Affordance perception and tool use for human-robot collaboration

Lorenzo Jamone[1], Giovanni Saponaro[1], Atabak Dehban[1], Alexandre Bernardino[1], José Santos-Victor[1]

***Abstract*— Humans can quickly perceive the affordances (i.e. action possibilities) of objects from vision and can predict the effects of the afforded actions on such objects. This ability is learned ecologically through the interaction with the environment and is exploited for action planning and problem solving. Clearly, endowing robots with similar capabilities is a fundamental challenge in cognitive and social robotics. We propose a framework in which a humanoid robot explores the environment and learns probabilistic dependencies between actions, objects visual properties and observed effects. By making inferences across the learned dependencies a number of cognitive and social skills are enabled: e.g. i) predicting the effects of an action over an object, ii) selecting the best action to obtain a desired effect, iii) emulating the action of a human. By exploring object-object interactions the robot can develop the concept of *tool* (i.e. a handheld "intermediate" object that allows to obtain a desired effect on another object), and eventually use the acquired knowledge to plan sequences of actions to attain a desired goal (i.e. problem solving); interestingly, this computational machinery can be also exploited for the execution of human-robot collaborative tasks.**

I. INTRODUCTION

Humans solve complex tasks on a routine basis, by choosing, amongst a vast repertoire, the most proper actions to apply onto objects in order to obtain certain effects. According to developmental psychology [1], the ability to predict the functional behavior of objects and their interaction with the body, simulating and evaluating the possible outcomes of actions before they are actually executed, is one of the purest signs of cognition, and it is acquired incrementally during development through the interaction with the environment. To reproduce such intelligent behavior in robots is an important, hard and ambitious task. One possible way to tackle this problem is to resort to the concept of affordances, introduced by Gibson in his seminal work [2]. He defines affordances as action possibilities available in the environment to an individual, thus depending on its action capabilities. From the perspective of robotics, affordances are powerful since they capture the essential object properties in terms of the actions that a robot is able to perform [3]. They can be used to predict the effects of an action, or to plan the actions to achieve a specific goal; by extending the concept further, they can facilitate action recognition and be exploited for robot emulation [4] (i.e. the robot reproduces the effects of an observed human action), they can be a basis to learn tool use [5], [6], and they can be used together with planning techniques to solve complex tasks [7]. We propose a probabilistic model of affordances that relates the visual shape properties of a hand held object (intermediate) and an acted object (primary) with the visual effects of the motor actions of the agent. We performed experiments in which the iCub [8] humanoid robot learns these object affordances by performing numerous actions on a set of objects displaced on a table (see Fig. 1).

[1]All authors are with the Instituto de Sistemas e Robótica, Instituto Superior Técnico, Universidade de Lisboa, Portugal. `ljamone,gasaponaro,adehban,alex,jasv at isr.ist.utl.pt`

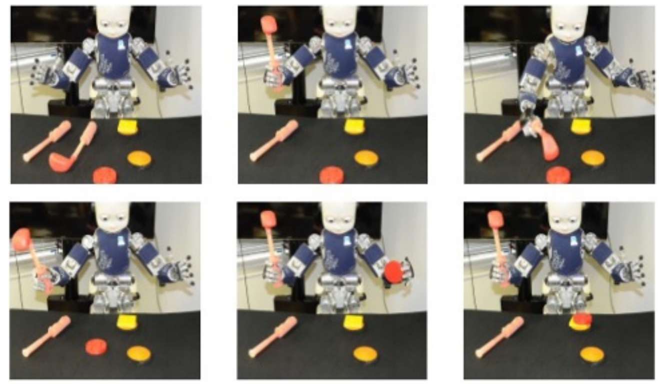

Fig. 1. The iCub humanoid robot standing in front of a table full of objects: affordance perception can be exploited for problem solving (e.g. to perceive that a rake can be used to pull an object closer).

II. A COMPUTATIONAL MODEL OF AFFORDANCES

Inspired by the framework of [9], we use Bayesian Networks (BN) to model the relationship between actions, target objects (acted upon), intermediate objects (hand-held), and the resulting effects; after learning the parameters of such a model, we can infer i) affordances of target objects, ii) affordances of intermediate objects, and iii) affordances of the interaction between intermediate and target objects (see Fig. 2, top image). Expert design or structure learning can be used to define one specific DAC (Directed Acyclic Graph) connecting the nodes (e.g. the black connections in Fig. 2, top image), among all the possible ones (i.e. the grey connections). Further details can be found in [5], [6].

A. A model for tool use

Tools can be typically described by three functional parts: a handle, an effector, and a body of a certain length L connecting the two (see right part of Fig. 2). These three parts are related to three different motor behaviors humans have to perform in order to successfully use a tool: grasping the handle, reaching for a desired pose with the effector and then executing an action over an affected object. We can therefore define three levels of tool affordances: i) *usage affordances*, what actions the tool affords, ii) *reach affordances*, what part

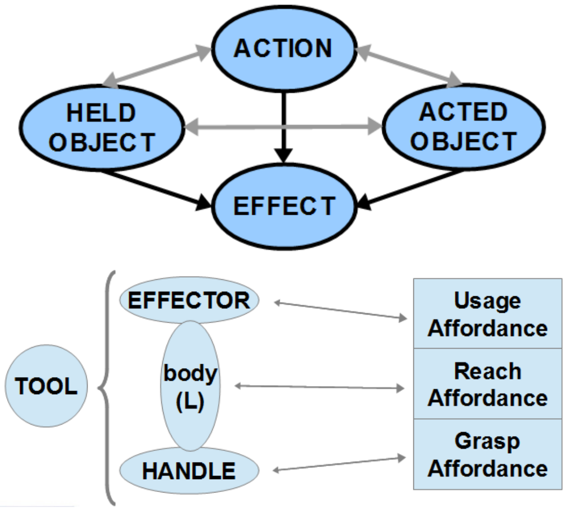

Fig. 2. Top image: Bayesian Network model of affordances, modeled as relations between actions, effects and objects (both held and acted). Bottom image: general model for tool use. The model in the top image corresponds to the Usage Affordances part of the model in the bottom image.

of the workspace the tool affords to reach for, and iii) *grasp affordances*, what grasps the tool affords (see right part of Fig. 2). The outcomes of these three reasoning processes are based on internal models that the robot can learn through motor exploration. The model of affordances in the top part of Fig. 2 represents the *usage affordances*. In previous work we proposed learning strategies that enable a robot to learn its own body schema [10]–[12], and to update it when tools are included [13], [14], and a representation of its own reachable space [15], [16]; these internal models are related to the *reach affordances*. Also, a number of models for learning and using *grasp affordances* have been proposed in the literature (e.g. [17], [18]).

B. Use affordances for problem solving

Interestingly, the predictions obtained from a probabilistic model of affordances can be used to ground the planning operators of a probabilistic planning algorithm (e.g. PRADA [19]). Through this computational machinery, a robot can plan the sequence of actions that has the higher probability to achieve the goals, given its own sensorimotor abilities and the perceived properties of the available objects. Moreover, we recently demonstrated in the context of the EU project Poeticon++ how the iCub can solve tasks requested by a human user through natural language in noisy unstructured environment (e.g. to make a sandwich!) [7]. The first step is to represent the human verbal request as a sequence of goals; then, the capability to perceive the object affordances and to predict the effects of the actions allows the robot to achieve such goals with flexibility (i.e. different actions might be used to achieve a given goal, depending on the circumstances) and robustness (i.e. relying on continuous perception and reasoning). Clearly, this kind of adaptive behaviors, which are supported by previous learning (i.e. learning how to perceive affordances through an ecological approach), are crucial for a robot to naturally collaborate with humans in real environments (e.g. home), since they permit to fill the gaps that are typically present in natural human verbal communication: for example, the robot would understand from its visual perception that he can use a hook-like elongated tool to retrieve an out-of-reach piece of tomato in order to grasp it, even if the human verbal request was simply “Can you take the tomato?” (see Fig. 1).

References

[1] C. von Hofsten, “An action perspective on motor development,” *Trends in Cognitive Sciences*, vol. 8, pp. 266–272, 2004.

[2] J. J. Gibson, *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin, 1979.

[3] L. Jamone, E. Ugur, A. Cangelosi, L. Fadiga, A. Bernardino, J. Piater, and J. Santos-Victor, “Affordances in psychology, neuroscience and robotics: a survey,” *IEEE Transactions on Cognitive and Developmental Systems*, 2016, early access.

[4] A. Dehban, L. Jamone, A. Kampff, and J. Santos-Victor, “Denoising auto-encoders for learning of objects and tools affordances in continuous space,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[5] A. Gonçalves, G. Saponaro, L. Jamone, and A. Bernardino, “Learning visual affordances of objects and tools through autonomous robot exploration,” in *IEEE ICARSC*, 2014.

[6] A. Goncalves, J. Abrantes, G. Saponaro, L. Jamone, and A. Bernardino, “Learning intermediate object affordances: Towards the development of a tool concept,” in *IEEE ICDL-Epirob*, 2014.

[7] A. Antunes, L. Jamone, G. Saponaro, A. Bernardino, and R. Ventura, “From human instructions to robot actions: Formulation of goals, affordances and probabilistic planning,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[8] G. Metta, L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. von Hofsten, K. Rosander, M. Lopes, J. Santos-Victor, A. Bernardino, and L. Montesano, “The icub humanoid robot: An open-systems platform for research in cognitive development,” *Neural Networks*, vol. 23, no. 89, pp. 1125–1134, 2010.

[9] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, “Learning object affordances: From sensory motor coordination to imitation,” *IEEE Transactions on Robotics*, vol. 24, no. 1, pp. 15–26, 2008.

[10] L. Jamone, L. Natale, K. Hashimoto, G. Sandini, and A. Takanishi, “Learning task space control through goal directed exploration,” in *IEEE ROBIO*, 2011.

[11] L. Jamone, L. Natale, G. Metta, F. Nori, and G. Sandini, “Autonomous online learning of reaching behavior in a humanoid robot,” *International Journal of Humanoid Robotics*, no. 9(3), pp. 1 250 017.1–1 250 017.26, 2012.

[12] B. Damas, L. Jamone, and J. Santos-Victor, “Open and closed-loop task space trajectory control of redundant robots using learned models,” in *IEEE IROS*, 2013.

[13] L. Jamone, B. Damas, J. Santos-Victor, and A. Takanishi, “Online learning of humanoid robot kinematics under switching tools contexts,” in *IEEE ICRA*, 2013.

[14] L. Jamone, B. Damas, N. Endo, J. Santos-Victor, and A. Takanishi, “Incremental development of multiple tool models for robotic reaching through autonomous exploration,” *PALADYN Journal of Behavioral Robotics*, no. 3(3), pp. 113–127, 2013.

[15] L. Jamone, L. Natale, G. Sandini, and A. Takanishi, “Interactive online learning of the kinematic workspace of a humanoid robot,” in *IEEE IROS*, 2012.

[16] L. Jamone, M. Brandao, L. Natale, K. Hashimoto, G. Sandini, and A. Takanishi, “Autonomous online generation of a motor representation of the workspace for intelligent whole-body reaching,” *Robotics and Autonomous Systems*, no. 64(4), pp. 556–567, 2014.

[17] L. Montesano and M. Lopes, “Learning grasping affordances from local visual descriptors,” in *IEEE ICDL*, 2009.

[18] R. Detry, D. Kraft, O. Kroemer, L. Bodenhagen, J. Peters, N. Kruger, and J. Piater, “Learning grasp affordance densities,” *PALADYN Journal of Behavioral Robotics*, no. 2(1), pp. 1–17, 2011.

[19] T. Lang and M. Toussaint, “Planning with noisy probabilistic relational rules,” *Journal of Artificial Intelligence Research*, vol. 39, pp. 1–49, 2010.