19th Portuguese Conference on Pattern Recognition (RecPad 2013)
Lisbon, Portugal, 1 November 2013

# Autonomous Learning of Tool Affordances

Afonso Gonçalves
agoncalves@isr.ist.utl.pt

Giovanni Saponaro
gsaponaro@isr.ist.utl.pt

Lorenzo Jamone
ljamone@isr.ist.utl.pt

Alexandre Bernardino
alex@isr.ist.utl.pt

Instituto de Sistemas e Robótica,
Instituto Superior Técnico

## Abstract

The autonomous use of tools by robots is a topic where studies are still rare even though much attention is given to the affordances of objects. In this paper we propose an extension to the learning object affordances model presented by Montesano *et al.* in [1] by including tools as an additional node into the Bayesian Network of affordances. We also present an automatic method for gathering the data required in the learning of such a network using the iCub simulator.

## 1 Introduction

The use of tools by an animal species is considered a major sign of intelligence, as it shows the ability to understand relations between object features and the resulting effects of interaction with the environment. Their use is of the utmost importance to humans as they are used in most of human activities. For a humanoid robot to be able to perform the same everyday tasks humans do, tool manipulation is imperative.

In this paper, similarly to Stoytchev [2] and Jain *et al.* [3], we explore the extension of reach that a tool can provide to alter the position of another object, which is based on the definition of tool given by Beck [4]: "*Tool use is the external employment of an unattached environmental object to alter more efficiently the form, position, or condition of another object, another organism, or the user itself when the user holds or carries the tool during or just prior to use and is responsible for the proper and effective orientation of the tool.*"

Closely associated with the concept of tool is the concept of affordances introduced by Gibson [5], which describes what an organism directly perceives of an object as doable (affordable actions) to said object. The understanding of such properties and relations can lead to emergence of planning behaviors and more complex activities.

In order to achieve a method which enables a robot to learn how to successfully use tools in the best way possible, this paper introduces a way to represent tool affordances, grounded to the individual experience of the robotic agent, and also an autonomous behavior for collecting the data required for learning such representation. However, deriving analytical or physical models of the effect that a tool produces on objects is hard due to uncertainty in contact interactions, friction, impact forces and sensor noise. Therefore we adopt a learning by exploration methodology to obtain causal models of tool affordances.

## 2 Tool Affordances Modeling

The affordances of a tool are tightly related not only with the tool itself but also with the presented environment (objects), the individual motor capabilities of an agent and the effects on the environment recognizable by said agent. In order to successfully use tools, there is a need to encode the relation between these 4 elements: Tool, Object (Environment), Action, and Effects.

In [1], Montesano *et al.* introduced a way of encoding the dependencies between actions, directly applied to objects and its resulting effects as a Bayesian Network. In this paper we propose the introduction of an additional node to the network which encodes the use of tools as shown in Figure 1.

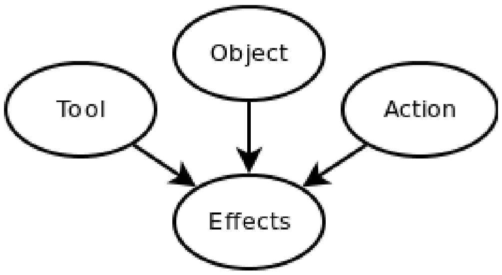


Figure 1: Proposed Bayesian Network model of tool affordances

This Bayesian Network graphical model of affordances expresses the joint probability of an event tuple P(T=tool, O=object, A=action, E=effect) as a factorization: P(T,O,A,E) = P(E|T,O,A).P(T).P(O).P(A). With this model it is possible to use the derived affordances to: predict the effects on a given object using a specific tool P(E|T,O,A); plan the best action which produces the desired effects on an object given a tool P(A|T,O,E); select the most likely object to produce the desired effects given a tool and action P(O|T,A,E); or select the best tool given an object, action and desired effect P(T|O,A,E).

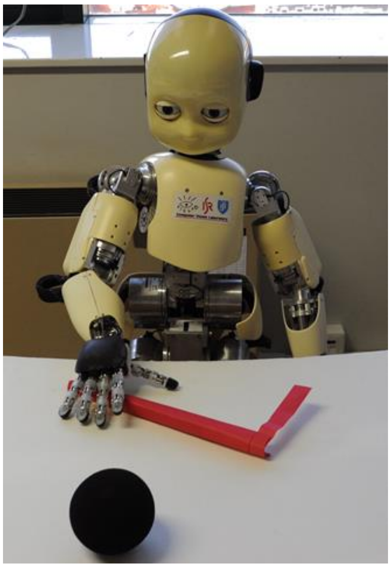

Figure 2: The proposed scenario, iCub humanoid robot with a tool and an actuated object

### 2.1 Visual Descriptors

A set of visual descriptors is used for tool and object recognition and classification. These describe shape, size, position and color of the segmented objects and tools.

The descriptors, calculated in the 2D camera image plane, are as follows:

- Normalized x and y coordinates of the center of the enclosing rectangle.
- Normalized width of the enclosing rectangle.
- Normalized height of the enclosing rectangle.
- Angle (orientation) of the enclosing rectangle.
- Hue normalized color histogram of the pixels inside the object's region.
- Area (number of pixels).
- Convexity - ratio between the perimeter of the object's convex hull and the perimeter of the object's contour.
- Eccentricity - ratio between the minor and major axis of the minimum-area enclosing rectangle.
- Compactness - ratio between the object area and its squared perimeter.
- Circleness - ratio between the object area and the area of its enclosing circle.
- Squareness - ratio between the object area and the area of its minimum-area enclosing rectangle.

By continuously tracking an object during an interaction with it and storing both the visual description before and after the interaction, the effects can be recognized and classified as change in position, color and/or shape of the object.

Additionally, we propose to divide a tool in 3 different segments, each with its own visual descriptors parameters:

- Handle, which is the graspable part of the tool.
- End-effector, the end extremity of the tool which will come into contact with the object.
- Body, the connecting part between the handle and the end-effector.

With this separated parameterization of each part of a tool we hope to give a better generalization capability to the learning of the affordances, as for example certain tools might not afford certain actions not because its end-effector is inappropriate but because the tool might be too short to reach the target object or simply not graspable by the robot.

## 3 Simulated Learning Environment

In order to learn the structure and parameters of the affordances Bayesian Network for a specific set of tools, objects, actions and effects, an experimental setup using the iCub simulator was developed and automated.

This simulated setup enables the iCub robot to repeatedly perform automated trials on a provided environment at a much higher rate than using the real robot, which will be posteriorly used to validate the results. Such a trial involves 5 steps:

1. Observation of the tool. A specific tool is presented to the robot and its visual descriptors are calculated and stored. Shown in Figure 3, Figure 4 (A) and (B).

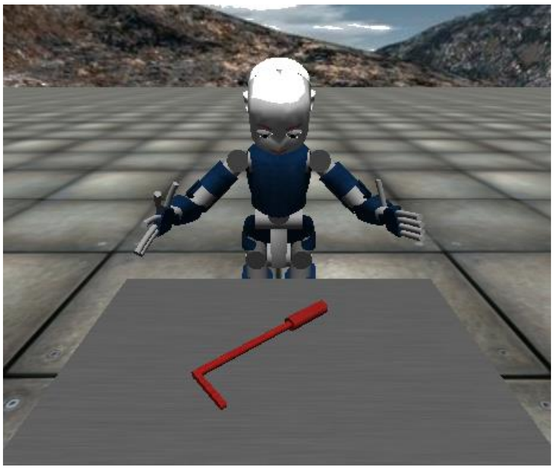

Figure 3: First step for learning the affordances of a tool in the simulated experiment, the visual descriptor of the tool are calculated and stored

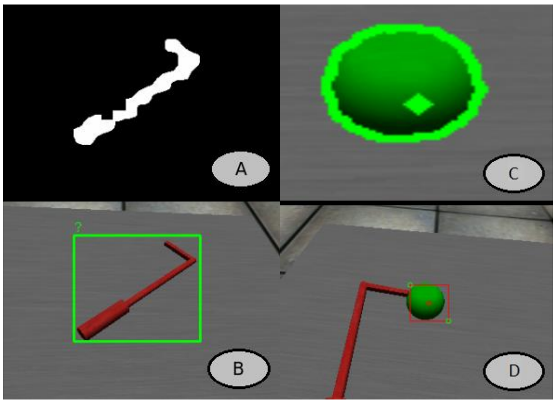


Figure 4: Example of visual tool recognition, detected blob of a tool (A), bounding box locating the tool (B); Example of the visual recognition of an object, active segmentation of the object to track (C), tracking of the segmented object during action with tool (D)

2. Grasping of the tool. The robot grasps the tool by the handle after it is automatically placed on his hand.
3. Observation of the object. A specific object is presented to the robot; its visual descriptors are calculated, stored and starts to be actively tracked. An example of a segmented object to track is shown in Figure 4 (C).
4. Act. The robot is commanded to perform an action on the object using the tool. The action and its parameters are stored. Figure 4 (D) and Figure 5 show an example of an action being performed by the robot.

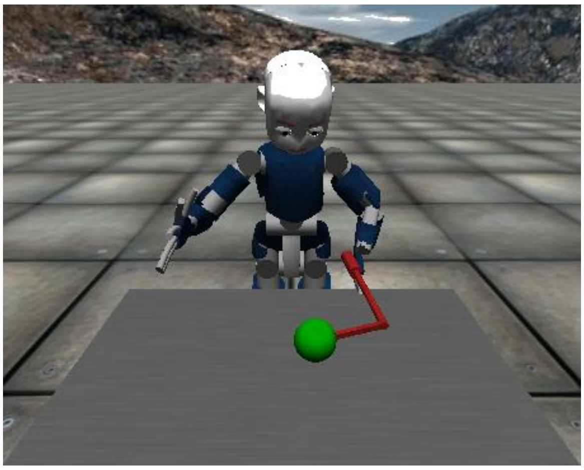

Figure 5: Fourth step for learning the affordances of a tool simulated experiment, acting on an object with a tool

5. Re-observation of the tracked object. The visual descriptors of the object are recalculated after the interaction with the tool and its parameters are stored.

Initially the set of available actions, tools and objects will be limited to the simple case of:

- One parameterized action: Push(x, y, z, θ, radius), where x, y and z are the 3D coordinates where to act with the tool end-effector, θ the angle of approach, and radius the radial offset from which the tool end-effector will approach the x, y ,z position.
- Three uniformly colored tools: Stick, L-stick and T-stick.
- Three uniformly colored objects: Box, Ball and Cylinder

These sets will later be expanded to include more complex and varied behaviors and objects.

## 4 Future Work

This paper gives an overview of a Bayesian Network model for tool affordances and an experimental setup for the iCub robot which enables the collection of data required to generate such network.

Although the outline of the experimental setup is established there is still work to be done in both the motor and visual elements, as additional actions and visual descriptors are being evaluated. More importantly multiple trials combining all the available tools, actions, and objects are still needed in order to generate the data that will be used in the continuation of this work which is to learn the actual structure and values of the affordances Bayesian Network.

## 5 Acknowledgements

This work was partially funded by the EU Project POETICON++ and FCT [PEst-OE/EEI/LA0009/2013].

## References

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

[2] J. Gibson. The ecological approach to visual perception. Boston: Houghton Mifflin, 1979.

[3] A. Stoytchev. Behavior-Grounded representation of tool affordances. In Proc. ICRA, pages 3060-3065, 2005.

[4] R. Jain and T. Inamura. Learning of tool affordances for autonomous tool manipulation. System Integration (SII), 2011 IEEE/SICE International Symposium on, pages 814-819, 20-22 Dec. 2011.