# Detecting Fine-grained Affordances with an Anthropomorphic Agent Model

Viktor Seib, Nicolai Wojke, Malte Knauf, Dietrich Paulus

Active Vision Group (AGAS), University of Koblenz-Landau
Universitätsstr. 1, 56070 Koblenz, Germany
{vseib, nwojke, mknauf, paulus}@uni-koblenz.de
http://robots.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 preliminary 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 in a home environment by employing an anthropomorphic body model. Affordances as defined by Gibson [1], [2] 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. [3] and Chemero et al. [4].

In the presented approach we concentrate on the complementary nature of an agent and its environment. The environment in this case is a typical home or office with typical furniture. Consequently, we propose to model the agent as an anthropomorphic body and define a set of parameterized affordance models. The idea behind these models is that a home or office environment for humans must reflect anthropomorphic, i.e. human body characteristics. A system equipped with these models is thus able to detect affordances for different body poses in the environment. We analyze the input data and search for affordances by applying the anthropomorphic agent model and segment the data supporting the affordance.

We present preliminary 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*. However, we aim at distinguishing affordances on a finer-grained scale.

Our results suggest that objects used by humans in a home environment provide distinct affordances on a fine-grained scale. Segmenting the fine-grained affordance data provides a first hint on the actual object behind the data.

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.

## 2 Related Work

There have been many approaches to apply ideas coming from the theory of affordances to robotics. Since affordances represent action possibilities that an environment offers to a specific agent, affordance related research in robotics mainly focuses on topics such as action planning, effect estimation and affordance recognition and learning. Different techniques are applied to reach these goals. Kjellström et al. [5] and Lopes et al. [6] propose to infer affordances by imitation. A robot observes humans using objects in different types of actions. The goal is to enable robots to use previously unknown object classes and apply this correctly with respect to the functionality. Affordance learning through interaction is addressed by Montesano et al. [7] and Ridge et al. [8]. These approaches have in common that a robotic arm interacts with the environment while monitoring the effects of the interaction with a camera.

Stark et al. [9] present a system for detection of functional object classes based on visual hints on object affordances. Affordance hints are acquired from observed human-object interaction, where the interaction region serves for feature extraction. A modified variant of the Implicit Shape Model [10] is used for training and functional object category detection.

Castellini et al. [11] propose to use affordances for object recognition. In their approach visual features of objects and kinematic features of a hand while grasping the object is recorded. Castellini et al. show that their results are enhanced when both, kinematic features of the hand and visual information, is included in the classification process.

Other approaches exploit only visual hints for affordance detection, but apply them in very different ways. Hinkle and Olson [12] 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 [13]. 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. The 3D data is cut into horizontal slices and 2D segmentation is performed

within each slice. The segmentation results are classified as object parts (planes) and matched to the object models.

Bar-Aviv and Rivlin [14] 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. For each object hypothesis and agent pose a score is computed. The object is assigned the label of the hypothesis with the highest score. Bar-Aviv and Rivlin discuss many examples, however they do not present a detailed evaluation.

Similar as Wünstel and Moratz [13] we use a plane segmentation approach in our method. However, we do not restrict the data to 2D. Further, 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 [14] 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 approach of detecting fine-grained affordances using an anthropomorphic agent model. Our approach is based on visual data only. We do not record kinematic data of an agent, neither do we detect affordances by interaction. As pointed out by Hinkle and Olson [12] detection by interaction would only reveal affordances that afford certain actions to the interacting agent. A robot not capable of sitting would thus not be able to discover sitting affordances. In our definition we rather employ the *observer's* view as introduced by Şahin et al. [3]. While the environment is being observed by a robot equipped with certain sensors, the system is looking for affordances that afford actions to a predefined model. In our case this predefined model is an anthropomorphic agent.

### 3.1 Agent and Affordance Models

There are several diverse definitions of affordances that differ in certain aspects. Usually, affordances are defined as relations between an agent and its environment [1], [3], [4]. Since these two entities are crucial for affordances, we start with their definitions.

For our purpose of affordance detection we define the agent as an anthropomorphic body model $\mathcal{H}$ (Fig. 1). The model is defined as a directed acyclic graph. In this graph, nodes represent joints in a human body and edges represent parameterized spatial relations between these joints. The spatial relations correspond to average human body proportions. The nodes contain information on how the joints can be revolved without harming the human.

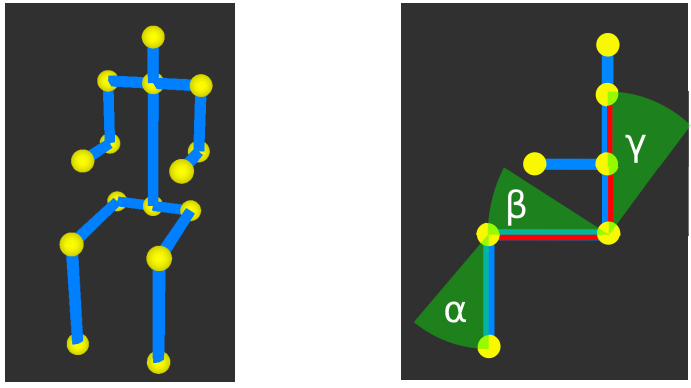


**Fig. 1.** The anthropomorphic agent model: nodes are depicted in yellow, edges in blue. A perspective view of the model in a sitting pose that serves as the initial body pose for affordances derived from sitting is shown on the left. Relevant joint limits (green) and control areas (red) for the *sitting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

The environment $\mathcal{E}$ is a set of features. A feature can be anything in the environment that is visually perceived. So far, we limit the features to arbitrarily oriented planes that are segmented from the input data. A plane consist of 3D points with an oriented bounding box determined from eigenvalue decomposition. Further, a normal vector is assigned to each plane by the eigenvector corresponding to plane's smallest eigenvalue.

If the combination of the agent and the environment affords a certain behavior $i$ than the affordance

$$\mathcal{A}_i = (\mathcal{H}, \mathcal{E})_i \tag{1}$$

is present. So far, we only provided own definitions of the agent $\mathcal{H}$ and the environment $\mathcal{E}$. In the next step, we extend this definition to fine-grained affordances.

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*. Other specializations of *sitting* include *sitting with armrests* and *sitting in front of a table*. The latter means that the agent's arms reach a plane to perform other tasks upon it. Note that some of these fine-grained affordances include others. For example, if the environment $\mathcal{E}$ affords *sitting with backrest* to the agent $\mathcal{H}$ it must necessarily afford *sitting without backrest* as well. That is to say, the agent can choose not to use the backrest while seated.

We thus extend the above definition by the concept of fine-grained affordances. The fine-grained affordances

**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 $\mathcal{P}_2$

```
    E ← segmentPlanes(P)
    S ← ∅
    for all horizontal planes p ∈ E do
      if supportsModels(p, H, f1) then
 5:     S ← S ∪ p
      end if
    end for
    for all s ∈ S do
      V ← vertical planes ∈ E close to s
10:   if supportsModels(v, H, f2) and v is biggest plane ∈ V that supports the models
      then
        P2 ← P2 ∪ v
        P2 ← P2 ∪ s
      else
        P1 ← P1 ∪ s
15:   end if
    end for
```

$$\mathcal{A}_{i|J} = (\mathcal{H}, \mathcal{E}, \mathcal{F}_i)_i \tag{2}$$

are present, if the environment $\mathcal{E}$ affords the behavior $i$ with specializations $J$ to the agent $\mathcal{H}$. Here, $\mathcal{F}_i$ is a list of fine-grained affordance models derived from $\mathcal{A}_i$. Further, $J$ is a list of unique identifiers for fine-grained affordances $f_j \in \mathcal{F}_i$ with $j \in J$. Each $f_j \in \mathcal{F}_i$ is a tuple $f_j = (C, L)$ where $C$ is a list of contact areas (Fig. 1) that need to be supported by the features in $\mathcal{E}$ and $L$ are joints in $\mathcal{H}$ and their allowed transformations.

Coming back to the example of the *sitting* affordance, a stools affords

$$\mathcal{A}_{\text{sitting}|\{\text{w/o backrest}\}} = (\mathcal{H}, \mathcal{E}, \{f_{\text{w/o backrest}}, f_{\text{with backrest}}\})_{\text{sitting}},$$

whereas a chair with a backrest affords

$$\mathcal{A}_{\textit{sitting}|\{\text{w/o backrest, with backrest}\}} = (\mathcal{H}, \mathcal{E}, \{f_{\text{w/o backrest}}, f_{\text{with backrest}}\})_{\text{sitting}}.$$

The initial pose for fine-grained affordances derived from the general affordance *sitting*, as well as relevant joint limits and control areas are presented in Fig. 1.

### 3.2 Detecting Affordances

Our method is applied to 3D depth data from an RGB-D camera. We do not use full 3D models, but rather single depth data views of a test scene. This data intuitively corresponds to what an agent perceives of the environment without

**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.

further knowledge about the structure of the observed objects. The algorithm used for affordance detection is outlined in Alg. 1. For better readability we replaced $f_{\text{w/o backrest}}$ by $f_1$ and $f_{\text{with backrest}}$ by $f_2$. First, plane segmentation on the input point cloud $\mathcal{P}$ is performed. In a subsequent step, all horizontal planes that represent an abstract view of the environment $\mathcal{E}$ containing all planes 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, apart from detection both point clouds, $\mathcal{P}_2$ and $\mathcal{P}_1$, 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 $\alpha$ and $\beta$ in the sitting affordance with the constraint that the agent's feet always touch the floor a valid range for the height

**Fig. 3.** Examples of negative data used in the evaluation. Top row: bed and two different dressers. Bottom row: shelf, heating element and a desk.

of the sitting plane is found. Similarly, for the plane $v$ the angle $\gamma$ is varied to check whether the sitting agent can make use of it.

The dimensions for both planes are directly derived from the anthropomorphic 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.

## 4 Experiments and Results

We conducted our experiments on data that was acquired in our lab. Data acquisition was performed with an RGB-D camera that was moved around an object and roughly pointed at that object's center. In total, we acquired data from 17 different chairs and 3 stools to represent the two 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. Example views of negative data is presented in Fig. 3. The whole evaluation dataset contains 403 scene views with 294 positive and 109 negative data examples.

**Table 1.** Parameter sets used for the first round of evaluation. Five different ranges for $D_{min}$, $D_{max}$ were chosen with decreasing constraints on allowed plane sizes. Each of the five ranges was combined with 7 different values for $\alpha$, $\beta$, $\gamma$

| parameter set | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 ... 14 | 15 ... 21 | 22 ... 28 | 29 ... 35 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$, $\beta$, $\gamma$ [degrees] | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 20 | 25 ... 50 | 20 ... 50 | 20 ... 50 | 20 ... 50 |
| $D_{min}$ | 0.8 | | | | | | | 0.7 | | 0.6 | 0.5 | 0.4 |
| $D_{max}$ | 1.1 | | | | | | | 1.2 | | 1.3 | 1.4 | 1.5 |

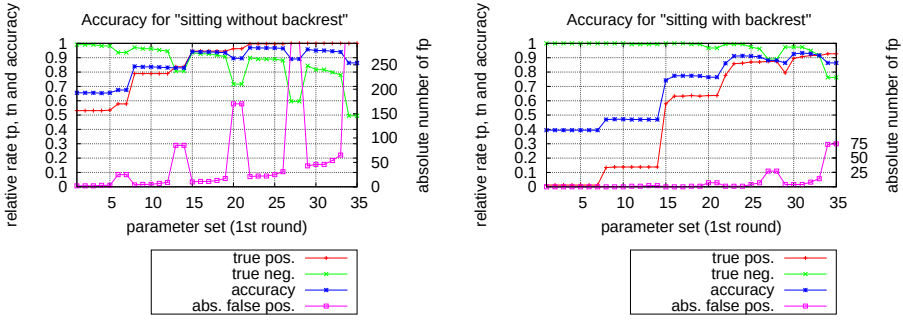


**Fig. 4.** Illustration of fine-grained affordance detection results for *sitting without backrest* (left) and *sitting with backrest* (right) over all parameter sets. In both plots, the relative number of true positive (tp) and true negative (tn) detections is shown. These measures are summarized in the accuracy measure. Additionally, each plot shows the absolute number of false positive (fp) detections on the right vertical axis. The axis is scaled to the absolute number of true positive scene views in the dataset: 297 (left) and 247 (right).

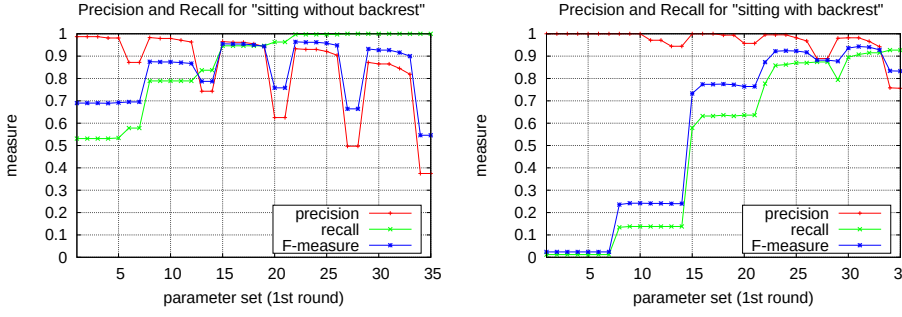


**Fig. 5.** Illustration of the precision and recall for both fine-grained affordances over all tested parameter sets. The F-measure is used to summarize precision and recall.

Our affordance models have five parameters: the angle parameters $\alpha$, $\beta$, $\gamma$ and the size range parameters $D_{min}$, $D_{max}$. We performed two rounds of experiments. 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. The purpose of the first round was to find reasonable parameters in general. In this case, all angle parameters were assigned the same values and all size range parameters change mutually by 0.1 and -1.0 respectively. In contrast, for the second round relevant parameters were changed individually.

The parameter sets for the first round are shown in Tab. 1. We computed precision and recall for both fine-grained affordances for each parameter set. As Hinkle and Olson [12] we additionally included the F-measure in our evaluation. The F-measure is defined as

$$\text{F-measure} = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

and represents a harmonic mean between precision and recall. These three measures are presented in Fig. 5. Additionally, we present the accuracy and relative number of true positives (tp) and true negative (tn) detections, as well as the absolute number of false positives (fp) in Fig. 4. Please note that the *sitting with backrest* affordance depends on the presence of the *sitting without backrest* affordance. Thus, the number of true positives for *sitting without backrest* also includes objects that expose the *sitting with backrest* affordance.

Best results in terms of accuracy and F-measure (i.e. over 0.9) for the affordance *sitting without backrest* were obtained with $\alpha$, $\beta$ and $\gamma$ between 20° and 40°, while the performance drops significantly at higher values. The variation of $D_{min}$ and $D_{max}$ revealed best results in a range between 0.6 and 0.4 for $D_{min}$ and between 1.3 and 1.5 for $D_{max}$. For *sitting with backrest* angle parameters between 25° and 40° seem appropriate, while the negative impact of higher values is not as severe as for the *sitting without backrest* affordance. However, high variations in the quality of the results are observed with varying size range parameters. Again, best values are similar to the *sitting without backrest* affordance: 0.5 and 0.4 for $D_{min}$ and 1.4 and 1.5 for $D_{max}$.

Building on these results, we selected 24 additional parameter sets. For this second round of experiments we focused on varying individual parameters. The main difference between the two fine-grained affordances investigated here is the inclusion or exclusion of the backrest with the corresponding parameter $\gamma$. We therefore varied $\gamma$ independently of the other two angle parameters. Further, we investigated the effect of a fixed $D_{min}$ while $D_{max}$ is varied and vice versa.

The chosen parameter configurations for the second round of experiments are shown in Tab. 2. Precision, recall and F-measure for these parameter sets are shown in Fig. 6.

The results of the second round of experiments show a strong influence of $D_{min}$ on the *sitting without backrest* affordance, while $D_{max}$ shows only little influence. The number of false positives increases dramatically $D_{min}$ below 0.5 causing a significant drop in precision. High values for parameter $D_{min}$ cause

**Table 2.** Parameter sets used for the second round of evaluation. *gamma* is always incremented by 5°, $D_{min}$ decremented by 0.1 and $D_{max}$ incremented by 0.1

| parameter set | 1 ... 4 | 5 ... 8 | 9 ... 12 | 13 ... 16 | 17 ... 20 | 21 ... 24 |
|---|---|---|---|---|---|---|
| $\alpha$, $\beta$ [degrees] | 30 | 35 | 40 | 30 | 35 | 40 |
| $\gamma$ [degrees] | 25 ... 40 | 25 ... 40 | 25 ... 40 | 25 ... 40 | 25 ... 40 | 25 ... 40 |
| $D_{min}$ | 0.6 ... 0.3 | 0.6 ... 0.3 | 0.6 ... 0.3 | 0.5 | | |
| $D_{max}$ | 1.5 | | | 1.4 ... 1.7 | 1.4 ... 1.7 | 1.4 ... 1.7 |

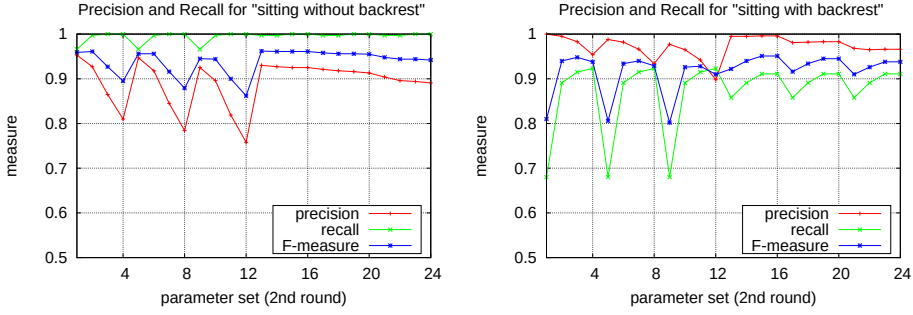


**Fig. 6.** Illustration of the precision and recall for both fine-grained affordances over all tested parameter sets in the second round of experiments. The F-measure is used to summarize precision and recall. Note the modified range of the vertical axis.

**Table 3.** Summary of best parameter sets in both evaluation rounds. F-measure values for both affordances as well as their mean is shown

| first round | param. set | sit. w/o backr. | sit. with backr. | mean |
|---|---|---|---|---|
| best w/o backr. | 22 | 0.964 | 0.873 | 0.918 |
| best with backr. | 31 | 0.927 | 0.943 | 0.935 |
| best mean | 24 | 0.962 | 0.924 | 0.943 |

| second round | param. set | sit. w/o backr. | sit. with backr. | mean |
|---|---|---|---|---|
| best w/o backr. | 13 | 0.962 | 0.922 | 0.942 |
| best with backr. and best mean | 15, 16 | 0.961 | 0.951 | 0.956 |

many false negatives for the *sitting with backrest* affordance and a steep drop of the recall. However, $\gamma$ and $D_{max}$ have only moderate influence. Further, choosing values higher than 1.6 for $D_{max}$ does not seem to affect the results.

Table 3 summarizes the best results achieved in both rounds of experiments. Parameter sets needed to obtain highest F-measure values for the detection of both fine-grained affordances are shown in Tab 4.

**Table 4.** This table shows affordance model parameters that result in highest F-measure values for the detection of both fine-grained affordances

| $\alpha$, $\beta$ | $\gamma$ | $D_{min}$ | $D_{max}$ |
|---|---|---|---|
| 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 in the first round of evaluation was best for $\alpha$ and $\beta$ between 20° and 40° (Fig. 4 and Fig. 5). 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, $\alpha$ and $\beta$ directly influence the allowed heights of the sitting planes. The observed drop of the performance for higher values than 40° is due to numerous planes in the datasets that are of low height, but otherwise would allow sitting. The effect of less restrictive parameters for allowed plane sizes ($D_{min}$, $D_{max}$) is not as severe as of the angle parameters. Again, an increase of false positives is observed in Fig. 4, however most of them are caused by the angle parameters. The second round of experiments focused on separate variations of the plane size parameters. It revealed a strong influence on the $D_{min}$ parameter. Again, if its value 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 (Fig. 6). On the other hand, $D_{max}$ has only a moderate effect.

The *sitting with backrest* affordance is additionally influenced by the parameter $\gamma$ for the inclination of the backrest. As for the *sitting without backrest* affordance, higher values than 40° cause many false positives. However, in this case their number is not as numerous. While in the former case the number of false positives forms two high peaks (Fig. 4) that go beyond the presented scale (297 and 489 compared to 297 positive data samples), in the case of *sitting with backrest* their number only reaches 75 (compared to 247 positive data samples). 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 beyond the value 1.6.

To sum up, the employed parameters influence the results in many different ways. Most significant is the negative effect of $\alpha$ and $\beta$ above 40° and of $D_{min}$ below 0.5. Luckily, as shown in Tab. 3 and Tab. 4 parameter ranges exist that allow high detection rates for both fine-grained affordances while at the same time limiting the number of false negative detections. The preliminary results are an F-measure of 0.961 for *sitting without backrest* and an F-measure of 0.951 for *sitting with backrest*. These values were achieved with the same parameters for both affordances and strongly support our approach of fine-grained affordance detection.

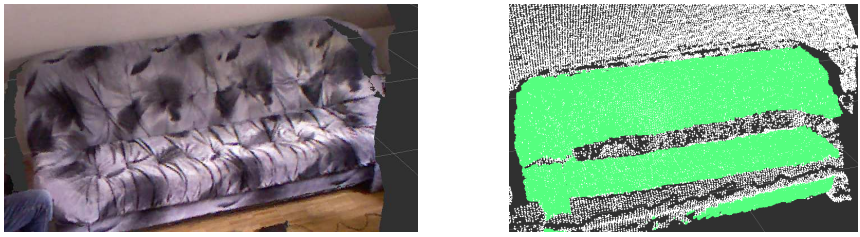

**Fig. 7.** Detection example of the fine-grained affordance *sitting with backrest* on a sofa. The parameters used were $\alpha = \beta = 30°$, $\gamma = 35°$, $D_{min} = 0.5$ and $D_{max} = 5$. Note the modified high value for $D_{max}$. The original image is shown on the left. The right image shows the detection result (green) on the input point cloud.

Our current detection algorithm is feature-centered. We first detect features (in this case planes) to create an abstract environment representation $\mathcal{E}$. However, we expect significant improvement if the model matching does not rely on segmented planes, but instead fits the agent $\mathcal{H}$ directly into the data (agent-centered approach). Although using segmented planes significantly decreases the search space for affordances, we plan to omit the plane segmentation in favor of more robust detection results. This would not only decrease the influence of the plane size parameters $D_{min}$, $D_{max}$, 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).

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. 7). While in principle it is possible to perform multiple subsequent detections with different parameters, a direct detection is preferable. Further, on soft furniture plane segmentation is more likely to fail due to the uneven surface of the sofa. This is another motivation for us to reformulate our approach to be agent-centered instead of feature-centered.

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 further formalize our approach of fine-grained affordances.

## 6 Conclusion and Outlook

In this paper we presented on 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. To define these models, we used several heuristic parameters as mentioned in Sec. 3.3. These values greatly influence the performance of the system as was shown in Sec. 4 with various parameter sets. The high values of the F-measure of 0.956 supports our approach of fine-grained affordance detection.

Our ongoing work includes improvements of the following aspects of the proposed approach. We plan to evaluate our approach on a larger test set including more pieces of furniture. Currently, we are also working on directly fitting the anthropomorphic agent model into the data without relying on plane segmentation. This will not only improve the detection results, but also make our system more robust to parameter variations. Further, we will include more fine-grained affordances that can be detected with a sitting pose of the agent. These include *sitting with armrest* and *sitting in front of a table*. We plan to investigate how our agent model can be generalized to find more affordances from different body poses. Finally, we will train the implicit shape model approach that we extended to work with 3D data to be able to classify the segmentation results of our algorithm to obtain object categories from affordances.

## References


1. Gibson, J.J.: The concept of affordances. Perceiving, acting, and knowing (1977) 67–82
2. Gibson, J.J.: The ecological approach to visual perception. Routledge (1986)
3. 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) (2007) 447–472
4. Chemero, A., Turvey, M.T.: Gibsonian affordances for roboticists. Adaptive Behavior **15**(4) (2007) 473–480
5. Kjellström, H., Romero, J., Kragić, D.: Visual object-action recognition: Inferring object affordances from human demonstration. Computer Vision and Image Understanding **115**(1) (2011) 81–90
6. Lopes, M., Melo, F.S., Montesano, L.: Affordance-based imitation learning in robots. In: Proc. of IROS 2007, IEEE (2007) 1015–1021
7. Montesano, L., Lopes, M., Bernardino, A., Santos-Victor, J.: Learning object affordances: from sensory–motor coordination to imitation. IEEE Transactions on Robotics **24**(1) (2008) 15–26
8. Ridge, B., Skocaj, D., Leonardis, A.: Unsupervised learning of basic object affordances from object properties. In: Computer Vision Winter Workshop. (2009) 21–28
9. Stark, M., Lies, P., Zillich, M., Wyatt, J., Schiele, B.: Functional object class detection based on learned affordance cues. In: Computer Vision Systems. Springer (2008) 435–444

10. Leibe, B., Leonardis, A., Schiele, B.: Combined object categorization and segmentation with an implicit shape model. In: Proc. of ECCV Workshop on Statistical Learning in Computer Vision. Volume 2. (2004)
11. Castellini, C., Tommasi, T., Noceti, N., Odone, F., Caputo, B.: Using object affordances to improve object recognition. IEEE Transactions on Autonomous Mental Development **3**(3) (2011) 207–215
12. Hinkle, L., Olson, E.: Predicting object functionality using physical simulations. In: Proc. of IROS 2013, IEEE (2013) 2784–2790
13. Wünstel, M., Moratz, R.: Automatic object recognition within an office environment. In: CRV. Volume 4. (2004) 104–109
14. Bar-Aviv, E., Rivlin, E.: Functional 3d object classification using simulation of embodied agent. In: BMVC. (2006) 307–316