Learning Grasping Affordance Using Probabilistic and Ontological Approaches

Carl Barck-Holst, Maria Ralph, Fredrik Holmar and Danica Kragic

Abstract— **We present two approaches to modeling affordance relations between objects, actions and effects. The first approach we present focuses on a probabilistic approach which uses a voting function to learn which objects afford which types of grasps. We compare the success rate of this approach to a second approach which uses an ontological reasoning engine for learning affordances. Our second approach employs a rule-based system with axioms to reason on grasp selection for a given object.**

I. INTRODUCTION

There has been increased interest in recent years towards developing robots that can be used within domestic environments where robots need to engage in learning and reasoning tasks within a short time frames. We are interested in exploring learning from an affordance perspective and examining how learning takes place within an ontological framework. The concept of affordances focuses on how objects can be used in the world. Although the definition of affordances varies among researchers, it relates what functionality an object provides to an agent. For instance a glass affords drinking to a human: person might not even perceive the glass, but rather the 'drinkability' affordance it provides.

Gibson [1] first coined the phrase and used the concept to explain that an agent is able to directly perceive affordances without the need for specific classification in order to perceive the function of an object, i.e. what it affords. For a robot, this means selecting features from sensory input that support the identification of an affordance [2]. We are often not interested in searching for a specific object, or type of object based on our representation of how the object appears. Rather we are more concerned with how the object can be used or what tasks it can be used for. If we cannot reach something placed high on a shelf for example, we search for something to stand on - an object that *affords* standing. In this case a box, a chair, or any similar type object may be chosen.

The concept of affordances has also been studied more recently from a number of perspectives. In [3], traversability affordance is explored through a robotic arm that learns how to traverse obstacles of unknown objects to reach a target object. Initial internal rehearsal using simulation is used to establish the likelihood for success or failure. In [4], an autonomous mobile robot also learns traversability affordance. In this case, the robot learns what objects can and cannot be traversed over. From an ontology perspective, in [5], an affordance-based ontology is proposed for semantic robots. A robot learns what certain objects, situations or contexts afford such as how robots with different embodiments can act in given situations.

In our work, we are interested in using the affordance concept to relate objects, actions and effects. Other work has been done previously on developing this type of relationship. Montesano et al. [6] for example model this concept using a Bayesian network approach. In their work, objects are defined according to shape, size and color. Actions are defined as grasp, tap and touch, and effects include contact duration between hand and object. They use imitation learning in a Bayesian network structure by observing what actions afford what effects when applied to specific objects. However, they do not examine success or failure and do not use their affordance model to reason when failures do arise. This approach also requires large amounts of training data and time to generalize well to new objects. As such, we will examine an alternative probabilistic approach in the form of a voting function, and compare this approach with an ontological reasoning engine to see how the result of grasping selections may differ.

A. Paper Contribution and Organization

This paper explores two approaches for learning affordances between actions and objects for a desired effect. In our case we are interested only in grasping actions, objects of different shapes and sizes, and the effect of either grasp success or failure. We also present how these approaches can be used for grasping of previously unseen objects.

The contribution of this paper is to examine how probabilistic and ontological approaches can be used to model affordances, the trade-offs between these two approaches, and how coupling these types of systems may produce more successful outcomes. It is well known that probabilistic approaches are limited particularly by their need for large amounts of training data, and their ability to grow and change as new knowledge about the world becomes available. In this paper we show how a probabilistic approach differs from a rule-based system by comparing and contrasting a voting function approach with an ontological reasoning engine used for grasp selection. An ontological approach which uses rules to link object-action pairs for successful outcomes typically requires less data and therefore time to update when new knowledge is found. As opposed to a probabilistic approach, rule-based systems need only update a rule, or add a new rule

Authors are with the Centre for Autonomous Systems, CVAP, KTH, Stockholm, Sweden. {barck,mralph,holmar,danik}@csc.kth.se. The work was supported by the EU project GRASP, IST-FP7-IP-215821 and the Swedish Foundation for Strategic Research.

rather than update the entire knowledge base, however they are deterministic and cannot easily accommodate uncertainty in the world. By analyzing these two approaches we can explore how these approaches may work together in order to develop a system that requires less amount of time to train but can also handle uncertainty when it arises.

This paper is divided into the following sections. Section I-B presents the experimental setup and Section II discusses a probabilistic approach to building affordance relations between objects, actions and effects. Section III outlines an ontological rule-based approach for representing affordances and also presents results from experiments conducted. Finally Section IV concludes our discussion and highlights future work.

B. System Design and Experimental Setup

For the two subsequent approaches presented experiments were conducted to examine the number of attempts required before a successful grasp was found. For both approaches a total of 1000 trials were conducted for 9 objects. The objects in the object set consisted of a coke bottle (simple cylinder), a salt box (simple rectangular-box/cuboid), and a "homer" toy object (complex object). These objects were classified with size delimiters small, medium and large totaling 9 objects in all. These are shown in Figure 2.

We define a grasp action according to the type (T), region (R), and force (F) applied to the object. Since the grasp type has been set to a virtual two-fingered grasp, a grasp action is denoted as $G = \{R, F\}$. The region and force are defined by a set of four possible values each,

$$R = \{Bottom, Middle, Top, Above\}$$
$$F = \{NoForce, Small, Medium, Large\}$$

The ranges for each grip force ($f_i \in F$) applied were arbitrarily chosen values between 1 and 5 (i.e. 1=no force, 3=small, 4 = medium, and 5=large or maximum grip force allowed). In order to create a grasp region ($r_i \in R$) for each object, the object's parameters (i.e. length, width, and height) were extracted by first generating a 3D point-cloud of the image then creating bounding-boxes in order to estimate the object's geometry as presented in [7], [8]. These values were used to decompose the object into 3 box segments corresponding to 4 regions on the object. These regions were defined as bottom, middle, top, and above areas of each object, as shown in Figure 1. Figure 2 shows the box regions and example point cloud representation for one of the objects used in the object set.

We have defined effect (E) as either grasp success or failure according to the contact duration between the object and the hand ($EObjHD$),

$$E = \{EObjHD\} \wedge EobjHD = short, long$$

If contact is maintained up to a height of 200 mm then the grasp is considered successful otherwise:

$$f(z_t, threshold) \begin{cases} 0 & \text{if success (EObjHD = long),} \\ 1 & \text{if failure (EObjHD = short).} \end{cases}$$

We also assume that successful grasps are stable grasps based on feedback from tactile sensors. A simple control architecture for applying a grasping cycle has been implemented in the GraspIt! simulator [9] and used to conduct our experiments. Below we provide the notation used in the remainder of the paper.

- Sh - denotes shape of the object
- Sz - denotes size of the objects
- R - denotes grasp region
- F - denotes grip force
- s - denotes grasp success

In our case, we are interested in learning what objects afford what grasping actions and the target function is:

$$f : < Sh, Sz > \rightarrow < R, F >$$

Data collected and used as the knowledge-base is defined according to object shape, size, grasp region, grip force and success or failure. An example of the format for the representation of the training data is shown in Table I.

TABLE I
EXAMPLE FORMAT FOR OUR TRAINING SET.

Shape	Size	Grasp Region	Grip Force	Error
cylinder	small	top	medium	success
cylinder	medium	above	medium	fail: no contact
cylinder	medium	above	large	success
rectangle	large	middle	large	success
cylinder	small	bottom	medium	fail: cannot reach object

The training sets created were built from experiments conducted in GraspIt!. The dataset was comprised of 48 instances where all combinations of grasping regions and grip forces are applied to all three objects of different shapes and sizes. This resulted in a total of 144 training instances. We also built a world model in the simulator which simulates the lab setup: a 6 DoF Kuka robot arm equipped with a Barrett hand. This arm/hand combination is used for each grasping trial. Corresponding success and failures for each grasp attempt are recorded during each training session.

During both approaches, which will be discussed in subsequent sections, the experimental setup was the same. Firstly, the knowledge-base is initially empty and is updated with each subsequent grasp attempt. The input data to each of the two reasoning systems (probabilistic, ontological) consisted of generating randomly ordered triples of objects with various size permutations. That is, the coke bottle, salt box and "homer" toy objects were used with sizes small, medium and large with unique permutations for these object combinations. For example, an input to the system may result in objects: coke bottle small, "homer" large, salt box small being used. This may be followed by "homer" small, coke bottle small, and salt box large, and so on, in a random fashion. Once the 1000 trials have been completed, the

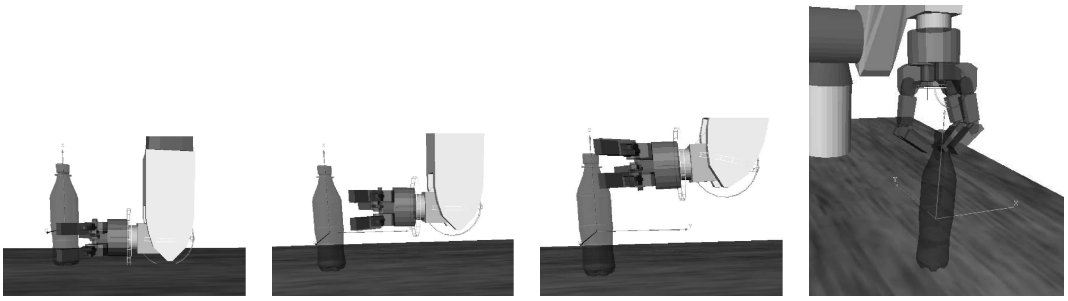
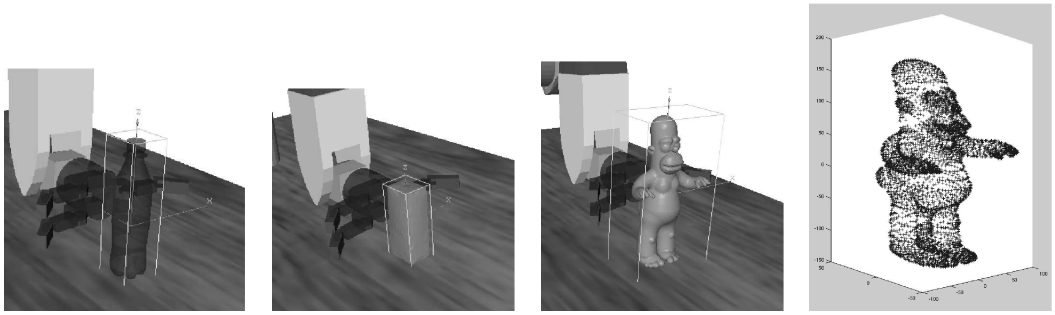


Fig. 1. Two grasp types (side grasp and top grasp) and four configurations.



(a) Cylindrical object

(b) Rectangular-box object

(c) A complex shape

(d) Point-cloud of a complex object used to estimate its bounding-box

Fig. 2. Objects used in experiments with their bounding-boxes for establishing grasp regions

average number of attempts before a successful grasp has been found are determined for each of the nine objects. These values are then used for comparison purposes between the two systems as shown in subsequent sections. If a failed grasp attempt has been established during testing, the object is repositioned back to its original pose and the new grasp attempted.

II. Approach I: A Probabilistic Approach to Learning Grasp Affordances

Our initial approach to grasp selection uses a probabilistic approach in the form of a voting function to infer the most likely successful grasp to select, as shown in Figure 3. In this approach the voting function is defined according to:

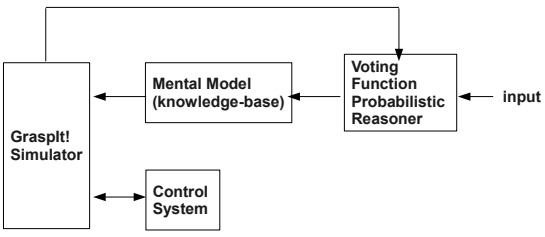
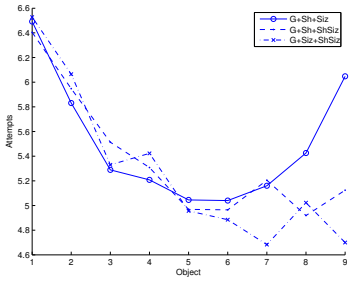


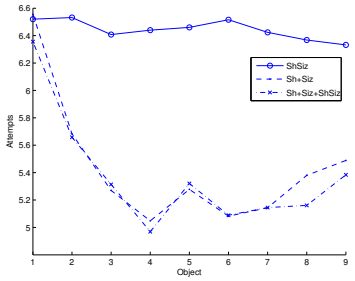
Fig. 3. The probabilistic approach is composed of four components.

$$V = \underbrace{P(s|R,F)}_{1} \quad \underbrace{P(s|R,F,Sh)}_{2} \quad \underbrace{P(s|R,F,Sz)}_{3} \quad \underbrace{P(s|R,F,Sh,Sz)}_{4} \tag{1}$$

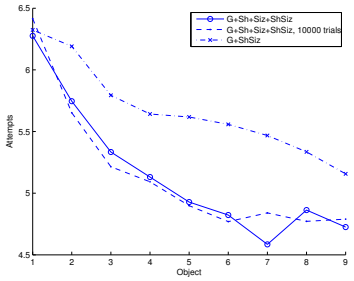
where the Grasp region (R) and Grip force (F) that maximize V for successful grasps (s) are selected as the next grasp configuration to try. Eq. 1 is used for the first grasp attempt and for all subsequent attempts if success is not achieved on the first try. Once the grasp has been attempted, this information is added to the knowledge base to update the coupling between objects and actions. If any probability within the voting function returns a value of zero, we use an estimation constant of m=0.001. If any probability within the equation returns an undefined value then we set m=0.5. An undefined case is a divide by zero situation where there is no evidence in the knowledge base of a particular grasp configuration for either a specific object shape or size. From a grasping perspective, we can take grasp configurations learned on a subset of objects and apply these to unknown objects with similar properties. The initial, empty knowledge-base is updated using the outcomes of each new trial - the instance that is attempted is added to the existing knowledge-base and the probability distributions are adjusted accordingly.



(a) O1: Some mixture of hypotheses are considered where: general case (G) is used along with a certain object size (Sz) is in the record searching, a certain object shape (Sh), or both specific object shape and size (ShSz) must be in the same record.



(b) O2:More specific cases where no general cases are considered: specific object size (Sz), specific shape (Sh) or specific shape and size (ShSz) must be in the same record



(c) O3:The last case which has a more comprehensive mixture of hypotheses to choose from: where all general and specific cases are considered

Fig. 4. Results from experiments run with variations of the equations outlined earlier in this section for Equation (1), where G corresponds to term 1, Sh corresponds to term 2, Sz corresponds to term 3, and ShSz corresponds to term 4. The combination of these equations in the graphs is represented with a "+" sign.

A. *Results*

We conducted experiments where either general grasps, only specific grasps, or a combination of both of these options were used to make a grasp selection. The voting function is run over 1000 randomly generated trials for most cases and over 10,000 randomly generated trials for the last case as shown in Figures 4(a), 4(b), and 4(c). These figures present the average number of grasp attempts made over the 9 objects used. The ordering of the objects is randomly generated from the 9 objects used and is not repeated for any of the trials.

As shown from the figures, most of the results for these set of experiments shows an improvement over time as the system progresses from one object to the next. This indicates that the system is learning what objects afford what grasps. However, findings suggest that results for runs with all probabilities, which corresponds to plot "G+Sh+Sz+ShSz" for Figure 4(c) and to sub-equations 1 to 4 as previously shown in Equation (1), produces the best outcome over time. To enhance this finding we ran the same selection of probability calculations over 10,000 trials and determined the average number of successful grasps over the same range of objects. This experiment is also shown in Figure 4(c) as a plot corresponding to "G+Sh+Sz+ShSz,10000 trials" and also covers the same set of equations just over a longer trial period. Observing the behavior of the system we see that running the data over a larger number of trials smooths out the graph and gives a clearer indication of how the system is performing over a longer period of time. This indicates that a good mixture of both general and specific knowledge is needed for appropriate grasp selection. Using a general grasp selection approach as shown in Figure 4(a) yields too broad a search space, producing more incorrect grasps chosen. Too specific a grasp selection approach or not enough of a mixture between general and specific cases, as shown in Figure 4(b), requires complete data to be present before a successful grasp is found, therefore constraining the search space too tightly. As such a combination of these two approaches with a good mixture of both general and specific cases appears to yield more effective results for these experiments, enabling the system to build more appropriate affordance links between certain objects and their grasping actions.

We are aware that limitations exist for this type of approach including the object descriptors chosen. In this case, a more rich object representation may be required in order to improve object-action affordance learning. This may result in choosing object shapes that reflect how close to the ideal the object really is. In other words, classifying an object as an exact cylinder, we can vary the degree to which other cylindrical-like objects may appear. The other issue that needs to be addressed is learning object shapes that are not present in the training data. In many cases objects are comprised of complex shape groupings which may not fit a particular shape classifier as accurately. One possible work-around for this has been presented in our previous work [7], [8] where a complex object is divided into several more simple shapes where each of these simple shapes can then be evaluated for a suitable grasp. We are also aware that the implementation of our current system may have other limitations since it employs a probabilistic structure with a rather small amount of data. This is a topic of our ongoing work. Here, we wanted to examine an alternative approach to affordance-based learning by exploiting a rule based approach in an ontological framework.

III. Approach II: An Ontological Approach to Learning Grasp Affordances

Here, we use a similar approach to that presented in [10]. Our approach to relating objects and actions using

affordances within an ontological framework is modeled using OWL and OWL-DL. This querying language is used to perform inferencing on a knowledge-base which is also updated in the presence of new knowledge in the same way as Approach I. Our Grasping Ontology is modeled as described in Figure 5, where OWL uses classes, object properties and individuals to model an ontology. Previously we have used five concepts (i.e. object shape, size, grasp region, grip force, and an outcome of success or failure) to characterize an object and afforded grasping parameters. Within the ontology framework we include two additional concepts: 1) object type and 2) action in order to use OWL in an appropriate manner. In OWL, the seven concepts are used as classes and the concept values as sub-classes in an appropriate hierarchy representing a minimal grasping taxonomy. The subclasses are defined to be disjoint from each other and correspond to the values that are provided from the perception and control systems. The class structure is defined as follows:

$$
\begin{aligned}
&\mathit{Shape} && \equiv \mathit{Cylinder} \sqcup \mathit{Box}\\
&\mathit{Size} && \equiv \mathit{LargeSize} \sqcup \mathit{MediumSize} \sqcup \mathit{SmallSize}\\
&\mathit{ObjectType} && \equiv \mathit{Coke} \sqcup \mathit{Homer}\\
&\mathit{GraspRegion} && \equiv \mathit{FromTop} \sqcup \mathit{HighRegion} \sqcup \mathit{MiddleRegion}\\
& && \quad \sqcup \mathit{LowRegion}\\
&\mathit{GraspForce} && \equiv \mathit{HighForce} \sqcup \mathit{MiddleForce} \sqcup \mathit{LowForce}\\
& && \quad \sqcup \mathit{NoForce}\\
&\mathit{Action} && \equiv \mathit{Grasp}\\
&\mathit{Outcome} && \equiv \mathit{Success} \sqcup \mathit{PositionError} \sqcup \mathit{GraspError}\\
& && \quad \sqcup \mathit{LiftError} \sqcup \mathit{SimulationError}
\end{aligned}
$$

The object properties were created such that each objectType has a shape (*hasShape*) and a size (*hasSize*), and affords (*affords*) none-to-many actions. Actions are afforded (*isAffordedBy*) by objects and lead to (*leadsTo*) some outcome. An action uses (*uses*) exactly one grasping region and one grasping force. In turn, every property has an appropriately named inverse property. In Figure 6 we illustrate the main building blocks of the system for Approach II. The

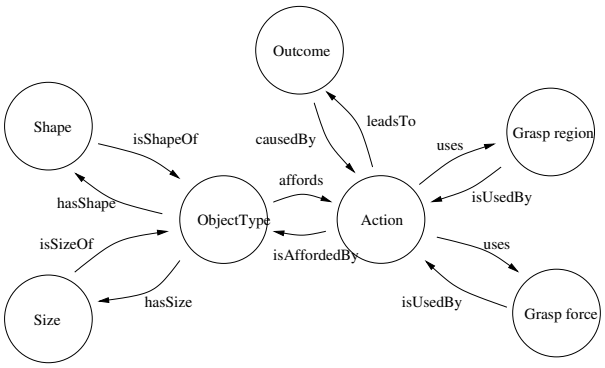


Fig. 5. The OWL ontology for grasping consists of seven classes and corresponding object properties. The main conceptual difference between this model and the probabilistic version is the inclusion of object type and action.

perception system extracts information from GraspIt! and performs the classification according to the grasping ontology. The knowledge-base (i.e. part of mental model) uses the controller and perception system to gather information from the environment. Reasoning is performed by querying the OWL reasoner (Pellet v1.5) and expanding the ontology as appropriate. The ontology presented was developed and tested using Protégé 4.

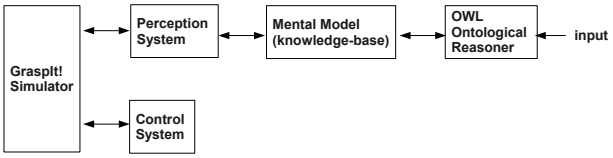


Fig. 6. The ontological approach is composed of five components.

A. Grasp inferencing

The key operation that a semantic reasoner is required to perform is to infer appropriate grasp parameters based on observations. From a semantic standpoint the operation is represented as:

$$
\begin{aligned}
&\mathit{object} && \equiv \mathit{objectType} \sqcap \mathit{hasShape}(\mathit{shape})\\
& && \quad \sqcap \mathit{hasSize}(\mathit{size})\\
&\mathit{action} && \equiv \mathit{actions} \sqcap \mathit{affords}(\mathit{object})\\
& && \quad \sqcap \mathit{leadsTo}(\mathit{Success})\\
&\mathit{graspRegion} && \equiv \mathit{graspRegion} \sqcap \mathit{usedBy}(\mathit{action})\\
&\mathit{graspForce} && \equiv \mathit{graspForce} \sqcap \mathit{usedBy}(\mathit{action})
\end{aligned}
$$

Potentially, there may exist several actions that the object affords and that lead to success. For the purpose of our experiments we randomly pick an action from the set of afforded actions. In the event that $\emptyset \equiv \mathit{objectType} \sqcap \mathit{hasShape}(\mathit{shape}) \sqcap \mathit{hasSize}(\mathit{size})$ we add a new sub-class to *objectType* and the object properties *newSubClass hasSize(size)* and *newSubClass hasShape(shape)* as well as the inverse properties, thus extending the ontology definition. The new object type will not afford any actions, so we will make a reasonable guess:

$$
\begin{aligned}
&\mathit{objectOfSize} && \equiv \mathit{objectType} \sqcap \mathit{hasSize}(\mathit{size})\\
&\mathit{actionSize} && \equiv \mathit{actions} \sqcap \mathit{affords}(\mathit{objectOfSize})\\
& && \quad \sqcap \mathit{leadsTo}(\mathit{Success})\\
&\mathit{objectOfShape} && \equiv \mathit{objectType} \sqcap \mathit{hasShape}(\mathit{shape})\\
&\mathit{actionOfShape} && \equiv \mathit{actions} \sqcap \mathit{affords}(\mathit{objectOfShape})\\
& && \quad \sqcap \mathit{leadsTo}(\mathit{Success})
\end{aligned}
$$

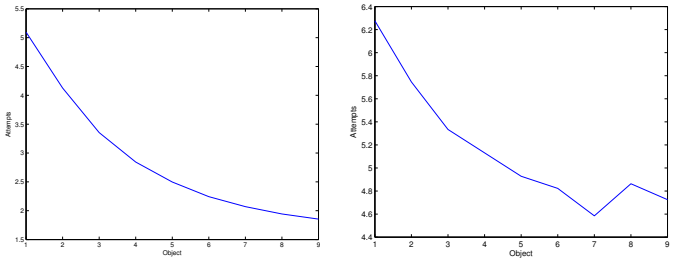
If $\mathit{actionSize} \sqcap \mathit{actionShape} \not\equiv \emptyset$ we simply pick one action and attempt to execute it, else we try $\mathit{actionShape} \not\equiv \emptyset$ and $\mathit{actionSize} \not\equiv \emptyset$ to find an action to attempt. After the action is performed we create a new sub-class for

action and define *newObject affords(newAction)*, *newAction uses(graspRegion)*, *newAction uses(graspForce)* and *newAction hasOutcome(outcome)*.

The ontology requires continuous updating. That is, new rules must be built in the presence of new situations not previously accounted for. This is particularly evident when grasp failures arise. Consider the case that we infer an action that we believe will lead to *success*, however the executed action produces the response *graspFail*. In this event the reasoning system must also be capable of understanding what went wrong during the grasping operation, and how to correct the failure. This requires the system to adapt and change/update itself as new information is made available. Although not considered here, diagnosing and correcting failures encountered following grasp execution is an important aspect of grasping research and it is a part of our ongoing work.

B. Results

We have conducted an initial set of experiments which uses our ontological reasoning engine. In this section we compare our findings with those presented earlier for Approach I which uses a voting function for selecting appropriate grasp choices. As shown in Figure 7, our ontology-based reasoning system appears to outperform results reported for Approach I for this set of experiments. This figure shows that over 1000 trials, an ontology-based reasoning system, that uses a developmental learning approach as was also used by Approach I, learns what objects afford what grasping actions for successful outcomes much sooner in comparison. This is indicated by the lower number of attempts recorded for Approach II over Approach I for the 9 objects grasped. This is particularly evident for later objects in the object set. As shown, by object 6, Approach II requires somewhere in the order of 2 attempts to learn a successful grasp choice whereas Approach I still requires approximately 4.5 attempts. These findings suggest that using a developmental learning approach with limited initial knowledge, a rule-based approach may be more feasible for modeling initial object-action affordances which are then updated in a probabilistic framework.



(a) O1: Results for number of successful grasps made over 9 objects for 1000 trials using our ontology-based reasoner.

(b) O2: Values previously reported for number of successful grasps made over 9 objects for 1000 trials using our voting function approach previously presented.

Fig. 7. Results from experiments run with voting function probabilistic approach, and ontological reasoner approach for 1000 trials.

IV. Discussion and Future Work

We have presented two learning approaches for modeling affordances between objects, actions and effects. We first discussed a probabilistic approach which employs a voting function to select grasping choices for a given object. Using variations in the combination of probabilities calculated we observe that a comprehensive mixture of both general and specific grasping options provides the most successful grasping outcomes. We then presented a second approach which uses an ontological framework. This approach employs a rule-based system that uses axioms to build relationships and to infer what grasps to choose given a set of objects. As shown from our experiments our ontological rule-based approach outperforms our initial probabilistic approach over time. However, coupling both of these approaches may improve overall performance for other experimental cases. For example, a rule-based system can be used within a probabilistic framework to diagnose and reason on grasp failures encountered. Rules can be used to search for possible failure causes and their corresponding corrective actions. Alternatively, we can also explore using a rule-based approach initially to make grasp choices. From here the system could then transition to a probabilistic approach when larger amounts of data become available. These cases will be considered for future work.

References

[1] J. Gibson, "The theory of affordance," *Percieving, Acting and Knowing*, vol. 25, p. 365, 1977.

[2] G. Fritz, L. Paletta, M. Kumar, G. Dorffner, R. Breithaupt, and E. Rome, "Visual learning of affordance based cues," *From Animals to Animats 9*, pp. 52–64, 2006.

[3] E. Erdemir, C. Frankel, K. Kawamura, S. Gordon, S. Thornton, and B. Ulutas, "Towards a cognitive robot that uses internal rehearsal to learn affordance relations," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nice, France, September 22-26 2008, pp. 2016–2021.

[4] E.Ugur, M. R. Dogar, M. Cakmak, and E. Sahin, "The learning and use of traversability affordance using range images on a mobile robot," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2007.

[5] S. Hidayat, B. Kim, and K. Ohba, "Learning affordance for semantic robots using ontology approach," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nice, France, September 22-26 2008, pp. 2630–2636.

[6] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, "Learning Object Affordances: From Sensory–Motor Coordination to Imitation," *IEEE Trans. on Robotics*, vol. 24, no. 1, pp. 15–26, 2008.

[7] K. Huebner and D. Kragic, "Selection of Robot Pre-Grasps using Box-Based Shape Approximation," in *IEEE Int. Conference on Intelligent Robots and Systems*, 2008, pp. 1765–1770.

[8] K. Hübner and D. Kragic, "Selection of Robot Pre-Grasps using Box-Based Shape Approximation," in *IEEE Int. Conference on Intelligent Robots and Systems*, 2008, pp. 1765–1770.

[9] A. T. Miller and P. K. Allen, "Graspit! A Versatile Simulator for Robotic Grasping," *IEEE Robotics and Automation Magazine*, vol. 11, no. 4, pp. 110–122, 2004.

[10] A. Cregan, M. Mochol, D. Vrandecic, and S. Bechhofer, "Pushing the limits of owl, rules and protégé," in *Proceedings of OWL Experiences and Directions Workshop*, Galway, Ireland, November 2005.