

An Evolutionary Active-Vision System

Toshifumi Kato

Institute of Robotic Systems
Swiss Federal Institute of Technology (EPFL)
CH-1015 Lausanne, Switzerland
Toshifumi.Kato@epfl.ch

Dario Floreano

Institute of Robotic Systems
Swiss Federal Institute of Technology (EPFL)
CH-1015 Lausanne, Switzerland
Dario.Floreano@epfl.ch

Abstract- We describe an evolutionary vision system capable of autonomously scanning through an image while zooming in and out and changing filtering strategy in order to perform shape discrimination. The system consists of a small artificial retina controlled by an evolutionary recurrent neural network without hidden units. We show that such a simple active-vision system can successfully recognize different shapes independently of their position and size by dynamically exploring relevant parts of the image. We also show that a standard feed-forward neural network trained with the back-propagation algorithm cannot perform the task, not even with hidden units added to the architecture. Given its compactness, computational requirements, and versatility, this evolutionary active vision system is a suitable solution for small-size and embedded vision systems with stringent energetic and computational requirements, such as micro-robotic systems. In addition, this approach provides a framework for studying emergent active-vision behaviors in autonomous systems.

Keywords: Active vision, evolution and neural networks, image discrimination.

1 Introduction

The conventional approach to artificial vision consists of passing an image through a sequential series of filters (an operation technically known as *convolution*) in order to progressively compress the original data and extract higher order features [18, 24]. This approach is based on the more or less explicit assumption that the goal of vision is to create representations that can be manipulated, memorized, and added to symbolic libraries. Similar mechanisms and assumptions are also found in models of biological vision where filters are represented by serially stacked layers of neurons whose functionalities are given by the patterns of local connectivity [16, 17, 21]. The main difference between artificial and bio-inspired systems is that the latter are often parallel, distributed, and capable of self-tuning the filter properties to the image statistics. In both cases, the vision systems require powerful computation and large memory resources.

The need for representations in vision has been recently criticized [5] because it implies the presence of an *homunculus* (“a little man in the head”) who understands the language used for representations and can relate those representations

to things in the image. Rooted in the ecological approach to vision championed by J.J. Gibson [10], these critics emphasize evidence that biological and machine vision can go a long way by relying only on simple and specialized mechanisms that exploit dynamical interactions between an organism and its environment [13, 23, 7].

Along similar lines, an increasing number of researchers emphasize the importance of *active vision*, a process by which an organism executes motor actions in order to select sensory information that makes a discrimination easier [1, 2].

The evolutionary system described in this paper falls within the representation-free and active-vision philosophy described above. It is an extremely compact system (without mechanisms that could support internal representations) that can perform complex visual discrimination by actively scanning an image while zooming in and out and changing filtering strategy. The large banks of parallel filters used in conventional vision systems are substituted here by a single small retina-like filter that is free to roam through the image and dynamically change its properties to carry out an otherwise conventional discrimination task. The way in which this small retina moves around the image and changes its resolution and filtering strategy is not predefined, but emerges out of an evolutionary process that selects individuals only for their ability to correctly make visual discriminations.

1.1 Related work

The literature on parsimonious, representation-free, and active vision systems is still relatively small and is to be found mainly in the domain of behavior-based and bio-inspired robotics (see, for example, the proceedings of the conferences *Simulation of Adaptive Behavior. From Animals to Animals* published by MIT Press).

Pioneering work on evolution of an active vision system on a mobile robot has been performed by the Sussex group [12]. The authors have incrementally evolved visually guided behaviors and sensory morphologies for a mobile robot expected to navigate towards a rectangle while avoiding a triangle. They started with a simple version of the task that selected controllers for their ability to navigate towards a wall covered by color paper. Subsequently, they narrowed the area of the color paper on the wall and resumed evolution from the last evolved generation. Finally, the rectangular shape was displaced and a triangular shape was positioned nearby. The fitness function was modified to encourage avoidance

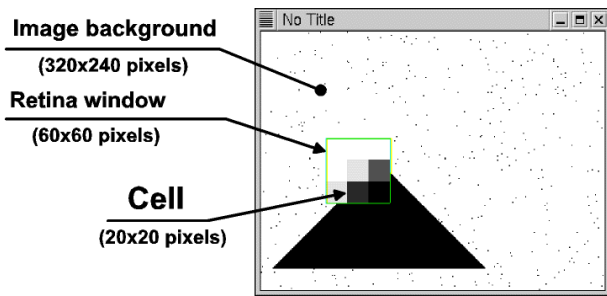


Figure 1: A snapshot of the active vision system over an image containing a triangle. The retina is composed of 9 cells.

of the rectangle and movement towards the triangle, and the previously evolved population was incrementally evolved in the new environment. The sensory morphology of the final evolved individual consisted of only two small receptive fields positioned along a diagonal line so that the combination of robot movements and image projection over the two receptive fields would cause navigation towards the correct shape.

Beer and colleagues [3, 23] evolved and analyzed small continuous recurrent neural network controlling a simulated agent equipped with simple vision systems where the dynamics of the interactions between the agent's actions and the visual signals allow the emergence of a complex set of vision-based abilities, such as discrimination between the self and other bodies, passing through openings suitable for the body size, and remembering object locations.

Nolfi and Marocco [20] evolved a neural network for a mobile robot equipped with a linear vision system capable of visually discriminating between ambiguous landmarks by exploiting the dynamic interactions between movements and photoreceptor activations.

The evolutionary active vision system described in this paper differs from those described above in that a) it does not attempt to evolve a visual morphology, but it assumes a geometric retina-like array of photoreceptors; b) it allows the neural mechanisms to change not only the position of the retina in the visual field (similar to the movement of a robot in three dimensions), but also the resolution of the retinal projection (similar to a zooming effect) and the characteristics of the receptive fields; c) it can be used both for conventional image discrimination tasks and for vision-base robot navigation; d) it is based on a simple behavioral fitness function that selects individuals for their ability to correctly recognize a shape. We will come back to the rationale and implications of these choices in the discussion section.

The experiments described in this paper are based on discrimination of shapes in static images. In order to compare the behavioral strategies of our active vision system with the above mentioned vision-based robot controller evolved by Harvey *et al.*, we test our system on images containing triangles and squares. In addition, the shapes can appear anywhere in the image and vary in size.

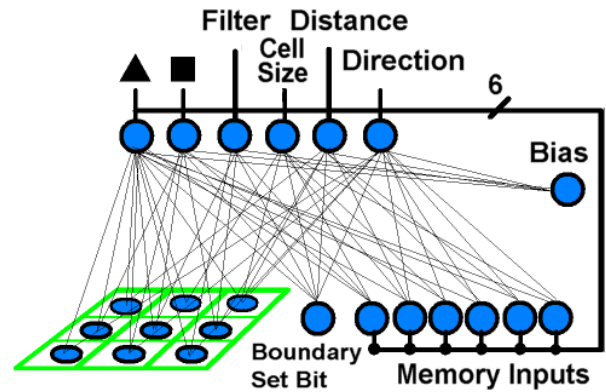


Figure 2: Neural architecture of the active vision system. Six output units receive signals from the retina cells and from a unit signalling whether the retina is against a border. The output units have recurrent connections, here represented as memory units that hold the activation of the output units at the previous time step [8].

2 System architecture

The active vision system consists of a small retina composed of a matrix of cells. In the experiments described here, the retina has a fixed size of 3 by 3 cells. Each cell has a receptive field mapping the values of underlying pixels in the image into a single value. Each pixel can take a value between 0 and 255 indicating a grey level. The retina can move across the image, zoom in and out, and change the property of the receptive fields (filtering strategy). Figure 1 shows the active vision system over an image containing a black triangle.

The outputs of the retinal cells are fed into a recurrent neural network without hidden units (figure 2). An additional input neuron becomes active whenever the retina hits a border of the image (this neuron would not be necessary in a mobile robot immersed in an environment and free to rotate). If the retina attempts to move over the border, this movement is suppressed.

The activations of the output units are passed through a sigmoid function. Two output units of the network encode the type of shape recognized by the system (triangle and square), the most active unit between the two being considered as the network response. A third unit encodes the type of filtering strategy used by all cells in the retina. In these experiments, we considered only two simple filtering strategies: sampling and averaging (figure 3). If the activation of this unit is above 0.5, the cell returns the value of the top leftmost pixel in the receptive field (sampling). If the activation is below 0.5, it returns the average of all pixel values. Therefore, the values returned by the retina can dynamically change depending on the value of this output unit.

A fourth unit encodes the resolution of cells in the retina, that is the receptive field area. The activation level of this unit is used to set the area of each cell to one of three possible resolutions (all cells in the retina have the same area). In

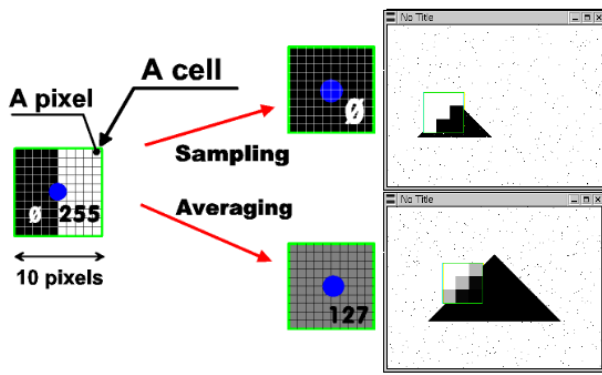


Figure 3: Two filtering strategies are available. Each cell can either return the value of the top leftmost pixel (black=0) or the average value of all pixels (gray=127). The first strategy is called sampling, the second is called averaging. The filtering strategy is set by one of network output units after every activation of the network. All cells in the retina use the same sampling strategy, as shown by the two snapshots on the right side of the figure.

the experiments described here, the side of each cell can be 5, 10, or 20 pixels corresponding to a retinal side of 15, 30, and 60 pixels respectively, as shown in figure 4. In addition, only at the first time step when a new image is presented, the cell side is set to 80 (corresponding to a retina side of 240 pixels) to allow a global view of the image and thus potentially reduce the time needed to search for the shape. A larger retinal size provides lower resolution and corresponds to a zoom out effect.

The remaining two output units encode the movement of the retina across the image in terms of distance (expressed in pixels) and angle with respect to its top leftmost corner. Distance and angle are a function of unit activations, the maximum distance being 50 pixels and the maximum angle being 359 degrees. Recurrent connections are implemented by adding a set of memory units that encode a copy of the output unit activations at the previous time step [8].

The connection strengths are evolved using the simple genetic algorithm described in [11]. Connection weights can take values in the interval $[-4.0, 3.0]$ and are encoded into 5 bits. The fixed architecture described above has 102 weights (including bias connections). An initial population of 100 individuals is evolved using truncated rank-based selection (the best 20 individuals make 5 copies each) and elitism (a randomly chosen individual of the new population is replaced by the best individual of the previous generation). Crossover probability per pair is 0.1 and mutation probability per bit is 0.01.

3 Shape Discrimination

In the experiments reported here the active vision system is evolved to discriminate between two shapes, an isosceles tri-

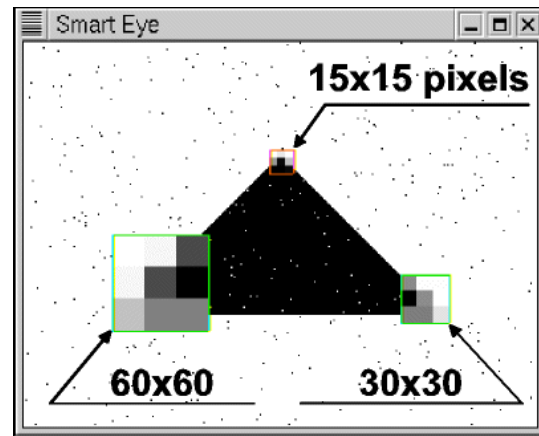


Figure 4: The retina can take on the three different sizes shown in the figure. The size is set by the corresponding output unit of the network after each activation. Larger sizes offer lower resolution and correspond to a zoom out effect.

angle and a square. We have chosen these two shapes in order to allow a comparison between the behavioral strategies evolved by our system and those evolved by the robot described by Harvey et al.

Each individual of the population is presented with 20 images, 10 containing a triangle and 10 containing a square. The entire image, or frame, is 320 pixels wide and 240 pixels high. Each image can have only one triangle or one square. Triangles and squares always appear at a random position in the image and take a random size between 20 and 100 pixels in height. The geometrical shapes are black (pixel value = 0) against a white background (pixel value = 255) (see figure 1). Since triangles are isosceles, the base is always set twice the height so that the total area is equal to that of a square of equal height. By doing so, the evolutionary vision system cannot recognize them by simply relying on the total area. In addition, some noise is added to the entire image by inverting the value of each pixel (black to white or vice versa) with a probability of 0.005.

Whenever a new image is presented to an individual, the memory units are set to zero, the retinal size is set to 240 by 240 pixels and to averaging mode, and the retina is positioned at the center of the image. This large size is available only at the first instant a new image is presented to let the vision system know more or less where the shape is instead of engaging in a random search with a small retinal size. Following that, the active vision system is let free to move 50 times and change resolution size and filtering strategy every time the output units decide to do so.

At each time step, the discrimination response (triangle or square) is recorded and used by the fitness function F to select individuals that respond correctly the highest number of times over all images, as follows

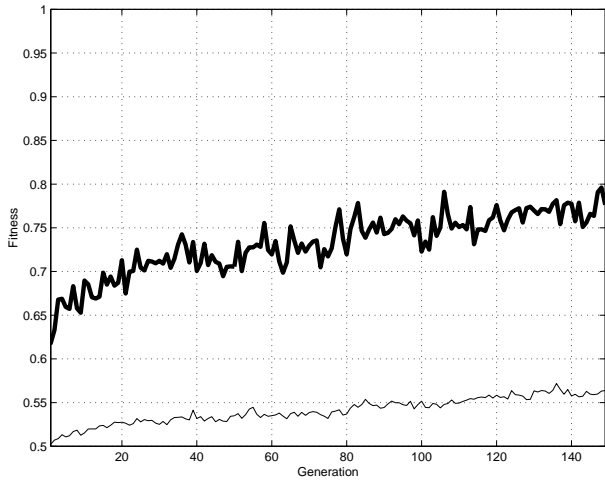


Figure 5: Fitness data for the evolutionary active vision system. Fitness values are percentages of correct response computed along the entire exploration phase. Thick line = best fitness. Thin line = average fitness. Given the binary nature of this discrimination task, 0.5 corresponds to chance level. Each data point is an average of five evolutionary runs. Notice that since the fitness is computed also before that the vision system has located the shape, 100% correct responses are not possible.

$$F = \frac{1}{I * S} \sum_{i=0}^I \sum_{s=0}^S R_i^s \quad (1)$$

where R_s^i is 1 if the vision system gives a correct response at step s for image i and 0 otherwise, I is the total number of images (20 in these experiments), and S is the number of steps per image (50 in these experiments).

Notice that since in these experiments the vision system is asked to give a response at every time step (the most active unit is taken as the response after every activation) and that the probability of encountering a triangle (or a square) is 0.5, it can obtain fitness values of 0.5 by fixing its response to either triangle or square irrespectively of what appears in the image.

3.1 Results

Five evolutionary runs were performed, each starting with a different random initialization. In all runs, both average and best fitness values gradually increased and reached a plateau after about 150 generations (figure 5) at about 80% correct responses for the best individuals. Notice that fitness 1.0 cannot be reached because the vision system is asked to provide a response also before having a chance to explore the shape.

The evolved behavioral strategies vary slightly across the five evolutionary runs, but all share some basic features. The vision system always starts with a fixed response (square or triangle, depending on the evolutionary run) and then moves

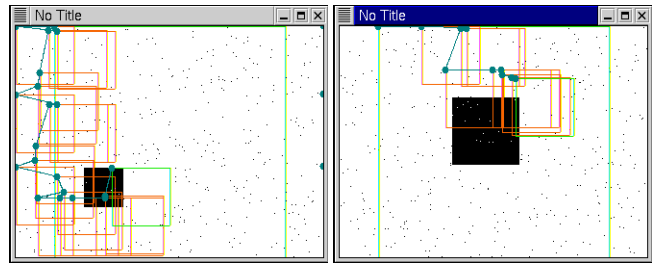


Figure 6: Examples of trajectories of an evolved individual. The retina always zooms and moves with respect to its top leftmost corner, here marked by a dot. The dots drawn after every retina movement are connected by a line. For graphical clarity, the values of the cells are not shown, only the retinal perimeter. *Left*: The retina starts with its initial size at the center of the image signalling “triangle”. It then shrinks to the top left corner and moves down towards the square where it slides along its left edge and starts signalling “square”. Finally, it explores the other three sides of the square maintaining the correct response. *Right*: The same individual begins signalling triangle and then moves towards the square where it visits the top and right edge changing the response into “square”.

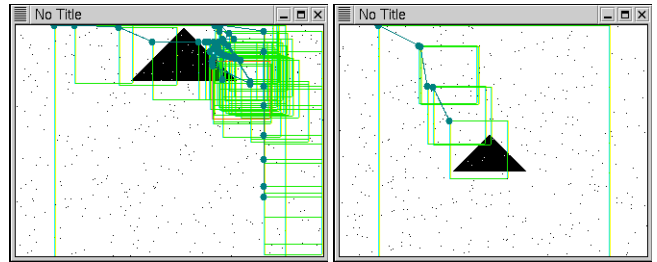


Figure 7: Examples of trajectories of an evolved individual, as in figure 6 above. *Left*: The recognition of a triangle is made by exploring its right corner and then drifting away while maintaining the correct response. *Right*: The recognition is performed by looking at the left edge of the triangle.

towards the shape. Once over the shape, the retina slides back and forth along one of its vertical edges. If the edge is straight, it sets its response to square, otherwise to triangle. Figure 6 shows the trajectories of the retina in the case of two squares and figure 7 shows the trajectories in the case of two triangles. A variation on this basic strategy consists of scanning the corners of the shapes instead of the edges. Once the shape has been recognized, sometimes the vision system moves away from the shape towards a border of the image maintaining the correct response (this behavior is made possible by the recurrent connections). The evolved vision systems are not always capable of moving towards the image area where the shape is located. In those cases, they use stereotypical movements across the image that give a high probability of encountering a shape.

Evolved vision systems use the sampling strategy most of

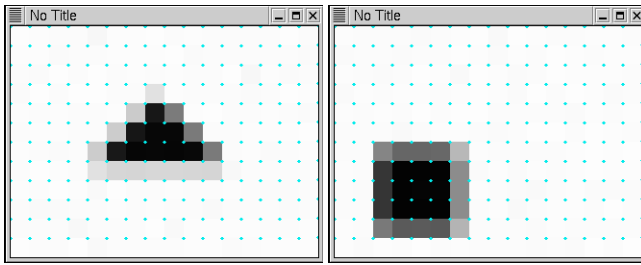


Figure 8: Examples of image preprocessing before presentation to the static neural network shown in figure 9 below. the image is divided in 192 cells of 20 by 20 pixels each, and the average value of the 400 pixels in each cell is taken as input value to the corresponding input neuron.

the times (61% on average), instead of the averaging strategy (see figure 3). Considering that the evolved discrimination strategy is based on the perimeter of shapes (edges and corners), the sampling method provides stronger contrast between the shape and the image background (the activation of input neurons is either zero or one) and consequently a more marked output.

At the same time, they resort almost always to the lowest available resolution, i.e. the largest retinal size. On the one hand, a large retinal size gives the vision system a better chance to locate the shape in the image. On the other hand, the shapes are big enough (minimum height is 20 pixels, maximum is 100, average measured size for the shapes seen by the best tested individuals is 60.58) to be discriminated correctly. In order to check the latter hypothesis, we performed five new evolutionary runs using shapes that can be much smaller (height ranges from 5 pixels to 100 pixels), so that they cannot be resolved at the lowest resolution. In these conditions the best evolved individuals almost 100 times more frequently use a higher retinal resolution (and consequently smaller retinal size), while still retaining the exploration strategies described above (watching edges and corners). Interestingly, these evolutionary runs display the top performance shown in figure 5 much earlier (after about 100 generations rather than 150). This result suggests that the ability to switch resolution more frequently helps also in the case of larger shapes. Indeed, the best evolved individuals change resolution more often also when presented with large shapes.

4 Stationary discrimination

In another set of experiments we attempted to train a “static” neural network to perform the same discrimination task on the complete image using the backpropagation algorithm [22].

The image is divided up into 192 cells, each measuring 20 pixels by 20 pixels (just like the size of a retinal cell of the active vision system at the lowest resolution), as shown in figure 8.

The average value of the 400 pixels in a cell represents the input of a corresponding neuron of a feed-forward neural

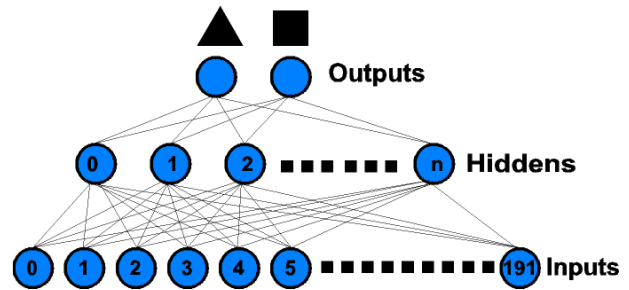


Figure 9: Architecture of the “static” neural network trained with the backpropagation algorithm on entire images. Different architectures have been studied, including one without hidden units.

network with two output units, each standing for one of the two shapes, triangle and square (figure 9).

The network is trained on a balanced set of images (half triangles and half squares) by randomly presenting a shape drawn at a random location with a random size (height range is between 20 and 100 pixels as in the first set of evolutionary experiments described above). The same computer code used for generating the images in the evolutionary experiments is used here too. The connection strengths are initialized to random values in the range $\pm 1/N$, where N is the number of connections in the network (including bias weights). The error between the network response and the correct shape type is computed and accumulated for each presentation of 10 squares and 10 triangles and is used to update the connection strengths.

Each training session consists of 15000 batches of 20 images (always created anew), corresponding to the number of individuals evaluated during an evolutionary run described above (150 generations with a population size of 100 individuals). We have trained networks without hidden units, with 5, 10, and 15 hidden units. Each network architecture has been trained 5 times, each time starting with new random weights. We have also tried several combinations of learning rates and momentum constants.

None of the network architectures has ever been capable of learning to discriminate between squares and triangles, their performances always oscillating around chance level.

5 Discussion

In this paper we have described an evolutionary active vision system that is capable of performing complex discrimination tasks with an extremely simple neural architecture without hidden units. We have also shown that such a discrimination task cannot be easily learned by a neural network trained on the whole image with the backpropagation algorithm.¹ Although somebody may claim that this is not a fair comparison

¹Although none of our networks could manage to learn the discrimination problem, we cannot rule out that with a different initialization, architecture, and connectivity a feed-forward neural network will learn to do it.

because the network does not have recurrent connections as in the case of the evolved network, the temporal dimension does not exist in this context. The presence of recurrent connections in the evolved retina may also induce one to argue that they represent a form of internal representation, contrary to what stated in the introduction section above. Even so, since the recurrent connections exist only at the output level, they do not represent a coding of the sensory information as in the conventional representational systems cited in the introduction.

Evolved vision systems can easily solve the problem because they actively select visual features that make the discrimination task easy. They search for vertical and inclined edges, or for rectangular and acute angles. When the image is large, they slide back and forth along an edge of the shape, probably to avoid making a wrong discrimination due to sampling of image noise. In other cases, they search for a shape angle and move over it for some time before drifting away while maintaining the correct answer.

The edge feature exploited by our evolved retina resembles in some way the strategy used by the evolved robotic vision system described by the Sussex group [12] and mentioned in the introduction section above. In that case, the vision system used only two receptive fields (similar to one of our retinal cells in averaging mode) placed along a diagonal line on the camera surface. The authors reported that the strategy used by the robot to approach the correct figure and maintain a proper trajectory consisted of concentrating on the edges of the shapes and exploiting the differential activation of the two receptive fields to trigger a rotation behavior or a straight motion. This strategy was made possible by the fact that the pixel activations were continuously updated as the camera swept over the image. Our system is different in that the motion is more similar to saccadic movements whereby no information is processed during eye movements.

The angle feature instead has been shown to be a very powerful indicator in computer vision because it is invariant to scaling and rotation. Angle detectors have indeed been built in control algorithms for mobile robots that must recognize objects or locate landmarks during map-based navigation [15]. The most important difference between human-designed and our evolved “angle detector” is that the former consists of modified Difference of Gaussian filters (analogous to so-called “complex” biological cells with off-center and on-surround receptive fields) whereas the latter are much cruder versions that exploit movement over the angle feature to do the discrimination between 90 degree and smaller angles.

Beside active exploration of the shapes, the evolved vision systems make also the best possible use of the filtering strategy and resolution type. As we have pointed out in the result section, although vision systems actively change resolution size while they explore the image, the sampling strategy is preferred because it provides a sharper contrast that does not require a fine tuning of the synaptic weights, as instead would

an averaging method. In other words, since the discrimination task does not require attention to lightness gradients (shapes are black against a white background), the sampling strategy allows a correct discrimination for a wider range of synaptic weights. Similarly, the fact that in the first set of experiments all vision systems almost always use the lowest resolution (i.e., the largest retinal size) is due only to the fact that shapes were sufficiently big to be resolved without zooming in and out. However, as soon as the shapes became smaller, the vision system switched much more frequently between low and high resolution. Low resolution was still maintained as a preferred mode though for two reasons. The first is that it allows a more efficient location of the shape, which can appear anywhere in the image. The second is that, although in our second set of experiments, the shapes could be much smaller, the average shape size (average height = 52 pixels) was still large enough to be resolved at the lowest resolution.

The retinal system used in these experiments is a square matrix. Although in principle we could have used any other morphology, or even let the vision morphology evolve as in the experiments performed by the Sussex group, a regular grid of photoreceptors may allow the vision system to exploit more easily correlations between neighbouring patterns of activations that may correspond to relevant (i.e., ecologically important) features in the environment. This is certainly the case in these experiments and it seems to have been the case in computer experiments performed to study co-evolution of male feather patterns and female visual preferences where symmetric patterns showed to be powerful indicators and detectors in presence of various distortions (perspective and distance) [9]. A simple feed-forward network in principle could not exploit location-specific correlations because the nodes can be permuted in any way without affecting the output [6]. However, this is not the case in our active vision system where active motion on a geometrical plane is determined by sensory input and affects the next sensory input. The presence of recurrent connections may exploit location-specific patterns of activation over time to support efficient exploration and feature detection.

The scanning patterns displayed by the evolved vision system resemble human patterns of eye movements of doctors while they explore images to perform binary decision tasks, such as the presence of breast cancer or of bone fracture (figure 10) [14]. Although our active vision system would most likely not be sufficient to perform these discrimination tasks in its current state, we plan to extend its structure to allow exploration of graded images and real pictures, such as those depicted in figure 10. Although it will undoubtedly be hard to match human performance, there will be several other aspects that can be studied with the evolutionary methodology described here, such as sequences and fixation times, relevant features, and even visual attention.

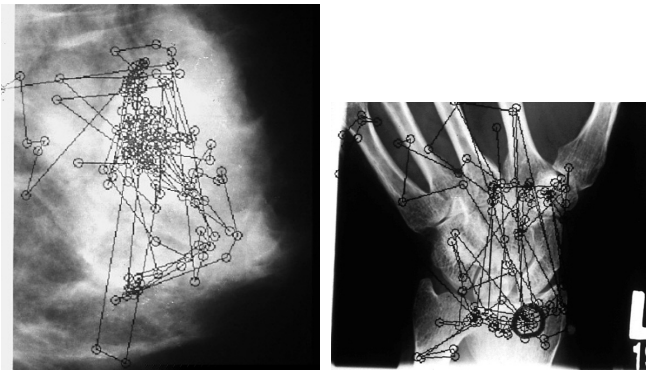


Figure 10: Patterns of eye movements of doctors scanning X-ray images for the presence of breast cancer (left) and of bone fracture (right). Dots represent fixation points [14]. These and other images of human eye scans are available at <http://www.radiology.arizona.edu/eye-mo/mainpage.htm>

6 Conclusion

We have described an evolutionary active vision system that can perform complex image discrimination tasks with very limited resources by actively exploring the image and selecting features that make the task easier to solve. Visual features, exploration strategies, filtering and resolution are autonomously evolved instead of being preprogrammed as in other active vision systems. This gives the active vision system the freedom to develop the strategies that are most suitable for the task and architecture available. It also allows a new series of experiments to study active biological vision within a synthetic approach [4, 19].

Current work aims at extending the features of the vision system while preserving the principle of simplicity in order to present it with more natural images. Another current direction aims at porting this system into a small mobile robot equipped with a mobile vision system and explore evolution of active vision for complex navigation tasks with limited computational resources.

Acknowledgments

T.K. has been supported by an exchange fellowship between EPFL and the University of Tokyo. D.F. acknowledges support from the Swiss National Science Foundation, grant nr. 620-58049. The authors thank the anonymous reviewers and Jesper Blynel for useful comments on this manuscript. T.K. wrote most of the software, carried out all experiments and analysis, and wrote part of this paper. D.F. conceived the evolutionary active vision system and the experiments, wrote part of the software and most of this paper.

Bibliography

- [1] R. Bajcsy. Active Perception. *Proceedings of the IEEE*, 76:996–1005, 1988.
- [2] D. H. Ballard. Animate vision. *Artificial Intelligence*, 48:57–86, 1991.
- [3] R. D. Beer. Toward the evolution of dynamical neural networks for minimally cognitive behavior. In P. Maes, M. Matarić, J-A. Meyer, J. Pollack, H. Roitblat, and S. Wilson, editors, *From Animals to Animats IV: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, pages 421–429. MIT Press-Bradford Books, Cambridge, MA, 1996.
- [4] V. Braitenberg. *Vehicles. Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA, 1984.
- [5] D. T. Cliff and J. Noble. Knowledge-based vision and simple vision machines. *Philosophical Transactions of the Royal Society of London: Series B*, 352:1165–1175, 1997.
- [6] N. D. Cook. Artefact or network evolution? *Nature*, 374:313–314, 1995.
- [7] F. J. Corbacho and M. A. Arbib. Learning to detour. *Adaptive Behavior*, 3(4):419–468, 1995.
- [8] J. L. Elman. Finding Structure in Time. *Cognitive Science*, 14:179–211, 1990.
- [9] M. Enquist and A. Arak. Symmetry, beauty, and evolution. *Nature*, 372:169–172, 1994.
- [10] J. J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, 1979.
- [11] D. E. Goldberg. *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley, Redwood City, CA, 1989.
- [12] I. Harvey, P. Husbands, and D. Cliff. Seeing The Light: Artificial Evolution, Real Vision. In D. Cliff, P. Husbands, J. Meyer, and S. W. Wilson, editors, *From Animals to Animats III: Proceedings of the Third International Conference on Simulation of Adaptive Behavior*. MIT Press-Bradford Books, Cambridge, MA, 1994.
- [13] I. Horswill. A simple, cheap, and robust visual navigation system. In J. Meyer, H. L. Roitblat, and S. W. Wilson, editors, *From Animals to Animats II: Proceedings of the Second International Conference on Simulation of Adaptive Behavior*. MIT Press-Bradford Books, Cambridge, MA, 1993.
- [14] E. A. Krupinski and R. M. Nishikawa. Comparison of eye position versus computer identified microcalcification clusters on mammograms. *Medical Physics*, 24:17–23, 1997.

- [15] S Lepretre, P. Gaussier, and J-P. Cocquerez. From navigation to active object recognition. In J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, and S. Wilson, editors, *From Animals to Animats VI. Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, pages 266–275. MIT Press, Cambridge, MA, 2000.
- [16] R. Linsker. Self-Organization in a Perceptual Network. *Computer*, 3:105–117, 1988.
- [17] H. A. Mallot. *Computational Vision*. MIT Press, Cambridge, MA, 2000.
- [18] D. Marr. *Vision*. Freeman, New York, 1982.
- [19] S. Nolfi and D. Floreano. *Evolutionary Robotics: Biology, Intelligence, and Technology of Self-Organizing Machines*. MIT Press, Cambridge, MA, 2000.
- [20] S. Nolfi and D. Marocco. Evolving visually-guided robots able to discriminate between different landmarks. In J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, and S. Wilson, editors, *From Animals to Animats VI. Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, pages 413–419. MIT Press, Cambridge, MA, 2000.
- [21] F. Rieke, D. Warland, R. van Steveninck, and W. Bialek. *Spikes. Exploring the neural code*. MIT Press, Cambridge, MA, 1997.
- [22] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning Representations by Back-Propagation of Errors. *Nature*, 323:533–536, 1986.
- [23] A. C. Slocum, D. C. Downey, and R. D. Beer. Further experiments in the evolution of minimally cognitive behavior: From perceiving affordances to selective attention. In J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, and S. Wilson, editors, *From Animals to Animats VI. Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, pages 430–439. MIT Press, Cambridge, MA, 2000.
- [24] R. Watt. *Understanding Vision*. Academic Press, London, 1991.