

Depth Based Object Detection from Partial Pose Estimation of Symmetric Objects

Ehud Barnea and Ohad Ben-Shahar

Dept. of Computer Science, Ben-Gurion University of the Negev, Beer Sheva, Israel
{barneaeh,ben-shahar}@cs.bgu.ac.il

Abstract. Category-level object detection, the task of locating object instances of a given category in images, has been tackled with many algorithms employing standard color images. Less attention has been given to solving it using range and depth data, which has lately become readily available using laser and RGB-D cameras. Exploiting the different nature of the depth modality, we propose a novel shape-based object detector with partial pose estimation for axial or reflection symmetric objects. We estimate this partial pose by detecting target's symmetry, which as a global mid-level feature provides us with a robust frame of reference with which shape features are represented for detection. Results are shown on a particularly challenging depth dataset and exhibit significant improvement compared to the prior art.

Keywords: Object detection, 3D computer vision, Range data, Partial pose estimation.

1 Introduction

The recent advances in the production of depth cameras have made it easy to acquire depth information of scenes in the form of RGB-D images or 3D point clouds. The depth modality, which is inherently different than color and intensity, is lately being employed to solve many kinds of computer vision problems, such as object recognition [12], object detection [7,26], pose estimation [24,1] and segmentation [25].

Owing to the growing research interest, several datasets that include depth information have also been publicly released [11,9,26], allowing to evaluate different algorithms requiring this kind of data. Like image databases for appearance-based detection (e.g., [5]), the Berkeley category-level 3D object dataset [9] contains a high variability of objects in different scenes and under many viewpoints. Together with available bounding box annotations this dataset is perfectly suitable for the task of object category detection.

Considering the available depth data, we propose to tackle the category-level object detection problem of *symmetric* objects. While this may seem limiting, a quick look around reveals the extent to which symmetry is present in our lives, and it may come as no surprise that most of the objects in the Berkeley category-level 3D object dataset [9] and in the Washington's RGB-D Object Dataset [11]

are symmetric as well. Our proposed detector therefore attempts to exploits symmetry to provide partial pose estimation and then constructs a representation that is based on this estimation. More specifically, once partial pose estimation is obtained from the symmetry, we construct a feature vector by processing surface normal angles and bins that are both calculated relative to the estimated partial pose. In what follows we discuss the relevant background (Section 2) followed by an elaborate description of our suggested algorithm (Section 3) and comparative results to previous methods (Section 4).

2 Background

2.1 Employing the Depth Modality

The depth modality has been applied to most kinds of computer vision problems, even more so following the release of the Kinect camera. This kind of data, in contrast to plain RGB, enables a much easier calculation of various things such as the separation of foreground from background (by thresholding distances), floor detection and removal (by finding the dominant plane in the scene [1]), segmentation of non-touching objects (by identifying and removing the floor [1]), or the estimation of surface normals and curvature (e.g., by fitting planes to 3D points in small neighborhoods [22]).

In previous studies, depth images were often treated as intensity images [26,9], allowing them to be used with previous algorithms requiring regular images. Exploiting surface *normals* as well, Hinterstoisser et al. [8] recently suggested a combined similarity measure by considering both image gradients and surface normals. This was done to combine the qualities of color with depth, as strong edges are prominent mostly *on* object silhouettes, whereas the normals make explicit the shape *between* silhouettes.

2.2 RGB-D Category-Level Object Detection

Object detection in color images has been a subject of research for many years. With the introduction of depth data to the field new challenges emerge such as how to properly use this kind of data, or how it may be used in conjunction with RGB data to combine the spatial characteristics of both channels.

A baseline study by Janoch et al. [9] employed the popular part-based detector by Felzenszwalb et al. [6]. Taking object's parts into account, this model is a variant of the HoG algorithm [4] that combines a sliding window approach together with a representation based on histograms of edge orientations. This detector was employed by running it on depth images while treating them as intensity images. The results indicated that applying this algorithm to color images always gives better results than applying it over depth images, though it is important to note that sometimes database objects have no depth information associated with them (see Figure 1), which gives a somewhat unfair advantage to color-based algorithms. Tang et al. [27] also uses the HoG formulation, but with histograms of surface normals that are characterized by two spherical angles.

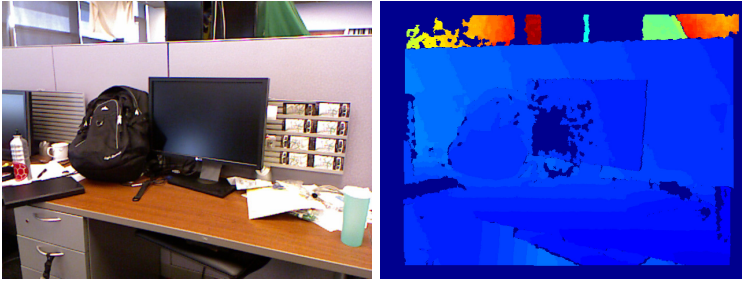


Fig. 1. An example of a color and depth image pair from the Berkeley category-level 3D object dataset [9]. The cup, bottle and monitor at the border of the color image are almost completely missing in the depth image due to a smaller field of view. Significant information is also missing from the monitor screen in the center (dark blue depicts missing depth information).

Since several previous algorithms require prior segmentation [2,21], Kim et al. [10] alleviate such requirement by detecting a small set of segmentation hypotheses. A part-based model generalized to 3D is used together with HoG features and other 3D features of the segmented objects. This scheme results in a feature vector containing both color features and 3D features, which gives better results in most categories. Depth information was also used for object size estimation [23,9] and to combine detector responses from different views [13].

Regardless of the color or depth features employed, an important issue of object detection is the treatment of object pose. A robust detector must be general enough to capture its sought-after object at different poses, and there are different ways of doing so. The naïve way, as shared by most detectors, is to rely on machine-learning classifiers to be able to generalize diverse training data. However, machine-learning methods have limitations (like any other method) and cannot always be expected to generalize well. In order to achieve better results some try and provide the learning algorithm with "easier" examples to learn from. This is done by estimating the object's pose prior to the classification phase [15], and using it to align the object to a canonical pose or to calculate features in relation to the estimation. For example, Tang et al.[27] sought detection by estimating the pose using the centroid of normal directions, but unfortunately with no improvement to detection results.

2.3 Symmetry Detection

Symmetry is a phenomenon occurring abundantly in nature. Extensive research has been done trying to detect all kinds of symmetry in both 2D [18] and 3D [17]. Complete symmetry detection (including the detection of multiple types of symmetry and across multiple objects in an image) is a hard problem due to the different types of symmetry found in nature. For this reason, research is usually focused on specific symmetries, ranging from rigid translation [29] and

reflection [16,19], to non-rigid symmetry [20] and curved glide reflection [14] (which is a combination of reflection and translation relative to a curve).

Proposed symmetry detection methods may also be classified as solving for either partial or global symmetries. While an image containing various symmetric objects is likely to contain a great deal of local symmetries, the image itself may not necessarily be globally symmetric. Partial symmetry detection entails finding the symmetric parts of the image, in contrast to global symmetry detection in which all of the image pixels are expected to participate. More formally, global symmetry, being a special case of partial symmetry [17], is characterized by a transformation that maps the entire data to itself, while for partial symmetry a sought-after transformation maps only part of the data to itself.

Considered somewhere between global and local, images of *segmented* symmetric objects may contain only object points, but under various viewpoints some of the data will have no visible symmetric counterpart. Working with this kind of data, a shape reconstruction algorithm by Thrun et al. [28] detects symmetry by employing a probabilistic model to score hypothetical symmetries. This way, several types of symmetry are found, including point reflection, plane reflection, and axial symmetry. The several types are found efficiently by taking into account the relations between symmetry types, as the existence of some symmetries entails the existence of others as well.

3 Partial Pose Invariant Detector

Considering the abundance of symmetric objects, a robust detection of symmetry may prove a valuable mid-level feature for many computer vision tasks. Here, we propose to use it for object detection in depth data, and the overview below is followed by a detailed account of the calculations.

We first observe that when a complete estimation of an object's pose is given, one may improve detector performance. The target can be aligned to match a given model, or alternatively a richer representation may be obtained using the pose estimation as a reference. Indeed, a quick and accurate estimation prior to the detection process is not an easy task. Still, seeking to benefit from such information, we propose to estimate the object's pose at least partially, by exploiting its symmetry properties.

Even though complete pose invariance is highly desired, in practice many objects rarely appear in all possible poses. For example, looking from a human's point of view, surfaces of tables are usually visible but bottoms of cars are rarely so. In such cases (and many others), even partial knowledge regarding symmetry may supply most of the pose information of objects observed from likely viewpoints. Figure 2 presents such cases for one selected class of objects.

As it happens, many objects, including almost all of the objects in both Berkeley's category dataset [9] and Washington's RGB-D Object Dataset [11], are highly symmetric. Thus, we limit our scope to working with objects with reflection symmetry over a plane and note that axial symmetrical objects are also plane symmetric [28]. Using the detected reflection symmetry plane a reference



Fig. 2. Typical examples of chair poses. If the symmetry of these chairs (in this case, about a reflection plane) is known, most pose variability in these examples may be accounted for. Indeed, when viewed from a typical viewpoint, the pose of a chair varies by rotation about an axis normal to the floor. This uncertainty is removed once the chair symmetry plane is known.

frame is constructed, and estimated surface normals [22] are represented as 2 angles based on the known partial pose. The normals are grouped inside bins calculated relative to the complete reference frame and a feature vector is constructed using this resultant histograms.

To conveniently leverage the advantages of depth data mentioned in section 2 we scan the space with a fixed-size 3D box sliding over the cloud of points in space. This is done in an efficient manner, without considering empty parts of space, places that are too far away from the camera or those containing just a small batch of nearby points. For each box the best symmetry plane passing through the box's center is found, and features are calculated using the data points inside the box. This is followed by a classification of the resultant feature vector using SVM. Since several boxes are usually classified as containing the same object, nearby detections are removed using a non-maximum suppression process. In the next sections we describe the calculations performed for every 3D box, consisting of the symmetry detection process and the computation of the features and feature vector used for classification.

3.1 Symmetry Plane Detection and the Reference Frame

Taking into account every relevant box in space greatly simplifies the symmetry plane detection task. If it contains the target object, most points inside the box are likely to belong to that object, while the number of outliers is usually not too great, a property that distinguishes range data from intensity/color data. More importantly, scanning the entire space, we are assured that some box will have its center point coincide with the sought-after symmetry plane. Since a plane can be represented with a point and a normal, only the normal remains to solve for. Be that as it may, we assume that points are generated from

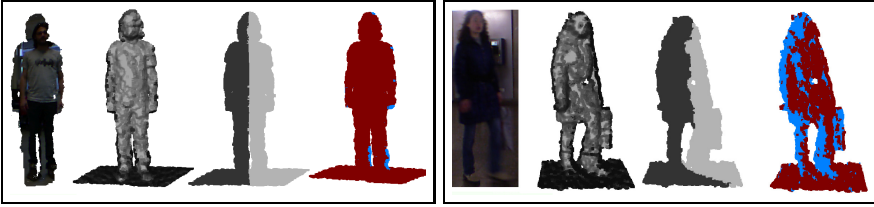


Fig. 3. Points with no visible symmetric partners are colored blue in the rightmost image of the two examples. The points are identified by the process described in the text, given the symmetry plane depicted in the 3rd image (in which points are colored according to their side of the plane). The RGB-D image for the right example is taken from the RGB-D People Dataset [26].

a perspective imaging device (from a single viewpoint), so the corresponding symmetric point of many inlier points (or even all of them) is simply not visible. These points are first identified following a scoring strategy that ranks every possible reflection plane normal. For that purpose, the 2 angles comprising the normal’s spherical representation¹ are quantized and the best pair is chosen using a score penalizing symmetric point pairs (relative to the candidate plane) with non-symmetric surface normals (in contrast to Thrun et al. [28]). The formal details follow below.

Prior to calculating a score for a candidate symmetry plane, we first deal with the inlier points without symmetric partners. Self occlusion dictates that when observing an object from one side of the symmetry plane, most of the visible inlier points that are observed will be the ones that share the same side with the camera. For the same reason, inlier points that are observed on the other side should all have visible symmetric points in the camera’s side (as portrayed by the “lady” object in Figure 3). Knowing this, we find these points on the closer side of the candidate plane with no partners on the farther side and disregard them from the score calculation. To do so we rely on surface normals, observing that a point \mathbf{x} on a surface with an estimated surface normal [22] \mathbf{n}_x is visible from viewpoint \mathbf{p}_v if [22]:

$$\mathbf{n}_x \cdot (\mathbf{p}_v - \mathbf{x}) > 0. \quad (1)$$

Consequently, we look for points whose symmetric partner has a surface normal that points away from the camera. A point \mathbf{x} with estimated normal \mathbf{n}_x is reflected over a candidate symmetry plane with center point \mathbf{p} and normal \mathbf{n}_p by:

$$\tilde{\mathbf{x}} = \mathbf{x} - 2 \cdot \mathbf{n}_p \cdot d_x, \quad (2)$$

where d_x is the signed distance between the point \mathbf{x} and the plane. Correspondingly, \mathbf{x} ’s normal is reflected as well by:

$$\tilde{\mathbf{n}}_x = \mathbf{n}_x - 2 \cdot \mathbf{n}_p \cdot d_n, \quad (3)$$

¹ A normal’s magnitude is always 1 and thus only its direction should be represented.

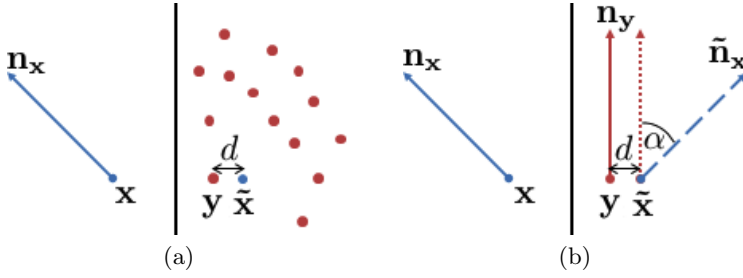


Fig. 4. Determining a reflection score for point \mathbf{x} using its reflection $\tilde{\mathbf{x}}$, the closest point \mathbf{y} is found (a), then the point is scored according to the distance d and normal difference α (b)

where \mathbf{n}_x is the normal we wish to reflect and d_n is the signed distance between the normal’s head and the candidate plane, centered at the camera’s axes origin with normal \mathbf{n}_p . Thus \mathbf{x} has no symmetric partner if:

$$\tilde{\mathbf{n}}_x \cdot (\mathbf{p}_v - \tilde{\mathbf{x}}) \leq 0. \tag{4}$$

Following that, each point is assigned a *point reflection score* that measures the "wellness" of its reflection. An observed point \mathbf{y} with normal \mathbf{n}_y closest to $\tilde{\mathbf{x}}$ (and in the same side) is found using a kd-tree data structure and the score of \mathbf{x} is determined by:

$$x_{score} = d + w \cdot \alpha, \tag{5}$$

where d is the distance between $\tilde{\mathbf{x}}$ and \mathbf{y} , α is the angle between $\tilde{\mathbf{n}}_x$ and \mathbf{n}_y (see Figure 4), and w is a weighing factor. A lower value implies good symmetry and the best plane is chosen as the one that minimizes the mean score of all the contributing points ².

In order to have a complete reference frame for the symmetry plane we endow \mathbf{n}_p with another unit length reference vector \mathbf{r} that lies *on* the plane. It is chosen to be on the verge of visibility according to equation 1, and to be directed upwards. Summarizing these constraints we get:

1. $\|\mathbf{r}\| = 1$ (\mathbf{r} is of unit length)
2. $\mathbf{r} \cdot (\mathbf{p}_v - \mathbf{p}) = 0$ (\mathbf{r} is on the verge of visibility)
3. $\mathbf{r} \cdot \mathbf{n}_p = 0$ (\mathbf{r} is on the symmetry plane)
4. $\mathbf{r} \cdot [0, 1, 0] \leq 0$ (\mathbf{r} points up³)

Therefore, we calculate \mathbf{r} with:

$$\mathbf{r} = \mathbf{n}_p \times \frac{(\mathbf{p}_v - \mathbf{p})}{\|\mathbf{p}_v - \mathbf{p}\|}, \tag{6}$$

² Points from the farther side to the camera, and closer side points for which equation 4 does *not* hold.

³ Working relative to the Kinect camera’s coordinates (in which the y axis points down) the "up" direction is defined as the negative y direction.

and complete the reference frame by calculating the third orthonormal vector:

$$\mathbf{i} = \mathbf{n}_p \times \mathbf{r}. \quad (7)$$

Examples of detected symmetry together with a complete reference frame may be seen in Figure 5 for two objects from the chair category.

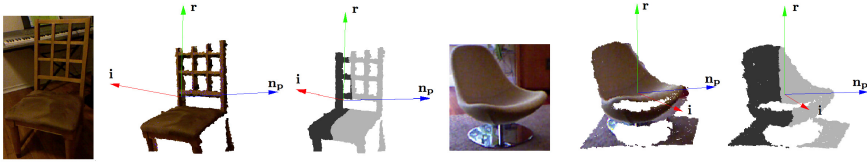


Fig. 5. The detected symmetry plane and reference frame of two chairs. The vector \mathbf{n}_p is the normal of the detected plane (blue), \mathbf{r} points up (green), and \mathbf{i} is the cross product of the two vectors (red).

3.2 Feature Vector Construction

Our feature vector for the purpose of detecting objects with varying pose is based on histograms of normals that are accumulated into angular bins that are created relative to the reference frame. Our representation of normal directions is relative only to the estimated symmetry.

For every point in the box we use a representation of the surface normal based on two angles. However, instead of representing normals with regular spherical coordinates (that will depend on \mathbf{r}) we use two angles that are independent of our choice of reference frame. Let \mathbf{x} be a cloud point with surface normal \mathbf{n}_x , and let $\bar{\mathbf{x}}$ and $\bar{\mathbf{n}}_x$ be their projections on the symmetry plane (\mathbf{p}, \mathbf{n}_p). The first angle $\theta \in [0, \pi]$ we use is the one between the normal \mathbf{n}_x and the plane normal \mathbf{n}_p :

$$\theta = \cos^{-1}(\mathbf{n}_x \cdot \mathbf{n}_p). \quad (8)$$

The second angle $\phi \in [0, 2\pi)$ in our representation is the signed angle between the projected normal $\bar{\mathbf{n}}_x$ and the vector connecting the box center \mathbf{p} with the projected point $\bar{\mathbf{x}}$:

$$\phi = \cos^{-1}\left(\frac{\bar{\mathbf{n}}_x}{\|\bar{\mathbf{n}}_x\|} \cdot \frac{\mathbf{p} - \bar{\mathbf{x}}}{\|\mathbf{p} - \bar{\mathbf{x}}\|}\right). \quad (9)$$

This is then followed by an addition of π depending on the direction of $\bar{\mathbf{n}}_x$ relative to direction vector:

$$direction = \mathbf{n}_p \cdot \frac{\mathbf{p} - \bar{\mathbf{x}}}{\|\mathbf{p} - \bar{\mathbf{x}}\|}. \quad (10)$$

These two angles supply us with a representation that depends only on the estimated symmetry plane. We selected this particular representation of surface normals because it is fixed under different poses of an object.

The normals, represented using the above two angles, are accumulated in histograms based on the spatial location of their points. In order to be robust to variations of the arbitrarily chosen \mathbf{r} vector a given box is divided into spatio-angular bins based on distance from the symmetry plane together with angular distance from \mathbf{r} , as can be seen in Figure 6. Each bin is associated with two 1D histograms for the angles described above, which are then normalized and concatenated to form the feature vector that is then classified using RBF-kernel SVM [3].

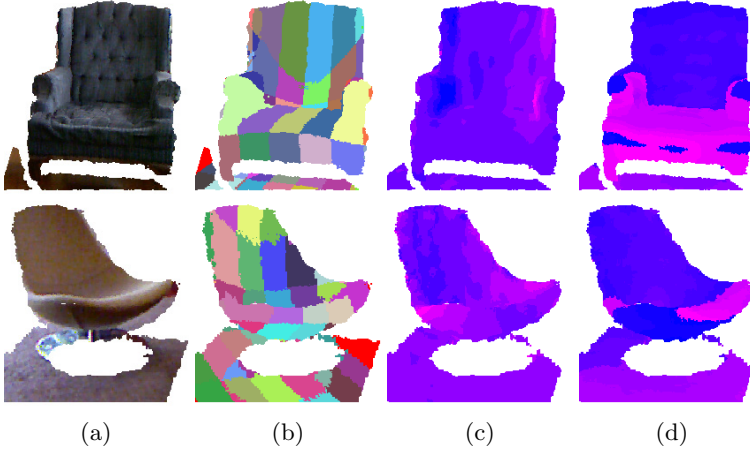


Fig. 6. Boxes are divided into spatio-angular bins (b). Each point is associated with two angles, depicted in (c) and (d).

4 Experimental Results

We evaluate our proposed approach over the Berkeley category-level 3D object dataset [9] that exhibits significant variations in terms of objects and poses. The dataset supplies basic object annotations in the form of 2D bounding boxes that we use for the evaluation of detector performance. However, our detector requires training examples in the form of fix-sized 3D bounding boxes in space. To this end we extended the provided annotations and added the 3D center point of every annotated object⁴.

Object detection algorithms incorporating depth information may be grouped in two categories, either working with depth only, or combining depth and color. The former allows to understand the strength of the depth modality in itself (for object detection), while the latter better leads to an understanding of the

⁴ Supplementary information, including our annotations, will be available at <http://www.cs.bgu.ac.il/~icvl>

role depth plays in relation to color (appearance) and the interaction between the two modalities. Seeking to properly compare the two groups of detectors, it is important to note that not only the latter contains more information, but also that depth-only detectors suffer an intrinsic disadvantage in this particular dataset, since objects' depth information is sometimes completely missing due to the smaller field of view of Kinect's depth camera compared to its color camera, or due to imaging artifacts resulting in pixels with no depth information (see Figure 1 again). For these reasons we compare our results to algorithms employing only depth information.

Detection results for the different categories in the dataset are presented in Figure 7 and compared to the depth-only baseline by Janoch et al. [9]. As can be seen, our proposed approach displays a significant improvement for all object categories other than monitors. We note that previous results obtained by algorithms using both color and depth achieve better performance but are not considered here for their reliance on different (and richer) sensory data.

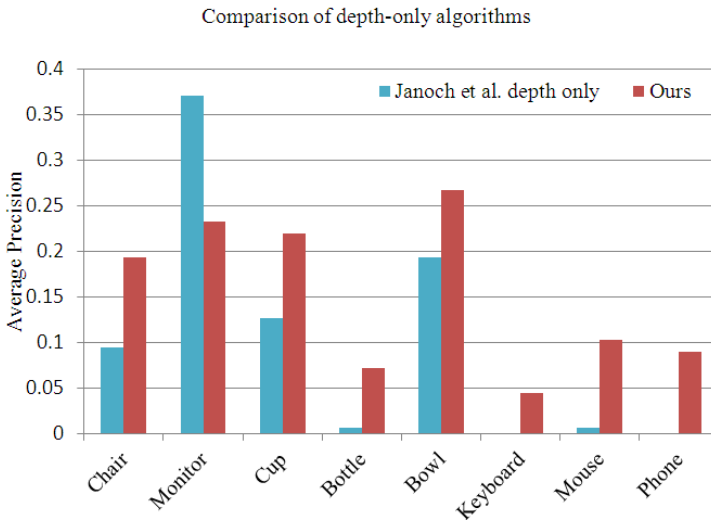


Fig. 7. Average precision of depth-only algorithms over the Berkeley category-level 3D object dataset [9]. Our method presents significant improvement for most categories except for the monitor class.

Finally, examples of true and false positives over selected categories are illustrated in Figure 8 (for chairs) and Figure 9 (for cups).

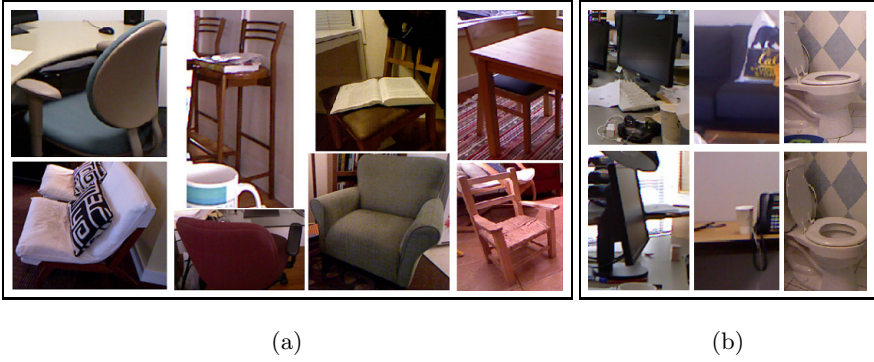


Fig. 8. Examples of true positives (a) and false positives (b) of chairs. As can be seen, chairs with different poses and different geometry are successfully detected, while scenes with chair-like geometry may lead to false detection. Such structures may be induced by couches, toilets, monitors placed on desks, or desks placed next to walls.

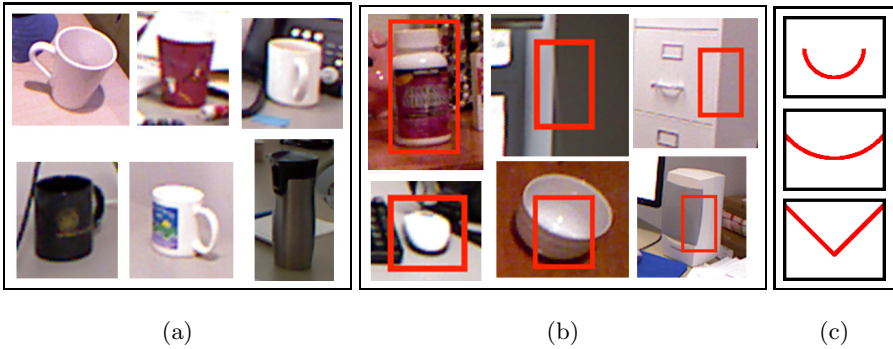


Fig. 9. Examples of true positives (a) and false positives (b) of cups. Cups with different shapes and poses are detected, and cup-like structures may lead to false detections. Red rectangles mark the precise locations of hallucinated cups, and as can be seen, may be induced by bottles, mice, bowl-parts, or corners of walls, drawers or speakers. The red curves in (c) illustrate a cross-sections of a cup, a bowl, and a corner (respectively, from top to bottom), portraying the large similarity between shapes.

5 Conclusions and Future Work

The current availability of depth cameras has made depth information easily accessible, with many possibilities for new and exciting directions of research. The unique properties of this kind of information raises the question of how to properly exploit it. Focusing on object detection from depth-only data, we addressed this question with an object detector involving pose estimation. Since a complete and accurate pose estimation would be too expensive for a sliding

window detector, we compute a partial pose based on symmetry, a property that is robust as a mid-level and global feature over object's points. The obtained symmetry, detected using surface normals, may account for most of the pose variations of many objects, and is leveraged by a specially crafted feature vector consisting of angular binning and histograms of surface normals. Our approach does not require registration and can be used when only depth information is present (as is the case for most depth cameras). Under such conditions it exhibits significant improvement in performance compared to previous depth-only detection algorithms.

Much work can be done to continue and build on this framework. The symmetry detection process can be made better and more suitable for this kind of data, in which most points are inliers, and outliers usually lie on the box's borders. Incorporating the proposed approach together with color gradients is likely to present further improvement for cases where a registered pair of color and depth images are available.

Acknowledgements. This research was funded in part by the European Commission in the 7th Framework Programme (CROPS GA no. 246252). We also thank the generous support of the Frankel fund and the ABC Robotics Initiative at Ben-Gurion University.

References

1. Aldoma, A., Vincze, M., Blodow, N., Gossow, D., Gedikli, S., Rusu, R., Bradski, G.: Cad-model recognition and 6dof pose estimation using 3d cues. In: *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 585–592 (2011)
2. Bo, L., Lai, K., Ren, X., Fox, D.: Object recognition with hierarchical kernel descriptors. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1729–1736. IEEE (2011)
3. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2(3), 27:1–27:27 (2011)
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 886–893 (2005)
5. Everingham, M., Van Gool, L., Williams, C., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. *International Journal of Computer Vision* 88(2), 303–338 (2010)
6. Felzenszwalb, P., McAllester, D., Ramanan, D.: A discriminatively trained, multi-scale, deformable part model. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2008)
7. Hinterstoisser, S., Holzer, S., Cagniard, C., Ilic, S., Konolige, K., Navab, N., Lepetit, V.: Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 858–865 (2011)
8. Hinterstoisser, S., Cagniard, C., Ilic, S., Sturm, P., Navab, N., Fua, P., Lepetit, V.: Gradient response maps for real-time detection of textureless objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(5), 876–888 (2012)

9. Janoch, A., Karayev, S., Jia, Y., Barron, J., Fritz, M., Saenko, K., Darrell, T.: A category-level 3-d object dataset: Putting the kinect to work. In: *Consumer Depth Cameras for Computer Vision* (2011)
10. Kim, B., Xu, S., Savarese, S.: Accurate localization of 3d objects from rgb-d data using segmentation hypotheses. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 886–893 (2013)
11. Lai, K., Bo, L., Ren, X., Fox, D.: A large-scale hierarchical multi-view rgb-d object dataset. In: *Proceedings of the IEEE International Conference on Robotics and Automation* (2011)
12. Lai, K., Bo, L., Ren, X., Fox, D.: Sparse distance learning for object recognition combining rgb and depth information. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 4007–4013 (2011)
13. Lai, K., Bo, L., Ren, X., Fox, D.: Detection-based object labeling in 3d scenes. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1330–1337 (2012)
14. Lee, S., Liu, Y.: Curved glide-reflection symmetry detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(2), 266–278 (2012)
15. Lin, Z., Davis, L.S.: A pose-invariant descriptor for human detection and segmentation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part IV*. LNCS, vol. 5305, pp. 423–436. Springer, Heidelberg (2008)
16. Loy, G., Eklundh, J.-O.: Detecting symmetry and symmetric constellations of features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3952, pp. 508–521. Springer, Heidelberg (2006)
17. Mitra, N., Pauly, M., Wand, M., Ceylan, D.: Symmetry in 3d geometry: Extraction and applications. In: *EUROGRAPHICS State-of-the-art Report* (2012)
18. Park, M., Lee, S., Chen, P., Kashyap, S., Butt, A., Liu, Y.: Performance evaluation of state-of-the-art discrete symmetry detection algorithms. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2008)
19. Podolak, J., Shilane, P., Golovinskiy, A., Rusinkiewicz, S., Funkhouser, T.: A planar-reflective symmetry transform for 3d shapes. *ACM Transactions on Graphics* 25(3), 549–559 (2006)
20. Raviv, D., Bronstein, A., Bronstein, M., Kimmel, R.: Symmetries of non-rigid shapes. In: *Non-rigid Registration and Tracking through Learning workshop (NRTL)*, *IEEE International Conference on Computer Vision (ICCV)*, pp. 1–7 (2007)
21. Redondo-Cabrera, C., López-Sastre, R., Acevedo-Rodriguez, J., Maldonado-Bascón, S.: Surfing the point clouds: Selective 3d spatial pyramids for category-level object recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3458–3465 (2012)
22. Rusu, R.: *Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments*. Ph.D. thesis, Technische Universitaet Muenchen, Germany (2009)
23. Saenko, K., Karayev, S., Jia, Y., Shyr, A., Janoch, A., Long, J., Fritz, M., Darrell, T.: Practical 3-d object detection using category and instance-level appearance models. In: *IEEE International Workshop on Intelligent Robots and Systems*, pp. 1817–1824 (2011)
24. Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., Moore, R.: Real-time human pose recognition in parts from single depth images. *Communication of the ACM* 56(1), 116–124 (2013)
25. Silberman, N., Fergus, R.: Indoor scene segmentation using a structured light sensor. In: *IEEE International Conference on Computer Vision Workshops, ICCV Workshops* (2011)

26. Spinello, L., Arras, K.: People detection in rgb-d data. In: IEEE International Workshop on Intelligent Robots and Systems, pp. 3838–3843 (2011)
27. Tang, S., Wang, X., Lv, X., Han, T.X., Keller, J., He, Z., Skubic, M., Lao, S.: Histogram of oriented normal vectors for object recognition with a depth sensor. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012, Part II. LNCS, vol. 7725, pp. 525–538. Springer, Heidelberg (2013)
28. Thrun, S., Wegbreit, B.: Shape from symmetry. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1824–1831 (2005)
29. Zhao, P., Quan, L.: Translation symmetry detection in a fronto-parallel view. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1009–1016 (2011)