Reinforcement learning in a self-organised representation of feature space

Kevin Brohan* **Alex Cope**** **Kevin Gurney**** **Piotr Dudek***
*The University of Manchester
Manchester, Uk
**University of Sheffield
Sheffield, UK

Abstract

We propose a method of integrating top-down and bottom-up attention, presenting a robotic implementation of a biologically inspired active vision system. Digits are classified with a self-organised map of features and working memories of rewarding features are maintained as neural activity in the same topological arrangement. From these maps, a retinotopic saliency map is generated, while inhibition of return encourages exploration of the scene.

1. Introduction

Behaving agents frequently need to choose one of a number of different possible actions at a point in time. In order to economically extract rewards from the environment, it is necessary to estimate the probability of receiving a reward for each potential action within the current behavioural context and to prioritize suitable actions. We investigate this problem with a robotic system implementation of biologically inspired visual attention, in which the goal is to orient gaze towards rewarding cues in order to accumulate the maximum number of rewards over time. Visual cues consist of seven-segment digits of equal size on a neutral background (Fig. 1, Fig. 2(a)). Internal representations of the cues are developed on a self-organised map (SOM) (Kohonen, 1982).

Rewards are given for directing gaze to a particular category of digit (i.e. a reward for looking at the digit *3* in any of its positions across the scene) Reinforcement learning is used to develop working memories of rewarding and unrewarding features in a neural layer, which is topologically mapped onto the SOM and provides top-down bias for visual attention. Inhibition of return (which prevents the eye from returning to recently attended locations) and stochastic action selection drive exploration of the visual scene. Previous biologically-inspired systems that address the control of visual attention include Butko and Movellan (2010) and the robotic implementation of Ognibene et al. (2008).

We have previously described a system in which the development of feature maps is guided by reward signals and where the reward expectation map, operating in the feature space, provides a top-down bias modulating the visual saliency map that guides action selection (Brohan et al., 2010). In this paper, we present a robotic implementation of a similar system and demonstrate that the robot learns to associate the relevant stimulus with the reward and that it adjusts its selective attention mechanism to adapt to a changing reward scenario.

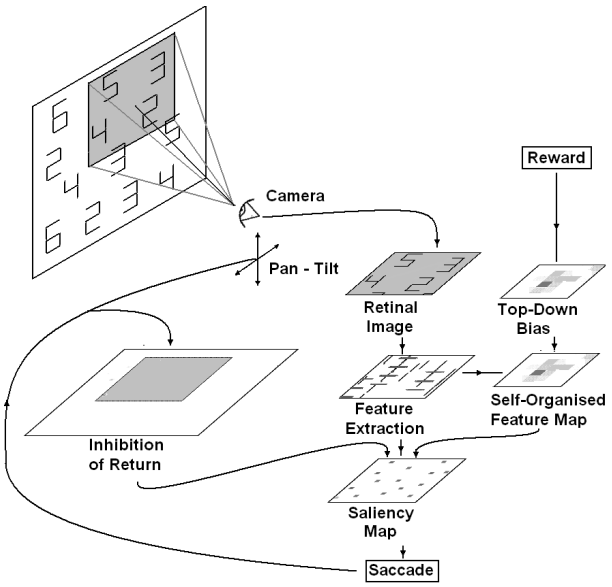
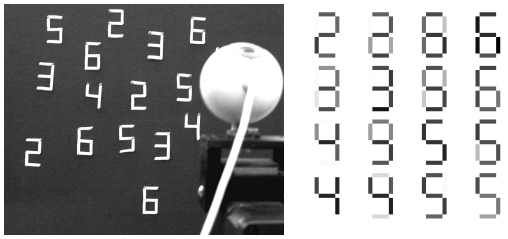


Figure 1: Overview of the model

2. Model Overview

The overview of the model is shown in Fig. 1. Bottom-up digit recognition is achieved by extracting feature vectors (line segments) from the retinal image and classifying them on a SOM. The SOM has 16 nodes with a 4 x 4 square topology with initially random weights. The SOM is trained offline using hand-selected feature vectors from the image. The trained SOM is shown in (Fig. 2(b)).

Top-down reward expectation is integrated with the SOM classification activity to form a combined



(a) (b)

Figure 2: (a) Photograph of the experimental setup with camera, pan-tilt unit and cues. (b) Image of the representations at each of the 16 SOM nodes after training

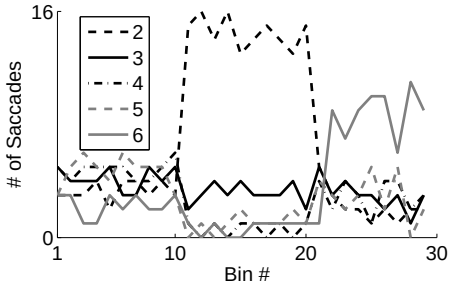


Figure 3: Number of saccades made to each feature in bins of 20 timesteps

feature saliency map. The combined feature saliency map activity is then expressed in retinotopic coordinates (using location information obtained at the feature classification stage) and saccades to recently attended locations are suppressed with an inhibition of return signal.

In order to encourage a balance between exploration/exploitation, a winning location is stochastically chosen from the retinotopic saliency map (by modulating the map activity with white noise) and a saccade is made to that location by moving the motors of the pan-tilt unit. A reward is given for making a saccade to a correct target (externally specified), which then generates activity in the region of the working memory map that corresponds to the attended feature while a saccade to an unrewarding location reduces the activity in the working memory. Finally, activity is stimulated in the inhibition of return map at the previous gaze location.

3. Results and Conclusion

The system was implemented on a desktop PC in APRON software (Barr and Dudek, 2009). The robot consists of a V-UAS14 camera (Logitech) mounted on a pan-tilt unit PTU-46-17.5 (Directed Perception) and a controller PT-D46r5C14S (Directed Perception).

To test our model, we present an experiment with three epochs, each containing 200 saccades, in which we test the ability of the working memory to bias saccades towards rewarding cues. In the first epoch, the robot explores the visual world without receiving any rewards, in the second epoch, rewards were given for making saccades to the digit *2*, and in the third epoch rewards were given for making saccades to the digit *6*. In both rewarding epochs the system learns to preferentially direct saccades to the rewarding digit. Fig. 3 shows the number of saccades made to each of the digits in bins of 20 time steps. The system does not have a significant preference for any of the digits in the unrewarding case (first 10 bins). In the second epoch, the number of saccades made to the rewarding digit (*2*) increased greatly due to the reinforcement learning of the rewarding features. Saccades to each of the other digits decreased in frequency. Saccades were preferentially made to the rewarding digit (*6*) in the third epoch, although with less success due to the more distributed representation of *6* on the SOM 2(b).

In summary, we have demonstrated that the model successfully drives a robot to seek rewarding cues in a visual search task by integrating bottom-up and top-down signals for feature-based attention.

Acknowledgements

This work was supported by EPSRC Grant no. EP/C516303.

References

Barr, D. R. W. and Dudek, P. (2009). Apron: A cellular processor array simulation and hardware design tool. *EURASIP Article ID 751687*.

Brohan, K., Gurney, K., and Dudek, P. (2010). Using reinforcement learning to guide the development of self-organising feature maps for visual orienting. In Illiadis, L. S., Diamantaras, K., and Duch, W., (Eds.), *International Conference on Artificial Neural Networks*, pages 180–189.

Butko, N. and Movellan, J. (2010). Infomax control of eye movements. *IEEE Transactions on autonomous mental development*, 2(2):91–107.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69.

Ognibene, D., Balkenius, C., and Baldassarre, G. (2008). Integrating epistemic action and pragmatic action: A neural architecture for camera-arm robots. In *From Animals to Animats 10: Proceedings of the Tenth International Conference on the Simulation of Adaptive Behavior (SAB2008)*, pages 220–229. Springer-Verlag, Berlin.