A Dynamic Color Perception System for Autonomous Driving in Unmarked Roads: Supplementary Document

Aparajit Narayan, Frédéric Labrosse, Muhanad H. Mohammed Alkilabi, and Elio Tuci

I. Evolutionary active vision approach

Due to lack of space in the main paper, we present a more detailed review of works using embodied neural network controllers with active vision properties in this supplementary document. One of the pioneering works in the use of artificial evolution to develop situated active robotic vision systems is [1]. In this study, a robot suspended from a gantry frame has stepper motors that allow translational movement in the X and Y directions, and a CCD camera pointing down at a mirror inclined at 45° to the vertical. The task of the robot is to distinguish a white isosceles triangle from a white rectangle fixed to one of the black gantry walls by navigating towards the triangle. The contribution of this study is in showing the significance of the movement of the robot in carrying out the discrimination task. The work described in [2] explores the possibility of evolving neuro-controllers for mobile robots that can use their visual perception to perform tasks which are difficult or impossible using only proximal sensing. The task chosen is to move to the center of a cylindrical arena, and stay there. The authors show that relatively small non-modular neural network controllers that process low pixel resolution images, can compensate for this deficiency by generating the actions that bring forth the most informative sensory stimulation, which in turn is used to generate task-effective actions.

Experiments to solve a perceptual task (differentiating between a rectangle and triangle) were also carried out in [3], with a network having only nine cells as its visual input or "retina". The network however had additional feedback units, which gave it the ability to zoom in and out of the image plane. It was also able to control the filtering strategy used to reduce pixels into the final nine values that were fed into the input layer. Thus each input neuron corresponded to an area on the image plane, the size of which was determined by the network's output neuron. A regular feed-forward neural network which remained static and considered the entire image plane in its input vector was also tried to solve the same task. The network was trained using back-propagation under different configurations (number of hidden layers, learning rate and momentum of the gradient descent algorithm). The results showed that under no circumstances could this "static" network succeed in the task of differentiating between triangular and rectangular shapes, despite its higher resolution visual receptive field. The robots were required to carry out vision based collision avoidance. The results of this work indicate that, as it is the case with biological agents (e.g., the "Kitten in the Gondol" experiment mentioned earlier), neural networks with active body movement out-performed their "passive" counterparts. The difference in development of neural networks receptive fields under active and passive vision conditions is further explored by the same lab (Laboratory of Intelligent Systems, Swiss Federal Institute of Technology (EPFL)) in [4] and [5]. In both cases, a recurrent neural network (with hidden units) was evolved to control a mobile robot required to travel collision-free in a walled arena. Networks were evolved for maintaining a straight trajectory during the trial. The network had a feedback unit similar to that in [3] controlling the filtering strategy (i.e., which pixels of the raw image are combined to form the final input vector). The network outputs are also used to set the pan and tilt camera orientation, as well as the speed of the robot wheels. In [4], it is shown that a network evolved in a simulated environment with active vision capabilities (and then ported to the robot) develops sensitivity to a different set of features as compared to a network which is developed by training on static snapshot images of the operational environment. Indeed, as mentioned in [6], neural networks with active vision capabilities develop sensitivity to fewer but more important features in the environment and successfully carry out tasks by maintaining a fix on them. In [5] the importance of active body control for the development of visual systems in neural networks evolved to control mobile robots is further demonstrated. A neural network controller was evolved for the same navigation task described in [4], and two sets of experiments were carried out. In one set an online learning rule (Hebbian plasticity) was applied and the weights of the neural network controller were updated as the network moved in the environment. In the other trials, the weights were also updated, but the robot was only free to move its camera (through pan and tilt outputs), not its body (motor outputs were ignored). After a number of iterations the learning was stopped and the networks with restricted movements were free to move. It was observed that only networks that had retained active body movement during the learning updates could complete the task successfully, while those that were unable to move the robot were unable to carry out the navigation task and avoid colliding with the arena walls.

Besides active vision abilities, it is also important to

Author e-mails: [apn3, ffl, mhm1]@aber.ac.uk (Aberystwyth University), elio.tuci@gmail.com (Middlesex University London)

This work is partially supported by Fujitsu and HPC Wales.

consider the effect of neural networks which can integrate information over time (recurrent networks) as opposed to being completely reactive (regular feed-forward). This is investigated in [7], where a recurrent network (similar to the architecture described in [4] and [5]) was required to differentiate between two patterns and to navigate a mobile robot to a goal destination based on the pattern discrimination task. The two patterns were on opposing walls, while the network's visual receptive field was too small to distinguish one from the other instantly in one update cycle. The network was able to solve the task by scanning the walls, sequentially searching for features and integrating this information to identify larger patterns. Based on these principles, it was demonstrated in [8] that complex machine vision tasks such as automatic driving and indoor navigation could be solved by simple low-resolution artificial neural network controllers when implemented as integrated action-perception control systems wherein information from one update cycle is retained to influence subsequent outputs. The automatic driving task consisted of a network driving a car in different courses in simulated mountain roads. Results indicated that the best evolved networks performed as well as or better than human drivers tested on this virtual platform. Similar to other experiments, this relatively simple network with active vision capabilities was able to successfully navigate these roads and even carry out sharp manoeuvres at bends by focusing on and maintaining a fix on one simple feature (such as the far edge of the road).

II. Testing SegNet

Convolutional neural networks trained on large datasets have outperformed other 'algorithmic' methods on several computer vision benchmark problems like ImageNet, MNIST etc. There have been a number of works mentioned in Section II of the main paper which have implemented them for road-detection. However, from our review there has been yet to be an implementation that has been proved to be applicable for the entire range of environments/scenarios that are to be encountered in the real world. [9] is one such recent implementation wherein the network produces pixel-wise segmentation of the input image. Pixels are classified into a variety of categories such as road, trees etc. The system was successfully evaluated for urban city environments. We tested the network on a few sample images representing more unstructured non-road scenes. The segmented output produced by the network is presented in figure 1. As can be seen from the figure, the output images have such a large number of incorrectly segmented pixels, that it could not be reliably used to steer a self-driving robot.

III. Additional Methodological Details

Figure 2 shows the architecture of the neural network controller. Output neurons number 32 and 33 (see figure 2) control the left motor and neurons 34 and 35 control the right motor. Outputs 36, 37 and 38 regulate the dynamic colour mixing properties of the network. Table I shows the colour distribution properties for textures used to render the 12 evolution environments. Snapshots of these environments are

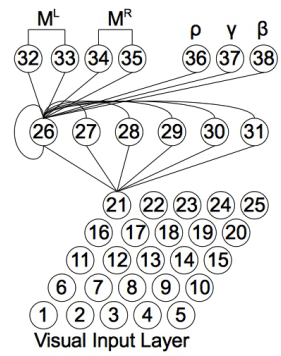


Fig. 2. The neural network. The lines indicate the efferent connections for only one neuron of each layer. Each hidden neuron receives an afferent connection from each input neuron and from each hidden neuron, including a self-connection. Each output neuron receives an afferent connection from each hidden neuron.

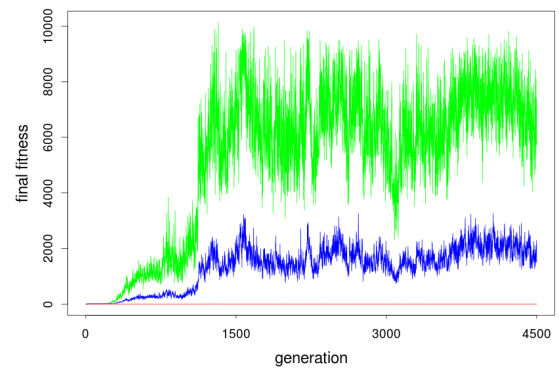


Fig. 3. Fitness graph for best evolutionary run. Green indicates the best, blue the average and red the lowest fitness in a generation.

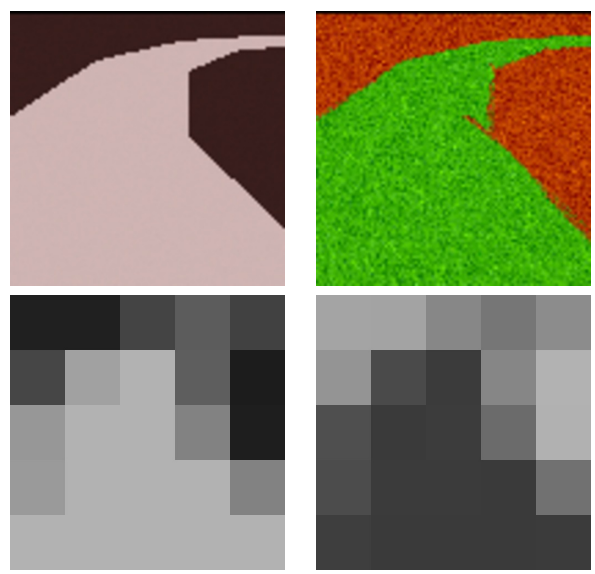
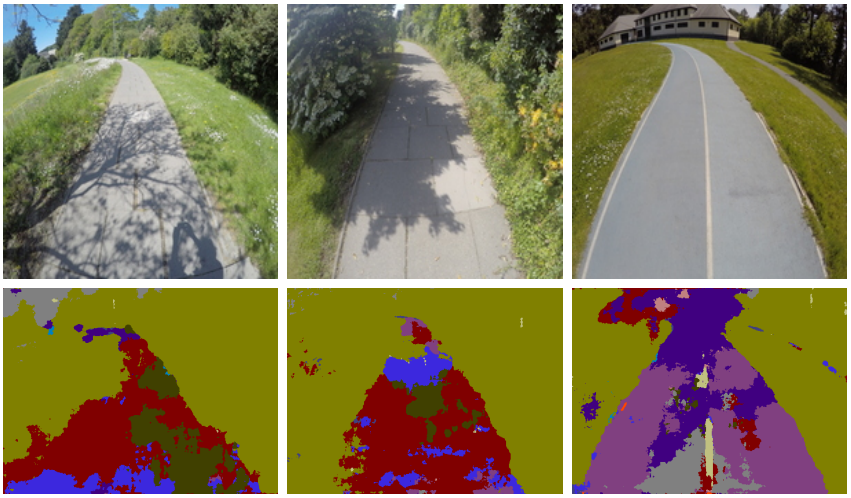


Fig. 5. Views from the 'virtual robot' looking at the road in simulation scenes and the corresponding input vector after the dynamic colour feedback has been applied to the raw image. Column (a) corresponds to scene 2 and column (b) corresponds to scene 7 (see table I).

shown in figure 4. The effect of the network's dynamic colour perception is shown in figure 5. Figure 6 shows examples of



(a) (b)

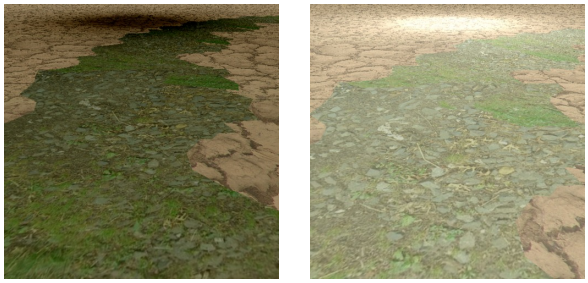
Fig. 1. The top row shows three images of environments where we conducted the experiments described in this study. The bottom row shows the corresponding pixel wise classification of the road images using the deep convolution neural network described in [9]. Purple pixels (which are almost entirely absent for 1.a and 1.b) correspond to the 'road' category.

TABLE I

R,G,B COLOUR DISTRIBUTION OF TEXTURES USED IN CREATING ROAD SCENES. 'R' INDICATES UNIFORM RANDOM DISTRIBUTION IN 0–255.

Scene	Road						Non-Road					
	Red		Green		Blue		Red		Green		Blue	
	Avg	Sd	Avg	Sd	Avg	Sd	Avg	Sd	Avg	Sd	Avg	Sd
1	57.4	1.5	R	R	R	R	207.4	1.6	R	R	R	R
2	207.4	1.6	R	R	R	R	57.4	1.5	R	R	R	R
3	R	R	29.0	6.3	R	R	R	R	179.0	6.3	R	R
4	R	R	179.0	6.3	R	R	R	R	29.0	6.3	R	R
5	R	R	R	R	47.2	4.1	R	R	R	R	197.2	4.1
6	R	R	R	R	197.2	4.1	R	R	R	R	47.2	4.1
7	59.9	20.8	173.4	21.5	R	R	178.5	20.4	53.4	21.3	10	10
8	178.5	20.4	53.4	21.3	R	R	59.9	20.8	173.4	21.5	R	R
9	56.4	21.9	R	R	168.2	21.7	178.8	21.9	R	R	59.6	20.0
10	178.8	21.9	R	R	59.6	20.0	56.4	21.9	R	R	168.2	21.7
11	R	R	185.9	25.7	65.3	25.0	R	R	66.3	26.1	185.9	24.8
12	R	R	66.3	26.1	185.9	24.8	R	R	185.9	25.7	65.3	25.0

rendered scenes used in Test 5 (see Table 1 in main paper) to test the effect of bright spotlights and shadows.



(a) (b)

Fig. 6. Images of simulated environments used in *Test 5*, in which the scenes presents (a) shadows, (b) bright spots. See main paper for further information.

IV. SUPPLEMENTARY RESULTS

Videos of the controller performing in outdoor and virtual environments are available to view/download from the following link https://www.aber.ac.uk/en/cs/research/ir/dss/#road-driving. To further demonstrate that the controller's color perception system is integral to for it's ability to navigate and not simply a by-product of the artificial evolution, we consider the network using the RGB color model but with the feed-back from it's color outputs ignored. Instead based on analysis of the color-properties of the textures used to render scenes in *Test 1* (see main paper section IV), we know the color channel carrying the highest/lowest contrast level for a

TABLE II

TABLE SHOWING THE RELATIVE PERFORMANCE IN TEST 1 OF THE DYNAMIC COLOUR MIXING NEURAL NETWORK PROPOSED BY THIS PAPER. IT IS COMPARED TO A CTRNN USING INPUTS WHEN THE DYNAMIC COLOUR-MIXING IS BYPASSED AND HIGH/LOW CONTRAST COLOUR CHANNELS ARE FED THROUGH.

Network	**Colour Input**	**Percentage Successful Trials**
CTRNN 3 Layers	RGB (Dynamic Mixing)	57.0
CTRNN 3 Layers	RGB (Highest Contrast forced)	41.5
CTRNN 3 Layers	RGB (Lowest Contrast forced)	35.5

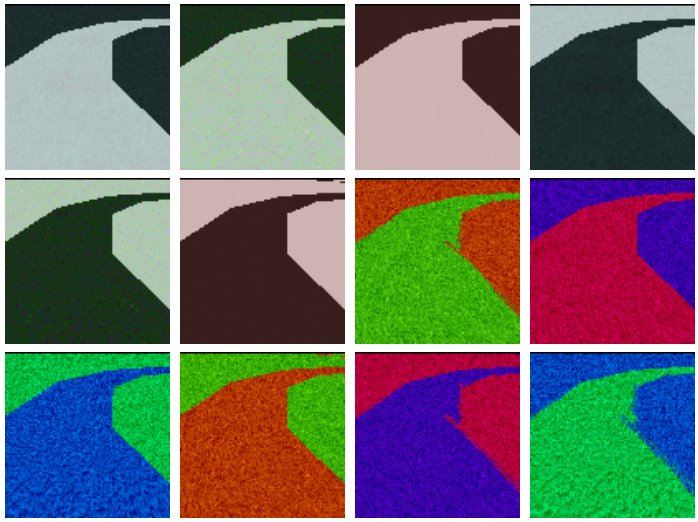


Fig. 4. Snapshots of the 12 virtual evolution environments. The colour distribution properties of the textures used to render these are described in table I.

Fig. 7. Outdoor environments. The black line refers to the robot's trajectory, green dot the starting position and red dot the end position of the trial. Images in the topmost, middle and bottom rows correspond to trials carried out with the ASH, USH and BUV colour models (respectively).

particular environment. We feed the relevant highest/lowest contrast channel to the network for all the trials instead and cut-off the network's color perception feed-back. From the results it can be seen that the network performs the best (57.0 % success) when it's feedback properties are not interfered with. Receiving the highest possible contrast values as it's final input vector deteriorates the performance to 41.5 %.

TABLE III
CENTROIDS OF THE THREE CLUSTERS RESULTING FROM THE MEAN-SHIFT SORTING ALGORITHM. VALUES IN THE CELLS ARE THE PERCENTAGE OF TIME EACH COLOR PARAMETER IS EITHER BELOW THE LOW THRESHOLD OF 0.2, OR ABOVE THE HIGH THRESHOLD OF 0.8.

	Percentages (%)					
Cluster	$\rho > 0.8$	$\rho < 0.2$	$\gamma > 0.8$	$\gamma < 0.2$	$\beta > 0.8$	$\beta < 0.2$
one	17.8	81.1	59.3	39.8	21.1	77.3
two	5.6	94.0	88.2	10.9	5.1	94.0
three	21.6	77.3	41.3	58.3	35.7	63.0

In the main paper, section IV.C deals with analysing the system's mechanisms that enable it to achieve adaptability to different environments. A piece of analysis that we carried out dealt with capturing activations of the three colour parameters ρ, γ, and β for all outdoor trials (carried out with the USH colour model). The activations for each of the 50 trials were compressed into 6 dimensional points. The points represented how many iterations ρ, γ, and β were above a high (0.8) and low (0.2) threshold. These points were then fed into an unsupervised clustering algorithm (mean-shift), which separated the points into three clusters as shown in table IV. These 50 points plotted in each dimension are shown in figure 8.

REFERENCES

[1] I. Harvey, P. Husbands, and D. Cliff, "Seeing the light: artificial evolution, real vision," in *Proc. of 3rd Int. Conf. on Simulation of Adaptive Behavior (SAB)*, D. Cliff, P. Husbands, J. Meyer, and S. Wilson, Eds. MIT Press/Bradford Books, Boston MA, 1994, pp. 392–401.

[2] D. Cliff, P. Husbands, and I. Harvey, "Explorations in evolutionary robotics," *Adaptive Behavior*, vol. 2, no. 1, pp. 73–110, 1993.

[3] T. Kato and D. Floreano, "An evolutionary active-vision system," in *Proc. of the Congress on Evolutionary Computation (CEC)*, vol. 1. IEEE, 2001, pp. 107–114.

[4] M. Suzuki, D. Floreano, D. Paolo, and A. Ezequiel, "The contribution of active body movement to visual development in evolutionary robots," *Neural Networks*, vol. 18, no. 5-6, pp. 656–665, 2005.

[5] M. Suzuki, D. Floreano, and E. A. D. Paolo, "Constraints on body movement during visual development affect behavior of evolutionary robots," in *Proc. of the Int. Joint Conf. on Neural Networks*, 2005.

[6] M. Suzuki and D. Floreano, "Enactive robot vision," *Adaptive Behavior*, vol. 16, no. 2-3, pp. 122–128, 2008.

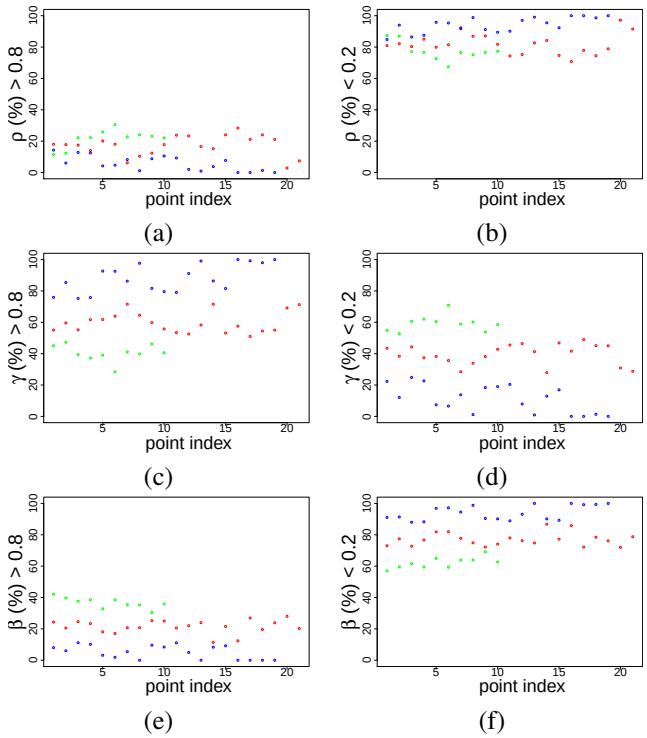


Fig. 8. Components of the 6-dimensional points considered for clustering. Points in red, blue and green correspond to clusters one, two and three, respectively (also see Table IV).

[7] ——, *Evolutionary Active Vision Toward Three Dimensional Landmark-Navigation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 263–273.

[8] D. Floreano, T. Kato, D. Marocco, and E. Sauser, "Coevolution of active vision and feature selection," *Biological Cybernetics*, vol. 90, no. 3, pp. 218–228, 2004.

[9] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv preprint arXiv:1511.00561*, 2015.

Fig. 9. To play the video, click on the image or use the following URL https://www.youtube.com/embed/EtgPU-mwn94. This video shows the 'evolved' neural network controlling the Pioneer 3-AT robot in 'Path 1' using the 'USH' colour model (refer to paper).

Fig. 10. To play the video, click on the image or use the following URL https://www.youtube.com/embed/6XBtsxax5xk. This video shows the 'evolved' neural network controlling the Pioneer 3-AT robot in 'Path 2' using the 'USH' colour model (refer to paper).

Fig. 11. To play the video, click on the image or use the following URL https://www.youtube.com/embed/MKvPLvQHcbM. This video shows the 'evolved' neural network controlling the Pioneer 3-AT robot in 'Path 3' using the 'USH' colour model (refer to paper).

Fig. 12. To play the video, click on the image or use the following URL https://www.youtube.com/embed/hyr5J47V_w0. This video shows the 'evolved' neural network controlling the Pioneer 3-AT robot in 'Path 4' using the 'USH' colour model (refer to paper).

Fig. 13. To play the video, click on the image or use the following URL https://www.youtube.com/embed/Mhki09BN0YM. This video shows the 'evolved' neural network controlling the Pioneer 3-AT robot in 'Path 5' using the 'USH' colour model (refer to paper).

Fig. 14. To play the video, click on the image or use the following URL https://www.youtube.com/embed/wn6YvTwEWCQ. This videos shows an 'evolved' neural network controller navigating different virtual simulated road environments.