A Bio-Inspired Robot With Visual Perception of Affordances

Oscar Chang

Universidad Central de Venezuela

Abstract. We present a visual robot whose associated neural controller develops a realistic perception of affordances. The controller uses known insect brain principles; particularly the time stabilized sparse code communication between the Antennal Lobe and the Mushroom Body. The robot perceives the world through a webcam and canny border openCV routines. Self-controlled neural agents process this massive raw data and produce a time stabilized sparse version, where implicit time-space information is encoded. Preprocessed information is relayed to a population of neural agents specialized in cognitive activities and trained under self-critical isolated conditions. Isolation induces an emergent behavior which makes possible the invariant visual recognition of objects. This later capacity is assembled into cognitive strings which incorporate time-elapse learning resources activation. By using this assembled capacity during an extended learning period the robot finally achieves perception of affordances. The system has been tested in real time with real world elements.

Keywords: Affordance Perception, Robotic Vision, Cooperative Neural Agents

1 Introduction

Affordance is a quality of an object or environment that allows (or suggests) an individual to perform an action [1]. The term is used in various fields including AI, cognition, perceptual psychology, industrial design, HCI, etc. Perceiving affordances has been related to infants development [2] and has opened vigorous research movement in AI [3] , artificial vision [4] and robotics [5]. Affordance demands the recognition of a class of objects (or environments) with no clear-cut differences, with many diffuse characteristics, with arbitrary boundaries, sizes and designations [1]. It may also trigger in the individual a complex response, such as moving toward the object and sitting on it. To initiate or not a real action will depend in the mediation of others agents. The execution of this excitation-response agreement, trivial for living creatures, combines difficult problems such as tracking and recognizing a moving object [6], the growth of cognitive abilities [7] and the formation of agents societies [8]. In this work we present a visual driven robot whose neural controller support expansible invariant object recognition [9]. In order to extend the robots learning period this paper incorporates a self-controlled grow algorithm in which the robot's available learning resources are distributed over an extended period of time. The combination of a long educational experience and the gradual release of learning capabilities finally develop in the robot a credible visual perception of affordances. The controller has an implicit biological structure where cooperative neural agents mimic two key insect brain elements: the antennal lobe (AL) and the Mushroom Body (MB).

2 Previous works

In previous works by Chang the following operative tools were established [9],[6]: 1) A neural artificial vision system which relies on the computer models of the AL and MB of insects. 2) A flow of circulating information defined by time stabilized sparse code. 3) An expansible learning capacity based upon isolated tunable agents (ITAs). In this paper we incorporate two new elements: 1) An operative unit called "cognitive string", formed by several ITAs ruled by a common time-released learning mechanism. 2) A "selective reward system" in which short-term learning events are aimed at specific cognitive string.

3 The Robot and its Multi-Agent Neural Controller

The used robot has one moving eye and two final effectors (servomotors) which handle the physical flags P and C. The robot watches the world through a two axis moving webcam and takes as visual input different classes of untailored 2D and 3D images (Fig 1). After training it develops affordance perception for some specific object classes which activate the final effectors.

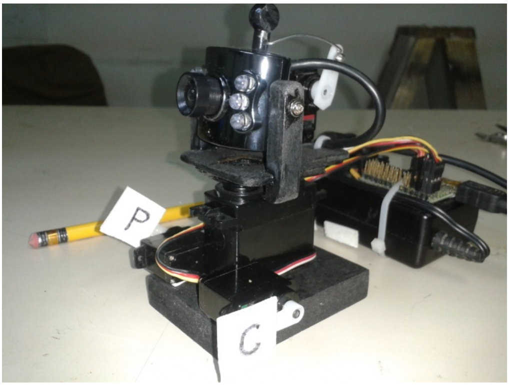

Fig. 1. Visual affordances perceiving robot. The robot observes the world through a two axis moving webcam. Some images afford "painting" and activate the effector P. Some others afford "cutting" and activate the effector C.

The used neural controller utilizes two key insect brain agents: the antennal lobe (AL) and the Mushroom Body (MB) (Fig 2).In the modeled AL primary receptors are pixels in 100x100 a moving region of interest (ROI) image, captured with a webcam and simplified with canny edge detection (a). These pixels feed an ANN pre trained as a crosshair reticle tracker (b). This ANN participates in a close loop feedback system (c) and becomes a generic tracking agent, producing a continuous flow of space-time related unstable code (d). An averaging agent (e) stabilizes this flow and passes a sparse version to the equivalent MB, formed by a set of isolated tunable agents (ITAs) composed by small ANNs (f) specialized in learning, recognition and memory formation. Through OR like operators ITAs' output are grouped into cognitive strings (g) which finally activate the physical effectors (h).

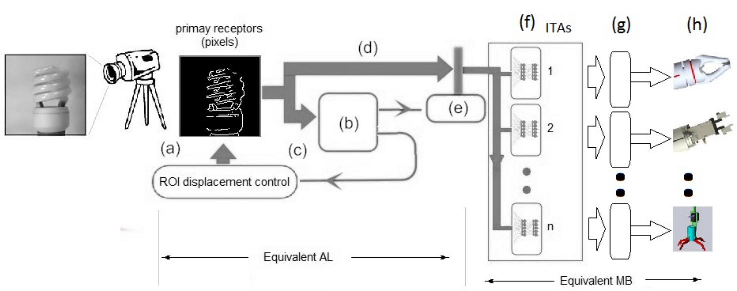

Fig. 2. The neural controller. A 100x100 canny image feeds a modeled Antennal lobe (AL) which generates time stabilized sparse code. This resource is passed to an artificial mushroom body (MB) where isolated tunable agents (ITAs) carry out cognitive duties. ITAs are assembled into cognitive strings which finally activate the effectors.

4 The artificial AL and MB

In insects the AL converts crude sensors data to a special form of space-temporal code essential for object recognition and relayed to the cognitive elements in the MB [10]. In our AL a backpro trained ANN operates in a closed loop mechanism where images from a video stream control image position [9]. This loop generates a flow of space-time related data which is subsequently stabilized and sparsed. The resulting *time stabilized sparse code* (TSSC) is relayed to cognitive agents in the simulated MB. The insects' MB serves as a large screen where objects can be much more easily discriminated [10]. In our MB cognitive agents called ITAs (Isolated Tunable Agents) are built with trainable three layers ANNs formed by 2500 inputs, 10 hidden and 5 outputs neurons. As in biology ITAs use as input neurons the TSSC coming from the AL (2500 signals).

5 Isolated Learning

When learning to recognize an object each ITA behaves as an auto-critical individual who uses the following learning rules: Rule 1 Look toward the outside world. See the object for a while and use backpro to: 1a) learn to fire with the object. 1b) partially forget what you have learned somewhere else. Rule 2 Look inside yourself. See your own noise source for a while. Use backpro to: 2a) learn not to fire with noise. 2b) partially forget what you have learned somewhere else. Using these rules a reward R is defined as a short term learning experience during which one selected ITA receives 100 consecutive backpro cycles watching the chosen object followed by 100 cycles watching white noise. Targets are properly set so that ITA's central output neuron learns to fire with the object and not to fire with noise. At 50 frames/sec a reward lasts 4 seconds and about 5 rewards are needed to memorize one object. Rewards shall not exceed a maximum number or the affected ITA will be degraded (overexposure). When trained under the above principles an ITA shows an emergent capacity to discriminate the learned object from many others, while absorbing a finite quantity (roughly 20%) of visual variances and white noise.

6 The time-released learning resources

Our next goal is to expand the number of ITAs dedicated to the learning of one object so that class recognition is attained. To this end ITAs are assembled into cognitive strings S^1, S^2,... S^n formed by m by consecutive ITAs numbered from 1 to m. To avoid overexposure a self-controlled time-released mechanism operates in each string distributing the received rewards as: The firs active ITA is the number 1. At any given time only the active ITA in the string receives rewards. Every active ITA i, which receives 15 (or so) consecutive rewards freezes its weight information and passes the active condition to the i+1 ITA. Once trained and for recognition purposes the ITAs' outputs in the same string are "ored" together. A selective reward R^i is now defined as a reward that only affects the active ITA in the cognitive string S^i. This seletive norm make possible to dedicate a whole string to the invariant recognition of one object thus expanding object recognition into class recognition.

7 Results

7.1 Experiment 1

The Emergence of Affordance Perception. In this experiment the robot develops perception of affordance for two classes of objects: class P represented by brushes which afford "painting" and class C represented by scissors which afford "cutting". These classes were chosen because physical samples of them were readily available and because they both represent difficult to recognize items, very sensitive to rotational translation. Two cognitive strings S^P and S^C in the MB are selected to develop affordances for painting (brushes) and cutting (scissors) respectively. Once trained the robot demonstrates its perceptions by activating its final effectors P and C. Each string comprises 20 ITAs which cover the rotational image variances of one full rotation per object. For training a human places objects (scissor or brushes) in the robot field of view and sends selective rewards R^P or R^S aimed at the respective strings. Since each trained ITA absorbs about 20 degrees of object rotation, 20 of them cover a full turn. In figure 3 (upper right) two trained ITAs process original images turned into canny images and TSSC. Time stabilized sparse features have been created in the ITAs' hidden layer (weights of one hidden neuron are shown). Using a Pentium Core i5 the learning time is about 8 minutes. Once trained the robot visually scans the shown landscape (left) and after three minutes correctly perceives the eight existing affordances. Some image zooming is tolerated and look alike objects such as pliers and relays are rejected.

Fig. 3. Affordance perception for multiple object visualization

7.2 Experiment 2

A non-easily distractible eye. To test the consistency of its perception of affordance the above trained robot is set to explore the whole Caltech 101_Object Categories data set. After examining the 9146 images in 5 hours the robot reports the 9 mistaken, look-alike elements shown in figure 4. It also recognizes 31 out of 39 true affordances in the "scissor" category.

8 Discussion and Conclusions

The robot shows a robust affordance perception capacity. For the whole Caltech dataset the error is limited to 0.098 %. More ITAs per cognitive string can be used as to cover full tilting and zooming for each desired affordance. A credible perception of affordance appears only after a prolonged learning period, which

Fig. 4. Searching for affordances in the 101_ObjectCategories Caltech dataset

in turn requires a space-time distribution of learning resources prepacked in cognitive strings. The used learning and feed forward mechanisms have a natural parallel structure so high speed operation could be expected when using parallel computing. We have developed and tested a robotic vision system capable of showing clear relations between affordances and perception-action under broad visual conditions. The proposed neural controller uses cooperative neural agents organized as the artificial versions of the AL and MB of living insects. In the proposed MB basic cognitive agents called ITAs, sensitive to short term learning experiences, are assembled into operative modules called cognitive strings. Inside the strings orderly activated ITAs store time stabilized sparse features of selected objects. The combination of a prolonged educational experience, time-elapse release of learning resources and the time stabilized sparse feature extraction finally develops in the robot a credible form of visual affordance perception. In extended images search the neural visual controller shows a good rejection of false affordances. This may be relevant for constructing efficient, no easily distractible robots. In principle the proposed techniques can expanded to higher pixel resolution and many affordance perceptions.

References

1. Gibson, J.: The Ecological Approach To Visual Perception. Taylor & Francis (2013)
2. Saccani, R., Valentini, N.C., Pereira, K.R., Müller, A.B., Gabbard, C.: Associations of biological factors and affordances in the home with infant motor development. Pediatrics International **55**(2) (2013) 197–203
3. Mateas, M.: Expressive AI: A semiotic analysis of machinic affordances. In: Proceedings of the 3rd Conference on Computational Semiotics and New Media, UK, University of Teesside (2003)
4. Hermans, T., Rehg, J.M., Bobick, A.: Affordance Prediction via Learned Object Attributes. In: IEEE International Conference on Robotics and Automation (ICRA): Workshop on Semantic Perception, Mapping, and Exploration. (May 2011)

5. Horton, T.E., Chakraborty, A., St. Amant, R.: Affordances for robots: a brief survey. AVANT. Pismo Awangardy Filozoficzno-Naukowej (2) (2012) 70–84
6. Chang, O.: Recognizing a moving object by using neural nets and ocular micro tremor. Revista de la Facultad de Ingenieria Universidad Central de Venezuela **28** (12 2013) 49 – 55 URL:http://ucv.academia.edu/OscarChang/Papers.
7. Goertzel, B., Bugaj, S.V.: Stages of cognitive development in uncertain-logic-based ai systems. In: Proceedings of the 2007 Conference on Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms: Proceedings of the AGI Workshop 2006, Amsterdam, The Netherlands, The Netherlands, IOS Press (2007) 174–194
8. Minsky, M.: The Society of Mind. A Touchstone Book. Simon & Schuster (1986)
9. Chang, O.: Reliable object recognition by using cooperative neural agents. (2014) URL:http://ucv.academia.edu/OscarChang/Papers IEEE WCCI 2014 in press.
10. Huerta, R.: Learning pattern recognition and decision making in the insect brain. AIP Conference Proceedings **1510** (2013) 101–19