A Moderately Large Size Dataset to Learn Visual Affordances of Objects and Tools Using iCub Humanoid Robot

Atabak Dehban
email:adehban@isr.tecnico.ulisboa.pt
Institute for Systems and Robots, University of Lisbon
Champalimaud Research

Lorenzo Jamone
http://lorejam.blogspot.pt/
Institute for Systems and Robots, University of Lisbon

Adam R. Kampff
http://www.kampff-lab.org/
Sainsbury Wellcome Centre for Neural Circuits and Behaviour (SWC), London, UK

José Santos-Victor
http://users.isr.ist.utl.pt/~jasv/
Institute for Systems and Robots, University of Lisbon

Abstract

Affordances are defined as action opportunities that an environment offers to an agent: relations between visually perceived properties of an object, the possible actions afforded, and the effects of such actions. This notion can be generalized to the concept of tools: i.e. the visual appearance of a tool suggests what actions it *affords* to do on an object. Inspired by the amount of trials human babies need to learn these kinds of relations, we have gathered a relatively big dataset using iCub humanoid robot. The robot performs different actions using easy-to-rebuild tools on various objects selected from the YCB objects set and observes the results of the actions. The dataset can facilitate research in the areas of sensorimotor learning, active perception and cognitive developmental robotics.

1 Introduction

A central theory in ecological psychology is that performing actions is essential in developing visual perception. One compelling evidence of this claim comes from the famous experiments of Held and Hein [9] in the 60s in which they showed kitten that had been exposed to sufficient visual stimuli become functionally blind if they were deprived of the ability to initiate movement while being exposed to this stimuli during some early periods of their visual development.

Actions are deeply entangled with visual perception. They can either guide perception like active vision [2] and active perception [4] or help the agent to pick up relevant visual features with respect to its motor capabilities to accomplish a goal. This latter notion is related to affordances (see [10] for a review).

By exploiting the direct relation between visual appearances of objects and the the object's respond to being subject to actions from the motor repertoire of an agent, researchers in the field of robotics have tried to solve the problem of interaction with the environment without the need to recognize the elements of it. One example of this approach is explained in [14] where the researchers have provided a data set of images annotated not by the object's classes but by the way different parts of objects *afford* useful functionalities to humans like *openable* or *supportable*.

Another major area of research where the concept of affordances has been extensively used is table top object manipulation [13, 16]. These studies explore how objects behave when a robot executes different actions on them with its body. For example in [6] a semantic web is developed which maps the co-occurrences of different concepts acquired through different sensory modalities (vision, tactile, auditory) while the the robot is interacting with objects by performing different actions such as moving, pushing etc.

More recently, similar concepts have been used by researchers to encompass object-object relations and to extend the robot capabilities by using tools [8, 15]. In [12], the robot tries to predict the effect of pushing action on an object given the visual features of different tools in hand. In another work [7], the robot attempts multiple actions with different tools on various objects and tries to learn the resulting motion of the object through several trials.

Despite the ubiquity of the scenario and even though different researchers are trying to solve different aspects of the problem, the absence of a data set related to this table top object manipulation with tools slows down the comparison of different techniques. Moreover, the advances of learning representations directly from images (see [11] for an overview) suggests that it is possible to directly infer the properties of interest from image pixels given sufficient trials.

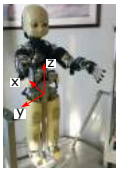

Figure 1: iCub main axis.

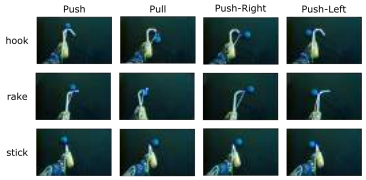

Figure 2: Initial relative position of each tool with respect to the object for different actions

To address these issues, we are introducing a new data set which is collected by the iCub humanoid while performing actions with different objects and tools. To the best knowledge of authors, this data set is unique as it gathers some desirable features which are important to various tasks. (1) It is collected by the robot itself and not a human demonstrator, thus the sensorimotor experience of the robot is authentic. (2) The data set is concerned with object-object relations related to tool use and as a result, it can be adopted to different bodies of robots as long as the robot can hold that tool. (3) By selecting accessible objects and easily fabricated tools, different researchers can reproduce the results or extend the data to include more trials, objects, etc.

2 Description of the data set

The data set contains information about the effect of performing different actions on various objects while the robot is holding one of the tools in its left hand. The tool is always held in the same way but the its perceived end-effector changes in correspondence with different actions.

Actions: Regarding each action, the robot places the tool's end-effector to predefined positions relative to the object and attempts the action. Four action classes are considered to be performed and the end-effector moves for 12 cm along one of the four relative directions (also refer to Fig. 1):

Push push the object against the x axis. Tool tip placed below the object.

Pull pull the object towards the x axis. Tool tip placed on top of the object.

Push-Right push the object against the y axis. Tool tip placed on the left side of the object.

Push-Left push the object towards the y axis. Tool tip placed on the right side of the object.

Fig. 2 shows where the robot tries to place the tool tip with respect to objects. This position is not always accurate due to the errors in measurements of the robot's joint angles. One thing to note is that because stick is not a useful tool to draw objects towards the robot, it is placed close to the object and the object essentially doesn't move (unless a small movement is triggered because of the errors in the initial placement of the tool on the table).

Objects: In order to make it easier to reproduce the results of the experiment, it was decided to select some objects from the Yale-CMU-Berkeley (YCB) Object and Model set [5]. The selected objects are light weight and cover a relatively vast variation of visual appearances. Such

Figure 3: Selected objects from the YCB object set

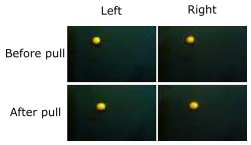

Figure 4: The result of pulling lemon with rake from left and right eyes

Figure 5: Tools view from robot perspective. a) stick b) rake c) hook

variations result in different motions each object would experience when being subject to different actions under different tools. Fig. 3 shows the set of objects from the view point of the robot.

Tools: Tools were also selected with similar intentions as objects: to be (l) light weight; (2) accessible and (3) visually different so as to offer different affordances. They were built from lightweight PVC pipes and can be more suitable or less for different action/object pairs. Fig. 5 shows each tool from the robot's view point.

Trial: During each trial, aforementioned objects are placed on a table in various positions and orientations. The robot is holding one of the previously introduced tools in hand and the transformation from tool tip to the center of the palm is provided by the experimenter. Afterwards, the robot calculates the desired initial end effector position with respect to the object according to Fig. 2 and the initial image of the object from left and right cameras together with the initial 3d position of the object are saved in a buffer before any action is taken into place. following these initial recordings, the robot attempts to put the tool tip in the correct position and perform one of the four actions. When the action is completed and the robot joints return to their predefined positions, the experimenter decides whether the action was successful or not. In case of successful actions all the data related to the initial and final view of the object together with its 3d displacement are saved to the disc. Otherwise the buffers are flushed and the robot prepares for the next trial. A video of the robot doing the trials is accessible in the address: https://youtu.be/pKa6GNeBfjk. Some of the visual features related to objects and in particular, the ones used in [8] are also provided in addition to the raw images of objects and tools.

Statistics: In total, there are 11 objects, 4 actions, 3 tools and at least 10 repetition of each trial which sums up to ~1320 unique trials and 5280 unique images of resolution 320*200 (the top 40 rows of pixels are cropped as they correspond to regions outside of the boundaries of the table). Fig. 6 shows the scatter plots of objects displacement on the x and y axis (according to Fig. 1) for each action and tool and in Fig. 4 the images from the robot's point of view of doing a pull action with the rake on the lemon is depicted.

Applications: Despite the differences in data types, actions, etc. the goals of this experiment are not far from the experiment introduced in [1] and their proposed architecture can be modified and adapt to the presented dataset. Moreover, the same experiments were used in [3] to learn the affordance network part of their proposed probabilistic planer. This data can also be used as an initial test to assess the aptitude of applying some algorithms on robot generated data before some other task specific data is gathered. It should also be noted that the iCub robot has a vibrant

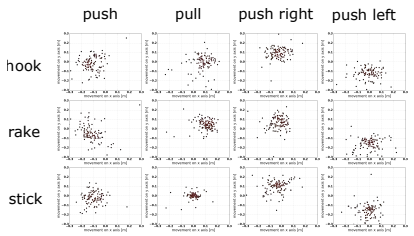

Figure 6: Object motion after applying the actions with different tools on the xy plane

community and the accessibility of tools and objects used in this study makes it possible for other researchers to augment/validate this dataset. This data itself can be accessed via http://vislab.isr.ist.utl.pt/datasets/

Acknowledgement

This work is partially supported by project FCT [UID/EEA/50009/2013]. A. Dehban is a doctoral candidate in Robotics, Brain and Cognition Ph.D. programme and is currently supported by the Portuguese FCT doctoral grant PD/BD/105776/2014.

[1] P. Agrawal, A. Nair, P. Abbeel, J. Malik, and S. Levine. Learning to poke by poking: Experiential learning of intuitive physics. In *2016 Neural Information Processing Systems (NIPS)*, 2016.

[2] John Yiannis Aloimonos, Isaac Weiss, and Amit Bandyopadhyay. Active vision. *International journal of computer vision*, 1(4):333–356, 1988.

[3] 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 *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5449–5454. IEEE, 2016.

[4] Ruzena Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8): 966–1005, 1988.

[5] Berk Calli, Aaron Walsman, Arjun Singh, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M Dollar. Benchmarking in manipulation research: The ycb object and model set and benchmarking protocols. *arXiv preprint arXiv:1502.03143*, 2015.

[6] Hande Celikkanat, Güner Orhan, and Sinan Kalkan. A probabilistic concept web on a humanoid robot. *IEEE Transactions on Autonomous Mental Development*, 7(2):92–106, 2015.

[7] A. Dehban, L. Jamone, A. R. Kampff, and J. Santos-Victor. Denoising auto-encoders for learning of objects and tools affordances in continuous space. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4866–4871, May 2016. doi: 10.1109/ICRA.2016.7487691.

[8] Afonso Gonçalves, João Abrantes, Giovanni Saponaro, Lorenzo Jamone, and Alexandre Bernardino. Learning intermediate object affordances: Towards the development of a tool concept. In *4th International Conference on Development and Learning and on Epigenetic Robotics*, pages 482–488. IEEE, 2014.

[9] Richard Held and Alan Hein. Movement-produced stimulation in the development of visually guided behavior. *Journal of comparative and physiological psychology*, 56(5):872, 1963.

[10] 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*, PP(99):1–1, 2016. ISSN 2379-8920. doi: 10.1109/TCDS.2016.2594134.

[11] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[12] Tanis Mar, Vadim Tikhanoff, Giorgio Metta, and Lorenzo Natale. Multi-model approach based on 3d functional features for tool affordance learning in robotics. In *Humanoid Robots (Humanoids), 2015 IEEE-RAS 15th International Conference on*, pages 482–489. IEEE, 2015.

[13] Luis Montesano, Manuel Lopes, Alexandre Bernardino, and José Santos-Victor. Learning object affordances: From sensory–motor coordination to imitation. *Robotics, IEEE Transactions on*, 24(1): 15–26, 2008.

[14] Abhilash Srikantha and Juergen Gall. Weakly supervised learning of affordances. *arXiv preprint arXiv:1605.02964*, 2016.

[15] V Tikhanoff, U Pattacini, L Natale, and G Metta. Exploring affordances and tool use on the icub. In *Humanoid Robots (Humanoids), 2013 13th IEEE-RAS International Conference on*, pages 130–137. IEEE, 2013.

[16] Emre Ugur, Erhan Oztop, and Erol Sahin. Goal emulation and planning in perceptual space using learned affordances. *Robotics and Autonomous Systems*, 59(7):580–595, 2011.