Omnidirectional Active Vision for Evolutionary Car Driving

Mototaka Suzuki [a,1], Jacob van der Blij [b], and Dario Floreano [a]

[a] *Laboratory of Intelligent Systems, Ecole Polytechnique Fédérale de Lausanne (EPFL)*
[b] *Artificial Intelligence Institute, University of Groningen (RuG)*

Abstract. We describe a set of simulations to evolve *omnidirectional active vision*, an artificial retina scanning over images taken via an omnidirectional camera, being applied to a car driving task. While the retina can immediately access features in any direction, it is asked to select behaviorally-relevant features so as to drive the car on the road. Neural controllers which direct both the retinal movement and the system behavior, i.e., the speed and the steering angle of the car, are tested in three different circuits and developed through artificial evolution. We show that the evolved retina moving over the omnidirectional image successfully detects the task-relevant visual features so as to drive the car on the road. Behavioral analysis illustrates its effective strategy in algorithmic, computational, and memory resources.

Keywords. Active Vision, Omnidirectional Camera, Artificial Evolution, Neural Networks, Mobile Robots

1. Introduction

The omnidirectional camera is a relatively new optic device that provides a 360 degrees field of view, and it has been widely used in many practical applications including surveillance systems and robot navigation [1,2,3]. However, in most applications visual systems uniformly process the entire image, which would be computationally expensive when detailed information is required. In other cases the focus is determined for particular uses by the designers or users. In other words, the system is not allowed to freely interact with the environment and selectively choose visual features.

Contrarily, all vertebrates and several insects – even those with a very large field of view – share the steerable eyes with a foveal region [4], which means that they have been forced to choose necessary information from a vast visual field at any given time so as to survive. Such a sequential and interactive process of selecting and analyzing behaviorally-relevant parts of a visual scene is called *active vision* [5,6,7,8]. Our previous work has demonstrated that it can also be applied to a number of real world problems [9].

In this article we explore *omnidirectional active vision* applied to a car driving task: Coupled with an omnidirectional camera, a square artificial retina can immediately ac-

---

[1]Correspondence to: Mototaka Suzuki, Ecole Polytechnique Fédérale de Lausanne, EPFL-STI-I2S-LIS, Station 11, CH-1015 Lausanne, Switzerland. Tel.: +41 21 693 7742; Fax: +41 21 693 5859; E-mail: Mototaka.Suzuki@epfl.ch; WWW homepage: http://lis.epfl.ch.

cess any visual feature located in any direction, which is impossible for the conventional pan-tilt camera because of the mechanical constraints. It is challenging for the artificial retina to select behaviorally-relevant features in such a broad field of view so as to drive a car on the road.

Omnidirectional active vision is not biologically plausible. But as claimed in [10] it is interesting to study visual systems from a broader point of view which contains those that have never been available in biology. Also, there are several engineering applications that could benefit from omnidirectional vision. Some promising applications are discussed in section 5.

A $1/10$ scale model car equipped with an omnidirectional camera and three different circuits are modeled in simulation. We show that the evolved retina moving over the omnidirectional image successfully detects the task-relevant visual features so as to drive the car on the road. Behavioral analysis illustrates its effective strategy in algorithmic, computational, and memory resources. In comparison to the results obtained with a pan-tilt camera mounted on the same car, we show that omnidirectional active vision performs the task very robustly in spite of more difficult initial conditions.

2. Experimental Setup

Figure 1 shows the real and simulated model cars as well as the views through the real and simulated omnidirectional cameras. The omnidirectional camera consists of a spherical mirror and a CCD camera. It is mounted on a $1/10$ scale model car (Kyosho^{TM}) which has four motorized wheels. We simulated the car and the circuits using Vortex[1] libraries, a commercially available software package that models gravity, mass, friction, and collisions. Additionally we used a vision software for modeling the view from the omnidirectional camera, which had originally been developed in the Swarm-bots project[2]. Figure 2 shows the three circuits; ellipse, banana, and eight shaped, used in the present evolutionary experiments.

An artificial retina actively moves over the omnidirectional view[3]. Figure 3 illustrates the unwrapping process from the polar view to the Cartesian view and the retina overlaid on each image. In order to evaluate the performance of the omnidirectional active vision system, we also simulate a pan-tilt camera mounted on the same car and compare the results obtained in the same experimental condition.

The neural network is characterized by a feedforward architecture with evolvable thresholds and discrete-time, fully recurrent connections at the output layer (Figure 4). The input layer is an artificial retina of five by five visual neurons that receive input from a gray level image of 240 by 240 pixels. The activation of a visual neuron, scaled between 0 and 1, is given by the average gray level of all pixels spanned by its own receptive field or by the gray level of a single pixel located within the receptive field. The choice between these two activation methods or filtering strategies, can be dynamically changed by one output neuron at each time step. Two proprioceptive neurons provide input information about the measured position of the retina with re-

---

[1]http://www.cm-labs.com
[2]http://www.swarm-bots.org/
[3]A similar approach has been taken for evolving flocking behavior of three simulated robots independently in [11], inspired by our previous work [9].
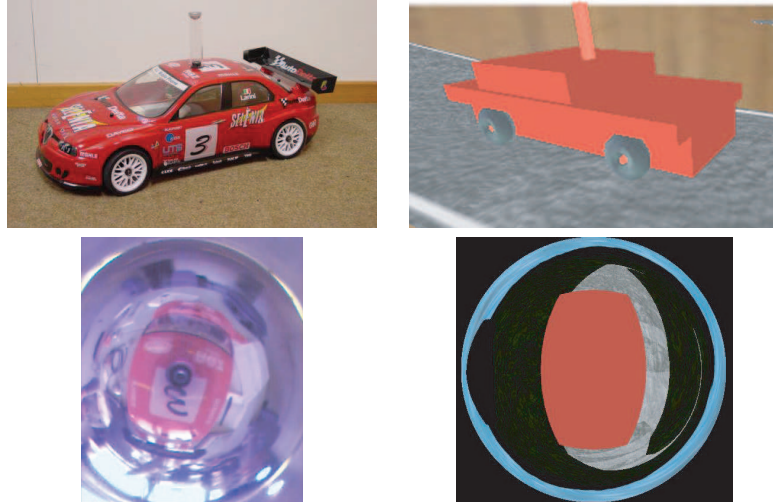
Figure 1. The real $1/10$ scale 4WD model car (Kyosho^{TM}) with an omnidirectional camera mounted on the roof of the car (top left) and the simulated one (top right). The car base is 19.5 cm (W), 43 cm (L), and 13.5 cm (H). View through the real omnidirectional camera (bottom left) and one through the simulated camera (bottom right).
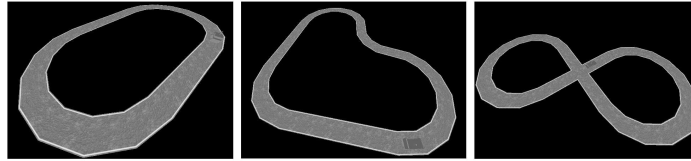

Figure 2. Three circuits where the robot car is tested: from left to right, ellipse, banana, and eight shaped circuits. Each circuit is designed such that an 8 m×8 m room accommodates it. The width of the roads in all circuits is 50 cm. In the present experiments the sidelines are omitted.

spect to the chassis of the car: the radial and angular coordinates for the omnidirectional camera; or the pan and tilt degrees for the pan-tilt camera. These values are in the interval $[retina\_size/2, radius - retina\_size/2]$ pixels ($retina\_size = 240$ pixels, $radius = 448$ pixels in these experiments) and $[0, 360]$ degrees for radial and angular coordinates respectively. The values for the pan-tilt camera are in the interval $[-100, 100]$ and $[-25, 25]$ degrees respectively. Each value is scaled in the interval $[0, 1]$ and encoded as a proprioceptive input. A set of memory units stores the values of the output neurons at the previous sensory motor cycle step and sends them back to the output units through a set of connections, which effectively act as the recurrent connections among output units [12]. The bias unit has a constant value of -1 and its outgoing connections represent the adaptive thresholds of the output neurons [13].

Output neurons use the sigmoid activation function $f(x) = 1/(1 + exp(-x))$ in the range $[0, 1]$, where x is the weighted sum of all inputs. They encode the motor commands of the active vision system and of the car for each sensory motor cycle. One neuron determines the filtering strategy used to set the activation values of visual neurons for the
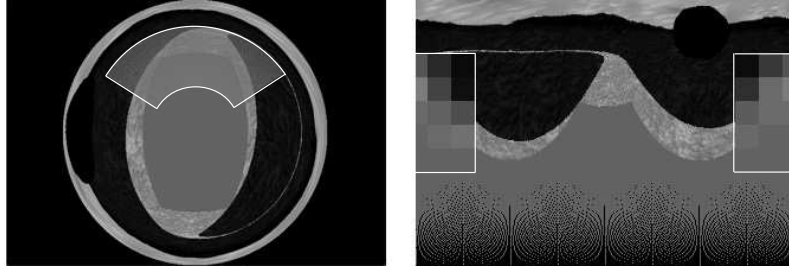
Figure 3. Left: The polar image taken by the omnidirectional camera and the overlaid retina. Right: The corresponding unwrapped image and the retina in Cartesian coordinate.

next sensory motor cycle. Two neurons control the movement of the retina (or camera), encoded as speeds relative to the current position. The remaining two neurons encode the directional and rotational speeds of the wheels of the car. Activation values above 0.5 stand for left (directional) and forward (rotational) speeds whereas activation values below 0.5 stand for right and backward speeds, respectively.
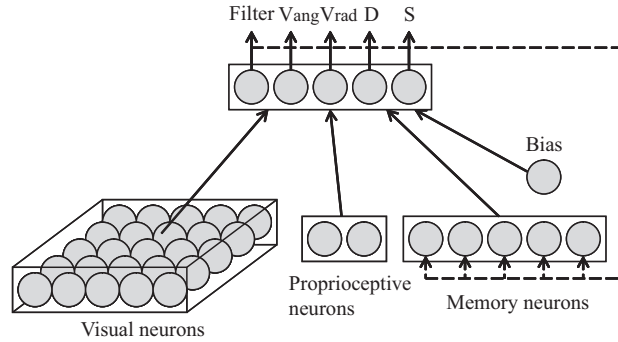

Figure 4. The architecture is composed of a grid of visual neurons with nonoverlapping receptive fields whose activation is given by the gray level of the corresponding pixels in the image; a set of proprioceptive neurons that provide information about the movement of the retina with respect to the chassis of the car; a set of output neurons that determine at each sensory motor cycle the filtering used by visual neurons, the new angular (V_{ang}) and radial (V_{rad}) speeds of the retina (or pan and tilt speeds of the pan-tilt camera), and the directional (D) and rotational (S) speeds of the wheels of the car; a set of memory units whose outgoing connection strengths represent recurrent connections among output units; and a bias neuron whose outgoing connection weights represent the thresholds of the output neurons. Solid arrows between neurons represent fully connected layers of weights between two layers of neurons. Dashed arrows represent 1:1 copy connections (without weights) from output units to memory units, which store the values of the output neurons at the previous sensory motor cycle step.

The neural network has 165 evolvable connections that are individually encoded on five bits in the genetic string (total length= 825). A population of 100 individuals is evolved using truncated rank-based selection with a selection rate of 0.2 (the best 20 individuals make four copies each) and elitism (two randomly chosen individuals of the population are replaced by the two best individuals of the previous generation). One-point crossover probability is 0.1 and bit-toggling mutation probability is 0.01 per bit.

3. Evolution of Neural Controllers

The fitness function was designed to select cars for their ability to move straight forward as long as possible for the evaluation time of the individual. Each individual is decoded and tested for three trials, each trial lasting 500 sensory motor cycles. A trial can be truncated earlier if the operating system detects a drop of the height of the center of the car, which corresponds to going off-road. The fitness criterion F is a function of the measured speed S_t of the four wheels and the steering direction D_t of the front wheels:

$$F = \frac{1}{E \times T \times S_{max}} \sum_{e=0}^{E} \sum_{t=0}^{T'} f(S_t, D_t, t) \tag{1}$$

$$f(S_t, D_t, t) = S_t \times \left(1 - \sqrt{|D_t|/D_{max}} \right) \tag{2}$$

where S_t and D_t are in the range $[-8.9, 8.9]$ cm/sec and $[-45, 45]$ degrees, respectively; $f(S_t, D_t, t) = 0$ if S_t is smaller than 0 (backward rotation); E is the number of trials (three in these experiments), T is the maximum number of sensory motor cycles per trial (500 in these experiments), T' is the observed number of sensory motor cycles (for example, 34 for a robot whose trial is truncated after 34 steps if the car goes off-road). At the beginning of each trial the position and orientation of the car as well as the position of the retina within the image are randomly initialized. We performed these replications of the evolutionary run starting with different genetic populations. In all cases the fitness reached stable values in less than 20 generations (Figure 5) which correspond to successful on-road driving. The fitness values both of the best individuals and of the population average obtained with the pan-tilt camera were higher than those with the omnidirectional camera in all three circuits. Notice that the fitness can never be one because the car must steer in corners so as to not go off-road.
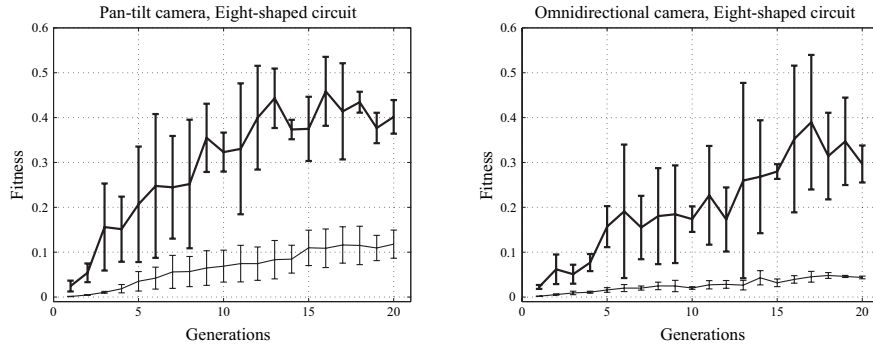

Figure 5. Evolution of neural controllers of the car with the pan-tilt camera (left) and the omnidirectional camera (right) in the eight shaped circuits. Thick line shows the fitness values of the best individuals and thin line shows those of the population average. Each data point is the average of three evolutionary runs with different initializations of the population. Vertical lines show the standard error.

4. Behavioral Analysis

The behavioral strategy of the best evolved car equipped with the pan-tilt camera is as follows: At the beginning of the trial the car points the camera downward and to its right (or left depending on the evolutionary run), and it steers so to maintain the edge between the road and the background within the retina.

Due to the lack of space, we show only the behavioral analysis of the best individual with the omnidirectional camera evolved in the eight shaped circuit because the circuit possesses all of the geometrical characteristics which the ellipse and banana shaped circuits have. It also has an intersection which would disturb the car's perception if simple edge detection is the strategy developed by the evolved individual, which is sufficient for driving in the banana and ellipse shaped circuits. Indeed, since the best evolved individuals with the pan-tilt camera all adopted this strategy, they were disturbed more largely at the intersection and went off-road in several trials. Our preliminary tests also confirmed that the best individuals evolved in the eight shaped circuit were general in the sense that they could drive in the other two circuits as well.

Figure 6 shows the strategy of the best individual evolved in the eight shaped circuit: During the starting period the evolved car moves back and forth, and then starts going forward at full speed once the retina finds the front road arena. Another effective strategy acquired by the best individuals in other evolutionary runs is that the car moves forward very slowly until the retina finds the road area, and starts driving at full speed immediately after finding it (data not shown). Both behaviors serve to spare time for the retina to find the road area during the most critical period at the beginning of the trial.

In the intersection although the perception of the car is disturbed by the crossing road, which corresponds to the slight deviation of the trajectory, the evolved car manages to find the straight road beyond the intersection by moving the retina upward in the radial direction, which corresponds to "looking farther", and corrects its trajectory (Figure 6, right). After passing the intersection, the retina moved downward again and maintained the straight road area in sight. The rotational speeds of the wheels and the angular position of the retina did not change significantly while passing the intersection.

5. Discussion

The slightly lower fitness values of the evolved individuals with the omnidirectional camera than those with the pan-tilt camera are due to two main reasons: 1) It is harder to find the front road area out of the omnidirectional camera view than out of the pan-tilt camera view; 2) Evolved individuals can find the area, but it takes them some time because during the road search the car does not move much.

Despite this more difficult initial condition, we have shown that artificial evolution selects the individuals capable of "waiting" until the retina finds the appropriate location and of driving at full speed after that. Therefore the slightly lower fitness values of the best evolved individuals with the omnidirectional camera do not mean that those individuals are inferior to those with the pan-tilt camera. The lower fitness is caused by the waiting behavior at the beginning of each test in order to robustly start driving at full speed afterward. Such a behavior has never been observed in any evolutionary run with the pan-tilt camera. Once the road area is detected, the best evolved car with the omnidi-
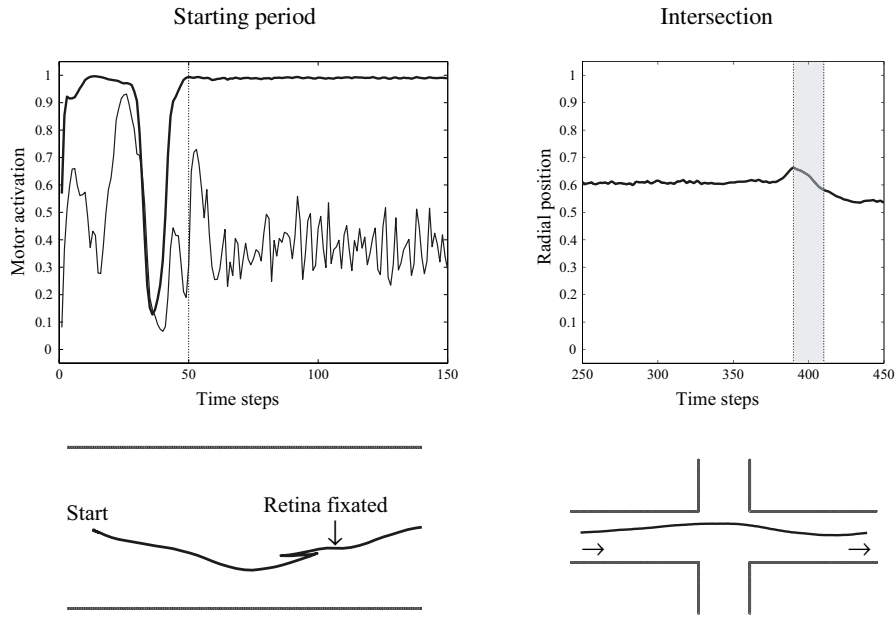
Starting period   Intersection

Figure 6. An example of the evolved car with the omnidirectional camera driving in the eight shaped circuit. Top left: Motor activations of the wheel speed (thick line) and the steering angle (thin line) during the first 150 time steps. Speed values above – and below – 0.5 stand for forward – and backward – rotational speeds. Angle values are mapped from 0.5 to 1 for rotations to the left of the car and from 0.5 to 0 for rotations to the right. Dotted vertical line denotes the moment when the retina fixated its position on the front road area. Bottom left: The corresponding trajectory of the car (shown only for the first 70 time steps). Top right: The radial position of the retina. Shaded period corresponds to while passing the intersection. The angular position of the retina remained stable and did not change even while passing the intersection. Bottom right: The corresponding trajectory of the car around the intersection.

rectional camera perfectly drives as that with the pan-tilt camera does. The advantage is lower algorithmic, computational, and memory resources.

For comparison with the results obtained with the pan-tilt camera, we did not implement the zooming function in the present experimental setup. However it enables the system to select visual features more specifically by choosing an appropriate resolution for each feature. Indeed, our previous work has demonstrated that in a non-panoramic view the active zooming function played a crucial role in their performance [9], which encourages us to further apply it to the current system.

The current setup also allows us to lay multiple retinas within a single omnidirectional camera image so that they each specialize in detecting separate features simultaneously and perform behaviors. In real circuits, there are a number of features to which car drivers must pay attention (e.g., sidelines, signals, other cars). Indeed, [14] developed a multi-window system with a conventional non-panoramic camera for detecting such features during real car navigation, but the position and shape of each window were predetermined by the designers. Active feature selection by multiple retinas which are moving independently over the omnidirectional view may display significant advantages over an active single retina or the fixed multi-window system in several tasks, e.g., navigation

in dynamic, unknown environments. Further investigations must be done to validate this hypothesis.

The present method may also offer a significant advantage in landmark navigation of mobile robots because of its fast, parallel searching for multiple landmarks in any direction. Instead of processing the 360 degrees field of view, the system could actively extract only small task-related regions from the visual field, which dramatically lightens the computation and the memory consumption.

6. Conclusions

In this article we have explored *omnidirectional active vision* applied to a car driving task. The present simulations have shown that the evolved artificial retina moving over the omnidirectional view successfully detects the behaviorally-relevant feature so as to drive the car on the road.

Although it costs time to find the appropriate feature in such a broad field of view during the starting period, the best evolved car overcomes the disadvantage by moving back and forth or moving slowly until the retina finds the appropriate feature. Then the car starts driving forward at full speed. The best evolved car equipped with the omnidirectional camera performs robust driving on the banana, eight, and ellipse shaped circuits in spite of the more difficult initial condition. The advantage of the present methodology is lower algorithmic, computational, and memory resources.

Currently we aim to implement an active zooming function of the retinas, to generalize the neural controllers in different-featured circuits (e.g., backgrounds, sidelines, textures), and to transfer the evolved neural controller into the real car shown in Figure 1.

Acknowledgments

References

[1] Y. Matsumoto, K. Ikeda, M. Inaba, and H. Inoue. Visual navigation using omnidirectional view sequence. In *Proceedings of International Conference on Intelligent Robots and Systems*, pages 317–322, 1999.

[2] L. Paletta, S. Frintrop, and J. Hertzberg. Robust localization using context in omnidirectional imaging. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 2072–2077, 2001.

[3] I. Stratmann. Omnidirectional imaging and optical flow. In *Proceedings of the IEEE Workshop on Omnidirectional Vision*, pages 104–114. IEEE Computer Society, 2002.

[4] M. F. Land and D.-E. Nilsson. *Animal Eyes*. Oxford University Press, Oxford, 2002.

[5] R. Bajcsy. Active perception. *Proceedings of the IEEE*, 76:996–1005, 1988.

[6] J. Aloimonos, I. Weiss, and A. Bandopadhay. Active vision. *International Journal of Computer Vision*, 1(4):333–356, 1987.

[7] J. Aloimonos. Purposive and qualitative active vision. In *Proceedings of International Conference on Pattern Recognition*, volume 1, pages 346–360, 1990.

[8] D. H. Ballard. Animate vision. *Artificial Intelligence*, 48(1):57–86, 1991.

[9] D. Floreano, T. Kato, D. Marocco, and E. Sauser. Coevolution of active vision and feature selection. *Biological Cybernetics*, 90(3):218–228, 2004.

[10] C. G. Langton. Artificial life. In *Artificial Life*, pages 1–48. Addison-Wesley, 1989.

[11] S. Lanza. Active vision in a collective robotics domain. Master's thesis, Faculty of Engineering, Technical University of Milan, 2004.

[12] J. L. Elman. Finding structure in time. *Cognitive Science*, 14:179–211, 1990.

[13] J. Hertz, A. Krogh, and R. G. Palmer. *Introduction to the theory of neural computation*. Addison-Wesley, Redwood City, CA, 1991.

[14] E. D. Dickmanns, B. Mysliwetz, and T. Christians. An integrated spatio-temporal approach to automatic visual guidance of autonomous vehicles. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(6):1273–1284, 1990.