Affordance of Object Parts from Geometric Features

Austin Myers, Angjoo Kanazawa, Cornelia Fermuller and Yiannis Aloimonos
University of Maryland College Park, MD, USA
{amyers, kanazawa, fer, yiannis}@umiacs.umd.edu

Extended Abstract

As robots begin to collaborate with humans in everyday workspaces, they will need to understand the functions of tools and their parts. To cut an apple or hammer a nail, robots need to not just know the tool's name, but they must localize the tool's parts and identify their functions. In this extended abstract, we give an overview of our work on localizing and identifying object part affordance. *We present a framework which provides 3D predictions of functional parts that can be used directly by a robot. We introduce the RGB-D Part Affordance Dataset with 105 kitchen, workshop, and garden tools. We analyze the usefulness of different features, and show that geometry is key for this problem. Finally, we demonstrate that the approach can generalize to novel object categories, so robots like PR2, Asimo, and Baxter could use tools never seen before.*

Imagine Baxter in a kitchen, trying to serve soup from a pot into a bowl. Baxter needs to find a ladle, grab the handle, dip the bowl of the ladle into the pot, and transfer the soup to the serving bowl. But what if the ladle in this kitchen has a different shape and color from the ladles that Baxter has seen before? What if Baxter has never seen any ladles at all? Today, computer vision allows robots to recognize objects from a known category, providing a bounding box around the ladle. However, in these situations Baxter needs to not just detect the ladle, but more importantly he needs to know which part of the ladle he can grasp and which part can contain the soup. As Gibson remarked, "If you know what can be done with a[n] object, what it can be used for, you can call it whatever you please" [3].

Gibson defined affordances as the latent "action possibilities" available to an agent, given their capabilities and the environment [3]. In this sense, for a human adult, stairs afford climbing, an apple affords eating, and a knife affords the cutting of another object. The last example is the most relevant to a collaborative robot, so we use the term *effective affordance* to differentiate the affordances of tools from those found in other settings. If robots could identify the effective affordances of parts, then they could use a variety of tools, even those they have never seen before.

Segmentation Based Affordance Identification. Man-made tools are typically composed of parts, where each part is a collection of *surfaces* that can provide some effective affordance. We define a surface's effective affordance by the way it comes in contact with the objects that they affect. For example, the inner surface of a coffee mug is "contain" since it contacts the liquid that it holds, while the surface of the handle is "grasp" because it can be held by a hand. Since we consider the surfaces that make up object parts, we take a segmentation based approach to affordance identification. We use a modified SLIC [1], which incorporates depth and surface normal information, to divide objects in an RGB-D image into a smaller surface fragments.

Geometric Features for Affordance Identification. Our goal is to predict the affordances of these surfaces from their visual features, such as color or depth from a Kinect sensor. We hypothesize that *there is a deep relationship between effective affordance and geometry* of a part, since the geometric and physical properties of object are closely tied to the ways they can interact with the environment. Therefore, we compare three geometric features: depth, normal, and curvature. We use a hierarchical sparse coding technique, M-HMP [2], to extract representations for each of the pixel-level feature types. Recent feature learing methods like [2, 8] have been shown to learn mid-level representations invariant to small deformations directly from data, and have achieved state-of-the-art performance on several computer vision tasks. Most importantly, a feature learning approach provides am equal footing to evaluate each of the different feature types.

Framework for Affordance Identification. Given an object in an RGB-D image we divide it into a collection of surfaces using superpixel segmentation. For each superpixel, we compute the M-HMP features of its pixels and aggregate them using max-pooling. This gives a feature vector for each surface that can be classified with a linear SVM. Finally, we refine the predictions and introduce pairwise information between segments by modeling the superpixel neighborhood graph as a conditional random field. [5].

Figure 1. Objects from the RGB-D Part Affordance Dataset. Each column shows example objects with parts that share the same affordance. The top and bottom rows show example training and testing objects for the novel category setting, respectively.

RGB-D Part Affordance Dataset. We developed a new dataset tailored to everyday tools and the affordances of their parts. The dataset contains 105 kitchen, workshop, and garden tools, and provides pixel-level affordance labels for more than 10,000 RGB-D frames covering a full 360° range of views. These objects were collected from 17 different object categories with 7 affordances: grasp, wrap-grasp, cut, contain, support, scoop, and pound. Examples of the five effective affordances are shown in figure 2. The dataset is also designed so that each affordance is represented by objects from several categories, which permits zero-shot or novel category test settings. The dataset will be available at `www.umiacs.umd.edu/~amyers/part-affordance-dataset`.

Results. We first analyze the effectiveness of different raw feature types in order to test our hypothesis that geometry is related to part affordance. As shown in figure 3, we found that geometric features significantly outperform appearance features for predicting most affordances. This differs from recent results for RGB-D object recognition, which found that visual features outperform geometric features in instance and category recognition [6].

Following these results, we evaluate our framework by testing on objects from known and novel categories. We can see from table 1 that *Geometry* (depth, normal, and curvature) is superior to *Appearance* (RGB and gray) for both known and novel settings. Even more telling, combining all features does not provide significant improvement, indicating that geometry is key for this task. While the CRF does not give a quantitative improvement, we found that it is an important step for producing an output useful to a robot.

Conclusions We introduced a novel problem of localizing and identifying part affordance, and a new dataset designed to address it. We then proposed a framework to predict the affordance of parts for objects of known and completely novel categories. Finally, we showed that geometry is critical for predicting affordance. This new dataset and the failures and success of the proposed method open many avenues for future research [7].

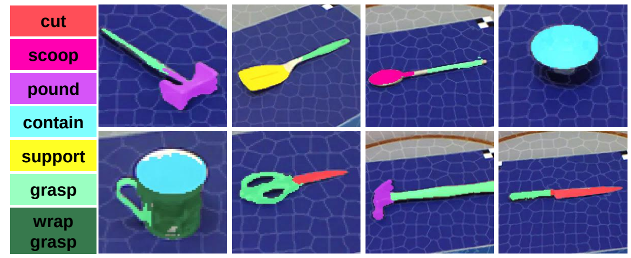

Figure 2. Example results from our framework for known category (top row) and novel category (bottom row) experimental settings.

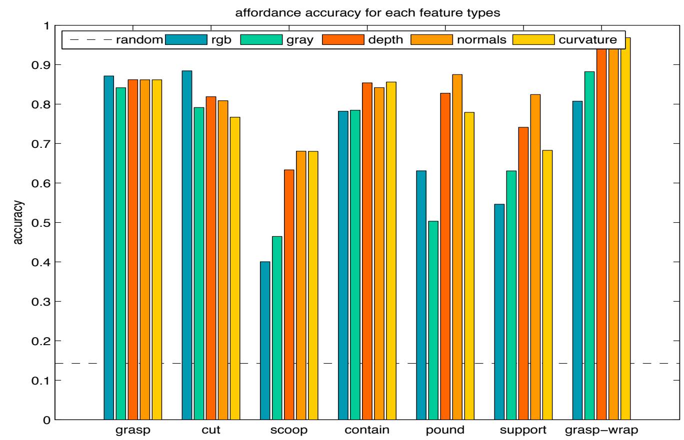

Figure 3. Comparison of different raw features for each affordance type in the known category setting.

	Appearance	Geometry	All	All + CRF
Known	73.2 ± 3.5	86.5 ± 6.6	86.2 ± 5.6	86.5 ± 5.0
Novel	46.0	63.6	64.8	64.8

Table 1. Results for known and novel settings using appearance, geometry, and all features, and finally the complete framework.

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *PAMI*, 34(11):2274–2282, 2012.

[2] L. Bo, X. Ren, and D. Fox. Multipath sparse coding using hierarchical matching pursuit. *CVPR*, 2013.

[3] J. J. Gibson. The theory of affordances. *Perceiving, Acting, and Knowing: Toward and Ecological Psychology*, 1977.

[4] A. Karpathy, S. Miller, and L. Fei-Fei. Object discovery in 3d scenes via shape analysis. *ICRA*, 2013.

[5] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *ICML*, 2001.

[6] K. Lai, L. Bo, X. Ren, and D. Fox. A large-scale hierarchical multi-view rgb-d object dataset. *ICRA*, 2011.

[7] A. Myers, A. Kanazawa, C. Fermuller, and Y. Aloimonos. Affordance of object parts from geometric features. *IROS*, 2014.

[8] R. Socher, B. Huval, B. Bhat, C. D. Manning, and A. Y. Ng. Convolutional-recursive deep learning for 3d object classification. *NIPS*, 2012.