2D and 3D functional features for tool affordance learning and generalization on humanoid robots.

Tanis Mar[1], Vadim Tikhanoff[1], Giorgio Metta[1], Lorenzo Natale[1]

Abstract— **Future humanoid robots are expected to be able to carry out a wide range of tasks for which they had not been equipped, or even designed for, much in the same way as humans are able to discover and learn new abilities. In order to achieve such desirable capability, they will need to learn how to take advantage of external environmental elements to help them perform the tasks for which their own manipulators are insufficient or inefficient; Tool use, in short, will allow robots to extend their capabilities beyond the limitations of their own bodies.**

A critical aspect in this problem, crucial for enabling knowledge generalization, is the choice of a proper tool representation. Ideally, the representation scheme applied should not only be able to discriminate among different tools, but also to consider the following aspects: i) different tools afford the same functionality if the share common geometrical features, ii) the effect that can be achieved with a tool depends as much on the action performed as on the way in which it is grasped. On the following, we introduce two sets of features used for describing tool's geometry on 2D and 3D respectively, and evaluate their performance on a simple affordance scenario.

INTRODUCTION

A practical approach to tackle the complex nature of tool use learning is to not consider objects in terms of their label or category, as is commonplace in most artificial intelligence or computer vision algorithms, but rather in terms of their functionality. To that end, roboticists can build upon the theoretical framework proposed by J.J. Gibson in the late 70's around the concept of affordances [1]. In its original formulation, affordances were defined as the potentiality for action of any element in the environment for a given agent. This abstract idea was later formalized as the relationship between an entity in the environment, the possible actions that a particular agent is capable of performing on it, and the effect that it generates on such entity or the environment [2]. Such approach is also in line with a growing body of evidence from neuroscientific studies which suggests that primates and humans perceive objects and tools in terms of their affordances, rather than only their category [3], [4], and even that context and hand posture modulate tool-object perception in the brain [5].

Based on this formalization and subsequent refinements, many authors have proposed diverse working models for robots to learn the effects of their actions on objects through interacting with them. This is a fundamental capability of any cognitive system, as it bridges the sensorimotor interaction layer (which deals with processing signals coming directly from sensors and actuators) with high-level layers that can reason on symbols. Some of these studies have implemented affordances as look-up tables [6], [7], Bayesian Networks [8], [9], [10], [11], or general purpose classifiers coupled with clustering methods to discover the available affordances' elements categories [12], [13], [14], [15].

When introducing tools, early approaches to affordance learning have exploited simplified representations in which tools were identified using labels (i.e. rake, screwdriver, etc) [16], [17], [18]. While efficient in contsrained scenarios, this approach leads to poor generalization performance. More recently, a few studies have proposed to apply geometry-based *functional features* to represent the tool, with the aim to match the tool's characteristics expressed by such functional features with its possible affordances. Some of them focused primarily on how the tool effector could match the object's shape to enable a certain functionality [19], [20], while others focused on the general properties of the tool independently of the target object [21], [22], [23], [24]. Yet, these studies apply small sets of broad descriptors (size, length, etc), which would not be able to discriminate subtler differences between tools that might actually change their functionality. Moreover, most studies on tool affordances assume that the tool is always grasped in a *canonical* pose (except in [25]), despite the fact that the way in which a tool is grasped greatly affects its functionality. The methods proposed below tackle the need for more descriptive while still generalizable functional feature sets, which additionally take implicitly into account the tool grasp when learning its affordances.

METHODS

A. Experimental Setup

The experiments were performed using the iCub Humanoid Robot as well as its simulator. On the former, we gathered affordance data from the interaction with a target object using four tools, while on the latter seven different tool models were used. In order to study how different tool grasps affect the tool's functionality, the iCub held each tool with the end-effector oriented in three different directions: right, front and left. Each of the possible grasp arrangements of a particular tool in a particular orientation is referred to as a tool-pose, following the nomenclature used in [25]. The effect of the tool use was measured as the displacement of a small cube placed on a table in front of the robot, whose position was being tracked.

B. Tool-pose based functional features

During these experiments, two different sets of features were applied to represent the tool-pose being handled by the

[1] T. Mar, V. Tikhanoff, G. Metta and L. Natale are with the iCub Facility, Istituto Italiano di Tecnologia, Via Morego 30, 16163, Genova, Italy

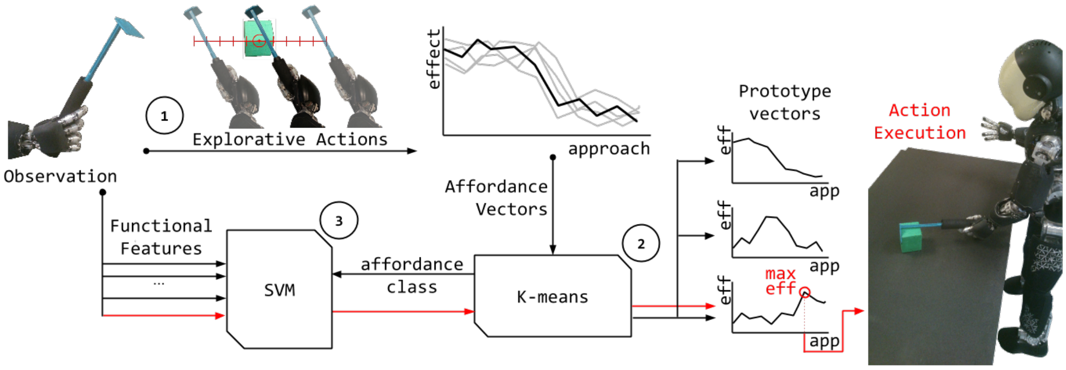

Fig. 1. Diagram of the proposed affordance learning and prediction method. Black and red arrow lines represent the flow of information in the training and prediction phases, respectively.

iCub:

a) The first set was extracted from the normalized 2D contour of the handled tool; it consisted of 75 geometrical features based on its general shape (such as area, perimeter, etc), convex hull, skeleton, moments, angle signature and transformed domains (wavelet and Fourier). In order to obtain the contour, first the tool was segmented from the background using the algorithm described in [26] seeded with the tooltip position. Then, the segmented blob was normalized with respect to the angle between the tooltip and the hand center, and cropped to retain only the end-effector part of the tool.

b) The second set was extracted from a 3D model of the handled tool that was either provided (on simulation), or obtained by the iCub by exploring the object from different perspectives and merging the point-clouds reconstructed using its stereo-vision capabilities. This set consists of a concatenation of Oriented Multi Scale Extended Gaussian Images, which describes the 3D geometry of the tool and retains information about the orientation of the tool with respect to the hand reference frame.

C. Action performance and effect measurement

Action was defined as a simple pull action parametrized by the approach to the target object, i.e., how much on top or to either side of the target object the end-effector of the tool was aimed. The effect was in turn defined as the displacement of the object as a consequence of each action, computed as the euclidean distance between the 3D location of the target object before the action and after it. However, instead of regarding action and effect as independent elements, they were merged into an *action-effect vector*, which measured the effect for all the range of values considered of the approach parameter.

D. Affordance discovery and self-supervised learning

The action-effect dataset gathered as described above was analyzed using K-means to discover if the individual action-effect vectors were distributed into clusters, which would represent affordance categories, and to observe whether these might be commonly shared among different tool-poses.

Test	Env.	Set 2D(%)	Set 3D(%)
Prediction	Sim.	81.9%	82.99 %
Prediction	Robot	64.1%	68.23 %
Leave-One-Out	Sim.	56.9%	57.3%
Leave-One-Out	Robot	53.9%	51.4%

TABLE I
PREDICTION PERFORMANCE FOR KNOWN (PREDICTION) AND UNKNOWN (LEAVE-ONE-OUT) TOOL-POSES.

Subsequently, a SVM classifier was trained by using the extracted functional features as an input, and the labels provided by the clustering procedure, that is, the corresponding affordance categories, as the teaching signal.

E. Affordance Prediction and Action Execution

By means of the training process, the model learned the mapping between the values of the tool-poses' functional features and the affordance type they offer, in terms of the action-effect vector cluster. Thereby, retrieving the prototype vector of the predicted affordance category from the K-means map, the model was able to predict the effect for each of the possible values of the approach parameter. Thus, the parameter for which the expected effect was maximum was chosen and the respective pulling action subsequently performed. A diagram representing all the steps and dataflow of the experiment described in this Section can be observed on Figure 1.

RESULTS

Two tests were carried out to evaluate the learning and prediction capabilities of the model with either set of features: For the first test the model was trained with data from all available tool-poses, and it had to predict the affordance category of a new example of one of the known tool-poses; In the second test, the robot had to predict the affordance category of a handled tool-pose which had been left out of the training set. Results of both tests when applying both feature sets can be observed on Table I

REFERENCES

[1] J. J. Gibson, "The Ecological Approach to the Visual Perception of Pictures," *Leonardo*, vol. 11, no. 3, p. 227, 1978.

[2] A. Chemero, "An Outline of a Theory of Affordances," *Ecological Psychology*, vol. 15, no. 2, pp. 181–195, 2003.

[3] S. T. Grafton, L. Fadiga, M. a. Arbib, and G. Rizzolatti, "Premotor cortex activation during observation and naming of familiar tools." *NeuroImage*, vol. 6, no. 4, pp. 231–236, 1997.

[4] P. O. Jacquet, V. Chambon, A. M. Borghi, and A. Tessari, "Object affordances tune observers' prior expectations about tool-use behaviors." *PloS one*, vol. 7, no. 6, p. e39629, Jan. 2012.

[5] N. Natraj, V. Poole, J. C. Mizelle, A. Flumini, A. M. Borghi, and L. a. Wheaton, "Context and hand posture modulate the neural dynamics of tool-object perception," *Neuropsychologia*, vol. 51, no. 3, pp. 506–519, 2013.

[6] P. Fitzpatrick, G. Metta, L. Natale, S. Rao, G. Sandini, L. Natalet, S. Raot, and G. Sandinit, "Learning About Objects Through Action - Initial Steps Towards Artificial Cognition," in *Procceedings of the IEEE Intrernational Conference on Robotics and Automation*, 2003, pp. 3140–3145.

[7] A. Stoytchev, "Toward Learning the Binding Affordances of Objects : A Behavior-Grounded Approach," in *AAAI Symposium on Developmental Robotics*, 2005, pp. 21–23.

[8] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, "Modeling affordances using Bayesian networks," in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, Oct. 2007, pp. 4102–4107.

[9] M. Lopes, F. S. Melo, and L. Montesano, "Affordance-based imitation learning in robots," *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1015–1021, Oct. 2007.

[10] 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, Feb. 2008.

[11] P. Osório, A. Bernardino, and R. Martinez-cantin, "Gaussian Mixture Models for Affordance Learning using Bayesian Networks," in *International Conference on Intelligent Robots and Systems*, 2010, p. 1'6.

[12] E. Ugur, E. Sahin, and E. Oztop, "Predicting future object states using learned affordances," *2009 24th International Symposium on Computer and Information Sciences*, pp. 415–419, Sep. 2009.

[13] T. Hermans, J. M. Rehg, and A. Bobick, "Affordance Prediction via Learned Object Attributes," in *IEEE International Conference on Robotics and Automation (ICRA 2011)*, 2011, pp. 1–8.

[14] S. Griffith, J. Sinapov, V. Sukhoy, and A. Stoytchev, "A BehaviorGrounded Approach to Forming Object Categories: Separating Containers from Non-Containers," *IEEE Transactions on Autonomous Mental Development*, vol. 4, no. 1, pp. 54–69, 2012.

[15] B. Akgun, N. Dag, T. Bilal, I. Atil, and E. Sahin, "Unsupervised learning of affordance relations on a humanoid robot," *2009 24th International Symposium on Computer and Information Sciences*, pp. 254–259, Sep. 2009.

[16] A. Stoytchev, "Robot tool behavior: A developmental approach to autonomous tool use," Ph.D. dissertation, Georgia Institute of Technology, 2007.

[17] J. Sinapov and A. Stoytchev, "Detecting the functional similarities between tools using a hierarchical representation of outcomes," in *2008 7th IEEE International Conference on Development and Learning*. Ieee, Aug. 2008, pp. 91–96.

[18] V. Tikhanoff, U. Pattacini, L. Natale, and G. Metta, "Exploring affordances and tool use on the iCub," in *Humanoids*, 2013.

[19] C. C. Kemp and A. Edsinger, "Robot Manipulation of Human Tools: Autonomous Detection and Control of Task Relevant Features," in *IEEE International Conference on Development and Learning*, 2006.

[20] T. E. Horton, "A Partial Contour Similarity-Based Approach to Visual Affordances in Habile Agents." Ph.D. dissertation, 2011.

[21] R. Jain and T. Inamura, "Learning of Tool Affordances for autonomous tool manipulation," *2011 IEEE-SICE International Symposium on System Integration SII*, pp. 814–819, 2011.

[22] ——, "Bayesian learning of tool affordances based on generalization of functional feature to estimate effects of unseen tools," *Artificial Life and Robotics*, pp. 1–9, Sep. 2013.

[23] A. Gonçalves, J. a. Abrantes, G. Saponaro, L. Jamone, and A. Bernardino, "Learning Intermediate Object Affordances : Towards the Development of a Tool Concept," in *IEEE International Conference on Development and Learning and on Epigenetic Robotics (ICDL-EpiRob 2014)*, no. October, 2014, pp. 13–16.

[24] A. Gonçalves, G. Saponaro, L. Jamone, and A. Bernardino, "Learning visual affordances of objects and tools through autonomous robot exploration," *2014 IEEE International Conference on Autonomous Robot Systems and Competitions, ICARSC 2014*, no. May, pp. 128–133, 2014.

[25] S. Brown and C. Sammut, "Tool Use Learning in Robots," in *Advances in Cognitive Systems*, 2011, pp. 58–65.

[26] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, Sep. 2004.