Detecting Fine-Grained Affordances with an Anthropomorphic Agent Model

Viktor Seib(✉), Nicolai Wojke, Malte Knauf, and Dietrich Paulus

Active Vision Group (AGAS), University of Koblenz-Landau,
Universitätsstr. 1, 56070 Koblenz, Germany
{vseib,nwojke,mknauf,paulus}@uni-koblenz.de
http://agas.uni-koblenz.de

Abstract. In this paper we propose an approach to distinguish affordances on a fine-grained scale. We define an anthropomorphic agent model and parameterized affordance models. The agent model is transformed according to affordance parameters to detect affordances in the input data. We present first results on distinguishing two closely related affordances derived from *sitting*. The promising results support our concept of fine-grained affordance detection.

Keywords: Affordances · Fine-grained affordances · Visual affordance detection · Object classification

1 Introduction

We address the task of detecting affordances on a fine-grained scale in a home environment. Affordances as defined by Gibson [3], [4] inherit the concept of direct perception and the complementary nature of an agent and its environment. Whether or not direct perception can be used in computer vision is still an open debate as discussed e.g. by Şahin et al. [6] and Chemero et al. [2].

In the presented approach we exploit the complementary nature of an agent and its environment. We propose to model the agent as an anthropomorphic body and define a set of parameterized affordance models. A home or office environment for humans must reflect human body characteristics. A system equipped with these models is thus able to detect affordances in the environment.

We present first results on two closely related affordances: *sitting without backrest* and *sitting with backrest* which stem e.g. from the objects stools and chairs, respectively. Traditionally, these two affordances would be both *sitting*. Our results suggest that objects used by humans in a home environment provide distinct affordances on a fine-grained scale.

The remaining of this paper is structured as follows. A brief overview on related work is given in Sect. 2 and a detailed explanation of our method is provided in Sect. 3. Section 4 presents and Sect. 5 discusses the results that we obtained from various test objects. Finally, Sect. 6 concludes this paper and gives an outlook to our ongoing work.

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part II, LNCS 8926, pp. 413–419, 2015.
DOI: 10.1007/978-3-319-16181-5_30

2 Related Work

There have been many approaches to apply ideas coming from the theory of affordances to robotics. We shortly review some approaches exploiting only visual hints for affordance detection. Hinkle and Olson [5] propose a method that uses physical simulation to extract an object descriptor. The simulation consists of spheres falling onto an object from above. A feature vector is extracted from each object depending on where the spheres come to rest. Subsequently, objects are classified as cup-like, table-like or sitable.

A method for office furniture recognition is presented by Wünstel and Moratz [7]. Object classes are modeled explicitly in a graph structure, where nodes represent the object's parts and edges the spatial distances of those parts. Affordances are used to derive the spatial arrangement of the object's components.

Bar-Aviv and Rivlin [1] use an embodied agent to classify objects. The object in question is moved to a virtual simulation environment where the compatibility of different agent poses with the object is tested. The object is assigned the label of the hypothesis with the highest score.

Similar as Wünstel and Moratz [7] we use a plane segmentation approach in our method. However, we encode the spatial information needed for affordance detection in an anthropomorphic agent model rather than creating explicit object models. Contrary to Bar-Aviv and Rivlin [1] who also use an embodied agent, our method operates directly on the data. We do not segmented and move the objects to a simulation environment where they are tested to belong to different classes. In our case, segmentation is a direct consequence of the detected affordances.

3 Model Definitions for Fine-Grained Affordance Detection

In this section we describe our method of detecting fine-grained affordances with an anthropomorphic agent model. Our approach is based solely on visual data.

3.1 Agent and Affordance Models

Our anthropomorphic body model is defined as a directed acyclic graph $\mathcal{H}$ (Fig. 1). In this graph, nodes represent joints in a human body and edges represent parameterized spatial relations between these joints. The nodes contain information on how the joints can be revolved without harming the human. The environment $\mathcal{E}$ is a set of features. So far, we limit the features to arbitrarily oriented planes that are segmented from the input data.

A fine-grained affordance is a property of an affordance that specializes the relation of an agent and its environment. We take the *sitting* affordance as an example. The affordance *sitting* is a generalization of more precise relations that an agent and its environment form. In this paper, we demonstrate our ideas by distinguishing between the fine-grained affordances *sitting without backrest* and *sitting with backrest*.

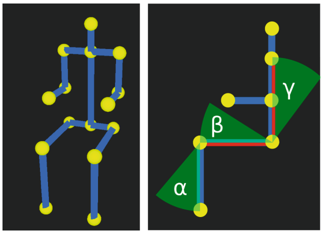

Fig. 1. The anthropomorphic agent model: nodes are depicted in yellow, edges in blue. A perspective view of the model in a sitting pose is shown on the left. This pose serves as the initial body pose for affordances derived from *sitting*. Control areas (red) that must be supported by features as well as relevant joint limits (green) for the *sitting with backrest* affordance are displayed on the right.

Fig. 2. The top row shows example objects from our evaluation. Chairs and stools served for the two fine-grained affordances. The bottom row presents affordance detection results: the *sitting with backrest* affordance is shown in green, whereas the *sitting without backrest* affordance is shown in blue.

3.2 Detecting Affordances

The algorithm used for affordance detection is outlined in Alg. 1. It operates on single scene views from an RGB-D camera. The affordance models f_1 and f_2 denominate the *sitting without backrest* and *sitting with backrest* affordances, respectively. First, plane segmentation on the input point cloud $\mathcal{P}$ is performed. Then, all horizontal planes from the abstract view of the environment $\mathcal{E}$ are tested to comply with the agent model $\mathcal{H}$ and the affordance model f_1 as described in Sec. 3.3. Every plane that affords sitting for the given agent is added to the set S of sitable planes. Then, for each sitable plane s vertical planes in close proximity are found. Each of the vertical planes is again tested to comply with the agent and the affordance models. If the sitable plane s and the vertical plane v together afford f_2 for the given agent, both planes are added to the output point cloud $\mathcal{P}_2$. Otherwise, the sitable plane s is added to the output point cloud $\mathcal{P}_1$ which contains points for the affordance f_1. Thus, the algorithm additionally provides a segmentation of the found affordances. Please note that in Fig. 2 the bounding box around s was extended to the ground plane and all points inside this bounding box were added to $\mathcal{P}_2$ and $\mathcal{P}_1$ for visualization purposes.

3.3 Checking Model Parameters

In Alg. 1 model checking is carried out in two cases. First, to assure that a plane p is sitable and second to assure that a plane v can support the agent's back while it is seated on p.

By varying the angle parameters α and β in the sitting affordance with the constraint that the agent's feet always touch the floor a valid range for the height of the sitting plane is found. Similarly, for the plane v the angle γ is varied to check whether the sitting agent can make use of it.

The dimensions for both planes are directly derived from the agent model. They are given by the body width, the length of the thigh and the height of the back, respectively. Since the size of the planes does not have to match the model proportions exactly to allow sitting or back support, the size is considered valid if it is between the D_{min} and D_{max} percentage parameters of the affordance. For example, for a model width of 0.4 m and $D_{min} = 0.7$ and $D_{max} = 1.3$, the allowed plane sizes would be between 0.28 m and 0.52 m.

Algorithm 1 Fine-grained Affordance Detection.

Require: Point cloud $\mathcal{P}$, Affordance models f_1, f_2, Agent model $\mathcal{H}$
Ensure: Point cloud with segmented affordances $\mathcal{P}_1$ and $\mathcal{P}_2$
  $\mathcal{E} \leftarrow segmentPlanes(\mathcal{P})$
  $S \leftarrow \emptyset$
  for all horizontal planes $p \in \mathcal{E}$ **do**
    if $supportsModels(p, \mathcal{H}, f_1)$ **then**
      $S \leftarrow S \cup p$
    end if
  end for
  for all $s \in S$ **do**
    $V \leftarrow$ vertical planes $\in \mathcal{E}$ close to s
    if $supportsModels(v, \mathcal{H}, f_2)$ **and** v is biggest plane $\in V$ that supports the models **then**
      $\mathcal{P}_2 \leftarrow \mathcal{P}_2 \cup v$
      $\mathcal{P}_2 \leftarrow \mathcal{P}_2 \cup s$
    else
      $\mathcal{P}_1 \leftarrow \mathcal{P}_1 \cup s$
    end if
  end for

4 Experiments and Results

For our experiments we acquired data from 17 different chairs and 3 stools to represent fine-grained affordances. From these data, we extracted 247 different views of the chairs and 47 different views of the stools. Example views of these objects are shown in Fig. 2. Additionally, negative data (i.e. data without the two affordances) from 9 different furniture objects was obtained and 109 views of these objects extracted. Negative data includes objects like a bed, desks, tables, dressers and a heating element. The whole evaluation dataset contains 403 scene views with 294 positive and 109 negative data examples.

The influence of the five parameters (the angle parameters α, β, γ and the size range parameters D_{min}, D_{max}) was tested with 59 different parameter sets. In the first round the parameters were varied systematically over a wide range to obtain 35 different configurations for evaluation. For the second round we inspected the best parameters from the first round and created 24 additional configurations close to the best configurations from the first round. As Hinkle and Olson [5] we included the F-measure, a harmonic mean between precision and recall, in our evaluation. Precision, recall and F-measure for the second round of experiments are shown in Fig. 3. Best results for both fine-grained affordances are shown in Tab. 1, while the best parameter values are presented in Tab 2.

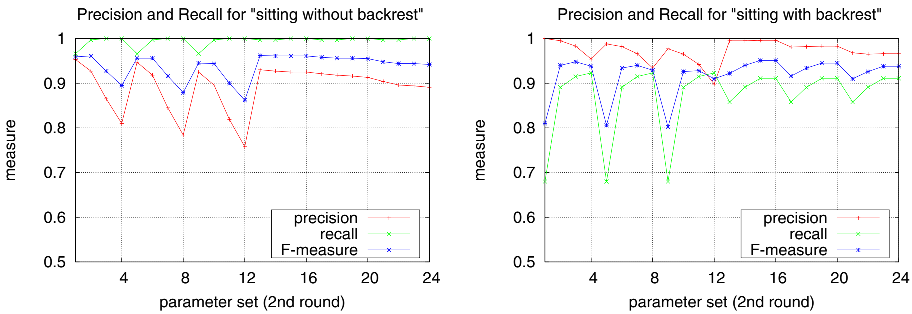

Fig. 3. Illustration of the precision and recall for both fine-grained affordances over all tested parameter sets in the second round of experiments.

Table 1. Best F-measure values for both affordances. In each line also the result of the other affordance is shown.

	sit. w/o backrest	sit. with backrest
best w/o backr.	0.962	0.922
best with backr.	0.961	0.951

Table 2. Affordance model parameters that result in highest F-measure values for the detection of both fine-grained affordances

α, β	γ	D_{min}	D_{max}
30°	35°-40°	0.5	1.4-1.6

5 Discussion

For the *sitting without backrest* affordance (in our test cases derived from the stool objects) the quality of the results was best for α and β between 20° and 40°. As is shown in Fig. 1 these parameters change the angles in the agent's legs. With the constraint that the agent's feet always touch the ground for comfortable sitting, α and β directly influence the allowed heights of the sitting planes. We observed a drop of performance for values higher than 40°. This is due to numerous planes in the datasets that are of low height, but otherwise would allow sitting. Also, if D_{min} is chosen to be only little restrictive (below 0.5) too many small planes and clutter are considered "big enough" for sitting, resulting in a drop of precision. On the other hand, D_{max} has only a moderate effect.

The *sitting with backrest* affordance is additionally influenced by the parameter γ for the inclination of the backrest. For γ, higher values than 40° cause many false positives. The additional effect of D_{min} and D_{max} include the valid dimensions for the size of the backrest that is compared with the agent's back. Again, D_{min} has more significant effects on the results, while D_{max} does not seem to have any effect at all for values higher than 1.6.

The employed parameters influence the results in many different ways. However, as shown in Tab. 1 and Tab. 2 parameters exist that allow high detection rates for both fine-grained affordances while at the same time limiting the number of false negative detections. These first results strongly support our approach of fine-grained affordance detection.

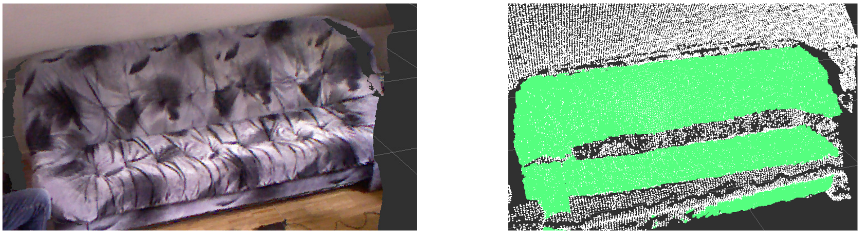

Fig. 4. Detection example of the fine-grained affordance *sitting with backrest* on a sofa with parameters $\alpha = \beta = 30°$, $\gamma = 35°$, $D_{min} = 0.5$ and $D_{max} = 5$. The original image (left) and the detection result in green (right) are shown.

The presented approach of fine-grained affordance detection originally stems from an algorithm to acquire hints to whether or not a stool or chair is present in the input data. Thus, our approach is tailored to this use case. However, the presented method needs to be further generalized to include the detection of fine-grained affordances present on other sitting furniture like sofas. To this point, to detect affordances on sofas, the model parameters need to be altered: D_{max} has to be set to higher values to support wider planes (Fig. 4).

6 Conclusion and Outlook

In this paper we presented an approach to detect affordances on a fine-grained scale by applying an anthropomorphic agent model and affordance models. In its current state our system is able to differentiate between two fine-grained affordances. The high values of the F-measure of 0.956 supports our approach of fine-grained affordance detection.

We continue our work in the two following aspects. First, our current algorithm is feature-centered as we initially detect features (planes) to create an abstract environment representation. However, we expect significant improvement if the agent model is directly fitted into the data (agent-centered approach). This would not only decrease the influence of the plane size parameters, but also allow detecting fine-grained affordances on mixed objects (e.g. a stool without backrest standing close to a wall that can support an agent's back while seated).

Second, we plan to evaluate our approach on a larger test set and include more fine-grained affordances that can be detected with a sitting pose of the agent (e.g. *sitting with armrest* and *sitting in front of a table*). An open question is also how an anthropomorphic agent model can be exploited to detect more fine-grained affordances from different body poses than sitting. As an example for a lying body pose the fine-grained affordances *lying flat* and *lying with pillow* can be distinguished. Fine-grained affordances without a body pose, but with similar actions include knobs attached to drawers and doors that can be *pulled open* or *pulled open while rotating* (about the hinge). We are currently looking for more examples for both cases (with and without body poses) to generalize and formalize our approach of fine-grained affordances.

References

1. Bar-Aviv, E., Rivlin, E.: Functional 3d object classification using simulation of embodied agent. In: BMVC, pp. 307–316 (2006)
2. Chemero, A., Turvey, M.T.: Gibsonian affordances for roboticists. Adaptive Behavior **15**(4), 473–480 (2007)
3. Gibson, J.J.: The concept of affordances. Perceiving, Acting, and Knowing, 67–82 (1977)
4. Gibson, J.J.: The ecological approach to visual perception. Routledge (1986)
5. Hinkle, L., Olson, E.: Predicting object functionality using physical simulations. In: Proc. of IROS 2013, pp. 2784–2790. IEEE (2013)
6. Şahin, E., Çakmak, M., Doğar, M.R., Uğur, E., Üçoluk, G.: To afford or not to afford: A new formalization of affordances toward affordance-based robot control. Adaptive Behavior **15**(4), 447–472 (2007)
7. Wünstel, M., Moratz, R.: Automatic object recognition within an office environment. In: CRV, vol. 4, pp. 104–109 (2004)