Towards Learning Basic Object Affordances from Object Properties *

Barry Ridge **Danijel Skočaj** **Aleš Leonardis**
Faculty of Computer and Information Science
University of Ljubljana, Slovenia
{barry.ridge, danijel.skocaj, ales.leonardis}@fri.uni-lj.si

Abstract

The capacity for learning to recognize and exploit environmental affordances is an important consideration for the design of current and future developmental robotic systems. We present a system that uses a robotic arm, camera systems and self-organizing maps to learn basic affordances of objects.

1. Introduction

The term *affordance*, coined by the psychologist J.J. Gibson (Gibson, 1986), refers to the interactive possibilites that are available to a cognitive system when confronted with particular objects or environments. Practical examples of such scenarios for robotic systems might include a mobile robot entering a room because it perceived that the room afforded being entered due to the door being open, or a system with a robotic arm picking up a cup because it perceived that the cup had a handle that afforded being grasped by the arm's gripper.

We present a cognitive vision system that learns basic object affordance properties by interacting with objects on a table surface using a robotic arm (Neuronics Katana 6M), and observing the result using camera systems (Point Grey Research Flea and Bumblebee). These devices are integrated over a distributed architecture, shown in Fig. 1. The experimental environment is shown in Fig. 2. The main idea is to allow the system to perform a variety of simple push actions on objects that are placed on the work surface, record video footage of the result, harvest this data for appropriate features and attempt to learn the similarities inherent in the behaviour of those objects that share physical properties that relate to what they afford when affected by such actions.

* This research has been supported by: Research program P2-0214 (Republic of Slovenia), EU FP6-004250-IP project CoSy and EU MRTN-CT-2004-005439 project VISIONTRAIN.

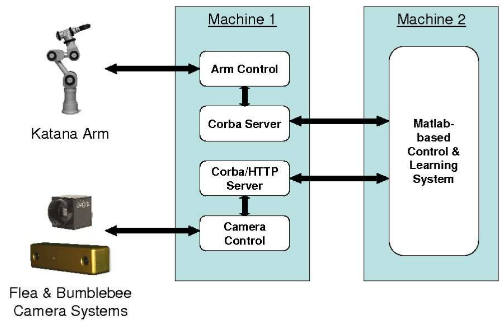

Figure 1: System architecture.

Figure 2: View of the workspace. The robotic arm holds a black plastic tool that is used to push objects on the work surface. Some sample objects are also shown.

2. Related Work

Researchers have previously tackled the issue of affordance learning in a number of different settings and emphasizing different aspects of the problem. Robotic systems have been developed to learn pre-specified binary affordance classes, e.g. rolling versus non-rolling objects (Fitzpatrick et al., 2003) or liftable versus non-liftable objects (Paletta et al., 2007). While such systems take on the difficult challenge of learning affordances in the real world, they would benefit from a more generalized form of learning that can discover affordance classes in the environment dynamically. Others (Cos-Aguilera et al., 2004) have used more sophisticated learning mechanisms to learn affordance properties, but in simulated environments.

In our case, we envisage using an unsupervised learning mechanism to form generalized affordance classes dynamically as the system interacts with objects using a repertoire of pre-specified actions.

3. Our System

In our research to date, we have been using *self-organizing maps (SOMs)* to accomplish this learning task. Labels indicating both action and object class are created during experiments, as well as an associated feature vector garnered from the video sequences that are taken after actions are performed on the objects. A SOM is incrementally trained to recognize what should occur when a particular action is performed on a particular object. Moreover, it generalizes effectively so as to estimate how similar a given action/object pairing is to its counterparts. This is useful, because it can then recognize that, e.g., pushing a round object from a certain direction will be similar to pushing another round object from the same direction, but different to pushing the same round object from a different direction, or a flat object from the original direction.

In (Ridge et al., 2008), we presented an experiment where three different pushing actions were provided to the system for interacting with four household objects (a toy cube, a round child's toy rattle, a Pepsi can and a mobile phone, as seen in Fig. 2). After an object was pushed with one of the possible actions, features such as mean velocity, total distance traveled, final object orientation, etc. were gathered from the resultant video sequence. These features, along with the associated action/object label, were used to train a SOM incrementally. As the system performed more interactions, it learned to distinguish between the objects that rolled in the workspace (toy rattle and Pepsi can) and those that did not (toy cube and mobile phone).

However, this experiment assumed that the object class and state were known in advance- potentially from a seperate perception module. Our current work involves changing this so that, instead of using object class labels, object properties such as surface curvature, object thickness, object shape, etc., are gathered from both regular RGB images and range data (from the Bumblebee stereo camera). These features can either be used as raw input to the SOM, along with the result features and the action label, or the object property feature space can be analyzed seperately to form atomic labels, e.g. "curved" or "non-curved". After training, when the system encounters novel scenarios, it can then take object property measurements and use its trained SOM(s) to judge how similar the current scenario is to previously encountered ones and predict the resultant outcome based on this past experience.

An important point is that such a SOM-based learning mechanism is agnostic to the types of objects used in the experiment, the types of actions that are used to interact with them, or the affordance classes that emerge from such interactions. If the feature set is diverse enough to capture the essential aspects of the affordances, then the map will organize in the feature space and effectively cluster action/object pairings with similar affordances.

4. Conclusion

In a cognitive system, the ability to learn continuously during a lifespan as more environmental experience is gained, is crucial for successful development. One aspect of this is the ability to learn affordance classes as they become apparent through interaction. While our system is capable of both predicting and generalizing resultant feature vectors based on object properties, in past experiments we have assumed that such properties are known. Our ongoing work involves adding perceptual capabilities to the system using, e.g., range data, so that it can perceive such object properties directly. In the future, we would like to implement some form of constructivist learning where the system uses learned basic affordances to construct more complex affordance concepts, e.g. this obstacle affords being avoided if the object is first pushed forward and then pushed to the right.

References

Cos-Aguilera, I., Canamero, L., and Hayes, G. (2004). Using a SOFM to learn object affordances. *Proceedings of the 5th Workshop of Physical Agents, Girona, Catalonia, Spain.*

Fitzpatrick, P., Metta, G., Natale, L., Rao, S., and Sandini, G. (2003). Learning about objects through action- initial steps towards artificial cognition. In *Proceedings of IEEE International Conference on Robotics and Automation.*

Gibson, J. (1986). *The Ecological Approach to Visual Perception.* Lawrence Erlbaum Associates.

Paletta, L., Fritz, G., Kintzler, F., Irran, J., and Dorffner, G. (2007). Learning to perceive affordances in a framework of developmental embodied cognition. *IEEE 6th International Conference on Development and Learning.*

Ridge, B., Skočaj, D., and Leonardis, A. (2008). A system for learning basic object affordances using a self-organizing map. In *Proceedings of the 2008 International Conference on Cognitive Systems*, Karlsruhe, Germany.