IEEE International Conference on Development and Learning and on Epigenetic Robotics (ICDL-EpiRob)
Lisbon, Portugal, 18-21 September 2017

Shape-Based Attention for Identification and Localization of Cylindrical Objects

Rui Figueiredo*, Atabak Dehban*†, Alexandre Bernardino*, José Santos-Victor* and Helder Araújo‡
*Institute for Systems and Robotics, Instituto Superior Tećnico, Universidade de Lisboa, Lisbon, Portugal
Email: {ruifigueiredo,adehban,alex,jasv}@isr.tecnico.ulisboa.pt
†Champalimaud Centre for the Unknown, Lisbon, Portugal
‡Institute for Systems and Robotics, Universidade de Coimbra, Coimbra, Portugal
Email: helder@isr.uc.pt

***Abstract*—In this paper, we propose a novel framework for detecting and identifying cylindrical shapes, commonly found in daily contexts, using multi-modal visual information provided by RBG-D cameras. The current state-of-the-art methods for cylinder detection are based on RANSAC and Hough transforms which estimate cylinder parameters using 3D point cloud information. However, the presence of distracting non-cylindrical shapes leads to time-consuming parametric fitting of wrong detections, compromising the efficiency of the whole method.**

We tackle the aforementioned problem by introducing a biologically plausible framework, incorporating a pre-attentive mechanism which learns in a supervised and data-efficient manner to selectively discard irrelevant shapes before further processing. A set of experiments with real data are conducted to assess the advantages of our framework. The results demonstrate that combining bottom-up 3D segmentation with top-down shape-based attention 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.

I. Introduction

In many tasks involving interaction with the surrounding environment, biological and artificial systems require accurate object recognition and pose estimation capabilities. These tasks are successful manipulation and grasping, obstacle avoidance and self localization with respect to known landmarks, to name a few.

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 be flexible enough to cope with inter and intra-class variability generalizing to never-seen objects. Furthermore, in order to cope with transmission bandwidth and computational processing capacity limitations, efficient and fast perception is a primordial requirement for real-time performance. Therefore, it is extremely important to design efficient perceptual systems that are not only robust to sensory noise and occlusion, but also to clutter and visual distractors.

In this work, we propose a novel biologically inspired 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), and whose identification plays an important role in children development and learning [1], [2], as well as in robotic grasping applications [3], [4].

*First two authors contributed equally to this work

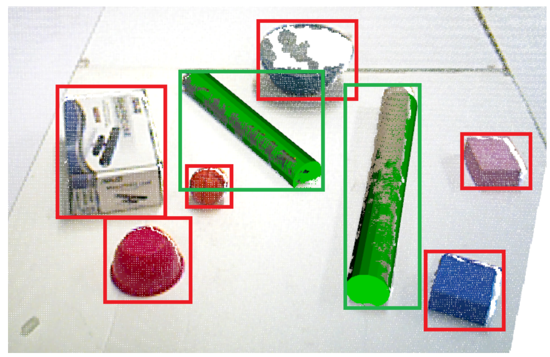


Fig. 1: A snapshot of a RGB-D point cloud and overlaid cylindrical (green) and non-cylindrical (red) shapes detected with our methodology. Figure best seen in color.

The proposed framework relies on the tabletop assumption, i.e., objects are placed on flat surfaces (Fig. 1). In order to deal with cluttered environments which are often populated with multiple non-cylindrical shapes, we take advantage of the recent advances in deep learning architectures to introduce an efficient recognition module that learns to filter out irrelevant visual distractors. More specifically, we introduce the use of pre-attentive shape-based selection mechanisms, that avoid the need of time-consuming, top-down cylinder parameter identification at an early stage, on irrelevant salient candidate objects.

The main contribution of this paper is a novel multi-modal framework which takes advantage of state-of-the-art 3D cylinder identification and image-based object recognition models to accurately and rapidly estimate pose and shape parameters of objects of interest. Unlike previous approaches that are solely based on 3D depth information, we combine a state-of-the-art [5], [6] cylinder fitting approach which is based on a robust and computationally efficient 2-step Generalized Hough Transform (GHT) with a 2D image-based top-down proposal rejection mechanism to increase the quality and speed of correct estimations. Since gathering a large dataset, required for recent recognition techniques is laborious and time consuming, we provide a semi-automatic data gathering procedure, using 3D information, which greatly facilitates the acquiring and labeling of relatively large amounts of data as our second contribution. Our ROS [7] and Caffee [8] C++ implementation runs in real-time on a GPU, allowing an easy and direct integration in robotics systems, e.g. in grasping and manipulation pipelines. The code and dataset of our experiments will be released when the final version of this manuscript is prepared.

The remainder of this paper is organized as follows: In section II we overview the related work on attention and object identification in the recent literature. In section III we describe in detail the various steps involved in the proposed cylinder identification and localization 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

Successful identification of objects in an environment requires not only the development of 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. At first we overview the state-of-the-art in visual attention with an emphasis on shape-based models of selective attention. Afterward, we analyse various object identification paradigms proposed in the literature, suitable for applications requiring identification and localization of parametric shapes.

A. Shape-based Selective Attention

Visual attention plays a central role in biological and artificial systems to control perceptual resources [9], [10]. The classic artificial visual attention systems use salient features of the image obtained from the information provided by hand-crafted filters. Recently, deep neural networks have been developed for recognizing thousands of objects and autonomously generate visual characteristics optimized by training with large data sets. Besides being used for object recognition, these features have been very successful in other visual problems such as object segmentation, tracking and recently, visual attention.

Evidence from neurophysiology studies [11] suggests that people consider shape as an important feature dimension among other low-level visual features (e.g. texture and color). In [12] 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 attentional mechanisms for shape features. Furthermore, infants rely more on shape than on color when learning new objects, which in turn allows them to generalise to other objects with similar visual features while interacting with them [13]. This fact motivates the need of developing more sophisticated, shape-biased and biologically plausible, attentional architectures [14].

B. Object identification in robotics

Object recognition and pose estimation in an important subject in computer vision with various applications in robotics [15]. 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 more commonly used [16]. These representations are often unsuitable, when flexibility and generalization to novel objects is a requirement. The dominant strategies rely in machine learning techniques that are able to generalize to similar objects using a set of sample images acquired by the robot sensors. In this paper we focus on cylindrical shapes and, thus, we will combine the generalization capabilities of state-of-the-art deep learning techniques with robust 3D model-based fitting approaches.

For the extraction of simple geometric shape primitives like planes, cylinders, cones and spheres, the two most common paradigms are the Hough transform [17] and Random Sample Consensus (RANSAC) [18], which are robust to outliers and noisy data.

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 [5], [6]. More specifically, they rely on a 2D Hough transform to estimate orientation followed by a 3D Hough transform to simultaneously detect radius and position. Although reducing the exponential complexity factor, these approaches still lacks speed in dense point clouds, being incapable of filtering at early step different object shapes which act as distractors. The lack of processes for fast hypothesis verification sets one of the main setbacks of the current approaches for real-time applications.

This work significantly differs from previous approaches as it incorporates a mediating shape-based pre-attention mechanism to reduce the space of possible cylindrical shapes [19] to a small subset of prominent objects in the field of view. The 2D image patches, comming from 3D segmentation are first classified and only if they belong to object classes of interest, they will be used for parameter identification which results in faster and more accurate estimates.

III. 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 Figure 2, is an integration of different cognitive blocks which are responsible for object segmentation and shape recognition, fitting and localization.

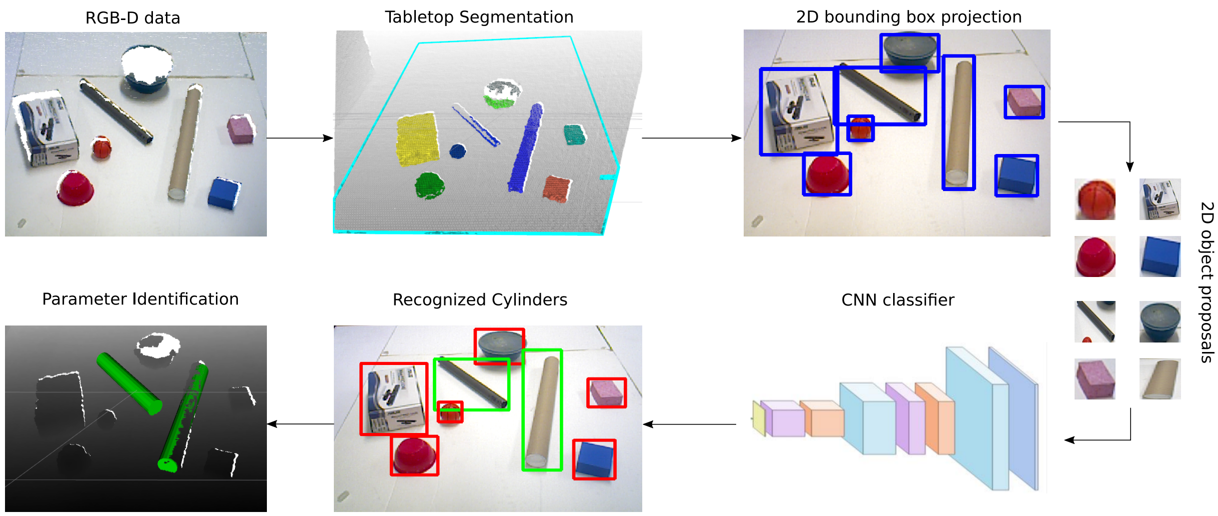


Fig. 2: General overview of our shape-based attention framework.

In the remainder of this section we describe in detail the multiple components of our pipeline.

A. 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 [20]. First, we use a RANSAC-based fitting approach, which efficiently operates on downsampled organized point cloud data [21], in order to detect planes on the scene and segment objects above these planes. We rely on Euclidean clustering [21] 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 biologically inspired 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.

B. 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, deep neural networks 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, their performance even at test time is relatively slow due to the large amount of their parameters.

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.

2) Cylindrical-shapes recognition: For the second problem, i.e. architecture design, we propose to use transfer learning [22]. More specifically, we have used a network previously trained on imagenet dataset [23] 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 *good* 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) Performance speed-ups: Since the tasks requires to have relatively fast performance at test time, we used a relatively small neural network called SqueezeNet [24]. 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.

C. Robust 3D cylindrical shape fitting

Our Hough-based cylinder fitting approach is based on the former work of Rabanni et al. [5] and divides the cylinder parametric fitting in two independent 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 the original work [5] the unit sphere of orientations is uniformly and deterministically sampled at a predefined number of points [25], 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.

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, 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)

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

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 where each $\mathbf{d}^i \in \mathcal{D}$ represents an orientation, allowing for efficient voting with estimated surface normals, using simple inner products (equation 3).

2) Fast 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 $P = \{\mathbf{p}^s, s = 1, ..., N_s\}$.

First, we estimate the surface normals at each scene point $\mathbf{p}^s \in \mathcal{P}$ using the Principal Component Analysis (PCA) of the covariance matrix created from its k-nearest neighbors. Let $\mathcal{N} = \{\mathbf{n}^s, s = 1, ..., N_s\}$ 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 $\mathbf{c}^s_{\text{max}} \in \mathbb{R}^3$ and the corresponding eigenvalue $k_{\text{max}} \in \mathbb{R}$. Let $\mathcal{C} = \{\mathbf{c}^s_{\text{max}}, s = 1, ..., N_s\}$ denote the set of principal curvature directions and $\mathcal{K} = \{k^s_{\text{max}}, s = 1, ..., N_s\}$ the set 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 principal curvature directions $\mathbf{c}^s_{\text{max}} \in \mathcal{C}$ to cast continuous votes in the accumulator according to the function

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

This voting function gives more weight to directions that are simultaneously, orthogonal to the the normal and the principal curvature directions. Furthermore, the eigenvalue k^s_{max} functions as a curvature high-pass filter, that suppresses low curvature candidates, since points belonging to flat surfaces have very low k^s_{max}. After determining the cylinder orientation we proceed with the estimation of the cylinder position and radius. For further details we refer the interest reader to [5].

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.

IV. RESULTS

In order to assess the behavior of the proposed framework with real data acquired from a low-cost consumer RGB-D sensing device, we created multiple tabletop scenarios, each containing various different shapes including cylindrical objects (see Fig. 3a for an example view). We quantitatively and qualitatively evaluated our attentional framework's computational time improvements, in the presence of salient visual distractors.

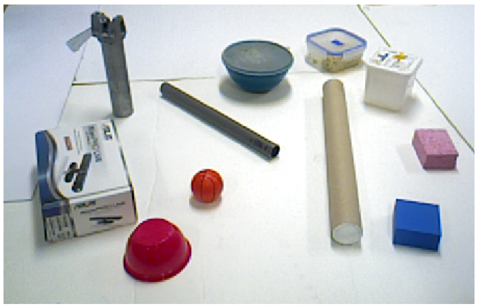
In all experiments, the selected parameters for the cylinder parameter estimation methodology where the following: The number of orientation sample points in the Hough accumulator space was set to $N_d = 450$. The radius Hough space was defined in the interval $[0.25, 0.35]$ and quantized into 10 bins, and the height was defined in the interval $[0.05, 1.0]\,m$ and discretized into 100 bins.

A. Classifier Performance Analysis

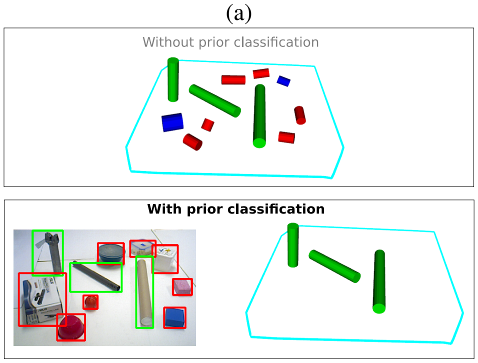
As described in the previous section, we fine-tune SqueezeNet with the newly gathered dataset which contains about 11000 train images and 1200 test images. Fig. 4 shows a few samples that were used to train the network. The original dataset contained less than 3000 samples and,

	Scene Objects Number		Avg. Processing Time (ms)			
	Cylinders	Distractors	Segmentation	Classification	Identification	Total
no classifier	3	8	100	-	213	313
with classifier				64	70	**234**

TABLE I: Quantitative analysis of the time performance of the proposed pipeline, for the specific episode of Fig. 3a



(a)



(b)

Fig. 3: Qualitative assessment of our framework with data data acquired with an Asus Xtion 3D camera. (a) Example testing scene. (b) Cylinder recognitions (Left): Good and bad classifications in green and red, respectively. Parameter identification (Right): 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.

in order to gain more robustness to different orientations, they were mirrored in vertical and horizontal directions, effectively quadrupling the amount of available data.

All the layers of the network were fine-tuned. This is justified because imagenet samples are not centered, as opposed to our dataset, thus requiring a higher capacity for domain adaptation. The learning rate for fine-tuning the network was emperically selected as 0.01 and we kept other parameters as their proposed values by [24]. Fig. 5 shows the performance of the classifier during training.

Our initial experiments with the neural network classifier suggests a generalization to unseen cylindrical and non-cylindrical objects. However, not surprisingly, it is more reliable in classifying seen cylinders. Introducing more unique cylinders can help mitigating this effect.In order to quantitatively evaluate the performance of the 2D image-based deep neural network classifier, it is compared with a baseline indicator of the fit quality criteria defined in section III-C3. Fig. 6 compares the precision recall curves of the two classifiers.

(a) Cylindrical samples;

(b) Non-cylindrical samples.

Fig. 4: Sample examples from the training dataset after rotation augmentation.

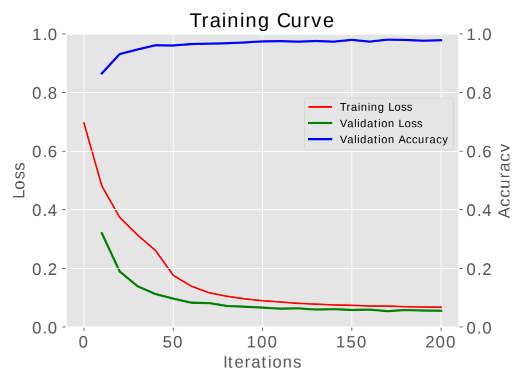


Fig. 5: Loss and accuracy evolution of the classifier on training and validation data.

B. Overall Framework Assessment

Figure 3 depicts the cylinder parameters estimation quality for the proposed cylinder fitting methodology in the presence of noisy 3D point cloud data. The use of prior classification results not only in temporal gains (see Table I), but also on early filtering of non-cylindrical distractors, hence improving the reliability of the 3D cylinder fitting approach. Overall, the incorporation of shape-based pre-attention mechanism, results in dramatic 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.

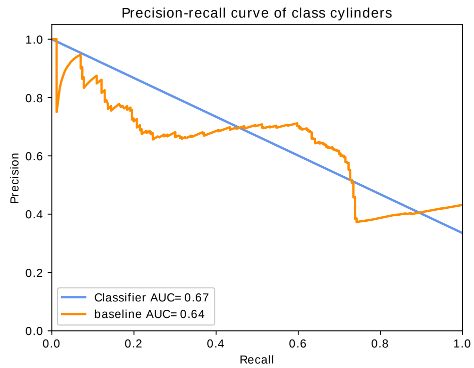


Fig. 6: Precision-Recall plots of the binary classifiers on the test data.

V. CONCLUSIONS

In this paper, we have proposed a biologically inspired robust and efficient cylinder detection framework. Unlike previous approaches that are solely 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 early stage. The results demonstrate significant detection and speed-up time improvements.

Currently, the bottleneck section of the presented pipeline is the generation of object proposals. Having an end-to-end mechanism, capable of directly predicting region proposals may significantly speed-up the whole process. 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. Last, utilizing a generic multi-label classifier paves the way to extend the current work for multiple shapes such as cubes, spheres, etc. without sacrificing the performance.

ACKNOWLEDGMENT

This work has been partially supported by the Portuguese Foundation for Science and Technology (FCT) project [UID/EEA/50009/2013]. Rui Figueiredo and Atabak Dehban are funded by FCT PhD grant PD/BD/105779/2014 and PD/BD/105776/2014, respectively. Helder Araújo would like to thank FCT grant [UID/EEA/0048/2013].

REFERENCES

[1] L. K. Samuelson and L. B. Smith, "Children's attention to rigid and deformable shape in naming and non-naming tasks," *Child development*, vol. 71, no. 6, pp. 1555–1570, 2000.

[2] ——, "They call it like they see it: Spontaneous naming and attention to shape," *Developmental science*, vol. 8, no. 2, pp. 182–198, 2005.

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

[4] A. T. Miller, S. Knoop, H. I. Christensen, and 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.

[5] 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.

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

[7] 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.

[8] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[9] D. Amso and G. Scerif, "The attentive brain: insights from developmental cognitive neuroscience," *Nature Reviews Neuroscience*, vol. 16, no. 10, pp. 606–619, 2015.

[10] R. Parasuraman and S. Yantis, *The attentive brain*. Mit Press Cambridge, MA, 1998.

[11] V. Gal, L. R. Kozák, I. Kóbor, E. M. Bankó, J. T. Serences, and Z. Vidnyánszky, "Learning to filter out visual distractors," *European Journal of Neuroscience*, vol. 29, no. 8, pp. 1723–1731, 2009.

[12] B. B. Stojanoski and M. Niemeier, "Late electrophysiological modulations of feature-based attention to object shapes," *Psychophysiology*, vol. 51, no. 3, pp. 298–308, 2014.

[13] R. W. Fleming, "Visual perception of materials and their properties," *Vision research*, vol. 94, pp. 62–75, 2014.

[14] S. Tek, G. Jaffery, L. Swensen, D. Fein, and L. R. Naigles, "The shape bias is affected by differing similarity among objects," *Cognitive development*, vol. 27, no. 1, pp. 28–38, 2012.

[15] R. M. Haralock and L. G. Shapiro, *Computer and robot vision*. Addison-Wesley Longman Publishing Co., Inc., 1991.

[16] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3d object recognition," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. Ieee, 2010, pp. 998–1005.

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

[18] 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.

[19] 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.

[20] M. Muja and M. Ciocarlie, "Table top segmentation package."

[21] R. B. Rusu, N. Blodow, Z. C. Marton, and 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.

[22] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, p. 9, 2016.

[23] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[24] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and $<$1mb model size," *arXiv:1602.07360*, 2016.

[25] 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.