Towards Robust Grasps: Using the Environment Semantics for Robotic Object Affordances

Paola Ardón and **Èric Pairet** and **Subramanian Ramamoorthy** and **Katrin Solveig Lohan**
Edinburgh Centre for Robotics
Heriot-Watt University and University of Edinburgh, UK.
{paola.ardon,eric.pairet,s.ramamoorthy}@ed.ac.uk;k.lohan@hw.ac.uk *

Abstract

Artificial Intelligence is essential to achieve a reliable human-robot interaction, especially when it comes to manipulation tasks. Most of the state-of-the-art literature explores robotics grasping methods by focusing on the target object or the robot's morphology, without including the environment. When it comes to human cognitive development approaches, these physical qualities are not only inferred from the object, but also from the semantic characteristics of the surroundings. The same analogy can be used in robotic affordances for improving objects grasps, where the perceived physical qualities of the objects give valuable information about the possible manipulation actions. This work proposes a framework able to reason on the object affordances and grasping regions. Each calculated grasping area is the result of a sequence of concrete ranked decisions based on the inference of different highly related attributes. The results show that the system is able to infer on good grasping areas depending on its affordance without having any *a-priori* knowledge on the shape nor the grasping points.

INTRODUCTION

Humanoid robots are playing increasingly important roles when it comes to indoor applications. Consider a robot assisting humans by finding, collecting and delivering an object. In such complex and dynamic environments, it is hard to provide the system with every possible representation of objects. This limitation can confuse the system into reaching very similar objects with completely different purposes, such as a candle for a glass full of liquid. Thus, the importance of a rich common sense library on object affordances that holds the start of robust robotics grasps.

Affordance is defined as "an opportunity for action" (Greeno 1994). Thus the interest in robotics on objects affordances and in artificial intelligence to investigate the best procedure to imitate the cognitive human development on how to interact with objects (Horton, Chakraborty, and Amant 2012). There is a wide range of theories that try to explain the human thinking, none of them taken as the ground truth one.

*The authors would like to acknowledge the support of the EPSRC IAA 455791 along with ORCA Hub EPSRC (EP/R026173/1, 2017-2021) and consortium partners.

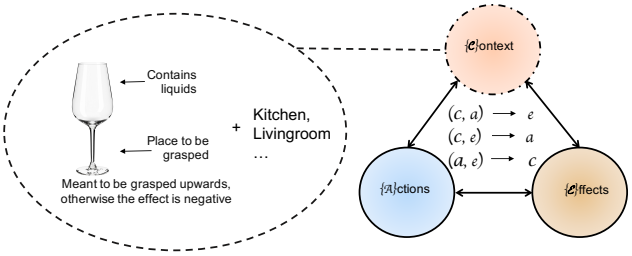


Figure 1: Affordances map model to create a correlation between the objects properties and their environment to improve on robotic grasps.

Thus, it is not surprising that the development of artificial intelligence is still a wide area of research. Humans heavily rely on shapes and environments to identify and categorise objects in order to infer an action (de Beeck, Torfs, and Wagemans 2008; Oztop, Bradley, and Arbib 2004). As a result, humans succeed at generalising an action towards objects of the same category with significantly different shapes, e.g. glasses: wine, tumbler, martini, etc., and differentiate how to manipulate objects with similar shapes but for different purposes, e.g. bowling pin vs water bottle.

In robotics, the most common approach to affordances is to learn direct mappings to labels (Bonaiuto and Arbib 2015; Hermans, Rehg, and Bobick 2011; Lenz, Lee, and Saxena 2015; Montesano et al. 2008). However, this mapping accuracy is constrained by the amount of data needed to learn the grasping areas in each of the affordance groups. These learning methods do not reveal *what are the features that encode the good object affordances?* Namely, these affordances do not strictly belong to the object itself. Instead, they are the result of the relationship established between them and the surroundings. Moreover, to engage in an interaction with humans, the robot has to be able to represent and reason with different sources of knowledge and decrease the already eminent uncertainty in the environment (Pairet et al. 2018b).

Studies on the development of human cognitive methods demonstrate that humans improve the interactive learning process with objects not only based on previous experience with them (or similar ones) but also by inferring in the context of the environment where these objects reside (Wertsh and Tulviste 1990). As a result, creating a relationship be-

tween the object, the scenario where it is more likely to be found, and the set of possible actions to interact with. Using the same analogy, in robotics, obtaining the grasp actions depending on the object affordances can be improved by integrating semantic attributes of the object and the environment in which these objects are usually found.

This paper summarises an architecture to address the challenges previously described. The presented solution builds upon the assumption that, the robot visual feedback represents a good source of information. Thus, the focus on the improvement of affordances reasoning and actions. The work establishes its foundations on the affordances map presented in Figure 1 (Montesano et al. 2008), particularly on its *context* element where the affordance identification resides. In this work the context $\mathcal{C} = \{c_1, c_2, ..., c_n\}$ is modified to be the set of semantic attributes of the object and the environment that build upon the affordance; while $\mathcal{A} = \{a_1, a_2, ..., a_n\}$ the set of available actions and $\mathcal{E} = \{e_1, e_2, ..., e_n\}$ the effects of performing those actions are kept as the original model.

The framework allows the system to model an unknown object and to reason on its affordance by correlating features from the target and its environment. This with the objective of calculating the best possible grasping region which is highly related to the object's affordance group. Each abstract grasping area is the result of a sequence of concrete ranked decisions based on the inference of different highly related attributes. The system combines object reconstruction methods based on geometrical approaches and deep learning techniques that delivers an efficient Knowledge Base (KB) for object affordances grasping behaviours useful in indoor environments.

RELATED WORK

Despite the wide range of methods for robotic grasps, this summary focuses on those that do not need *a-priori* information about the object in order to reconstruct it and methods that focus on the object affordance independently of the object grasp action.

Object Modelling for Grasping Based on Geometry

There are works that profit from superquadric modelling to then extract the possible grasps of an object using classifiers, (Goldfeder et al. 2009; Vezzani, Pattacini, and Natale 2017). (Goldfeder et al. 2009) integrates shape primitives and superquadrics, but the object representation is a multi-level superquadric tree. This tree is created using a decomposition of the initial model, which contains the shape primitives. After a pruning routine, a subspace containing a set of suitable grasps is obtained. (Vezzani, Pattacini, and Natale 2017) uses the superquadric modelling for both the object and the end-effector showing the method to be successful at computing the grasping area of the object and the desired pose of the end-effector.

Object Affordances for Grasping

Many methods extract viable grasping points on the objects, independently if the object is known or novel to the system, thus not explicitly considering the target's affordance. Examples of such works are (Ardón, Dragone, and Erden 2018; Lenz, Lee, and Saxena 2015; Zech and Piater 2016), to mention some. Others focus on learning the robot's control and dynamic models to achieve a grasp, such as (Stoytchev 2005; Bonaiuto and Arbib 2015). The latter learn grasp affordances from motor parameters to plan grasps using trial-and-error reinforcement learning. (Stoytchev 2005) follows psychology theories such as the ones presented in (Greeno 1994) to learn from exploratory behaviours the invariants in the resulting set of observations for the grasps.

There are also those who focus on the object affordances themselves without taking into account the grasping region. An example is (Moldovan et al. 2012) that implement a Bayesian network probabilistic method to learn to differentiate affordances models among two objects. Their proposed method shows good results under uncertainty.

In the vast repertoire of learning methods connecting affordances, not necessarily limited to objects, some works try to mimic the human reasoning by building a KB of actions based on tasks built upon reinforcement learning (Zhu, Fathi, and Fei-Fei 2014; Sridharan 2017). Instead, (Montesano and Lopes 2009; Kraft et al. 2009; Madry, Song, and Kragic 2012) learn the visual descriptors of the objects using classifiers, such as support vector machine (SVM), to categorise the objects and obtain the possible grasps. Others such as (Nguyen et al. 2017; Do, Nguyen, and Reid 2018) use classifiers alone to build a model using deep Convolutional Neural Networks (CNN) based on the visual objects features, resulting in a plausible generalised method given the robustness of their data.

PROPOSED SOLUTION AND SYSTEM INTEGRATION

The proposed framework is divided into two sub-stages as shown in Figure 2. This method focuses on modelling the object and extracting the valuable features of the target and its surrounding environment. These sets of features allow the system to deduce the target's affordance to improve the grasping actions.

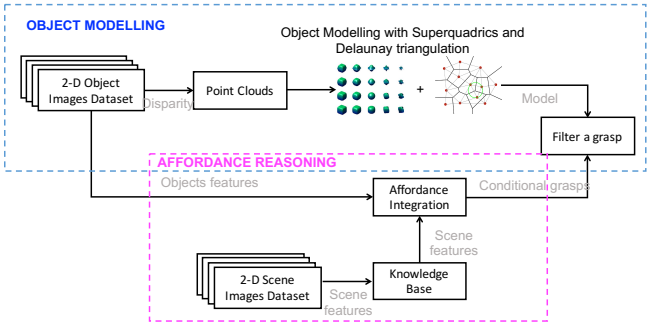


Figure 2: (a) Proposed solution to a grasping framework using affordances theory, where the context consists not only of the object but also of the environment features.

Object Modelling

Learning techniques are part of the state-of-the-art when it comes to extracting the grasping points on objects. However, they bring some limitations, such as to collect or find a suitable dataset that maps the two-dimensional (2-D) images to the labelled three-dimensional (3-D) grasping points. This approach models the object using a combination of superquadric modelling and Delaunay triangulation allowing the system to grasp novel objects without any *a-priori* information. Superquadrics are a family of geometric shapes similarly defined as ellipsoids and other quadrics, except that the squaring operations are replaced by arbitrary powers that are the ones that adapt the shape to the surface of the perceived object. The framework starts by approximating the object to a superquadric model (Jaklic, Leonardis, and Solina 2013):

$$F(x,y,z,\boldsymbol{\lambda}) : \left(\left(\frac{x}{\lambda_1} \right)^{\frac{2}{\lambda_5}} + \left(\frac{y}{\lambda_2} \right)^{\frac{2}{\lambda_5}} \right)^{\frac{\lambda_5}{\lambda_4}} + \left(\frac{z}{\lambda_3} \right)^{\frac{2}{\lambda_4}}, \tag{1}$$

where (x, y, z) is a 3-D point in the superquadric model and $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_5]$ defines the superquadric shape. Equation 1 provides a simple test whether a given point lies inside or outside a superquadric:

$$P(x,y,z) = \begin{cases} F < 1, & inside \\ F = 0, & on\ surface \\ F > 0, & outside \end{cases} \tag{2}$$

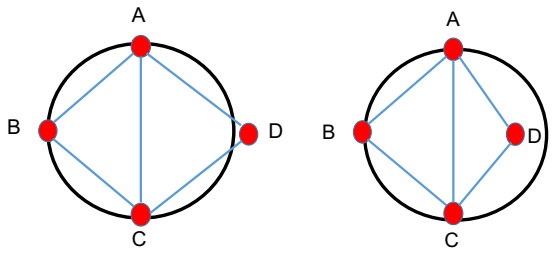
Nonetheless, one of the known problems of superquadrics is that it samples more points around the curvatures of the perceived shape (Jaklic, Leonardis, and Solina 2013).

Thus, in order to extract grasping points along the whole surface of the object, the superquadric is combined with a Delaunay triangulation. A Delaunay triangulation considers a set **P** of points in the (D-dimensional) Euclidean space. An example is shown in Figure 3. For a triangulation to be Delaunay no point in **P** should be inside the circumcircle shaped by the D-dimensional triangulation DT, with the angle vectors composed by the points in **P**, DT(**P**), formed by four chosen points inside **P** (Lee and Schachter 1980). In two dimensions, one way to detect if a point D lies in the circumcircle of A, B, C is to evaluate the determinant

$$\begin{vmatrix} A_x & A_y & A_x^2 + A_y^2 & 1 \\ B_x & B_y & B_x^2 + B_y^2 & 1 \\ C_x & C_y & C_x^2 + C_y^2 & 1 \\ D_x & D_y & D_x^2 + D_y^2 & 1 \end{vmatrix} > 0, \tag{3}$$

where A, B and C are sorted counterclockwise. This determinant is then positive, if and only if, D is inside the cirumcircle. The vertices of the Delaunay triangulation are the ones extracted as the grasping points of the object.

Figure 4 shows the process of visualising the grasping region on the object. A superellipsoid is matched using the dimensions of iCub humanoid robot end-effector (Metta et al. 2008). This superellipsoid and the robot's hand model are portrayed in Figures 4(a) and 4(b), and an example of a modelled object with the obtained grasping region is shown in Figures 4(c) and 4(d) respectively.

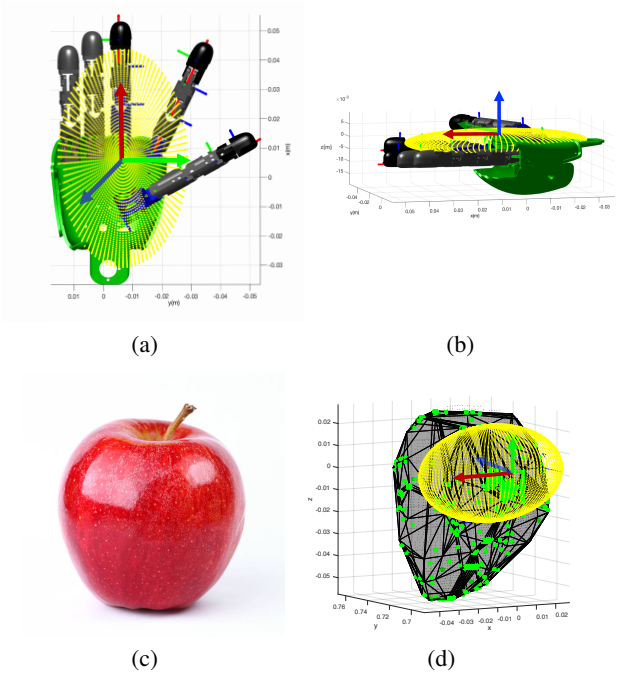


(a) (b)

Figure 3: Delaunay Triangulation example. (a) Delaunay triangulation, (b) not a Delaunay triangulation

Building the Knowledge Base

While the previous module does not need any *a-priori* information on the object to obtain a model, reasoning about the object affordance needs a library of features that gives some background about its correct affordance group. Knowledge Base (KB) methods are growing in artificial intelligence. They learn a set of general rules and features that allow the system to infer about an object or an action. Moreover, this method is not restricted to the output task, but it also allows



(a) (b) (c) (d)

Figure 4: End-effector and object modelling. (a)-(b) show iCub humanoid Robot end-effector CAD model with its superellipsoid in yellow (axis colors: x is red, y is green and the z is blue); (c) target object used for the sample reconstruction; (d) point cloud reconstruction using superquadrics and Delaunay triangulation, the detected grasping points are shown in green and the final location of the end-effector in yellow.

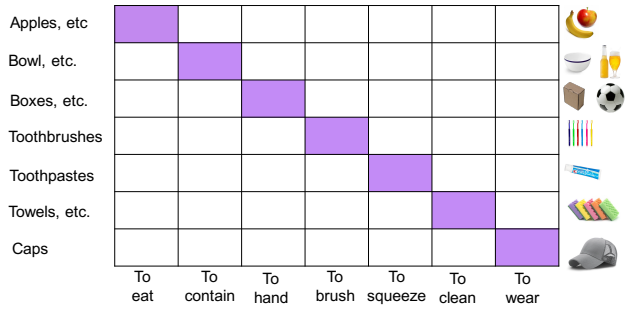


Figure 5: Objects used for our framework from the Washington dataset and the different affordances groups.

the system to query a larger array of questions regarding the features involved in the process.

In this work, a KB graph is used as a predictive model to an object affordance. The system collects a set of attributes about the objects and the environment, to then connect them in a graph style based on a set of general rules that defines the relationship among these attributes. Consequently, allowing the system to reason about the affordance group and the previously calculated grasping points. This KB consists of two steps: collecting data and learning this data relationship to reason on the affordance for grasping.

Collecting data: This is the repository of images collected from two datasets that are finally organised in the affordance categories shown in Figure 5. The first one is the Washington-RGB dataset that contains 300 objects and 51 different classes, providing the point clouds and the 2-D images for each one of the instances (Lai et al. 2012). The second dataset is the MIT Indoor scene recognition that contains 15620 different images of 67 different indoor environments (Quattoni and Torralba 2009).

Both datasets are split into 70% for training and the remaining 30% for testing. These subsets are used to train and test a battery of classifiers that help with defining good object affordances features.

Learning the knowledge base using the environment: A KB is visualised as a graph representation as illustrated in Figure 6 where the entities (nodes) are connected by general rules (edges). In this proposed solution, the entities include the target object, the object attributes and the resulting affordances groups. The general rules are the attribute to attribute relation. Weights define this relation, where the higher the weight, the higher the correlation between the two entities. The previously collected repertoire of images is used to define the attributes portrayed in Table 1 about the object:

- Shape attributes: This is defined as the set of visual attributes that describe the objects geometrical appearance,
- Texture attributes: Are a set of categories based on visual characteristics of the objects materials,
- Categorical attributes: Reflecting the semantic understanding of the object. For example, an apple is food, and

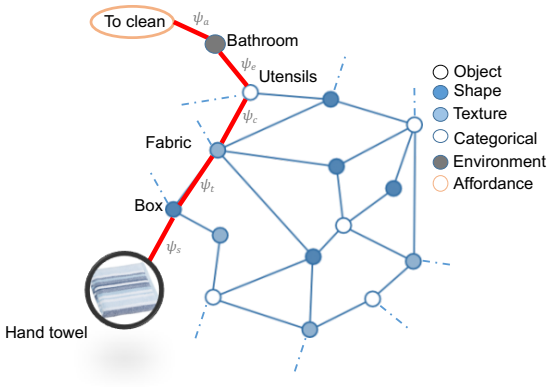


Figure 6: Example of a cleaning object and the extracted attributes used to build the KB graph learning the positive weights $\mathbf{\Psi}$ (shown in red) that result in an affordance group.

- Environment attributes: The scenarios in which the objects are more likely to be found in. This attribute is added with the purpose of facilitating the object affordances reasoning.

Figure 7 illustrates the hierarchical inference procedure followed in the KB, to arrive to an affordance group. This KB is built using four different deep learning neural networks that, through the pre-trained CNN resnet50, (He et al. 2016), extract features from the perceived images. These four different deep learning CNN correspond to the different attributes that define an entity set of the graph.

The KB is then a predictive model based on the hierarchical information obtained from the different attributes of the object (visualised as nodes in Figure 6) and the defined general rule that correlates attributes (the edges in Figure 6 from now on referred as weights). From each of the attributes, a set of weights $\Psi_{A_i} = \{\psi_1, \psi_2, ..., \psi_n\}$ is extracted hierarchically to infer on the next best entity candidate, where $\{A_i\}$ is an attribute and n the total number of entities in that attribute. The higher the ψ_n the higher the probability that the connected entities result in a better affordance inference. These weights are proportional to the posterior probability distribution obtained from the classification task. Such that the posterior probability distribution is defined as the Bayes rule:

$$\widehat{P}(a|x) = \frac{P(x|a)P(a)}{P(x)}, \tag{4}$$

Table 1: Used attributes and entities of the KB graph.

Attribute	**Attribute Categories**
Shape	box, cylinder, irregular, long, round
Texture	aluminium, cardboard, coarse, fabric, glass, plastic, rubber, smooth
Categorical	container, food, personal, miscellaneous, utensils
Environment	bathroom, bedroom, children room, closet, kitchen, livingroom, office

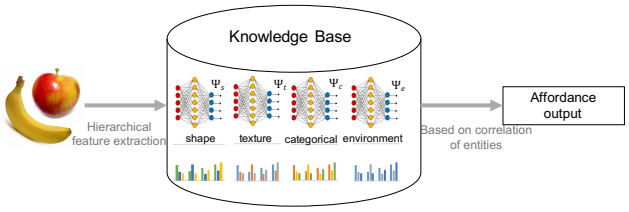


Figure 7: KB representation used for the object affordance inference. Given an image, the model estimates the attributes features in a hierarchical manner following the stated inference rule. These attributes are then accessible information on the KB. A predictive model is then applied to select the object affordance.

where x is an image belonging a class a, $P(a)$ is the posterior distribution and $P(x)$ is a normalisation constant that consists of the sum over a of the multivariate normal density. Figure 6 depicts an example of a cleaning or to hand over object, where the weights deduce the best path (shown in red) to the *to clean* affordance.

The collected information from each of the deep CNN is then learned using a decision tree as a predictive model,

$$(\boldsymbol{y}, Z) = (y_1, y_2, y_3, ..., y_n, Z), \tag{5}$$

where Z is the affordance group that the system is trying to infer and the vector $\boldsymbol{y}$ is the set of features $\{y_1, y_2, y_3, ..., y_n\}$, described as attributes categories in Table 1, used for the inference task.

Selecting the new grasping points: Once the object is classified into an affordance category, the grasping region is limited accordingly. The system selects from the set of grasping points obtained in the object reconstruction module and limits the grasps depending on the affordance action-effect of the object in the following manner:

- The grasping region should be in the middle and up-wards for objects that are meant to contain edibles.
- For the rest of objects, it is considered as the grasping region those areas where the density of grasping points is higher, given that the affordance action-effect is not critical (i.e., hand over, to brush, etc.).

SYSTEM RESULTS AND DISCUSSION

The results of the presented KB for object affordances including the environment features are presented in this section. As a reminder, the proposed framework is able to reason on the object affordance. In this work, affordance is understood as an action-effect relation of an object, with the purpose of discerning on the best possible grasps.

The current literature in affordances for grasping behaviours uses labelled grasping regions on the targets to train on the object affordance. Given that this approach aims to reason on the object grasping affordance without having any *a-priori* knowledge about its grasping regions, the presented evaluation of the results is done qualitatively.

Table 2: Each of the deep CNN accuracy performance.

Classifier	Accuracy
Shape	95.71%
Texture	98.83%
Categorical	99.91%
Environment	76.50%

Reasoning on the Object Affordance

The first tests are done individually on each of the deep learning CNN that build up the KB. 30% of the images from the Washington-RGB dataset were used for testing the battery of classifiers. Table 2 presents a summary of their accuracies, whereas exhaustively presented in literature, the environment recognition is the hardest classification to boost. Even though the aim of the proposed framework is not exclusively to improve the performance of the individual classifiers, these illustrated results match the state-of-the-art results shown in (He et al. 2016; Lai et al. 2012). In order to evaluate the overall performance of the KB the accuracy and probabilities distributions before and after adding the environment features were collected.

Figures 8 and 9 show the data for both cases. Not including the environment in the affordances has lower accuracy than adding these features to the KB, as illustrated in Figures 8(a) and 8(b). Furthermore, Figure 8(a) also shows a slightly higher spread among different affordance classes. For example, the case of affordances which objects have a general semantic categorical attribute such as "miscellaneous" or "container". A percentage of objects get confused among the *to contain, to brush, to eat,* and *to squeeze* categories. Regarding grasping, this miscue represents a significant negative effect, especially for objects which real affordance is *to contain* and its misclassification results in the system ignoring the lifting-up orientation of the object, thus dropping the food or liquid inside the object. This case is reduced by 4.24% when adding the environment features, as portrayed in Figure 8(b).

The posterior probability distribution of the objects among each category is also improved. Figures 9(a) and 9(b) show the overall increase in the median probability of the objects in the different affordances categories. Further, there is a notable decrement in the distribution of categories such as *to contain, to hand, to brush*, and *to eat* meaning that the model is more confident about the classification.

Obtained Grasping Points

The final goal is to obtain a system that, without any *a-priori* knowledge about the grasping regions of the objects, is able to reason on the affordance category and calculate the best possible grasping region. Figure 10 shows examples of different objects from which the grasping areas were extracted, before and after, inferring on the affordance of the target. These grasp regions are analysed qualitatively according to

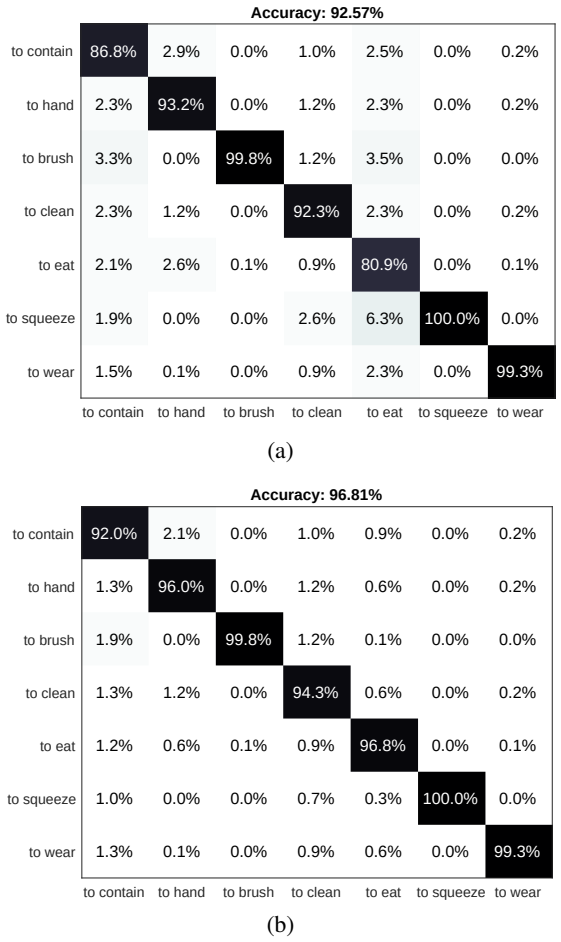


Figure 8: Affordance category classification performance. (a) Before adding environment features, showing an average diagonal accuracy of 92.57%; (b) After including the environment, showing an average diagonal accuracy of 96.81%.

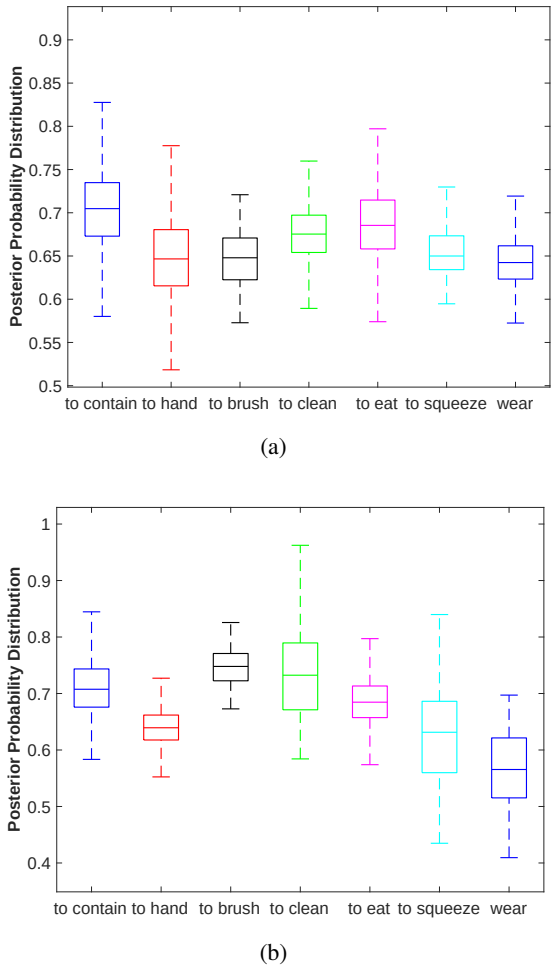


Figure 9: Distributional posterior probabilities per class of the knowledge base. (a) Distribution before the environment inclusion, and (b) after the environment features are included.

the most likely action that a human would take in order to obtain the less negative effect.

For example, Figures 10(a) and 10(b) show the obtained model from a water bottle. In these images, the achieved grasp before deducing the affordances, results on being placed on the lid of the bottle, which would result in a negative effect if the bottle contained liquid. On the other hand, Figure 10(b) shows the calculated grasping area after the affordance has been inferred, which shows to be a more plausible solution given the risk of the object being full. The same case can be pleaded for Figures 10(c) and 10(d). In a slightly different case, Figures 10(e) and 10(f) show two different grasping regions for an object which affordance has been determined as *hand over*. Thus both grasping choices seem acceptable given that there is no critical effect involved.

FINAL REMARKS AND FUTURE WORK

Past research has presented approaches to the grasping problem extensively. However, grasping behaviours depending on the object affordances is still an open challenge due to the large variety of object shapes and robotic platforms. Furthermore, the current approaches need large amounts of data to train a model without being able to generalise among different classes of objects successfully, nor to distinguish the best grasp area depending on the object's purpose of use.

Thus, in this work, the base of a cognitive grasping framework that is able to identify and encapsulate the good affordance features of an object is presented. This task is not only limited to the relationship that can be built between the target object and the agent but also considers the surrounding environment. The results show that without any *a-priori* awareness on the grasping area of the object, the designed KB is able to induce on the object's affordance. These results are further improved by the incorporation of the environment in

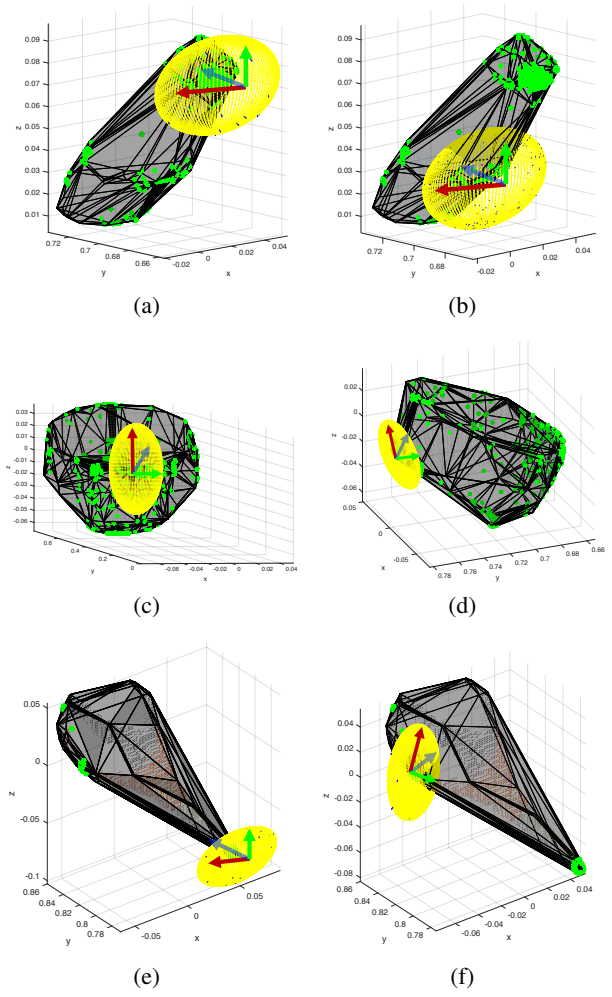


Figure 10: Objects modelling and grasping points before (left column) and after (right column) affordance reasoning, the tested objects are: first row, water bottle; second row, bowl; and third row, scissors. The extracted grasping points are shown in green while the region of the corresponding grasp is shown in yellow.

which these objects likely reside. Thus, allowing the system to have a better chance at deducing the grasping area of the object. Likewise, the presented framework has room for improvement, which is facilitated by its modularity. Overall, the performance of the KB can be increased by adding more attributes to the base, as well as modifying the predictive model in order to deal with uncertainty. Furthermore, the dynamics and system control schemes of the humanoid robot are considered out of the scope of this work. Nonetheless, (Pairet et al. 2018a) offers a learning-based framework that combines relative and absolute robotic skills for dual-arm manipulation suitable for dynamic environments tasks such as grasping objects that together with the semantics of the object offer a complete manipulation platform for humanoid robots.

References

Ardón, P.; Dragone, M.; and Erden, M. S. 2018. Reaching and grasping behaviours by humanoid robots through visual servoing. In *Haptics: Science, Technology and Applications, Springer International Publishing AG*, 353–365. Springer Nature.

Bonaiuto, J., and Arbib, M. A. 2015. Learning to grasp and extract affordances: the Integrated Learning of Grasps and Affordances (ILGA) model. *Biological cybernetics* 109(6):639–669.

de Beeck, H. P. O.; Torfs, K.; and Wagemans, J. 2008. Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. *Journal of Neuroscience* 28(40):10111–10123.

Do, T.-T.; Nguyen, A.; and Reid, I. 2018. Affordancenet: An end-to-end deep learning approach for object affordance detection. In *International Conference on Robotics and Automation (ICRA)*.

Goldfeder, C.; Ciocarlie, M.; Peretzman, J.; Dang, H.; and Allen, P. K. 2009. Data-driven grasping with partial sensor data. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, 1278–1283. IEEE.

Greeno, J. G. 1994. Gibson's affordances. *Psychological Review* 336–342.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

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

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

Jaklic, A.; Leonardis, A.; and Solina, F. 2013. *Segmentation and recovery of superquadrics*, volume 20. Springer Science & Business Media.

Kraft, D.; Detry, R.; Pugeault, N.; Baseski, E.; Piater, J. H.; and Krüger, N. 2009. Learning objects and grasp affordances through autonomous exploration. In *ICVS*.

Lai, K.; Bo, L.; Ren, X.; and Fox, D. 2012. Detection-based object labeling in 3d scenes. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, 1330–1337. IEEE.

Lee, D.-T., and Schachter, B. J. 1980. Two algorithms for constructing a delaunay triangulation. *International Journal of Computer & Information Sciences* 9(3):219–242.

Lenz, I.; Lee, H.; and Saxena, A. 2015. Deep learning for detecting robotic grasps. *International Journal of Robotics Research* 34(4-5):705–724.

Madry, M.; Song, D.; and Kragic, D. 2012. From object categories to grasp transfer using probabilistic reasoning. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, 1716–1723. IEEE.

Metta, G.; Sandini, G.; Vernon, D.; Natale, L.; and Nori, F. 2008. The icub humanoid robot: an open platform for research in embodied cognition. In *Proceedings of the 8th workshop on performance metrics for intelligent systems*, 50–56. ACM.

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

Montesano, L., and Lopes, M. 2009. Learning grasping affordances from local visual descriptors. In *Development and Learning, 2009. ICDL 2009. IEEE 8th International Conference on*, 1–6. IEEE.

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

Nguyen, A.; Kanoulas, D.; Caldwell, D. G.; and Tsagarakis, N. G. 2017. Object-based affordances detection with convolutional neural networks and dense conditional random fields. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Oztop, E.; Bradley, N. S.; and Arbib, M. A. 2004. Infant grasp learning: a computational model. *Experimental brain research* 158(4):480–503.

Pairet, È.; Ardón, P.; Brox, F.; Mistry, M.; and Petillot, Y. 2018a. Learning and generalisation of primitives skills towards robust dual-arm manipulation. In *AAAI Fall Symposium. Artificial Intelligence for Human-Robot Interaction*. AAAI Press.

Pairet, È.; Hernández, J. D.; Lahijanian, M.; and Carreras, M. 2018b. Uncertainty-based Online Mapping and Motion Planning for Marine Robotics Guidance. In *Intelligent Robots and Systems (IROS), 2018 IEEE/RSJ International Conference on*. IEEE.

Quattoni, A., and Torralba, A. 2009. Recognizing indoor scenes. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 413–420. IEEE.

Sridharan, M. 2017. Integrating knowledge representation, reasoning, and learning for human-robot interaction. In *AAAI Fall Symposium. Artificial Intelligence for Human-Robot Interaction*, 69–76. AAAI Press.

Stoytchev, A. 2005. Toward learning the binding affordances of objects: A behavior-grounded approach. In *Proceedings of AAAI symposium on developmental robotics*, 17–22.

Vezzani, G.; Pattacini, U.; and Natale, L. 2017. A grasping approach based on superquadric models. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, 1579–1586. IEEE.

Wertsh, J. V., and Tulviste, P. 1990. Apprenticeship in thinking: Cognitive development in social context. *Science* 249(4969):684–686.

Zech, P., and Piater, J. 2016. Active and transfer learning of grasps by sampling from demonstration.

Zhu, Y.; Fathi, A.; and Fei-Fei, L. 2014. Reasoning about object affordances in a knowledge base representation. In *European conference on computer vision*, 408–424. Springer.