Evolving an Artificial Vision System: Initial Considerations

Mark C. Sinclair
Electronic Systems Engineering,
University of Essex,
Colchester, Essex, CO4 3SQ, UK
mcs@essex.ac.uk

Adrian F. Clark
Electronic Systems Engineering,
University of Essex,
Colchester, Essex, CO4 3SQ, UK
alien@essex.ac.uk

Abstract

Current machine vision systems have many limitations and cannot match the elegant complexity of the human vision system, comprising both the eye and its associated neural processing. However, engineering artificial vision systems by simulating natural ones is impractical, as the latter are not fully understood. Instead, the authors propose using directed evolution of software agents in an artificial environment thereby following a simplified natural evolutionary pathway leading to the creation of an artificial vision system.

1 INTRODUCTION

This paper describes the authors' initial thoughts on evolving an artificial vision system, consisting of both the image sensor and subsequent image processing. The research is motivated by the limitations of current machine vision systems, which require manual tuning and lack generality. In the natural world, the evolution of a well-developed lens eye necessarily progressed alongside that of the associated neural processing. Moreover, such a complex system would only have evolved when the whole ecological environment had progressed to the point where visually-guided behaviour conferred a significant advantage (Nilsson & Pelger, 1994). Nevertheless, it would appear that "serviceable image-forming eyes have evolved between 40 and 60 times, independently from scratch, in various invertebrate groups" (Dawkins, 1994; Salvini-Plawen & Mayr, 1977). Moreover, even a pessimistic estimate[1] gives the effort "to evolve a fish eye from flat skin [as] under 400,000 generations" (Dawkins, 1994; Nilsson & Pelger, 1994). Consequently, the intention is to use directed evolution of software agents in an artificial environment to attempt to follow a simplified natural evolutionary pathway leading to the creation of an artificial vision system. The hope is that such an approach, if feasible, will result in novel and robust architectures that mirror some of the elegant complexity of natural vision systems. In effect, this work will be building on the ancient Greek proverb: "In the land of the blind, the one-eyed man is king".

[1] Although, as these authors admit, their estimate ignores the effort to evolve the associated neural processing.

2 EVOLVING AN ARTIFICIAL VISION SYSTEM

When people started researching computer vision roughly 40 years ago, it was thought that it would not take long to achieve: after all, humans do it with such ease. It is now known, of course, that the problem is vastly more difficult than was initially envisaged; indeed, it is fair to say that the discipline is still in its infancy. Nevertheless, in this time, some impressive vision-based systems have been designed and engineered. However, notwithstanding these successes, it is fair to say that existing machine vision systems are far from matching the performance of the human vision system for almost any task, and certainly unable to achieve the required generality. As the only examples of adequate vision systems currently come from the natural world, obviously there must be lessons to learn from them. However, computer vision researchers cannot progress by producing a detailed simulation of, say, the human vision system, because its operation is only superficially understood. Consequently, it is suggested here that the answer lies not in such a detailed simulation, but rather in the *evolution* of an artificial vision system.

In recent years, researchers in both artificial life and evolutionary robotics have addressed several related areas, including Reynolds (1992), who used genetic

Table 1: Function Descriptions (after Sinclair & Shami, 1997)

FUNCTION	ACTION
`forage`	Move either 1 or 2 cells depending on the first argument, collect (up to) either 1 or 2 food units depending on the second argument, and then rotate either left or right depending on the third argument.
`run`	Move either 2, 4 or 6 cells depending on the first argument.
`pigout`	Collect (up to) either 3 or 5 food units depending on the first argument.
`wander`	Move either 1 or 2 cells depending on the first argument, rotate either left or right depending on the second argument, and then move either 1 or 2 cells depending on the third argument.

programming (GP) (Koza, 1992) to evolve obstacle avoidance behaviour in software agents with limited visual capabilities. These occupy a 2-dimensional (2D) environment, and the evolved programs specify both sensor placement and agent behaviour. He discovered that the resulting sensor placements were closely tied to the evolved agent behaviours, and also that it was difficult to obtain robust, general-purpose solutions that did not have sensitive dependencies on aspects of the environment. Terzopoulos *et al.* (1996) worked with artificial fishes in a physics-based virtual marine environment. They describe the development of an active vision system for pursuing colour-coded prey fish. Their vision system, presented in more detail in (Terzopoulos & Rabie, 1997), has foveated retinal imaging, target-centering eye movements and image stabilisation to compensate for fish swimming motions. It was 'hand-crafted', although simulated annealing was subsequently applied to discover the low-level motor control required for prey pursuit. Hutt & Keating (1998) used a genetic algorithm (GA) (Goldberg, 1989) to evolve visually-guided foraging behaviour of agents in a simple 2D environment. They evolved not only a neural network to control the agent, but also the number, position, size and sensitivity of component cells on the agent's retina. Finally, Mark *et al.* (1998) also used a GA to evolve eye parameters and obstacle avoidance behaviour in Braitenberg vehicle-like agents. The agents evolved in a progressively more difficult 2D environment, and the GA encoded eye number, eye width and a neural network based controller. An important aspect of Mark *et al.*'s work, as for Reynolds and Hutt & Keating, is that their fitness function was task-related rather than depending on some abstract measure of visual acuity.

In evolving an artificial vision system, the authors' intention is to start with experiments to establish simple environments in which vision provides an evolutionary advantage, using pre-constructed vision systems. Then the aim is to evolve simple vision systems, initially evolving only the image-detection structures themselves, with pre-constructed image processing capabilities, but subsequently tackling the far more difficult aspect of evolving a full vision system. An incremental approach will necessarily be used, with more complex capabilities evolved by building on simpler, more basic ones. Throughout, it will be important to encourage generality in the evolved systems through careful selection of both environments and fitness functions, including the introduction of appropriate levels of noise (Jakobi *et al.*, 1995). It is expected that some of the early experiments, such as those discussed in this paper, will require environments that are little more than 2D, but subsequent work will probably necessitate richer environments, approaching full 3D. While it is anticipated that a single workstation platform will be sufficient for the early experimental work, a far more powerful coarse-grained parallel platform will be required as the research progresses.

3 THE LAND OF THE BLIND

For the initial experimental approach, it was decided to develop the simple food-gathering agents and environment first described by Maskell & Wilby (1996) and subsequently re-visited by Sinclair & Shami (1997). The former authors evolved their agents using a GA approach, whereas the latter compared both GA and GP performance.

As described by Sinclair & Shami (1997), the environment consists of a 10×10 grid of cells in which a single food-gathering agent is placed. The environment is toroidal, and filled with an even distribution of food units. The agent is controlled by an evolved program, composed either as a bit-string (GA) or parse tree (GP), and drawn from two sets of (almost identical) functions for agent movement and food manipulation. The agent is blind, and it is given 25 timesteps to gather as much food as possible, with the invocation of each function requiring one timestep. In their initial experimental results, Sinclair & Shami found the GP

Table 2: Argument Interpretation & Return Values (after Sinclair & Shami, 1997)

FUNCTION	ARG 1			ARG 2			ARG 3			RETURN
	0	1	2	0	1	2	0	1	2	
forage	0	1	2	0	1	2	nop	Left	Right	arg2
run	2	4	6							arg1
pigout	0	3	5							arg1
wander	0	1	2	0	1	2	0	1	2	arg2

Table 3: Eye Selection Function Descriptions

FUNCTION	ACTION
progn2	Simply evaluates its two arguments in turn.
wait	Waits (sleeps) for 1, 5 or 10 timesteps depending on the first argument.
patch	A macro that evaluates its first argument during the day and its second during the night.
pinhole	A macro that evaluates its first argument if there is food in the cell immediately ahead of the agent; otherwise its second argument.
eye	A macro that evaluates its first argument if there is food in the cell one step ahead of the agent; its second if there is food two steps ahead; its third if there is food three steps ahead; otherwise its fourth argument.

approach to be superior to GA; however they also discovered that, in both approaches, fit individuals were using only a subset of the available functions. By limiting the function set to only those that had been found useful, the performance of both approaches improved and became almost identical.

As the basis of this work, the GP approach and final function set adopted are the more focused ones used by Sinclair & Shami (Table 1, above). To satisfy the GP closure property (Koza, 1992), only three terminals, {0, 1, 2}, were provided, and the functions all handle integer arguments in the range 0–2; the mapping of these to meaningful values is shown in Table 2.

4 SIMPLE EYE SELECTION

For the first experiments, the grid environment was developed in an attempt to provide an evolutionary advantage to agents with better visual capability.

The environment now has periods of day and night, each lasting 10 timesteps, and the simulation time was extended to 200 timesteps, allowing the agent to experience 10 diurnal cycles. Food is no longer distributed evenly throughout the grid, but 12 units placed in every third cell, offset each row to avoid simply filling every third column! In addition, food only appears at the beginning of each day; any food remaining at the end of the day is completely removed before night; and, to avoid 'lazy' agents, no food appears in the cell in which the agent is located.

Table 4: Eye Selection Argument Interpretation & Return Values

FUNCTION	ARG 1			RETURN
	0	1	2	
progn2				0
wait	1	5	10	arg0
patch				0
pinhole				0
eye				0

Rather than simply accumulating food, each agent starts with 10 food units, and is required to expend food units whenever certain of the functions are invoked. If an agent's accumulated food falls to zero, it dies. Fitness is defined as the number of timesteps survived plus the number of food units remaining at the end; clearly the latter is always zero except for those agents that last the maximum of 200 timesteps. The additional functions introduced are described in Table 3, and their argument mappings in Table 4. Although `progn2`, `patch`, `pinhole` and `eye` take zero timesteps to invoke, and cost zero food units, all the other functions required both time and food. The *active* functions (Table 1) all last one timestep and consume two food units, whereas the single *passive* function, `wait`, required 1, 2 or 3 food units for waits of 1, 5 or 10 timesteps respectively.

Experiments were performed with four function/terminal sets. The *blind* function set comprises

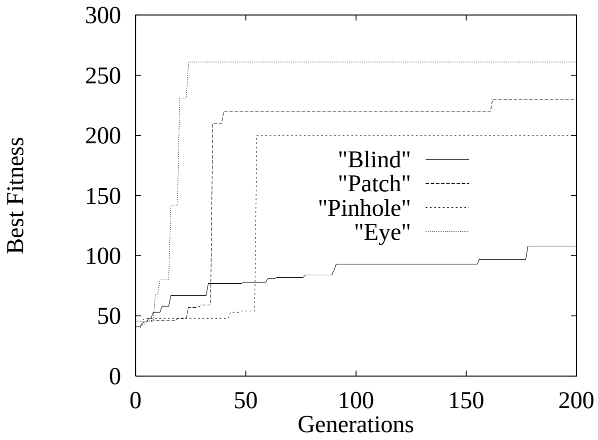

Figure 1: Typical Runs for Eye Selection

{`progn2`, `forage`, `run`, `pigout`, `wander`, `wait`}, plus the three simple terminals, {`0`, `1`, `2`}. The other three sets, *patch*, *pinhole* and *eye*, each simply add the single macros, `patch`, `pinhole` and `eye`, respectively to the blind function set.

The progress of four runs (for the same random seed) are shown in Fig. 1, with populations of 1,000 and parameters after (Koza, 1992). The advantage of the eye function set over pinhole, patch and the basic blind function sets can be clearly seen. With the latter, even after 200 generations, the best agents are still dying out just over half way through their 200-timestep simulation, whereas the eye function set has rapidly produced an agent that not only survives to the end, but accumulates a reserve of over 60 food units. Interestingly, for these particular runs, the patch function set obtained a better final fitness than the pinhole.

To allow statistical comparison, ten runs with different pseudo-random number seeds were made for each function set. The median best fitness values are recorded in Table 5, which shows a clear improvement in fitness from blind, through patch and pinhole, to eye function sets. However, it was not possible to establish the median differences as significant using a non-parametric median test (Sprent, 1992) due to the high variability between individual runs. It would appear that, in some cases, partial knowledge of the near-static food distribution is being inherited by agents, thereby reducing the intended evolutionary advantage of the sighted ones. Additional work will thus be required in refining the environment to give a clearer separation between the four function sets. It is anticipated that this will be achievable through greater variation in the food distribution, either day-to-day within each simulation, or generation-to-generation as evolution progresses. However, it is hoped that a fully pseudo-random food distribution, which would be computationally expensive, will not be required.

Table 5: Median Best Fitness for Different Eye Selection Function Sets (Groups of Ten Runs)

FUN. SET	BLIND	PATCH	PINHOLE	EYE
Med. Fit.	120.5	128.5	141	165.5

As further illustration, additional details of the best agent evolved using the eye function set in the example run of Fig. 1 are given in Fig 2. This shows, against timesteps, the average food in the environment, the agent's accumulated food (scaled down by a factor of 10), and the agent's diurnal activity factor. The passage of day and night can be seen in the average food level, which starts at 4 units each morning, is reduced by agent food gathering during the day, and then falls abruptly to zero at the onset of night. The agent's activity factor is taken to be one if the agent is either active in the day or passive in the night, otherwise it is zero. An agent which 'sleeps' at night (by executing `waits`) and then gathers food in the day is believed to be well adapted to this particular environment. As can be seen, this evolved agent's behaviour follows such a pattern, with periods of inactivity in the nights and bursts of food gathering during the days. The inactive periods can be seen in Fig 2 as unchanging agent food during `waits` of 10 timesteps; the food gathering by rapidly changing agent food levels as food is both accumulated and energy expended. There are only a few instances when the agent 'stays up late', seemingly always followed by 'sleeping in' the following morning! These are particularly visible in Fig 2, as in both cases the agent's diurnal activity factor is shown as zero. Overall, the agent is sufficiently well-adapted to its environment so that it almost exactly follows the expected diurnal pattern, with food conserved during the night (by waiting) and then actually increasing during the day, despite the expenditure required by its food gathering.

5 FUTURE WORK

The next experiments planned will build on the first by using a similar environment and function/terminal set. However, rather than providing ready-made vision macros, agents will have to evolve their own. Thus, instead of the `patch`, `pinhole` and `eye` functions, agents will be required to evolve their own eye structure. This will be represented by a second tree within each individual agent, which describes what, if any, visual capabilities the agent has. In their function set, the

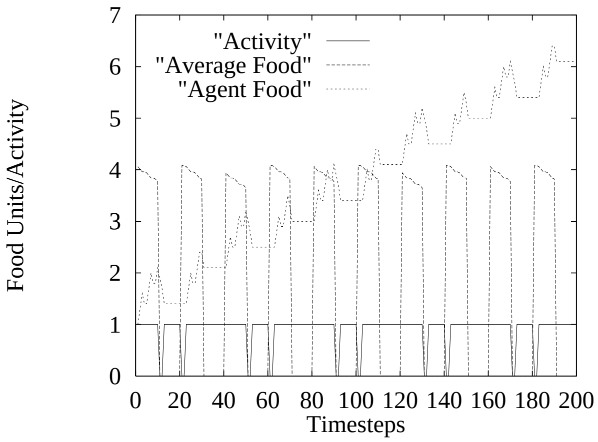

Figure 2: Execution of Best Eye Selection Agent

agents will be provided with a four-argument `struct` function. If an individual agent invokes this from its main parse tree, then its second tree is examined, and vision equivalent to blindness, `patch`, `pinhole` or `eye` macros made available.

It is anticipated that subsequent steps in the research will involve developing a richer visual environment for the agents to explore although, for computational reasons, this will be limited to the simplest necessary to stimulate the desired evolutionary progress. From the current single-agent environment, it is expected that a cooperative multi-agent and even competitive co-evolutionary environments will be needed in providing the necessary evolutionary pressure. The agents themselves will be given access to appropriate collections of both image detection and image processing functions from which they must evolve the increasingly complex vision systems sought. Throughout, the intention is to allow the environment itself to provide the stimulus for artificial vision system evolution, rather than constraining the evolution to the authors' limited preconceptions of suitable architectures and processes.

6 CONCLUSIONS

The authors believe that following a simplified natural evolutionary pathway to the creation of an artificial vision system is a vital step in overcoming some of the limitations of current machine vision systems, particularly when used as part of an autonomous decision-making process. In this paper, the authors have outlined the overall direction of their proposed programme of research, as well as outlining some simple initial experiments. At least at this early stage, such an approach seems promising, although inevitably much yet remains to be accomplished.

References

Dawkins, R. (1994), "The eye in a twinkling", *Nature*, v368, 21 Apr 94, pp. 690–691

Goldberg, D.E. (1989), *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley

Hutt, B. & Keating, D. (1998), "Artificial evolution of visually guided foraging behaviour", Proc. Artificial Life VI, Los Angeles, USA, June 1998, pp. 393–397

Jakobi, N., Husbands, P. & Harvey, I. (1995), "Noise and the reality gap: The use of simulation in evolutionary robotics", Proc. 3rd Eur. Conf. on Artificial Life (ECAL'95), Granada, Spain, June 1995, pp. 704-720

Koza, J.R. (1992), *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press

Mark, A., Polani, D. & Uthmann, T. (1998), "A framework for sensor evolution in a population of Braitenberg vehicle-like agents", Proc. Artificial Life VI, Los Angeles, USA, June 1998, pp. 428–432

Maskell, B. & Wilby, M. (1996), "Evolving software agent behaviours", Proc. IEEE Global Commun. Conf. (GLOBECOM'96), London, UK, Nov. 1996, pp. 90–94

Nilsson, D.E. & Pelger, S. (1994), "A pessimistic estimate of the time required for an eye to evolve", *Proc. of the Royal Society of London*, vB256, pp. 53–58

Reynolds, C.W. (1992), "An evolved, vision-based model of obstacle avoidance behaviour", Proc. Artificial Life III, Santa Fe, New Mexico, USA, June 1992, pp. 327–346

Salvini-Plawen, L.v. & Mayr, E. (1977), "On the evolution of photoreceptors and eyes", *Evolutionary Biology*, v10, pp. 207–263

Sinclair, M.C. & Shami, S.H. (1997), "Evolving simple software agents: Comparing genetic algorithm and genetic programming performance", Proc. IEE/IEEE Intl. Conf. on Genetic Algorithms in Engineering Systems: Innovations and Applications (GALESIA'97), Glasgow, UK, Sept. 1997, pp. 421–426

Sprent, P. (1992), *Applied Nonparametric Statistical Methods, 2nd Ed.*, Chapman & Hall

Terzopoulos, D., Rabie, T. & Grzeszczuk, R. (1996), "Perception and learning in artificial animals", Proc. Artificial Life V, Nara, Japan, May 1996, pp. 313–320

Terzopoulos, D. & Rabie, T.F. (1997), "Animat Vision: Active vision in artificial animals", *Videre: Journal of Computer Vision Research* v1 n1, Sept. 1997, pp. 2–19