Detecting Fine-grained Sitting Affordances with Fuzzy Sets

Viktor Seib, Malte Knauf and Dietrich Paulus
Active Vision Group (AGAS), University of Koblenz-Landau, Universitätsstr. 1, 56070, Koblenz, Germany

Keywords: Affordances, Fine-grained Affordances, Visual Affordance Detection, Object Classification, Fuzzy Sets.

Abstract: Recently, object affordances have moved into the focus of researchers in computer vision. Affordances describe how an object can be used by a specific agent. This additional information on the purpose of an object is used to augment the classification process. With the herein proposed approach we aim at bringing affordances and object classification closer together by proposing *fine-grained* affordances. We present an algorithm that detects fine-grained sitting affordances in point clouds by iteratively transforming a human model into the scene. This approach enables us to distinguish object functionality on a finer-grained scale, thus more closely resembling the different purposes of similar objects. For instance, traditional methods suggest that a stool, chair and armchair all afford sitting. This is also true for our approach, but additionally we distinguish sitting without backrest, with backrest and with armrests. This fine-grained affordance definition closely resembles individual types of sitting and better reflects the purposes of different chairs. We experimentally evaluate our approach and provide fine-grained affordance annotations in a dataset from our lab.

1 INTRODUCTION

Reasoning about an object's purpose is an important area in today's research on robotics. While object classification is a widely studied topic no solution exists yet as how to reflect the multitude of different object categories and relate them to possible actions of a robot or a person. Indeed, object classification approaches struggle with classes exhibiting large intra class shape variations. On the other hand, objects belonging to the same class share a certain functionality. At this point, affordances (Gibson, 1986) seem to provide a beneficial solution. While shape features are often acquired locally (i.e. around salient points) and might therefore be misleading, detecting a functionality of an object facilitates categorization. Additionally, recognizing affordances of objects instead of the object classes, allows objects and tools to be applied even without the precise knowledge of the class the object belongs to. Even objects of different classes can be applied according to a certain affordance required by the agent. For example, if an agent (e. g. a robot) needs to hammer, it would pick a heavy object providing enough space for grasping and a hard surface to hit on another object. This works without knowing the category *hammer* or having a hammer available by e. g. using a stone instead.

Approaches to detect and learn affordances in robotics often propose to infer affordance by imitation from observing humans (Stark et al., 2008), (Kjellström et al., 2011), (Lopes et al., 2007), or to learn affordances through interaction (Montesano et al., 2008), (Ridge et al., 2009). Other approaches focus on augmenting the performance of object recognition methods by recognizing affordances (Hinkle and Olson, 2013). In their approach, Castellini et al. (Castellini et al., 2011) record kinematic features of a hand while grasping objects. They show that visual features together with kinematic information help augmenting the object recognition. In contrast to these approaches using *interactive* affordances we do not record kinematic data of an agent, neither do we detect affordances by interaction. In contrast, visual perception is mostly common in today's robots and it is thus plausible to rely on that data. Thus, the approach proposed in this paper relies on visual data only.

In our approach we employ the *observer's* view on affordances as introduced by Şahin et al. (Şahin et al., 2007). 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 representing a humanoid. In recent work (Jiang and Saxena, 2013), (Grabner et al., 2011) this *observer's* view is often referred to as *hallucinating interactions*.

In the proposed method we focus on the comple-

Seib, V., Knauf, M. and Paulus, D.
Detecting Fine-grained Sitting Affordances with Fuzzy Sets.
DOI: 10.5220/0005638802890298
In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - Volume 4: VISAPP*, pages 289-298
ISBN: 978-989-758-175-5

Figure 1: Example furniture objects corresponding to the different fine-grained affordances detected by the presented approach. Top row: a stool representing the *sitting without backrest* affordance and two chairs representing the *sitting with backrest* affordance. Bottom row: three chairs representing the *sitting with backrest* and *sitting with armrest* affordances. Additionally, the rightmost chair also supports the *sitting with headrest* affordance.

mentary nature of an agent and its environment. We use indoor or home environments that are considered as artificial environments specifically designed to suit the needs of humans. Therefore, the complementary agent to the investigated environment is an anthropomorphic, i. e. human, body. Thus, for the purpose of this work affordances shall be informally defined as *action possibilities that the environment offers to an anthropomorphic agent*.

Related approaches in the literature (Hinkle and Olson, 2013), (Sun et al., 2010), (Hermans et al., 2011) distinguish affordances on a coarse scale. The considered affordances often include sitting (chairs), support for objects (tables) and liquid containment (cups). We propose looking closely at the individual affordances and distinguishing their functional differences an a fine-grained scale. We already introduced the concept of *fine-grained* affordances in (Seib et al., 2015) to closely resemble the functional differences of related objects. Although good results could be obtained, our previous work was a proof-of-concept with several limitations. The algorithm could be used to distinguish only 2 fine-grained affordances. Additionally, it relied on planes segmented from the scene that had to be oriented in a certain way. Further, without a fuzzy set formulation it relied on fixed values for important thresholds.

In the presented work, we concentrate on fine-grained affordances derived from the affordance *sitting*. We present a new algorithm for fine-grained affordance detection that exploits fuzzy sets and differentiates between 4 typical functionality characteristics of the *sitting* affordance. We divide the coarse affordance *sitting* into the fine-grained affordances *sitting without backrest*, *sitting with backrest*, *sitting with armrest* and *sitting with headrest*, whenever the sitting functionality is supported by additional environmental properties that can be exploited by the considered agent. Further, the presented approach no longer relies on features like segmented planes from the environment, but rather uses the whole input data for processing.

A system that is able to find affordances either encounters only those objects that were specifically designed to support the affordance in question or environmental constellations that afford the desired action. Our algorithm takes point clouds from a RGB-D camera as input. The input data is directly searched for affordances (and thus functionalities) without prior object segmentation. In the core of the algorithm, the agent model is transformed and checked for collisions with the environment. Specific goal configurations of the agent model represent different types of fine-grained affordances. The encountered affordances are segmented from the input point cloud. This segmentation can serve as an initial segmentation for a subsequent object classification step (not further explored in this work). Since the found affordances (especially on a fine-grained scale) provide many hints on the possible object class, categorization can be performed with fewer training objects or simpler object models. The presented fine-grained affordances correspond to objects such as a stool, chair, armchair and a chair with head support (Figure 1).

Specifically, an affordance-based categorization system can be exploited as outlined in the following. Affordances enable the detection of *sittable* objects even without knowing object classes as *stool*, *chair* or *couch*. Following the idea of fine-grained affordances, a stool standing close to a wall can even provide both affordances: sitting with and without backrest (in the former case the back is supported by the wall). This intuitively corresponds to the way a human would utilize an object to obtain different functionality.

The remainder of this work is structured as described in the following. Related work on affordances in robotics is presented in Section 2. Section 3 introduces the model definitions applied in our algorithm and Section 4 explains our approach for fine-grained affordance detection in detail. The proposed algorithm is evaluated in Section 5. Finally, a discussion is given in Section 6 and Section 7 concludes the paper and gives an outlook to our future work.

2 RELATED WORK

Affordances provide a new way to look at objects. Rather than determining the object's category by learning shapes of object classes, the visual information is used to detect the object's functionality. For example, objects of the class *chair* have a large intra class shape variation, imposing great challenges for object recognition systems. At the same time, all instances of class *chair* share the same functionality offering opportunities to augment object recognition systems.

Hinkle and Olson (Hinkle and Olson, 2013) use physical simulation to predict object functionality. The simulation consists of spheres falling onto an object from above. A feature vector is extracted from each object depending on where and how the spheres come to rest. The objects are classified as cup-like, table-like or sitable.

Research especially focusing on sitting affordances has been conducted over the past years. Office furniture recognition (chairs and tables) is presented by Wünstel and Moratz (Wünstel and Moratz, 2004). Affordances are used to derive the spatial arrangement of the object's components. Objects are modeled as graphs, where nodes represent the object's parts and edges the spatial distances of those parts. The 3D data is cut into three horizontal slices and within each slice 2D segmentation is performed. The segmentation results are classified as object parts and matched to the object models. Wünstel and Moratz' approach detects sitting possibilities also on objects that do not belong to the class *chair*, but intuitively would serve a human for sitting. Unlike the approach of Wünstel and Moratz, we encode the spatial information needed for affordance detection in an anthropomorphic agent model and affordance models, rather than creating explicit object models.

Approaches more similar to the one proposed in this paper use simulated interaction of an agent and the environment. Bar-Aviv and Rivlin (Bar-Aviv and Rivlin, 2006) use an embodied agent to classify sittable objects. Starting with an initial agent pose, the compatibility of different semi-functional agent poses with the object is tested. For each object hypothesis and agent pose a score is computed and most probable poses are further refined. The object is assigned the label of the hypothesis with the highest score. Contrary to Bar-Aviv and Rivlin (Bar-Aviv and Rivlin, 2006) who also use an embodied agent, our method operates directly on the whole input data. We do not need to segment the object prior to affordance detection. In our case, the segmented part of the scene is a result of the detected affordances on the input data.

Especially in design theory approaches of hierarchical affordance modeling were proposed (Maier et al., 2007), (Maier et al., 2009). Their goal is to divide objects into different functional parts that represent different affordances. This allows a designer to identify desired and undesired affordances in early stages of product design. Note however that this hierarchical affordance modeling is conceptually different from the fine-grained affordances applied in this paper. We do not separate objects in different parts with different affordances. Rather, our object independent approach separates an affordance (in this case the *sitting* affordance) into different sub-affordances on a fine-grained scale.

More recently, Grabner et al. (Grabner et al., 2011) proposed a method that learns sitting poses of a human agent to detect sitting affordances in scenes to classify objects. For training, key poses of a sitting person need to be placed manually on each example training object. In detecting chairs, their approach achieves superior results over methods that use shape features only. However, as pointed out by Grabner et al. their approach has difficulties in detecting stools, since they were not present in the training data. Consequently, the approach of Grabner et al. does not allow to detect affordances per se, but rather affordances of trained object class examples.

In the present paper we follow a different approach. Our goal is to directly detect sitting affordances in input data, independently of any possibly present object classes. Further, if a sitting affordance is detected, it will be categorized on a fine-grained scale according to the characteristics of the input data at the position where the affordance was found. Consequently, our approach does not rely on examples of sitting furniture, but only on the anthropomorphic agent model encoding (in our case) comfortable sitting positions. Our fuzzy function formulation encodes expert knowledge to connect the input data with the desired functionality with respect to the given agent model. Therefore, sitting affordances are detected on the data as it is, independently of the presence of actual object classes. Additionally and similar to Grabner et al., our approach suggests a pose how the detected object can be used by the agent.

Note that our approach is ignorant of any object categories. However, our fine-grained affordance formulation allows for a more precise object categorization as a consequence of affordance detection. Due to the fine-grained scale on which affordances are detected, object categories can be easily linked to the detection result (e.g. if a backrest could be detected or not). Our approach thus suggest as which kind of object the detected object exhibiting the affordance can be used. However, in this paper we concentrate on introducing the concept of fine-grained affordances. The detailed analysis of detected objects and their classification is left for future work.

3 MODEL DEFINITIONS

Usually, affordances are defined as relations between an agent and its environment (Gibson, 1986), (Şahin et al., 2007), (Chemero and Turvey, 2007). Since these two entities are crucial for affordances, we start with their definitions. Then, a definition of fine-grained affordances is provided.

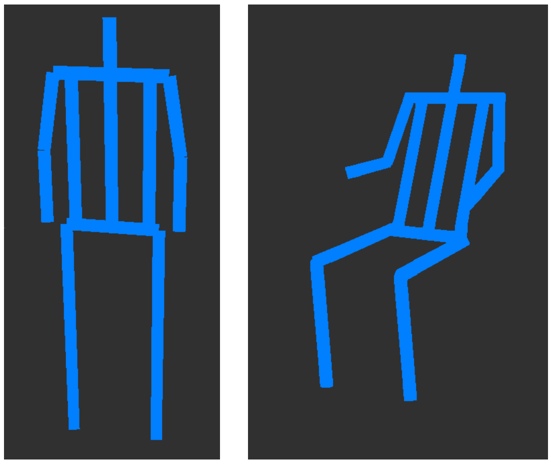
3.1 Agent and Environment

Our anthropomorphic agent model is defined as a directed acyclic graph $\mathcal{H}$ representing a human body (Figure 2). 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 while maintaining an anatomically plausible state (i.e. without harming a real human if the same state would be applied). When fitting the agent into the environment during affordance detection, the edges of the graph are approximated by cylinders for collision detection. Contrary to our previous work (Seib et al., 2015), we do not need an explicit environment model $\mathcal{E}$. Rather, $\mathcal{E}$ is simply the point cloud data of a scene where affordances should be detected.

3.2 Fine-grained Affordances

A fine-grained affordance is a property of an affordance that specializes the relation of an agent $\mathcal{H}$ and its environment $\mathcal{E}$. In the presented work, we take the *sitting* affordance as an example. The affordance *sitting* is a generalization of more precise relations that an agent and its environment can take. In this paper, we demonstrate our ideas by distinguishing between the fine-grained affordances *sitting without backrest*, *sitting with backrest*, *sitting with armrest* and *sitting with headrest*.

Note that some of these fine-grained affordances depend on others. For example, if the environment $\mathcal{E}$ affords *sitting with backrest* to the agent $\mathcal{H}$ it must



(a) (b)

Figure 2: The humanoid model: in an upright standing pose (a) and in a sitting pose as used in our experiments (b).

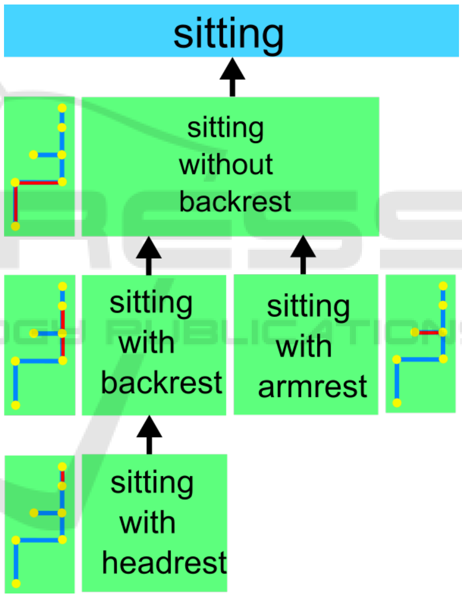
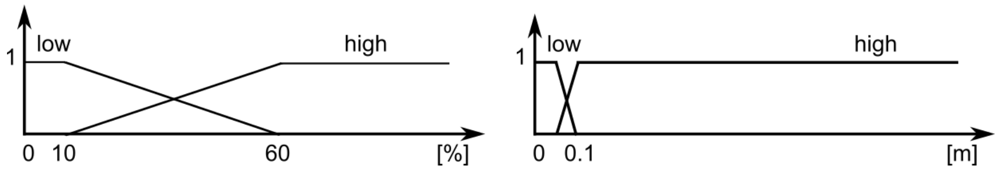


Figure 3: The presented affordance model specializes the sitting affordance into fine-grained affordances. The arrows indicate the dependencies between the fine-grained affordances. An agent pose for each fine-grained affordance is displayed. The edges shown in red are used for collision tests during detection.

necessarily afford *sitting without backrest* as well, because the agent can choose not to use the backrest while seated. The dependencies as defined in our models are depicted in Figure 3.

For each affordance $\mathcal{A}$, an initial pose of the agent needs to be defined (so far, we use only one affordance, namely *sitting*). Thus, every fine-grained affordance F_i specializing the same affordance $\mathcal{A}$ has the same initial pose. The initial pose refers to the



(a) discontinuity (b) roughness

(c) height (d) sitability

Figure 4: Membership functions used to find valid positions for sitting affordances. The functions in (a), (b) and (c) are used to evaluate the rule, while the function in (d) is used for defuzzyfication.

joint states of the simulated agent prior to any transformations and collision tests. Further, each fine-grained affordance is defined by a number of relevant body parts of the agent (i.e. edges in the graph $\mathcal{H}$) that are tested for collisions, as well as the corresponding goal angles for transformation.

In the context of this paper we define an affordance $\mathcal{A}$ as a set $\mathcal{A} = \{F_0 \dots F_j\}$ of fine-grained affordances F_i. The function $\mathrm{Aff} : \mathcal{H}_\mathcal{A} \times \mathcal{E} \times \mathcal{A} \rightarrow \{(\mathcal{F}, P, \mathcal{H}_g)_i\}$ determines for a given environment $\mathcal{E}$, affordance $\mathcal{A}$ and initial agent configuration $\mathcal{H}_\mathcal{A}$ a set of tuples. Each tuple contains $\mathcal{F} \subseteq \mathcal{A}$, a set of fine-grained affordances present at position P in the environment with a goal agent configuration $\mathcal{H}_g$. The algorithm presented in the next Section is an implementation of the above function Aff.

4 FINE-GRAINED DETECTION OF SITTING-AFFORDANCES

The algorithm for fine-grained affordance detection is essentially based on dropping an agent model in its initial pose into a point cloud at appropriate positions. These positions need to be found beforehand. The joints of the model are then transformed to achieve maximum contact with the point cloud. Only joints relevant for a certain affordance are considered. The initial pose of the agent, as well as the joint transformations are determined by the affordance models. Further, only the agent model and the affordance models determine the current functionality of the detected object. This means that the presented approach also finds objects that might not have been designed to fulfill a certain functionality. However, based on visual information and their position in the scene they afford the desired actions. We confined the evaluation to an agent representing an average human adult and to fine-grained affordances derived from the affordance *sitting*. The algorithm is described in the following. Additionally, it is outlined in Algorithm 1.

4.1 Extracting Positions of Interest

Before fitting the agent into the scene, the search space needs to be reduced to the most promising positions. We therefore create a height map of the scene (Figure 5 (b)). The point cloud is subdivided into cells of size c. In our experiments a size of $c = 0.05$ m provided a good balance between precision and calculation time. The highest point per cell determines the cell height. We decided in favor of the highest point instead of the average to avoid implausible values at borders of objects, where a cell may contain parts of the object and e.g. the floor.

Subsequently, a circular template, approximating the agent's torso, is moved over the height map to test whether a cell is well suited for sitting. The diameter of this template corresponds to the width of the agent as defined in the model. The decision for each cell is based on fuzzy sets as introduced by Lotfi Zadeh (Zadeh, 1965). We define 3 membership functions: discontinuity, roughness and height (Figure 4). Discontinuity is a measure defined in percent of invalid cells or holes within the current position of the circular template. Roughness is the standard deviation of the height of all cells within the circular template. Finally, the membership function height is used to include only cells in a certain height that allow comfortable sitting with bent knees, while the feet still touch the ground. However, this function can be disabled in the algorithm configuration to allow for valid sitting positions on the ground or on higher planes like tables. One single rule is enough to decide whether a position is suited for affordance detection or not. We use the intersection of these membership functions to obtain the following rule: *IF roughness is low AND discontinuity is low AND height is comfortable THEN the position is suited for sitting*. Of course, with more affordances, more rules will be needed. The fuzzy value obtained from these functions is defuzzyfied on the sitability function depicted in Figure 4 (d) using the *first of maximum* rule. The test is performed for both fuzzy sets of this rule, obtaining a crisp value for *sittable* and *not sittable* and deciding in favor of the fuzzy set with the higher crisp value. The positions obtained in this manner are used as possible sitting positions in further algorithm steps (Figure 5 (c)).

4.2 Agent Fitting

On every extracted position, $\frac{360^\circ}{w}$ agent models in the initial pose are dropped from a small height (we use a height of 0.1 m). The total number depends on the parameter w determining the angular rotation difference about the vertical axis between 2 subsequently tested models. We test several models in this step since the initial circular template was an approximation of the agent's torso, while in this step also the corresponding rotation needs to be found to provide enough room for the agent's legs. Dropping the agents is simulated by stepwise lowering the model until a collision is detected. If a collision occurs before any lowering of the model, the position is discarded. The affordance model is applied to all remaining positions.

The relevant joints for the affordance are gradually transformed from their initial pose to maximum allowed goal pose. The affordance is detected if a collision with the scene is encountered during the transformation. For instance, the fine-grained affordance *sitting with backrest* is detected during the transformation of the agent's torso, comparable to the agent's movement of leaning backward against a backrest. If a joint reaches its maximum goal pose without a collision the algorithms assumes that the affordance is not present.

We use the open source library *Flexible Collision Library (FCL)* (Pan et al., 2012) for collision detection. FCL detects collisions between 2 objects and returns the exact position at which the collision occurred. As input for FCL we convert the point cloud of the scene to the OctoMap representation (Hornung et al., 2013) and approximate the individual body parts of the agent by cylinders. The scene and agent are thus iteratively tested for collisions, by first transforming the corresponding joint and then performing the collision test. This procedure is repeated until the goal angle is reached or a collision occurs.

Note that in contrast to normal affordances, a fine-grained affordances might depend on the existence of another fine-grained affordance (Figure 3). In the presented work the *sitting without backrest* affordances is checked first, as other affordances depend on it. *Sitting with backrest* and *sitting with armrests* are

Algorithm 1: Fine-grained Detection of Sitting-Affordances.

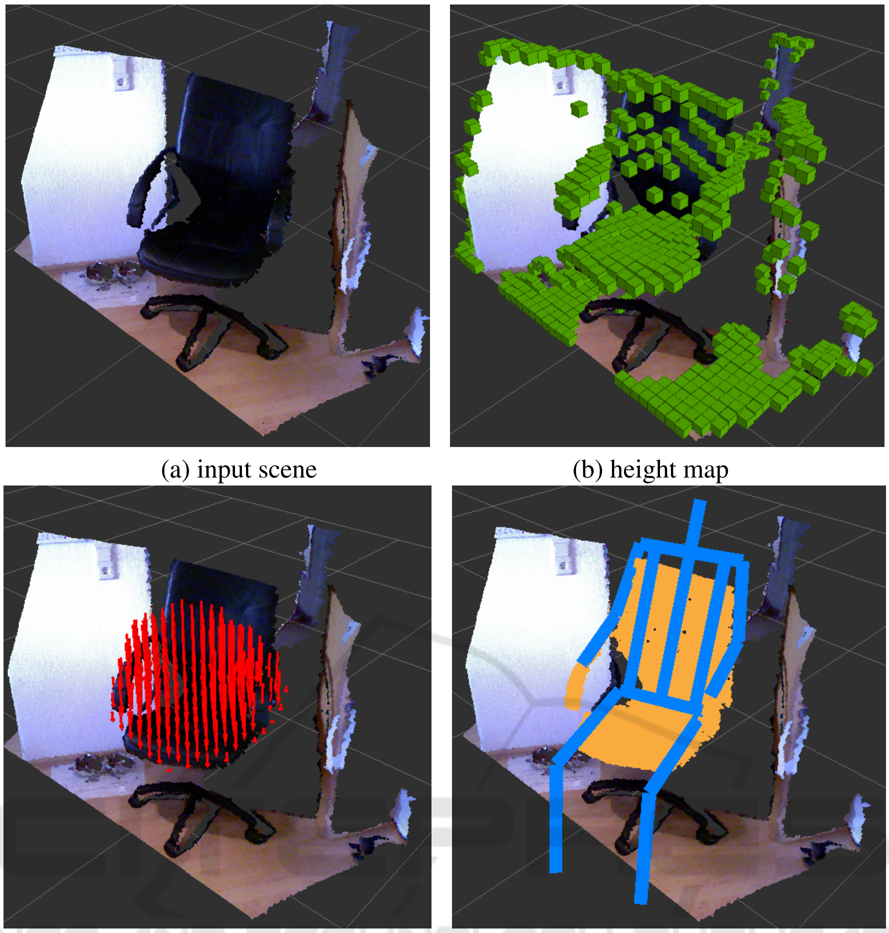
Require: Agent model $\mathcal{H}_\mathcal{A}$, Point cloud (environment) $\mathcal{E}$, Affordance models $\mathcal{A} = \{F_0 \dots F_j\}$,
Ensure: List of tuples $L = \{(\mathcal{F}, P, \mathcal{H}_g)_i\}$ with $\mathcal{F} \subseteq \mathcal{A}$, a set of fine-grained affordances, the position P in the environment and a goal agent configuration $\mathcal{H}_g$.

{find candidate positions}
$H \leftarrow createHeightMap(\mathcal{E})$
$C \leftarrow \emptyset$
for all cells $c \in H$ **do**
  if $roughness(c)$ is low **and** $discontinuity(c)$ is low **and** $height(c)$ is comfortable **then**
    $C \leftarrow C \cup c$
  end if
end for

{test whether sitting is possible}
$L \leftarrow \emptyset$
$\mathcal{F} \leftarrow \emptyset$
for all $c \in C$ **and** $\frac{360}{w}$ orientations of $\mathcal{H}_\mathcal{A}$ **do**
  place agent $\mathcal{H}_\mathcal{A}$ over cell c
  if not $collides(\mathcal{H}_\mathcal{A}, \mathcal{E})$ **then**
    lower $\mathcal{H}_\mathcal{A}$ until collision
    $\mathcal{F} \leftarrow \mathcal{F} \cup F_0$
    $P \leftarrow getPositionOfCell(c)$
    $\mathcal{H}_g \leftarrow \mathcal{H}_\mathcal{A}$

    {check all fine-grained affordances}
    for all $F_i \in \mathcal{A}$ **do**
      $\mathcal{H} \leftarrow \mathcal{H}_\mathcal{A}$
      while $isNotInGoalPose(\mathcal{H})$ **do**
        $transformJoints(\mathcal{H})$
        if $collides(\mathcal{H}, \mathcal{E})$ **then**
          $\mathcal{F} \leftarrow \mathcal{F} \cup F_i$
          $\mathcal{H}_g \leftarrow \mathcal{H}$
        end if
      end while
    end for
  end if

  {save results}
  if $\mathcal{F} \neq \emptyset$ **then**
    $L \leftarrow L \cup (\mathcal{F}, P, \mathcal{H}_g)$
  end if
end for



(a) input scene (b) height map

(c) possible positions for sitting (d) segmented object

Figure 5: Illustration of different algorithm steps. The input scene is shown in (a) and the corresponding height map in (b). Image (c) shows the possible positions for sitting affordances found by our fuzzy set formulation. The length of the red arrows corresponds to the defuzzyfied value from the sitability function. The final agent pose as well as the object segmentation is shown in (d) for the fine-grained affordances *sitting with backrest* and *sitting with armrest*.

checked subsequently. The *sitting with headrest* affordance is checked as the last one, since it depends on the presence of a backrest. The output of this step is the final pose of the agent (position P and joint states $\mathcal{H}_g$), as well as a set of detected fine-grained affordances $\mathcal{F}$.

4.3 Pose Selection and Object Segmentation

So far, we have obtained the position and the list of detected fine-grained affordances. However, we still have several hypotheses per position, since the agent was dropped at different rotations. To select the best pose for each position we use an assessment function. It is based on the total number of collisions detected for a pose. The assumption behind this is that a higher number of collisions indicates a more comfortable pose (a person sitting comfortably on a chair touches the chair at more points than a person sitting on the edge of a chair). We thus select the pose with the most collision points for each hypothesis position. This collision based rule favors affordances that involve a higher number of joints. However, this does not limit the validity of the output since more specialized affordances (i.e. using more joints) depend on other affordances which are also present at the detected positions.

After obtaining the affordances and highest rated poses, the partition of the scene exhibiting that affordance is segmented. We use a region growing algorithm where the position of the detected affordance serves as seed point. Each point below a certain Euclidean distance is added to the segmented scene part. A low value is well suited to close small gaps in the point cloud, but at the same time limit the segmenta-

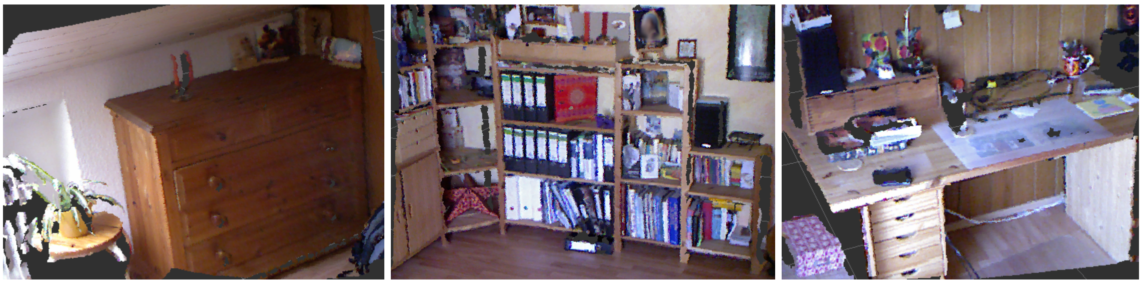


Figure 6: Example scenes without sitting affordances in the evaluation dataset.

tion result to one object. Further, points close to the floor are ignored. The segmentation result is shown in Figure 5 (d).

5 EVALUATION

We conducted our experiments on real-world 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 fine-grained affordances. From these data, we extracted 248 different views of the chairs and 47 different views of the stools. Example views of these objects are shown in Figure 1. Additionally, negative data (i.e. data without the fine-grained affordances) from 9 different furniture objects was obtained and 109 views of these objects extracted. Negative data includes objects like desks, tables, dressers and a heating element. Example views of negative data are presented in Figure 6. The whole evaluation dataset contains 404 scene views with 295 positive and 109 negative data examples. The dataset was annotated with expected positions and rotations of the agent for the individual fine-grained affordances. All this data is provided online[1] and is the first publicly available dataset with fine-grained affordance annotations.

We applied our affordance detection to these data. From each scene, the best scored affordance was extracted and compared to ground truth. If the detected position was within 0.2 m and within a rotation of 20° to the ground truth, a sitting affordance was correctly found and is further analyzed for fine-grained affordances. Examples of correct affordance detections are presented in Figure 7.

The results of the evaluation are shown in Table 1. Our algorithm is able to find almost all sitting possibilities, while making only little mistakes, as indicated by the results for the *sitting without backrest* affordance. While the recall for the *sitting with backrest* affordance is below the recall of the first affordance, it is still high at 83%. The ability of our algorithm to detect these two specialized affordances at the presented high rates speaks in favor of the presented approach. Note that these results were achieved without any training. All the knowledge required for detection is encoded in the simple agent and affordance models.

[1] Test dataset available at https://dl.dropboxusercontent.com/u/6693658/affordance_dataset.zip

Table 1: Evaluation results for each of the fine-grained affordances.

sitting affordance	precision	recall	f-score
w/o backrest	0.89	0.97	0.93
with backrest	0.89	0.83	0.86
with armrest	0.84	0.57	0.68
with headrest	0.98	0.39	0.58

The results for the fine-grained affordances involving an armrest and a headrest are below the aforementioned ones. F-scores of 68% (armrests) and 58% (headrest) indicate that our algorithm successfully differentiates between closely related object functionalities and is able to detect the corresponding fine-grained affordances in RGB-D data. However, the lower values indicate that the agent model might need more degrees of freedom during collision detection to better find differently shaped chairs. On average, in its current state our algorithm takes 4.3 seconds to process one scene (single thread execution).

6 DISCUSSION

The main idea of this paper is to introduce the concept of fine-grained affordances and overcome the limitations of our previous approach (Seib et al., 2015) that relied on plane segmentation and, thus, was not general enough. Here we have shown that our algorithm is able to differentiate affordances on a fine-grained scale without prior object or plane segmentation. Thus, the presented approach is more general

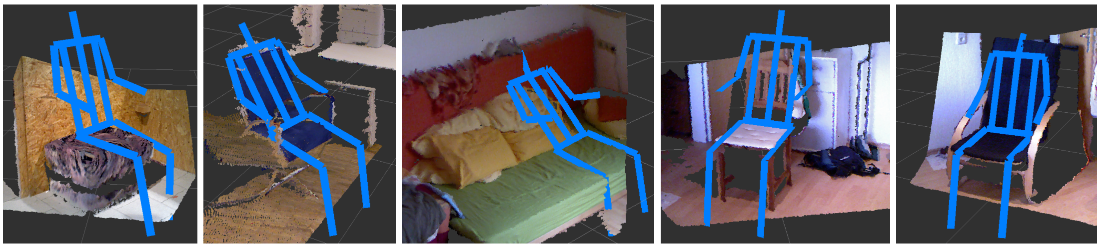


Figure 7: Resulting agent poses for some of the scenes in the evaluation dataset.

and can be applied to the input data directly.

To our best knowledge, no similar approaches exist in the literature that are able to differentiate affordances on a fine-grained scale. This makes it hard (if not impossible) to assess the quality of our approach and compare it to related work. We therefore want to give a discussion on certain properties of our algorithm and give a detailed outlook to our ongoing work in that field.

Apart from introducing the notion of fine-grained affordances the biggest difference to related work such as the approach of Grabner et al. (Grabner et al., 2011) is that we detect affordances directly. In contrast, Grabner et al. learn affordances as properties of objects which allows them to augment the classification ability of their approach. However, our approach is ignorant of any object categories.

While we believe that our approach will also benefit from machine-learning techniques (e.g. by learning the membership functions for the fuzzy sets), at this point we have completely omitted the learning step. This comes at the cost of manually defining "reasonable" values for the fuzzy sets (low effort) and a deformable human model (medium effort). Additionally, this raises the question on the extensibility of the approach. An initial agent pose needs to be provided for any new affordance that is included. However, if an agent model is already available (here for sitting) new poses can be added by simply transforming joint values in the corresponding configuration file, as we have done for illustration in Figure 2 (a). As a second step, the joints of interest that are involved in the new affordance description, need to be provided with a minimum and maximum angle for transformation.

A more complex extension of the algorithm would be to include a different agent, e.g. a hand for grasping. While the hand itself can be modeled again as an acyclic graph of joints, the initial hypotheses selection step must be changed completely. Instead of finding potential sitting positions in the height map, for a hand a different hypotheses selection needs to be applied (e.g. finding small salient point blobs). However, as soon as these hypotheses are found, the rest of the algorithm is the same: transform joints of the agent and evaluate a cost function that reflects the quality of affordance detection. We thus believe that the presented approach is generalizable and well suited for extension.

7 CONCLUSION AND OUTLOOK

In this paper we have further refined the term *fine-grained affordances* to better distinguish similar object functionalities. We presented a novel algorithm that is based on fuzzy sets, to detect these affordances. The algorithm has been evaluated on 4 specializations of the sitting affordance and we have shown that the presented approach is able to differentiate affordances on a fine-grained scale. For comparable state of the art approaches, these 4 fine-grained affordances would all have been the same affordance: sitting.

Apart from the ability to distinguish similar object functionalities, fine-grained affordances can be applied as a filtering or preprocessing step for object classification. The segmented object that results from the affordance detection is constrained to object classes that provide the detected affordance. Thus, if this object needs to be classified, it does not have to be matched against the whole dataset, but only against object classes exhibiting the found affordance.

The presented algorithm is ignorant of any object classes, since our goal is to detect affordances. This is evident from the leftmost image in Figure 7, where the agent is sitting with a backrest although the object it is sitting on does not have one. Clearly, here the environmental constellation (object and wall) provided the detected affordance. This demonstrates a strength of the concept of fine-grained affordances that we will further explore in our future work.

Further, we will investigate how an anthropomorphic agent model can be exploited to detect more fine-grained affordances from other body poses. As an example for a lying body pose, the fine-grained affordances *lying flat* and *lying on a pillow* can be distinguished. Fine-grained affordances can also be defined for other agents, e.g. a hand. In that case, *grasping* *with the whole hand* and *grasping with two fingers* could be distinguished, e.g. for grasp planning for robotic arms. Additionally, fine-grained affordances for grasping actions can include drawers and doors that can be *pulled open* or *pulled open while rotating* (about the hinge). We are currently looking for more examples for fine-grained affordances for different agents, to generalize our approach of fine-grained affordances.

REFERENCES

Bar-Aviv, E. and Rivlin, E. (2006). Functional 3d object classification using simulation of embodied agent. In *BMVC*, pages 307–316.

Castellini, C., Tommasi, T., Noceti, N., Odone, F., and Caputo, B. (2011). Using object affordances to improve object recognition. *Autonomous Mental Development, IEEE Transactions on*, 3(3):207–215.

Chemero, A. and Turvey, M. T. (2007). Gibsonian affordances for roboticists. *Adaptive Behavior*, 15(4):473–480.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Routledge.

Grabner, H., Gall, J., and Van Gool, L. (2011). What makes a chair a chair? In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1529–1536.

Hermans, T., Rehg, J. M., and Bobick, A. (2011). Affordance prediction via learned object attributes. In *International Conference on Robotics and Automation: Workshop on Semantic Perception, Mapping, and Exploration*.

Hinkle, L. and Olson, E. (2013). Predicting object functionality using physical simulations. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 2784–2790. IEEE.

Hornung, A., Wurm, K. M., Bennewitz, M., Stachniss, C., and Burgard, W. (2013). OctoMap: An efficient probabilistic 3D mapping framework based on octrees. *Autonomous Robots*. Software available at http://octomap.github.com.

Jiang, Y. and Saxena, A. (2013). Hallucinating humans for learning robotic placement of objects. In *Experimental Robotics*, pages 921–937. Springer.

Kjellström, H., Romero, J., and Kragić, D. (2011). Visual object-action recognition: Inferring object affordances from human demonstration. *Computer Vision and Image Understanding*, 115(1):81–90.

Lopes, M., Melo, F. S., and Montesano, L. (2007). Affordance-based imitation learning in robots. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 1015–1021. IEEE.

Maier, J. R., Ezhilan, T., and Fadel, G. M. (2007). The affordance structure matrix: a concept exploration and attention directing tool for affordance based design. In *ASME 2007 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pages 277–287. American Society of Mechanical Engineers.

Maier, J. R., Mocko, G., Fadel, G. M., et al. (2009). Hierarchical affordance modeling. In *DS 58-5: Proceedings of ICED 09, the 17th International Conference on Engineering Design, Vol. 5, Design Methods and Tools (pt. 1), Palo Alto, CA, USA, 24.-27.08. 2009*.

Montesano, L., Lopes, M., Bernardino, A., and Santos-Victor, J. (2008). Learning object affordances: from sensory–motor coordination to imitation. *Robotics, IEEE Transactions on*, 24(1):15–26.

Pan, J., Chitta, S., and Manocha, D. (2012). Fcl: A general purpose library for collision and proximity queries. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 3859–3866.

Ridge, B., Skocaj, D., and Leonardis, A. (2009). Unsupervised learning of basic object affordances from object properties. In *Computer Vision Winter Workshop*, pages 21–28.

Şahin, E., Çakmak, M., Doğar, M. R., Uğur, E., and Üçoluk, G. (2007). To afford or not to afford: A new formalization of affordances toward affordance-based robot control. *Adaptive Behavior*, 15(4):447–472.

Seib, V., Wojke, N., Knauf, M., and Paulus, D. (2015). Detecting fine-grained affordances with an anthropomorphic agent model. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., editors, *Computer Vision - ECCV 2014 Workshops*, volume II of *LNCS*, pages 413–419. Springer International Publishing Switzerland.

Stark, M., Lies, P., Zillich, M., Wyatt, J., and Schiele, B. (2008). Functional object class detection based on learned affordance cues. In *Computer Vision Systems*, pages 435–444. Springer.

Sun, J., Moore, J. L., Bobick, A., and Rehg, J. M. (2010). Learning visual object categories for robot affordance prediction. *The International Journal of Robotics Research*, 29(2-3):174–197.

Wünstel, M. and Moratz, R. (2004). Automatic object recognition within an office environment. In *CRV*, volume 4, pages 104–109. Citeseer.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8(3):338–353.