Object Affordances by Inferring on the Surroundings*

Paola Ardón Ramírez[1], Subramanian Ramamoorthy[2] and Katrin Solveig Lohan[3]

***Abstract*— Robotic cognitive manipulation methods aim to imitate the human-object interactive process. Most of the of the state-of-the-art literature explore these methods by focusing on the target object or on the robot's morphology, without including the surrounding environment. Most recent approaches suggest that taking into account the semantic properties of the surrounding environment improves the object recognition. When it comes to human cognitive development methods, these physical qualities are not only inferred from the object but also from the semantic characteristics of the surroundings. Thus the importance of affordances. In affordances, the representation of the perceived physical qualities of the objects gives valuable information about the possible manipulation actions. Hence, our research pursuits to develop a cognitive affordances map by (i) considering the object and the characteristics of the environment in which this object is more likely to appear, and (ii) achieving a learning mechanism that will intrinsically learn these affordances from self-experience.**

***Index Terms*— Humanoid robot, affordances, object recognition, learning, grasping**

I. Introduction

A. Motivation

Humanoid robots are playing increasingly important roles when it comes to indoor applications, for which object affordances are vital to succeed in the human-robot interaction task. Some of these applications include assisting humans in daily activities such as cooking, cleaning, shopping, among others, thus the importance of improving robotic grasp affordances, especially in dynamic environments.

Affordance is defined as "an opportunity for action", [7]. In robotics, we are interested in object affordances; investigating the best procedure to imitate the cognitive human development on how to interact with objects, [9]. There is a wide range of theories that try to explain the human thinking, none of them taken as the ground truth one, thus it is not surprising that the development of robotic cognitive techniques is still a wide area of research. Humans heavily rely on shapes and environments to identify and categorize objects in order to infer an action ([4], [13], [6]). As a result, we succeed at generalizing an action towards objects of the same category with significantly different shapes, e.g, glasses: wine, tumbler, martini, etc., and to differentiate how to manipulate objects with similar shapes but for different purposes, e.g, bowling pin vs. water bottle or a candle vs. a glass full of liquid. In robotics, the most common approach to affordance learning is to learn direct mappings from sensor measurements to affordance labels ([2], [3], [8], [10], [12]). However, the accuracy of this mapping is constrained by how good the perception and reconstruction of the object is, not to mention the robotic morphology constraints.

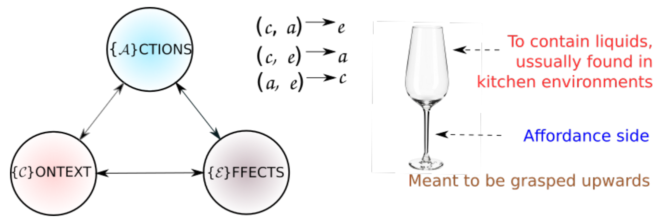

Fig. 1. Affordances model originally presented in [12], which creates a correlation between the objects and their properties as being detected by the robot sensors. We consider a slightly modified setting using reinforcement learning where: $\mathcal{C} = \{c_1, c_2, ..., c_n\}$ will be the set of semantic attributes of the object and the environment, $\mathcal{A} = \{a_1, a_2, ..., a_n\}$ the set of available actions and $\mathcal{E} = \{e_1, e_2, ..., e_n\}$ the effects of performing those actions as detected by the sensors. In this model, the relationship among components of two sets infers on the best match component from the third set.

B. Problem Statement and Hypothesis

In order to achieve cognitive grasping processes, there are two main approaches in the literature. On one hand, some of the methods focus on extracting viable grasping points on the objects, independently if the object is known, familiar or novel to the system. Examples of such works are [10], [3], [1], [16], [5], among many others. These data-driven methods use these extracted features to improve their grasping success rate. However, because of the need to constantly keep learning they require large amounts of data and are not well generalized among objects belonging to different categories. On the other hand, some works focus on learning the grasping task based on the robot's morphology using simple object primitive shapes such as spheres and boxes ([2], [8], [11]). These two different procedures consider an isolated target or many objects on a planar surface, which do not reflect real-world scenarios. Additionally, these two different approaches perform well independently, however, the literature does not put together *what are the features that encode the good object affordances?* These affordances do not belong strictly to the object nor to the robotic agent, instead, they are the result of the relationship established between them.

Social research studies on the development of human cognitive methods demonstrate that we humans improve our interactive learning with objects not only based on our previous experience with them (or similar ones) but also by inferring in the context of the environment where these objects reside ([15], [14]). Thus, we create a relationship

*Thanks to ORCA Hub EPSRC (EP/R026173/1, 2017-2021) and consortium partners

[1]Paola Ardón is with the School of Mathematical and Computer Science at Heriot-Watt University and with the School of Informatics at University of Edinburgh, Edinburgh, UK paola.ardon@ed.ac.uk

[2]Subramanian Ramamoorthy is with the School of Informatics, University of Edinburgh, Edinburgh, UK s.ramamoorthy@ed.ac.uk

[3]Katrin Solveig Lohan is with the School of Mathematical and Computer Science, Heriot Watt University, Edinburgh, UK k.lohan@hw.ac.uk

between the object, the scenario where is more likely to find it, and the set of possible actions to interact with it. Using the same analogy, in robotics, the object affordances can be improved by integrating semantic attributes of the object and the environment in which these objects are usually found, which is an approach not yet seen in the current literature.

C. Objectives

This research project aims to investigate object affordances to improve the manipulation success rate by including the context of the environment when building the relationship map between the target object and the agent, e.g, humanoid robot. For this purpose, we want to create a learning mechanism based on previous experience that intrinsically generates the reward of a successful grasp, with the purpose of avoiding the use of external datasets. Figure 1 is the common used affordances model, [12], modified for our proposal along with a toy affordances example. In our case, the set of semantic properties will be composed by the object and the environment. And, the set of actions and effects will be the result of the robot's own experience.

II. Method

The project comprises the following sequential stages:

A. Visual Features

This stage will explore how to improve object recognition, by correlating it with the environment it is most likely located in. It will be based on early cognitive vision (ECV) descriptors containing information about shape, texture and categorical classification of the objects, as well as to give valuable information on segmenting the foreground (unknown object) and the background (environment). Thus, it is twofold: (i) the robot first interacts (visually) with the object in order to acquire a model, and (ii) once the model has been obtained it can be used for segmenting the background and learn the relationship affordances map.

B. Affordances Learning

For learning affordances, we will explore the use of reinforcement learning techniques. Instead of relying on extrinsic reward signals we will explore the usage of intrinsic ones in order for the system to experience the *success of grasping*, just as living creatures learn the skill hierarchies [14]. This approach aims to overcome the large number of samples needed for the same task using methods such as Bayesian networks [11] and learning by demonstration [3], [8].

C. Reach and Grasp Planning

This stage will be achieved by using a motion planner that will guide the end-effector towards the automatically computed grasping point. Using an on-hand camera will allow readjusting the grasping point, which will lead to a motion planner with online capabilities able to work in a dynamic environment.

D. Testing our Method

This stage aims to answer the following questions: (i) can the system identify the right object? We will use object recognition benchmarking metrics to address this question (ii) does it choose the right action for the object? for which we will measure the grasp success based on the grasp stability.

III. Final Remarks

Past research has presented approaches to the affordance problem extensively. Nonetheless grasping is still an open challenge due to the large variety of object shapes and robotic platforms. The current state of the art methods is limited to specific robot manipulator, grasping scenarios, and objects. Further, the current approaches need a large amount of data to train the learning model without being able to successfully generalize among different classes of objects. Thus we aim to build a cognitive grasping framework that is able to identify and encapsulate the good features of an object that give valuable information about its affordances while learning from its own experience. This task should not only be limited to the relationship that can be built between the target object and the agent but also considering the environment surrounding the object.

References

[1] R. Ala, D. H. Kim, S. Y. Shin, C. Kim, and S.-K. Park, *A 3D-grasp synthesis algorithm to grasp unknown objects based on graspable boundary and convex segments*, Information Sciences, 295 (2015), pp. 91–106.

[2] J. Bonaiuto and M. A. Arbib, *Learning to grasp and extract affordances: the Integrated Learning of Grasps and Affordances (ILGA) model*, Biological cybernetics, 109 (2015), pp. 639–669.

[3] H. Dang and P. K. Allen, *Robot learning of everyday object manipulations via human demonstration*, IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings, (2010), pp. 1284–1289.

[4] H. P. O. de Beeck, K. Torfs, and J. Wagemans, *Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway*, Journal of Neuroscience, 28 (2008), pp. 10111–10123.

[5] C. Dune, E. Marchand, C. Collowet, and C. Leroux, *Active rough shape estimation of unknown objects*, in Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on, IEEE, 2008, pp. 3622–3627.

[6] L. Fadiga, L. Fogassi, V. Gallese, and G. Rizzolatti, *Visuomotor neurons: Ambiguity of the discharge or motorperception?*, International journal of psychophysiology, 35 (2000), pp. 165–177.

[7] J. G. Greeno, *Gibson's affordances.*, American Psychological Association, (1994).

[8] T. Hermans, J. M. Rehg, and A. Bobick, *Affordance prediction via learned object attributes*, in IEEE International Conference on Robotics and Automation (ICRA): Workshop on Semantic Perception, Mapping, and Exploration, Citeseer, 2011, pp. 181–184.

[9] T. E. Horton, A. Chakraborty, and R. S. Amant, *Affordances for robots: a brief survey.*, AVANT. Pismo Awangardy Filozoficzno-Naukowej, 2 (2012), pp. 70–84.

[10] I. Lenz, H. Lee, and A. Saxena, *Deep learning for detecting robotic grasps*, International Journal of Robotics Research, 34 (2015), pp. 705–724.

[11] B. Moldovan, P. Moreno, M. van Otterlo, J. Santos-Victor, and L. De Raedt, *Learning relational affordance models for robots in multi-object manipulation tasks*, in Robotics and Automation (ICRA), 2012 IEEE International Conference on, IEEE, 2012, pp. 4373–4378.

[12] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, *Learning object affordances: From sensory–motor coordination to imitation*, IEEE Trans. Robotics, 24 (2008), pp. 15–26.

[13] E. Oztop, N. S. Bradley, and M. A. Arbib, *Infant grasp learning: a computational model*, Experimental brain research, 158 (2004), pp. 480–503.

[14] J. Piaget and M. Cook, *The origins of intelligence in children*, vol. 8, International Universities Press New York, 1952.

[15] J. V. Wertsh and P. Tulviste, *Apprenticeship in thinking: Cognitive development in social context*, Science, 249 (1990), pp. 684–686.

[16] P. Zech and J. Piater, *Active and transfer learning of grasps by sampling from demonstration*, (2016).