing points using hough transform, IEEE Trans. Pattern Analy. Mach. Intell. 16 (4) (1994) 430–438.

[41] M.E. Muller, A note on a method for generating points uniformly on n-dimensional spheres, Commun. ACM 2 (4) (1959) 19–20.

[42] K.P. F.R.S., On lines and planes of closest fit to systems of points in space, Phil. Mag. Ser. 6 2 (11) (1901) 559–572.

[43] D.J. Kerbyson, T.J. Atherton, Circle detection using hough transform filters, in: Fifth International Conference on Image Processing and its Applications, 1995., 1995, pp. 370–374.
[44] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, in: C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, R. Garnett (Eds.), Advances in Neural Information Processing Systems 28, Curran Associates, Inc., 2015, pp. 91–99, [Online]. Available: http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf.
[45] R.P. de Figueiredo, P. Moreno, A. Bernardino, 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.
[46] J. Borrego, R. Figueiredo, A. Dehban, P. Moreno, A. Bernardino, J. Santos-Victor, A generic visual perception domain randomisation framework for gazebo, in: IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC). Torres Vedras, Portugal, 237–242.
[47] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, P. Abbeel, Domain randomization for transferring deep neural networks from simulation to the real world, in: Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on, IEEE, 2017, pp. 23–30.
[48] G. Heitz, Graphical Models for High-Level Computer Vision, Stanford University, 2009.

Rui Figueiredo received a M.Sc. degree in Electrical and Computer Engineering from Instituto Superior Técnico (IST), Lisbon, Portugal, in 2012. He has been a Research Assistant member of the Computer Vision Laboratory (VisLab), Institute for Systems and Robotics (ISR), Lisbon. His work as has been related with 3D geometry processing, computer vision and robotics, with a strong emphasis on the subject of object recognition. He has been responsible for software development, implementation and maintenance within two EU projects (First-MM and HANDLE) which were mainly directed towards robot grasping and in-hand manipulation applications. He is currently a Ph.D. candidate, within the Robotics, Brain and Cognition (RBCog) doctoral program, at the Institute for Systems and Robotics, Instituto Superior Técnico (ISR/IST), Universidade de Lisboa. His research interest lies at the intersection of machine learning, computer vision and robotics with the goal of developing efficient visual perception algorithms for robots, inspired by attention mechanisms existent in humans.