Embodied Object Recognition: When moving helps to recognise

Edgar Bermudez Contreras
School of Informatics
Univerisity of Sussex
Falmer, Brighton UK
Email: ejb21@sussex.ac.uk

***Abstract*— This work is a preliminary exploration of modulatory interaction of motor control signal and visual processing for embodied agents. The first results show that complex dynamics of evolved controllers can help to understand the interaction between motor and visual signals and maybe can shed some light about top-down attentional modulation in neuro-controllers for visually guided tasks.**

I. Introduction

Visual systems in nature are embodied. The fact that animals are able to interact with the environment while performing visually guided tasks suggests that vision is not only an isolated information processing system [1], [2], [3]. Given that visual systems in nature deal with a great amount of complex visual information, additional mechanisms like foveation and attention are also exploited. However, the design of artificial visual systems has not explored this characteristic exhaustively [4], [5]. Experiments have been reported where agents with simple visual systems can perform behavioural discrimination and other object recognition tasks employing complex visual information [6], [7].

The study of conditions where simple controllers can be helped to perform visually guided tasks by the exploitation of embodiness and active vision mechanisms can help to understand the role of attention and modulation of motor and visual processes in visually guided agents.

In this work, a preliminary study of the dynamics of evolved controllers employing an active vision approach is presented. The agent has a simple visual system (described in detail in [8]) to perform visually guided tasks. The visual system employed emulates the early areas in the visual cortex V1 but also top-down mechanisms such as attentional and foveation processes are considered. A brief analysis of the evolved neural controllers is followed by a future plan of my research following these ideas.

II. Methods

The experiments were carried out in InQubator, a simulator that renders 3D graphics (using OpenGL) and provides libraries for neural networks and genetic algorithms (GA) (see figure 1).

The visual field of the agent consisted of a rectangular region defined by 80×60 pixels with grey values between 0 and 255.

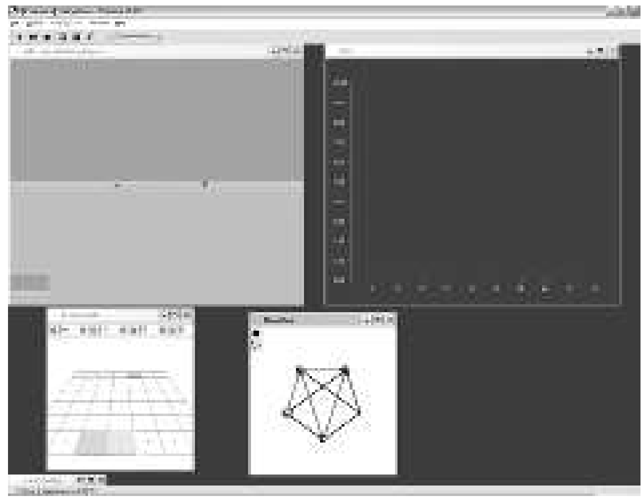

Fig. 1. InQubator

The visual system employed for this work was based on the RBF model (see a detailed description in [8]). This model consists of the application of low-pass filters with different orientations and sizes over the incoming visual information, emulating the primary visual cortex, V1 [9].

The filters employed were derivatives of Gaussians with four orientations, 0, 45, 90 and 135 degrees at scale sizes of 7×7, 15×15 and 29×29 pixels. The output of this model is a vector with all the outputs of the different filters applied over the input image. This vector is classified in different known categories according to the response of a Radial Basis Function Network (RBF). Each view tuned unit (VTU) of the RBF network is centered in one of the 8 views of the two known objects (see figure 2).

An attentional mechanism was employed based on blob detection. The criteria for blob detection was the intensity of grey for blobs bigger than 30 pixels in area. The blobs detected were scaled to the standard size of the pictures stored in the training phase of the visual system (80×60).

A. Controller

The controller is a fully connected Continuous Time Recurrent Neural Network (CTRNN) with five neurons. Three of these were sensor neurons measuring the activity of the

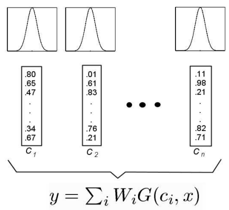

Fig. 2. VTU: each view vector c_i is the centre of a Gaussian function. The more similar a vector x is to a centre, the stronger the response of the unit. The output of the VTU, $y = \sum_i W_i G(c_i, x)$

visual system and two were output neurons connected to the motors of the agent (see figure 3). The state y of neuron i changes in time according to the differential equation:

$$\tau_i \dot{y}_i = -y_i \sum_j w_{ij} \phi(y_j + \beta_j) + g \cdot I$$

That is, the state is the integration of the weighted sum of the all incoming conections plus an a gained input I for input neurons. The time constant ϕ is the sigmoid activation function, $\tau \in [0.2, 2.0]$ and the bias $\beta \in [-10, 10]$ and all the weights $w_{ij} \in [-5, 5]$ are shaped by the GA.

The inputs of the network corresponded to the RBF units $RBF_i \in [0, 1]$ and to the centroid of the object detected $c \in [0, 1]$ where 0 corresponded to the furthest left pixel in the visual field and 1 to furthest right pixel.

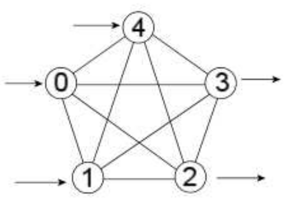

Fig. 3. Neural controller: fully connected CTRNN with 5 neurons. Neurons 0 and 1 are the input neurons with the activation of the RBF units, neuron 4 is the input neuron with the centroid of the object detected and neurons 2 and 3 are the output neurons connected to the motors of the agent.

The outputs of the network were defined by the output of two neurons of the network, connected directly (in the range of [0, 1]) to the motors of the agent.

B. Genetic Algorithm

A distributed GA was employed to evolve the neural controllers. A population of 49 individuals was evolved with mutation probability of 80% and an amount of 15% of mutation for each component. Also, there was an elitism probability of 80%.

The genome of each individual was given by a real vector of 35 elements, 5 for the time constants of each neuron, 5 for the bias of each neuron and 25 for the weights. Each element was coded as a real number in $[0, 1]$ and scaled according to the parameters described previously.

III. Experiments

The experiments carried out are a preliminary study of the visual system and evolvability of controllers for simple tasks. These tasks were selected based on works like [10].

Two objects, a kettle and a torus were placed in fixed positions in an arena of unlimited extension (see figure 1). The objects employed were positioned at the same distance from the agent. The kettle was at the coordinates (-1, 2), the torus was at (1, 2) and the agent was at (0, -2) at the beginning of each trial. The task consisted of approaching the kettle and avoiding the torus.

The fitness function for this task was defined as:

$$F = \frac{1}{d_1} - \frac{1}{d_2}$$

where d_i is the distance from the agent to the object i. Object 1 is the kettle and object 2 is the torus.

Each agent had three trials, always starting from the same position and with 0 degrees of orientation (i.e. facing the direction where the objects were). At the beginning of each trial the centroid information was reset to the centre of the visual field. As well, the RBF units activation was set to zero after every step. The state of the rest of the neurons was left with the activation of the previous trail.

A. Results

As this is only a preliminary study, I present the first and most interesting results of the experiments for the scenario described.

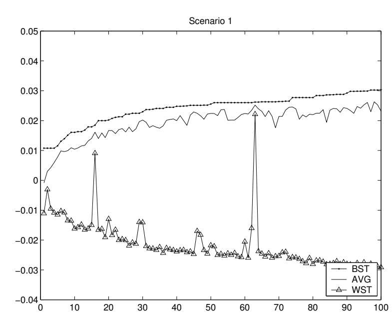

Fig. 4. Fitness function of evolved agents: the best, average and worst agents during 100 generations.

The performance of evolved controllers (described previously above) is presented in figure 4 for 100 generations. Evolved controllers under the scenarios described above were tested for 55 time steps and by randomising the initial orientation of the agent (within an arbitrary range) with the position of the objects remaining the same.

The evolved controllers show interesting dynamics. For instance, when the agent is able to centre the kettle in the visual field, the difference in the response of the RBF units is bigger than when it is not able to centre the object at the end of the trial (see figure 5). Also, the motor neurons show a smooth

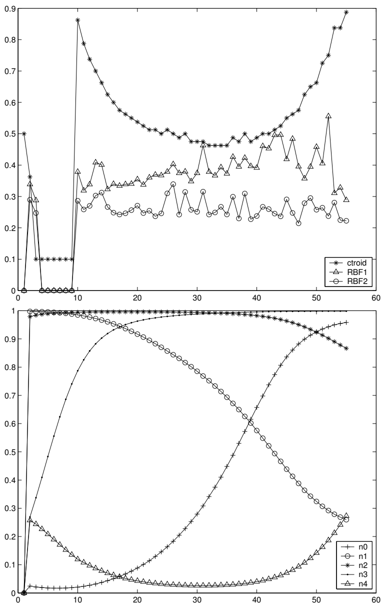

Fig. 5. Neural activity: inputs and outputs

modulation that represents the approaching behaviour and the activity of the neurons 0 and 1 (corresponding to RBF modules for object 1 and 2 in the visual system) show a discriminating behaviour around the time step 40 (see figure 5).

IV. CONCLUSION

More experiments are being considered to extend this work. To study the complex dynamics of evolved controllers using extended time of simulation (longer than in evolutionary phase), more experiments are going to be carried out. So far, the results suggest that an "evolutionary scaffolding" (incremental) approach could be useful to evolve controllers. First, starting with controllers that are able to fixate objects in the visual field and then, randomising the initial orientation and position of the objects. Also, a study of a possible interaction of attention mechanisms and tuning of the filters in the visual system and the motor system could be an interesting place to look at to try to understand the reliable and robust capabilities of animal vision.

V. FUTURE WORK

In the following months, my research work is going to focus mainly on the exploration of neural controllers and the analysis of the interaction and modulation between the visual information and motor system by top-down mechanisms like attention and bottom-up processes like filter tuning and foveation.

An incremental evolutionary approach is going to be applied to explore more complex tasks to study and analyse the evolved controllers.

As the final part of my project, downloading the evolved controllers into a real robot to test how the visual system tackles problems in the real world is still an interesting possibility.

ACKNOWLEDGMENT

The author would like to thank Thomas Buehrmann for useful discussions and help with InQubator. Also thanks to Emmet Spier for discussions about the ideas for this work.

REFERENCES

[1] D. H. Ballard, "Animate vision," *Artificial Intelligence*, 1991.
[2] Y. Aloimonos, Ed., *Active Perception*. Erlbaum, Hillsdale, NJ, 1993.
[3] H. Borotschnig, L. Paletta, M. Prantl, and A. Pinz, "Appearance-based active object recognition," *Image and Vision Computing*, vol. 18, pp. 715–727, 2000.
[4] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature*, vol. 2, pp. 194–203, 2001.
[5] D. Walther, L. Itti, M. Reisenhuber, T. Poggio, and C. Koch, "Attentional selection for object recognition – a gentle way," in *Second IEEE International Workshop*, ser. Biologically Motivated Computer Vision, S.-W. Lee, H. Buelthoff, and T. Poggio, Eds., 2002, pp. 387–397.
[6] E. Spier, "Behavioural categorisation: Behaviour makes up bad vision," *Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems*, pp. 133–139, 2004.
[7] D. Floreano, T. Kato, D. Marocco, and E. Sauser, "Coevolution of active vision and feature selection," *Biological Cybernetics*, vol. 90, pp. 218–228, 2004.
[8] E. B. Contreras, H. Buxton, and E. Spier, "Attention can improve a simple model for visual object recognition," submitted to Image and Vision Computing.
[9] J. Howell and H. Buxton, "Receptive fields functions for face recognition," in *Proc. 2nd International Workshop on Parallel Modelling of Neural Operators for Pattern Recognition*, November 1995, pp. 221–226.
[10] I. Harvey, P. Husbands, D. Cliff, A. Thompson, and N. Jakobi, "Evolutionary robotics: the sussex approach," *In Press*, 1996.