A synthetic vision system using directionally selective motion detectors to recognize collision

Shigang YUE[↓] and **F. Claire Rind**
Ridley Building, School of Biology and Psychology
Faculty of Science, Agriculture and Engineering
University of Newcastle, Newcastle upon Tyne, NE1 7RU UK

Abstract: *Reliably recognizing objects approaching on a collision course is extremely important. In this paper, a synthetic vision system is proposed to tackle the problem of collision recognition in dynamic environments. The synthetic vision system combines the outputs of four whole-field motion-detecting neurons each receiving inputs from a network of neurons employing asymmetric lateral inhibition to suppress their responses to one direction of motion. An evolutionary algorithm is then used to adjust the weights between the four motion-detecting neurons to tune the system to detect collisions in two test environments. To do this a population of agents each representing a proposed synthetic visual system were either shown images generated by a mobile Khepera robot navigating in a simplified laboratory environment or were shown images videoed outdoors from a moving vehicle. The agents had to cope with the local environment correctly in order to survive. After 400 generations, the best agent recognised imminent collisions reliably in the familiar environment where it had evolved. However when the environment was swapped only the agent evolved to cope in the robotic environment still signalled collision reliably. This study suggests that whole-field direction-selective neurons, with selectivity based on asymmetric lateral inhibition, can be organised into a synthetic vision system which can then be adapted to play an important role in collision detection in complex dynamic scenes.*

Keywords: *dynamic scene, direction selectivity, asymmetric lateral inhibition, collision, evolution, synthetic vision*

↓ Corresponding author
Dr. Shigang YUE
Brain Mapping Unit, Sir William Hardy Building
c/o Experimental Psychology, Downing Site
Downing Street, University of Cambridge
Cambridge, CB2 3EB United Kingdom
E-Mail: shigang.yue@ieee.org or sy262@cam.ac.uk
Tel: (0044) 1223 764672
Fax: (0044) 1223 764675

1. Introduction

For autonomous mobile robots, the ability to avoid collision is a very important basic skill. Mobile robots have used several kinds of sensors, such as visual, ultrasound, infra-red, laser, and mini-radar, for object detection (for example, [10, 1, 42, 4, 22]). However, it is still very difficult for a robot to run autonomously without collision in complex, outdoor environments without human intervention. Many traditional artificial robotic vision systems have not yet been able to rapidly extract the relevant data from the wealth of information in visual scenes [20, 9].

On the other hand, many animals can successfully avoid collision in the real world relying on their visual systems – the eyes. Biology has provided a rich source of inspiration for artificial visual systems and insects in particular, with their rapid reactions to dynamic scenes using only a small amount of neural hardware, are very attractive as sources of inspiration (for example, [18, 14, 19, 41], reviewed by [11]). In the locust, an identified neuron, the lobula giant movement detector (LGMD), and its postsynaptic partner, the DCMD (descending contralateral movement detector), have been found to respond vigorously to looming stimuli [24, 33, 29, 30, 35, 36]. The input circuitry of the LGMD neuron has been used as the basis for an artificial visual system for collision avoidance in robots and more recently in cars [5, 6, 28, 31, 32, 34, 43, 44]. However, the LGMD based neural network, as adapted for use as a VLSI circuit in a sensor for the car, also responds briefly to nearby fast translating objects and needs other forms of inputs to suppress these unwanted responses [43, 37].

Several feature selective neurons may be combined to provide a robust collision detecting visual system. Direction selective neurons, for example, have been found in animals for decades, for example, in insects such as the locust [26, 27], beetle and fly [15, 7], also in vertebrates such as the rabbit ([2, 3, 38], reviewed by [40]) and the cat (for example, [25, 21]). Such directionally selective neurons could be used to signal looming (for example, [17]) and recently they have been combined with an artificial LGMD neuron to improve reliability of a sensor for signalling collision in traffic scenes [37, 46]. Self-organizing neural networks can develop direction sensitivity (for example, [23, 39]). Recently, asymmetric inhibition has been found to underlie the direction selectivity of ganglion cells in the rabbit's retina [12]. Compared to those neurons with symmetric lateral inhibition, neurons with asymmetric lateral inhibition are able to capture directional motion cues as inhibition suppresses the response to motion in a particular direction while letting other direction pass through the system. The interactions of these direction selective neurons to guide behaviour in animals is still a subject

of speculation. However, modelling studies may provide chances to explore possible mechanisms for information processing between these neurons.

Inspired by this asymmetric lateral inhibition mechanism [12], we propose a synthetic vision system consisting of several asymmetric lateral inhibition based whole-field direction selective neurons to detect colliding objects in dynamic scenes. These direction selective neurons have a similar neural input structure to a LGMD neuron but with asymmetric lateral inhibition. The synthetic vision system is expected to respond to imminent collision in dynamic scenes but not to respond to fast translating, and other objects not on a collision course.

Once the neuronal structure of the vision system has been decided, another vital thing for the system is the tuning or learning process. Traditional searching methods may be difficult to apply because of the highly nonlinear nature of the synthetic vision system. Genetic Algorithms (GAs) [16, 13, 8, 43] are adaptive heuristic search algorithms based on evolutionary ideas of natural selection and genetic mutation. They have shown particular strength in searching for global solutions and are superior to many other traditional methods [13], especially in tuning visual neural networks [e.g., 43]. Therefore, it is ideal to explore the optimum performance of the bio-inspired synthetic vision system via an evolutionary process.

To evaluate the potential capacity of the synthetic vision system, a population of agents with the proposed vision system were put into two different environments: a mini robot in the laboratory, and a car driving on roads, and each underwent evolution in that environment for 400 generations. We then investigated if the evolved best agents in the two environments could cope with the events in their local environments, i.e., can the synthetic vision system recognise imminent collision after evolving? Can this consistency be transferred or generalized to a new environment? The following experiments were designed to answer these questions.

2. Methods

The synthetic vision system is based on four whole-field direction-selective neurons (Figure 1 *L* (left), *R* (right), *U* (up), *D* (down) see also [46] for additional details). Each of these direction selective neurons has one specified inhibitory direction generated by asymmetrically spreading lateral inhibition. Details of the synthetic vision system are given in the following section, 2.1. In each generation there were 40 agents representing the proposed

synthetic vision systems differing in some of the neural connection weights/threshold. The evolution process consisted of 400 generations, (details illustrated in section 2.2) and took place in one of two different environments (details in section 2.3). The whole evolutionary process was repeated three times for each environment and the best agent was chosen for further investigation and challenged with three new sets of test stimuli. The first from the environment they evolved in, the second from the novel environment and the third from a set of simple computer generated stimuli used in previous studies [29].

2.1. The synthetic vision system

The synthetic vision system is based on direction-selective neurons. There are many ways to form a computational model of a direction-selective neuron (for example, [23, 7, 39, 12]). As shown in Figure 1a, the direction selectiveness of the whole-field neuron in this study was achieved by asymmetric lateral inhibition which spread to only one direction. As an example, this *L* neuron responded to moving edges in all directions except the inhibited left direction. The four direction-selective neurons *L*, *R*, *U* and *D* cells were then organised by several neurons (*a b c d* and *e*) and their connections (Figure 1b). The excitation and inhibition described below are indicated by a value in arbitrary units.

2.1.1. The direction selective neuron L

P layer The first layer of the neural network consists of the photoreceptor *P* cells which are arranged in a matrix form; the luminance L_f of each pixel in the input image is captured by its corresponding photoreceptor cell, the change of luminance P_f between two successive frames of the image sequence is calculated and forms the output of this layer. The output of a cell in this layer is defined by equation:

$$P_f(x, y) = \sum_i p_i P_{f-i}(x, y) + abs(L_f(x, y) - L_{f-1}(x, y)) \tag{1}$$

where $P_f(x,y)$ is the change of luminance corresponding to pixel (x, y) at frame *f*, *x* and *y* are pixel coordinates, L_f and L_{f-1} are the luminance, subscript *f* denotes the current frame and *f-1* denotes the previous frame, the persistence coefficients p_i can be varied between 0-1.0. To increase the computing efficiency of the system, p_i was set to zero in the following evolution process. The absolute value operation in the above equation means the differences between successive images are all positive. If there is no difference between successive images, the *P* cells will not be excited.

S layer The output of the *P* cells form the two types of inputs to the cells in the next layer-*S* layer (Figure 1). One type is excitatory and the other one is inhibitory. The excitatory output of a *P* cell passes directly to its retinotopic counterpart in the S layer whereas the inhibitory output from a *P* cell is delayed by one image frame and then fed in one direction to some of its retinotopical counterpart's neighbouring cells. In the case shown in Figure 1, inhibition extends leftward. The inhibition can extend over a maximum of eight cells in any one direction. The total inhibitory input to a cell in this *S* layer is

$$I_f(x,y) = \sum_{i=1}^{N_{lay}} P_{f-1}(x+i,y) w_I(i) \tag{2}$$

where $I_f(x,y)$ is the inhibition to the cell in *S* layer at position (x, y) at frame f, N_{lay} is the number of cells that feed lateral inhibition to this cell and $w_I(i)$ is the local inhibition weight**.** The value of the local inhibition weight was set to be 5.5 based on previous trials to ensure direction selectivity. With this strong inhibition from the right side with one frame delay, the excitation caused by left moving edges can be weakened sharply. Inhibition with more than one frame delay would increase computing cost and its effect on the overall direction selectiveness would be negligible with the current strong inhibitory weights. With the spread of lateral inhibition N_{lay} is set to 8, the system can cope with image motion slower than 8 pixels per frame, which is equivalent to 120 degrees per second angular velocity using a camera working at 25 frames per second and with a 100X80 pixel 60 degrees field of view[↓]. At the edge of a layer, the inhibitory input to a cell may be limited because there may be no neighbouring cell on its outer side. The excitatory flow gathered in the *S* cell will be

$$E_f(x,y) = P_f(x,y) - I_f(x,y) W_I \tag{3}$$

where W_I is the coefficient controlling the global strength of inhibition and this was set to be 1.5 in this paper based on a previous study of lateral inhibition in the locust [28, 6].

L cell The excitations exceed a threshold (set to be 12 based on previous trials) in the *S* cells are summed by the left inhibitory cell *L* and the excitations below the threshold will be ignored in the summing operation. The summed excitation of the left inhibitory cell *L* is,

$$S_f^L = \sum_{x=1}^{n_h} \sum_{y=1}^{n_v} \Theta(E_f(x,y)) \tag{4a}$$

[↓] The angular velocity is calculated in this way: 60(25/(100/8)) =120 degrees per second

Since cells in this layer are also organised in a matrix, n_h is the number of cells counted in a row , n_v is the number of cells in a column and Θ denotes the threshold operation:

$$\Theta = \begin{cases} 1 & if\ E_f(x, y) > 12 \\ 0 & otherwise \end{cases} \tag{4b}$$

The excitation of the *L* cell is then sigmoid as,

$$s_f^L = (1 + e^{-S_f^L n_c^{-1}})^{-1} \tag{5}$$

where n_c is the total number of the cells in *S* or *P* layer. The number of cells in *S* layer is the same as that in *P* layer. Since each cell in P or S layer corresponds to a pixel in an input image, the total number of cells n_c is determined by the size of input images. For this paper the input images are 100 by 80 pixels. Therefore, n_c is 8,000 in *P* or *S* layer. Since S^L_f is greater than or equal to zero according to equation (4), the sigmoid excitation s^L_f varies within (0.5~1). The L cell has one inhibitory direction, left, and is not expected to respond to left translating movements in a certain velocity range. An L cell and its presynaptic networks could be made to respond to only one specified direction with inhibition that spreads to all directions apart from the specified one; however, it would be computationally very expensive. The robustness of these direction-selective cells has been demonstrated previously [46].

2.1.2. Combination of L, R, U and *D outputs*

The four direction selective neurons: *L*, *R*, *U* and *D* are further organised as shown in Figure 1 (b). The cells *a, b* and *c* gather information from the four neurons,

$$\begin{pmatrix} E_f^a \\ E_f^b \\ E_f^c \end{pmatrix} = \begin{pmatrix} w_{L-a} & w_{R-a} & w_{U-a} & w_{D-a} \\ w_{L-b} & w_{R-b} & w_{U-b} & w_{D-b} \\ w_{L-c} & w_{R-c} & w_{U-c} & w_{D-c} \end{pmatrix} (s_f^L\ s_f^R\ s_f^U\ s_f^D)^T \tag{6}$$

where E^a_f is the excitation in the *a* cell at frame f, $w_{L\text{-}a}$ is the weight of the connection between the cell *L* and the cell *a*, and the other symbols are named in a similar way. The weights are allowed to vary within specified domains in this study, as listed below in the matrix,

$$\begin{pmatrix} w_{L-a} & w_{R-a} & w_{U-a} & w_{D-a} \\ w_{L-b} & w_{R-b} & w_{U-b} & w_{D-b} \\ w_{L-c} & w_{R-c} & w_{U-c} & w_{D-c} \end{pmatrix} \sim \begin{pmatrix} (0.5 \sim 1.5) & (-1.5 \sim -0.5) & (-0.5 \sim 0.5) & (-0.5 \sim 0.5) \\ (-0.5 \sim 0.5) & (-0.5 \sim 0.5) & (0.5 \sim 1.5) & (-1.5 \sim -0.5) \\ (0.5 \sim 1.5) & (0.5 \sim 1.5) & (0.5 \sim 1.5) & (0.5 \sim 1.5) \end{pmatrix} \tag{7}$$

With these predefined adaptable weights in equation (7), cell *c* gathers excitation from the direction selective cells, and cell *a* and *b* compares the weighted excitations they get from the direction selective cells.

The cell *d* compares information from the cells *a* and *b*,

$$E_f^d = abs(w_{a-d} abs(E_f^a) + w_{b-d} abs(E_f^b)) \tag{8}$$

where the weights are limited within: $w_{a\text{-}d} \in [0.5 \sim 1.5]$, $w_{b\text{-}d} \in [-1.5 \sim -0.5]$. Finally, the "spiking cell" *e* compares information from the cells *c* and *d*,

$$E_f^e = abs(w_{d-e} E_f^d + w_{c-e} E_f^c) \tag{9}$$

The weights are bounded in the similar way: $w_{d\text{-}e} \in [0.5 \sim 1.5]$ and $w_{c\text{-}e} \in [-1.5 \sim -0.5]$. Once the excitation in cell *e* exceeds a given threshold, a spike is produced. If spikes are produced in all of five successive frames then this is defined as detection of an imminent collision. The optimal spike number for collision decision making has been investigated previously [43]; using multiple consecutive spikes instead of one can avoid false alarms triggered by visual perturbations such as camera shake. The threshold for cell *e* is allowed to vary within the range from 0.0 to 4.0.

The connectivity and the values of the weights of the input of the direction selective neurons (*L*, *R*, *U* and *D*) are fixed (see Figure 1). In contrast the weights of their output connections (see Figure 1, network on right hand side) can be altered during the evolutionary process. With the current structure, the different visual cues extracted by opposite direction selective neurons are compared by cell *a* and *b*, and further compared by cell *d*; the overall excitation of the four direction selective neurons are summed by cell *c*; the outputs from cell *c* and cell *d* are then compared by cell *e* to form the final outputs of the synthetic vision system.

In the whole synthetic vision system, there are 8000 *P* cells shared by the 4 direction selective neural networks, 32000 *S* cells in total (because each of the direction selective neural network has 8000 S cells), 4 whole-field direction-selective neurons and 5 further organised cells (namely, *a b c d* and *e*).

2.2. Algorithm for evolution

Genetic Algorithms [16, 13, 8, 43] are adaptive heuristic search algorithms based on the evolutionary ideas of natural selection and genetic mutation and have been proved to be efficient in tuning visual neural networks [e.g., 43]. In this paper, we also explore the overall performance of the bio-inspired synthetic vision system via an evolutionary process.

The population of 40 agents in each generation representing the synthetic vision system are processed via a genetic algorithm [8]. The first generation has one recommended agent with predefined connection weights/threshold (Table 2, Start) and 39 agents with randomly produced connection weights/threshold equally distributed in the corresponding bounding intervals. The connection weights/threshold of each agent are then coded in binary numbers as its chromosome. The precision of the binary representation of each weights/threshold is 10 bits. To form a new generation, the worst performing agents (20% of the whole population in a generation) will be replaced, i.e., the generation refreshing rate is 0.2. New agents (20% of a whole population) are produced by the best performing parents in the last generation through crossover. The stochastic universal sampling method is used in selecting parents for breeding [8]. Mutation is conducted to the binary chromosomes of these new produced agents with a mutation rate 0.2. This means 20% of these binary bits in the chromosomes will be randomly chosen to flip, either from 1 to 0 or from 0 to 1, during a mutation process. The generation refresh rate and mutation rate are set based on our previous experiments and studies [43].

The Agents' fitness value, namely weighted success rate, is evaluated based on their behaviour. In each generation, an agent that responds to all visual events correctly, i.e. recognises imminent collisions and makes no mistakes on translating scenes or other challenges, scores a fitness value (success rate) of 100%; an agent that fails in all events scores a fitness value 0%; an agent that fails in a non-colliding challenge scores a lowered fitness value (reduced success rate); an agent that fails in a colliding event gets a sharp reduction in fitness value since a collision event is much more important in scoring than a non-collision one. In the robotic laboratory, one undetected collision event receives a penalty equivalent to four false alarmed non-collision events; in the car driving environment, one undetected collision receives a penalty equivalent to eight false alarms in non-collision events. The fitness of an agent may be formulated as the following,

$$F_k = (1 - \frac{\sum_{i=1}^{N_v} f_{event}^{i}}{M_{nb}}) \times 100\% \tag{10}$$

where F_k is the fitness value of the kth agent in the population, f^i_{event} is the score for the ith event in the total N_v events, M_{nb} is the summation of all the possible high scores, and f^i_{event} is different for a collision and a non-collision event, and failure and success,

$$f^i_{event} = \begin{cases} K_{col} & if\ \ collision\ event, \quad failure \\ 0 & if\ \ collision\ event, \quad success \\ K_{non} & if\ non-collision\ event, \quad failure \\ 0 & if\ non-collision\ event, \quad success \end{cases} \tag{11}$$

Where K_{col} is the score for failure in a collision event, K_{non} is the score for failure in a non-collision event. For a collision event, failure means no collision signal is sent out by the agent 30~3 frames before collision. K_{col} is several times bigger than K_{non} , so that an agent that fails in all collision events and an agent that fails in all non-collision events will have the same fitness value: 50%. In the robotic laboratory, N_v = 20 (including 4 collision events), K_{col} is 4 and K_{non} is 1; in the driving environment, N_v = 18 (including 2 collision events), K_{col} is 8 and K_{non} is 1. Therefore, M_{nb} is 32 in both of the environments.

2.3. Environmental setting

Evolution environments As illustrated in Figure 2 and 3, the two different environments are represented by two groups of video sequences. Each sequence represents one type of scene that can cause strong excitation in the photoreceptor layer. One of the environments is the output of a video camera mounted on the turret of a Khepera robot (http://www.k-team.com) interacting with a black ball in a robotic laboratory. The ball is either moving/bouncing at a velocity of 0.3~1.25m/s across the robot's field of view in a translating way or approaching it on direct collision course at different speeds and from different sides. In the fast translating events, the ball is moving across the field of view within 10~14 frames, corresponding to 0.40~0.56 seconds. The robotic scenes also include turning as this often happened as one of the robot's motor activities. A total of 20 of these robotic scenes were used to construct a virtual environment for the agents to evolve in, (sample images are shown in Figure 2). The robot moves only in the turning scenes. The second environment consists of car driving scenes which are more complex, and include normal driving on a highway, collision with a car, turning, pedestrians, road symbols and high speed translating cars/vans, (sample images are shown in Figure 3). The car used for collision is actually an inflatable car. A total of 18 car driving video sequences were used in this second evolution environment.

It is important to note that the robot is usually moving in an arena with objects 0~50 cm away from it; for example the ball can almost touch its camera in a collision course. For a car, the camera is set behind the windscreen; a collision usually happens at least a car bonnet away from the camera.

Test sequences Three groups of sequences for testing the evolved agents are shown in Figure 4 (a~c). These test sequences include some challenging scenes. For the agent in the robotic laboratory, the test includes: ball looming, translating nearby, translating at high speed, bouncing from left to right, near miss (ball approaches the robot but misses it), and robot turning. For the agent in the driving environment, the test includes: collision with a car, high speed translating car at different distance, turning, translating human figures and driving into and out of a tunnel. We also use computer generated visual stimuli, which are often used by neurobiologist to challenge the animal's visual system, to test the evolved best agents. The stimuli include: expanding square and translating bar with different expanding/translating speed. The best evolved agents are not only tested by using sequences from the environment they evolved in but also tested by using sequences from new environments.

Special test sequences To test the capability of the directional selective neurons and the best agents, two special test sequences from the robotic laboratory were also recorded as shown in Figure 5.1. One of the sequences is a rapidly translating black ball, which crosses the field of view within six frames; its speed is equivalent to an angular velocity of 250˚/s, as shown in Figure 5.1 (a). This high angular velocity visual scene is quite rare in driving scenarios. However, in the robotic laboratory, it can be achieved by deliberately throwing the ball near the robot's camera. Another sequence is a white slowly translating ball against white and/or grey background. The ball is not resolved when moving on a white background until frame 36 as shown in Figure 5.1 (b).

Diverse test sequences To test the capability of the best agents in dealing with different collision scenes, twelve test sequences from the robotic laboratory were recorded and samples given in Figure 5.2. Each sequence contained a particular alteration or combination of differences from those encountered in the environment the system was evolved in. In addition to these test sequences, responses to systematic shifts in object approach trajectory away from a direct collision were recorded. As shown in Figure 5.3, these objects are: a small black ball 40mm in diameter; a large black ball 110 mm in diameter; a white textured ball 95 mm in diameter; a blue cylinder 130mm in length and 95 mm in diameter; an irregular shaped object consisting of black/white blocks assembled to a maximum 90 mm in height, 100 mm in width and 70 mm in depth; two sheets of 110 mm long 210 mm high paper with periodically

distributed stripes, (stripe intervals are 12.5 mm and 25 mm respectively); a small coloured plastic block 95 mm in height 31 mm in width; a sheet of cardboard 600 mm in width 460mm in height with irregular spots; a bundle of RS323 colour wires consist of eight thin wires each about 1 mm in diameter. These collisions were caused either by the robot approaching objects or by objects rolled towards the robot. The speeds of the robot for each situation are given in Figure 5.2. Additional video clips of single approaching objects but with different approaching speeds were also recorded and will be used in the test.

All the input images to the synthetic vision system are resized to 100 pixels horizontally by 80 pixels vertically; images are grey scale ranging in intensity from 0 to 255. The video images were all taken at 25 frames per second (fps).

3. Results and discussions

During evolutions agents in the driving environment were shown images videoed outdoors from a moving vehicle; agents in the robotic environment were shown images generated by a mobile Khepera robot navigating in a simplified laboratory environment. Three evolutions have been done for each evolving environment respectively. The success rate versus generation was plotted in Figure 6.1.

In the driving environment, the best performing agent in the 1st round beat the best agents from the other two rounds of evolution. We will therefore use the 1st round results in the later illustrations and employ the best agent in the 400th generation from the 1st round to represent the best driving agent in our comparisons. For the three different rounds of evolution in the robotic environment, the three best agents achieved the same highest score by the 400th generation. We randomly picked the 1st round of evolution for later comparisons. We also noticed that the weights/threshold of the three best robotic agents at the 400th generation were quite different though they achieved a 100% success rate. Therefore, if not stated otherwise 'the best agents' in the next parts of the paper means the best agents from the 1st round of evolution in the driving or the robotic environment .

The performances of the best agents in their original environment, together with the performance of the start guess agent, are listed in Table 1. The start guess agent with a balanced weight value scored very low success rates- 41% and 59% (Table 1). However, the best agents' performance improved in subsequent generations during the evolution processes.

By the 400th generation, the best agents in each environment had reached a 100% success rate (Table 1).

The changes in the weights of the connections as the best agents evolved are schematically illustrated in Figure 6.2 and detailed in Table 2. Many of the connections between the cells have been altered, and some have been strongly modified. For example, the connection *R-a* in the agent evolved in the robotic environment has been strengthened from -1.0 to -1.4892. The connection *R-a* in the agent evolved in the driving environment has not been changed to any significant degree. This reflects the different importance of information gathered in these cells in different evolution environments. The connections *U-c* and *D-c* have been weakened for both of the best agents to accommodate the asymmetrical movement in both of the environments (most of the activity happened on the ground for both the robot and the car).

The best agent from each evolution environment was then challenged with the three groups of test sequences and results are shown in the Table 3. In the test sequences, as shown in the Table 3, the synthetic vision system achieved a 100% success rate in its familiar environment. The best agent (from three evolutionary runs) evolved in the robotic laboratory environment could cope with all the three groups of test sequences; however, the best agent (from three evolutionary runs) evolved in the driving environment failed in coping with the collision event in robotic laboratory and with the computer generated expanding squares. The fitness value of the best agents in different generations tested in their swapped (i.e., non-native) environment is given in Table 4.

Details of the excitation levels and spikes produced by the two best agents evolved in the two environments when challenged with some of the test sequences are compared in Figures 7~11. The evolutionary changes have made the two best agents behave differently from one another when challenged with the test stimuli. For example, when challenged by a fast translating ball (Figure 4 (a), video sequence 2) or a rightward moving computer-generated black bar (Figure 4 (c), sequence 6), one responds with peaks in excitation and sparse spikes, whereas the other responds with continued high levels of excitation, although excitation remained below the threshold for spike production (horizontal dotted line), as shown in Figures 8 and 10.

The patterns of output excitation of the two best agents are of similar form when challenged with collision events; however, the output behaviour of the synthetic vision systems or agents is often quite different as shown in Figure 7 and Figure 10. One reason for this lies in the differences in their thresholds for spike production (as shown in Table 2, one is

2.3268 and the other is 3.0873). It is also found that the best agent evolved using images from the robotic laboratory produced trains of spikes in response to an accelerated translating bar (Figure 10(b)). This response is similar to that shown by the locust LGMD neuron to a moving bar that accelerates as it moves. The LGMD also responds strongly and persistently to an accelerating edge but only responds briefly to one that moves with a constant speed [29].

The best agents are also capable of distinguishing real collisions from those of 'collision-like' (Figure 4 (b), sequence no.6) or near missing (Figure 4 (a), sequence no.5) visual scenes as detailed in Figure 9. Driving into a tunnel is a typical 'collision-like' scenario (Figure 4 (b), sequence no.5) that may confuse the agents. When driving into a tunnel, however, the best agent evolved in the robotic laboratory responded with sparse spikes and the best agent evolved in driving responded with no spikes at all. Since the physical size of the tunnel is big, its expanding pattern in visual field is quite different to that of a colliding car, i.e., the expanding rate of the tunnel is slow and excitation passing through inhibition layer is not enough to exceed threshold. The evolved synthetic vision system has made use of these differences. When challenged with the near missing ball (Figure 4 (a), sequence no.5), both of the evolved best agents dealt with it correctly though their excitation patterns are slightly different as shown in Figure 9.

The success of the best agents relies on the responses of direction selective neurons. When visual movement is extremely fast and beyond the direction selective neurons ability to respond selectively, the best agents will not be capable of making a good decision. For example, the four direction selective neurons respond to the rightwards rapidly translating ball (Figure 5.1, (a)) similarly when the angular speed of the ball is extremely high (250˚/s in this case); the best agents evolved in the robotic laboratory made an incorrect decision (Figure 11, right). The current asymmetrical lateral inhibition only spreads 8 cells away, therefore, the system can only cope with visual motion slower than 120˚/s (with current input images and camera setting: 60˚ field of view, 100 pixels in horizontal and 25 frames per second). The ability of these direction selective neurons also relies on contrast in the input images. For the white ball on a white background (Figure 5.1, (b)), the four direction selective neurons did not respond to the moving white ball very much until after frame no.36 when contrast became available (Figure 11, left).

In swapped environments, the best agents evolved in the robotic laboratory failed to respond correctly to the fast translating car and the zebra line/arrows (Figure 12, (a)); the best agents evolved using driving scenarios failed in all the collision scenes in the robotic laboratory (Table 4, an example is shown in Figure 12, (b), left), partly due to the high

spiking threshold. When the threshold was lowered deliberately (to investigate its effect) from 3.0873 to 2.3268 (the same as the best agent evolved in robotic laboratory), its performance was improved from only 50% to 72%, so that it failed in one collision scene and five non-collision scenes. For example, it may fail in the turning scenario (as shown in Figure 12, (b), right) if the spiking threshold was lowered.

The two evolved best agents have also been challenged with collision scenes with a wide range of size, shape, texture and speed of colliding objects (Figure 5.2). The responses of the two evolved agents are shown in Figure 13.1, together with the excitation plot from the four direction selective neurons. Although the size, shape texture and approach speed of the colliding objects differ, the robotic agent detected these imminent collisions several frames in advance. These tests taken together, have demonstrated the reliability of the evolved best agents in detecting different imminent collisions in a familiar environment.

The best agents were tested with different strips which represent different spatial frequencies (Figure 5.2, (h), (i), (k) and (l)). When challenged with objects with stripes (Figure 5.2, (h) and (i)), the robotic agent detected the two striped paper walls with distinct successive spikes, despite of acute perturbations in excitation curves at the early stage of the approaching where the robot was further from the striped paper walls (Figure 13.1, middle column). For the non-collision turning scenes, i.e., turning within striped cylinders at about 63 degrees per second (Figure 5.2, (k) and (l)), the robotic agent responded with sparse spikes to the thin stripes (Figure 13.2, top of the middle column); but it sent out a false alarm to the thick stripes (Figure 13.2, bottom of the middle column). Our experiments also showed that when the turning speed was doubled (at 126 degrees per second), the agent did not respond to these stripes with high excitation as the direction selective neurons are not able to work at that high speed of visual motion. The above experiments showed the ability of the best evolved robotic agent in detecting colliding objects with different spatial frequencies.

To understand the effect of different approaching speeds on the robotic agent's ability in detecting collision, we tested the agent with video clips in which the robot approaches the textured white ball at different speeds. Results were shown in Figure 14. The robotic agent detected these collisions 1~7 frame before they occurred; no obvious correlation was observed between the approach speed and the number of frames before collision when collision was first detected. This means when a collision is detected, the distance from the camera to the approaching object is further if the approach speed is greater. Similar responses have also been observed in an LGMD model [44, 45].

As shown in Figure 13.1 (right column), the best driving agent was not able to detect these collisions. One important reason is that this is a new environment for the driving agent. However, it did respond to these imminent collisions with higher excitation level. Its performance would have been better if its threshold had been lower in these cases.

A further test was carried out to see if its response was selective for approaching objects that were on a collision course. We challenged the robotic agent with a ball rolling towards it at 370 mm/s along an approximately 1,000 mm long track. The approach angle of the ball was systematically moved from a direct collision course to *5.7* degrees off a collision course, as shown in Figure 15 (a). For each approach angle the excitation levels of the final spiking cells in the best robotic agent were plotted as before (Figure 15 (b)). A line shows the time at which the image of the approaching ball began to move out of the robot's field of view. In the frame prior to this time excitation levels were lower as the deflection angles became larger although more trials would be required to quantify this effect. The robotic agent was able to correctly predict the collisions in the first two video clips. This experiment demonstrated the ability of the robotic agent to distinguish real collisions from near-misses.

From these results, we find that the synthetic vision system based on whole-field direction-selective neural networks can be evolved for recognising imminent collision in simple or very complex environments. This study suggests that asymmetric lateral inhibition based whole field direction-selective neurons can be organised and adapted to play an important role in tasks such as collision detection particularly in complex dynamic scenes.

We also found that, in a novel environment, the best agent evolved in the relatively simple robotic environment can perform quite well in the new driving environment; however, the best agent evolved in the complex driving environment, cannot perform as well as it does in its familiar environment (Table 3 and Table 4). Generation by generation, the performance of the best agents evolved in the complex driving environment did not get better if tested in a swapped environment (Table 4) though their performance improved rapidly in their evolving environment (Table 3).

This is probably because the visual motion pattern in some of the non-collision sequences in the driving environment is very complex and the best agent has to develop specialized connections/threshold to deal with it. It is also the case that visual motion is biased in the driving scenes because of the Highway Code. In this case, the synthetic vision system evolved in the driving environment needs to adapt to its new environment again in order to cope with it correctly. The above results may also imply that, for animals (or artificial agents) needing to

live in different environments, redundant visual systems which are ready with only slight adaptation for new environments may be needed. In this way, the capability evolved in a complex environment may not been diminished when the agents enter a simple environment.

The overall organisation of these direction selective neurons in animals have not been fully understood although these neurons have been found in many different animal species [2, 3, 7, 12, 15, 21, 25~27, 38, 40]. Our research provides an alternative way to investigate the roles of these neurons in animals. As shown and discussed above, these direction selective neurons can be easily organised to cope with very complex visual tasks. Our modelling results may suggest that interactions between these neurons could also occur in animals driven by complex visual tasks. On the other hand, the evolved best synthetic vision system can be a functional sensor for artificial intelligent robots/agents.

In the synthetic vision system, the visual motion patterns are extracted simultaneously from complex backgrounds by the early image processing neurons, the direction-selective neurons. Since these neurons are robust in extracting visual cues [46], the whole vision system is reliable. The synthetic vision system can process video sequences at 30 frames per second in a 2.80 GH Dell INSPIRON 5150 computer in the experiments. This system can therefore be used in robots for real time autonomous navigation in complex environments, where the conventional robotic vision techniques are difficult to apply [9].

In the above experiments, the agents were challenged with recorded video images in which visual events are the same for every agent. It will be interesting to see how the synthetic vision system evolves if the agents are put into a real/virtual 3D environment and allowed to move around. Additionally, to cope with visual events in different environments, more visual motion cues such as diagonal motion may need to be extracted. In the future, other specialized structures may be integrated into the synthetic vision system so that the agents can interact with different environments efficiently.

4. Conclusions

In the above sections, the asymmetric inhibition based direction selective neurons have been organized by artificial evolution into a synthetic vision system for collision detection in simple or complex environments. The synthetic vision system has significant potential for adapting to different environments and recognizes collision reliably. We showed that the evolved agents with the proposed vision system can recognize imminent collision reliably in

their particular environments. However in a swapped environment, only the agent evolved in the simple environment performs well in the new complex environment; the agent evolved in the complex environment does not perform as well when swapped to a simple one. Further diverse and systematic tests demonstrated the reliability of the best agent evolved in a simple environment in detecting different colliding objects. This study suggests that whole-field direction-selective neurons, with selectivity based on asymmetric lateral inhibition, can be organised into a synthetic vision system which can then be adapted to play an important role in collision detection in complex dynamic scenes.

Acknowledgement

This work is supported by EU IST-2001-38097. We thank M. Soininen of Volvo Car Corporation for providing the driving video footage used in this paper. We thank the anonymous reviewers for their invaluable comments and suggestions in revising the manuscript. We also thank the comments raised in the internal review and Mr. M. Bendall for proof reading the manuscript.

References

[1]. Adams, M. D. (1988). Sensor modelling, design and data processing for autonomous navigation. River Edge, NJ, World Scientific.

[2]. Barlow, H. B., & Hill, R. M. (1963). Selective sensitivity to direction of movement in ganglion cells of rabbit retina. *Science*, 139, 412–414.

[3]. Barlow, H. B., & Levick, W. R. (1965). Mechanism of directionally selective units in rabbits retina. *Journal of Physiology (London)*, 178, 477-504.

[4]. Benet, G., Blanes, F., Simo, J. E., & Perez, P. (2002). Using infrared sensors for distance measurement in mobile robots. *Robotics and Autonomous Systems*, 40, 255-266.

[5]. Blanchard, M., Verschure, P. F. M. J., & Rind, F. C. (1999). Using a mobile robot to study locust collision avoidance responses. *International Journal of Neural Systems*, 9, 405-410.

[6]. Blanchard, M., Rind, F. C., & Verschure, P. F. M. J. (2000). Collision avoidance using a model of the locust LGMD neuron. *Robotics and Autonomous Systems*, 30, 17-38.

[7]. Borst, A., & Haag, J. (2002). Neural networks in the cockpit of the fly. *Journal of Comparative Physiology*, 188, 419-437.

[8]. Chipperfield, A. J., & Fleming, P. J. (1995). The Matlab genetic algorithm toolbox. *IEE Colloqium on Applied Control Techniques Using MATLAB*, Digest No.1995/014.

[9]. DeSouza, G. N., & Kak, A. C. (2002). Vision for mobile robot navigation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2), 237-67.

[10]. Everett, H. R. (1995). Sensors for mobile robots: Theory and application. AK Peters, Wellesley, MA.

[11]. Franceschini, N. (2004). Visual guidance based on optic flow: A biorobotic approach. *Journal of Physiology Paris,* 98, 281-292.

[12]. Fried, S. I., Muench, T. A., & Werblin, F. S. (2002). Mechanisms and circuitry underlying direction selectivity in the retina. *Nature*, 420, 411-414.

[13]. Goldenberg, D. E. (1989). Genetic algorithms in search, optimization and machine learning. Addison- Wesley, Reading, Mass.

[14]. Harrison, R. R., & Koch, C. (2000). A silicon implementation of the fly's optomotor control system. *Neural Computation,* 12, 2291-2304.

[15]. Hassenstein, B., & Reichardt, W., (1956). Systemtheorische analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. *Zeitschrift für Naturforschung*, 11b, 513–524.

[16]. Holland, J. H. (1975). Adaptation in natural and artificial systems. Ann Arbor: The University of Michigan Press.

[17]. Horridge, G. A. (1992). What can engineers learn from insect vision? *Philosophy Transactions of Royal Society London*, B, 337, 271-282.

[18]. Huber, S. A., Franz, M. O., & Buelthoff, H. H. (1999). On robots and flies: modelling the visual orientating behaviour of flies. *Robotics and Autonomous Systems*, 29, 227-242.

[19]. Iida, F. (2003). Biologically inspired visual odometer for navigation of a flying robot. *Robotics and Autonomous Systems,* 44/3-4, 201-208.

[20]. Indiveri, G. & Douglas, R. (2000). Neuromorphic vision sensors. *Science*, 288, 1189-1190.

[21]. Livingstone, M.S. (2005). Direction inhibition: A new slant on an old question. *Neuron*, 45, 5-7.

[22]. Manduchi, R., Castano, A., Talukder, A., & Matthies, L. (2005). Obstacle detection and terrain classification for autonomous off-road navigation. *Autonomous Robots*, 18, 81-102.

[23]. Marshall, J. A. (1990). Self-organizing neural networks for perception of visual motion. *Neural Networks*, 3, 45-74.

[24]. O'Shea, M., Rowell, C. H. F., Williams, J. L. D. (1974). The anatomy of a locust visual interneurone: The descending contralateral movement detector. *Journal of Experimental. Biology*, 60, 1–12.

[25]. Priebe, N. J., & Ferster, D. (2005). Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex. *Neuron,* 45, 133-145.

[26]. Rind, F. C. (1990a). A directionally selective motion-detecting neurone in the brain of the locust: physiological and morphological characterization. *Journal of Experimental Biology*, 149, 1-19.

[27]. Rind, F. C. (1990b). Identification of directionally selective motion-detecting neurones in the locust lobula and their synaptic connections with an identified descending neurone. *Journal of Experimental Biology,* 149, 21-43.

[28]. Rind, F. C., & Bramwell, D. I. (1996). Neural network based on the input organization of an identified neuron signalling impending collision. *Journal of Neurophysiology*, 75, 967– 985.

[29]. Rind, F. C., & Simmons, P. J. (1992). Orthopteran DCMD neuron: A reevaluation of responses to moving objects. I. Selective responses to approaching objects. *Journal of Neurophysiology*, 68, 1654–1666.

[30]. Rind, F. C., & Simmons, P. J. (1999). Seeing what is coming: Building collision sensitive neurons. *Trends in Neurosciences*, 22, 215-220.

[31]. Rind, F. C. (2002). Motion detectors in the locust visual system: from biology to robot sensors. *Microscopy Research and Technique*, 56, 256-269.

[32]. Rind, F. C. Santer, R. D., Blanchard J. M., & Verschure, P. F. M. J. (2003). Locust's looming detectors for robot sensors. In Barth, F.G., Humphrey, J.A.C., & Secomb, T.W. (Eds.), *Sensors and Sensing in Biology and Engineering*. New York, Spinger-Verlag, Wien.

[33]. Rowell, C. H. F., O'Shea, M., Williams, J. L. (1977). The neuronal basis of a sensory analyser, the acridid movement detector system .IV. The preference for small field stimuli. *Journal of Experimental Biology,* 68, 157-185.

[34]. Santer, R. D., Stafford R., & Rind, F.C. (2004). Retinally-generated saccadic suppression of a locust looming detector neuron: Investigations using a robot locust, *Journal of Royal Society London Interface*, 1, 61-77.

[35]. Schlotterer, G. R. (1977). Response of the locust descending contralateral movement detector neuron to rapidly approaching and withdrawing visual stimuli. *Canadian Journal of Zoology*, 55, 1372–1376.

[36]. Simmons, P. J., Rind, F. C. (1992). Orthopteran DCMD neuron: A reevaluation of responses to moving objects. II. Critical cues for detecting approaching objects. *Journal of Neurophysiology*, 68, 1667–1682.

[37]. Stafford, R., Santer R. D., & Rind, F. C. (2005). A bio-inspired visual collision detection mechanism for cars: combining insect inspired neurons to create a robust system. *Biosystems: Sixth International Workshop on Information Processing in Cells and Tissues (IPCT2005)* (in press)

[38]. Stasheff, S. F., & Masland, R. H. (2002). Functional inhibition in direction-selective retinal ganglion cells: spatiotemporal extent and intralaminar interactions. *Journal Neurophysiology*, 88, 1026-1039.

[39]. Tversky, T., & Miikkulainen, R. (2002). Modelling direction selectivity using self-organizing delay-adaptation maps. *Neurocomputing*, 44-46, 679-684.

[40]. Vaney, D. I., & Taylor, W. R. (2002). Direction selectivity in the retina. *Current Opinion in Neurobiology*, 12, 405-410.

[41]. Webb, B. & Reeve, R. (2003). Reafferent or redundant: integration of phonotaxis and optomotor behaviour in crickets and robots. *Adaptive behaviour*, 11 (3), 137-158.

[42]. Wichert, G. (1999). Can robots learn to see. *Control Engineering Practice*, 7, 783-795.

[43]. Yue, S., Rind, F.C., Keil, M.S., Cuadri, J., & Stafford, R. (2006). A bio-inspired visual collision detection mechanism for cars: Optimisation of a model of a locust neuron to a novel environment. *Neurocomputing*, 69(13-15), 1591-1598.

[44]. Yue, S., & Rind, F. C. (2005). A Collision detection system for a mobile robot inspired by locust visual system. *Proceeding of IEEE International Conference on Robotics and Automation*, Barcelona, Spain, 3843-3848.

[45]. Yue, S., & Rind, F. C. (2006). Collision detection in complex dynamic scenes using a LGMD based visual neural network with feature enhancement. *IEEE Transactions on Neural Networks*, 17(3), 705-716.

[46]. Yue, S., & Rind, F. C. (2006). Visual motion pattern extraction and fusion for collision detection in complex dynamic scenes. *Computer Vision and Image Understanding*, 104(1), 48-60.

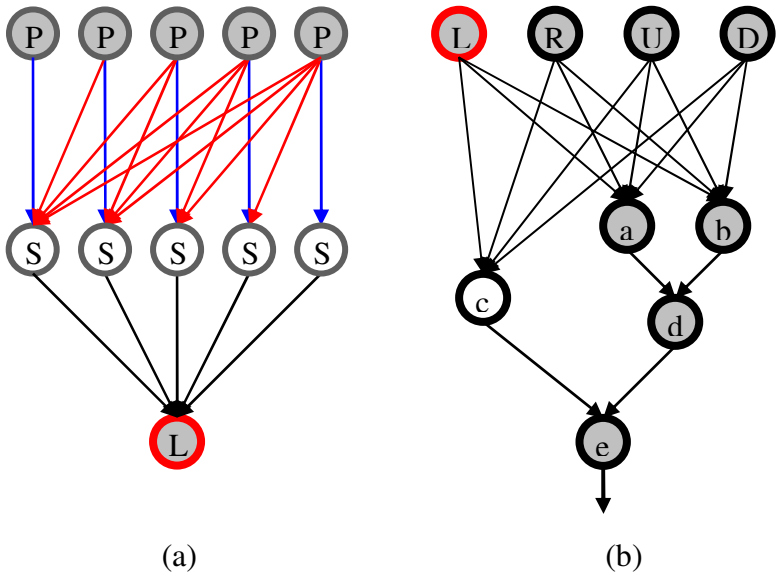


Figure 1. Schematic illustration of the synthetic vision system for collision detection. (a), A direction selective neural network with left sided lateral inhibition. It has three types of cells, i.e., photoreceptor cells P, summing cells S and direction-selective cells. Photoreceptor cells in P layer perceive the luminance change; feed excitation directly to its retinotopically corresponding cell in the summing cell (S) layer; and feed inhibition, delayed by one frame, to its left side neighbours in the S layer. Summing cells in S layer sum excitation and inhibition and pass excitation/inhibition to different cells in the next layer. Direction selective cell, in this example, is a left (L) cell. The L cell is one of four different cells, selective for leftward (L), rightward (R), upward(U) or downward(D) motion. The connectivity and weights of a direction-selective cell do not change in the evolutionary process. Correspond to the size of input images, there are a total of 8,000 P cells, 8,000 S cells and 1 L cell in this left inhibitory neural network. (b), The organisation of the synthetic vision system consisting of a combination of four direction-sensitive neural networks each with asymmetric inhibition projecting to the left(L), right(R), up(U) or down(D). The L, R, U and D neurons pass their excitation/inhibition to all the cells in the next layer, that is, cells a, b and c. Cell a and b then feed their information to cell d; the final cell e receives excitation/inhibition from cell c and d; spikes will be produced in the e cell if its threshold is exceeded. Five successive spikes means a collision is recognised by the network. Note in the figures, circles filled with grey means excitation and inhibition are summed and absolute value taken. In this example all but the S neurons have their outputs rectified. There are total 80 X 100 P cells shared by the 4 direction- selective neural networks, 4 x 80 x 100 S cells, 4 direction-selective neurons, 4 intermediate cells and 1 spiking cell in this synthetic vision system.

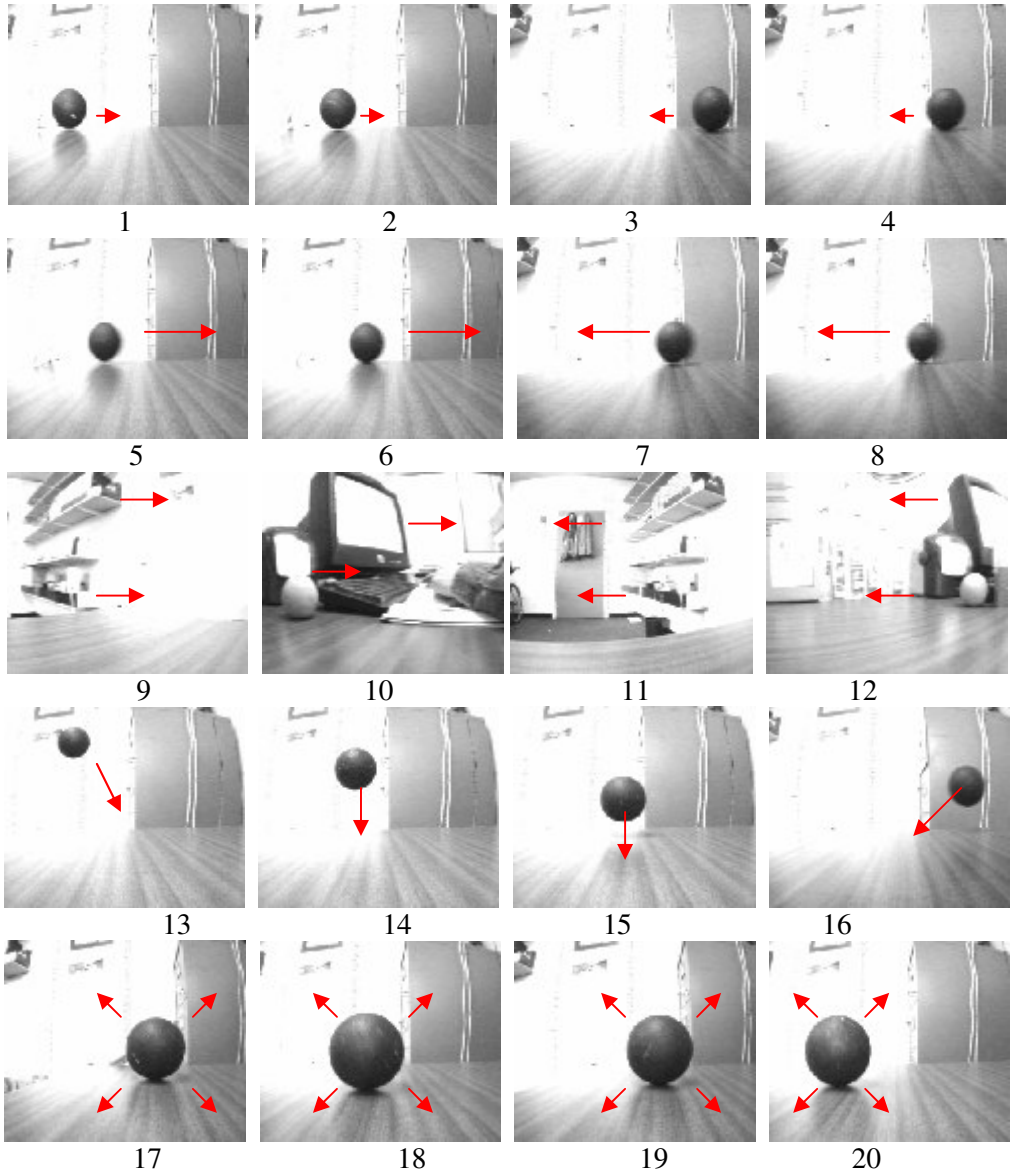


Figure 2. Samples from video sequences making up the ***robotic laboratory environments*** *in which the agents are evolving. The number under each image is the number of its corresponding video sequence. The arrows in the images are added for schematically indicating the visual motion direction. The black ball is 95mm in diameter. In video sequences 1 and 2, the ball is moving across the field of view from left to right at intermediate speed, taking 18 and 27 frames respectively; in video sequences 3 and 4, the ball is moving across the field of view from right to left in intermediate speed, taking 19 and 20 frames respectively; in video sequences 5 and 6, the ball is moving fast across the field of view from left to right, taking 11 and 14 frames respectively; in the video sequences 7 and 8, the ball is moving fast across the field of view from right to left, taking 11 and 13 frames respectively; in video sequences 9 and 10, the robot is turning anticlockwise at about 50°/s while moving forwards, at 3.2cm/s; in video sequences 11 and 12, the robot is turning clockwise at about 50°/s while moving forwards, at 3.2cm/s; in video sequences 13, the ball is bouncing to right; in video sequence 14 and 15, the ball is bouncing up and down; in video sequences 16, the ball is bouncing to left; in video sequences 17, the ball is approaching the robot at 0.4~0.5m/s from right side; in video sequence 18 and 19, the ball is approaching the robot at 0.4~0.5m/s from the central area; in video sequence 20, the ball is approaching the robot at 0.4~0.5m/s from left side. There are 60 frames in each video sequence. The collision sequences are 17~20. The robot's field of view is 60°.*

Figure 3. Sample images from video sequences making up the ***car driving environment*** *in which the agents are evolving. The number under each image is the number of its corresponding video sequence. In video sequences 1~6, the agent faces fast translating cars or figures, while driving at very low speed; in video sequences 7~8, there are road symbols, while driving at high speed; in video sequences 9~11, another car is cutting in or approaching but is not going to collide, while driving at normal speed on a motor way; in video sequences 12~13, the agent faces slowly walking pedestrians, while driving at low speed; in video sequences 14~16, turning and driving at low speed; in video sequences 17~18, collision at low and high speed respectively. The collision sequences are 17 and 18. All the sequences were recorded at 25 frames per second.*

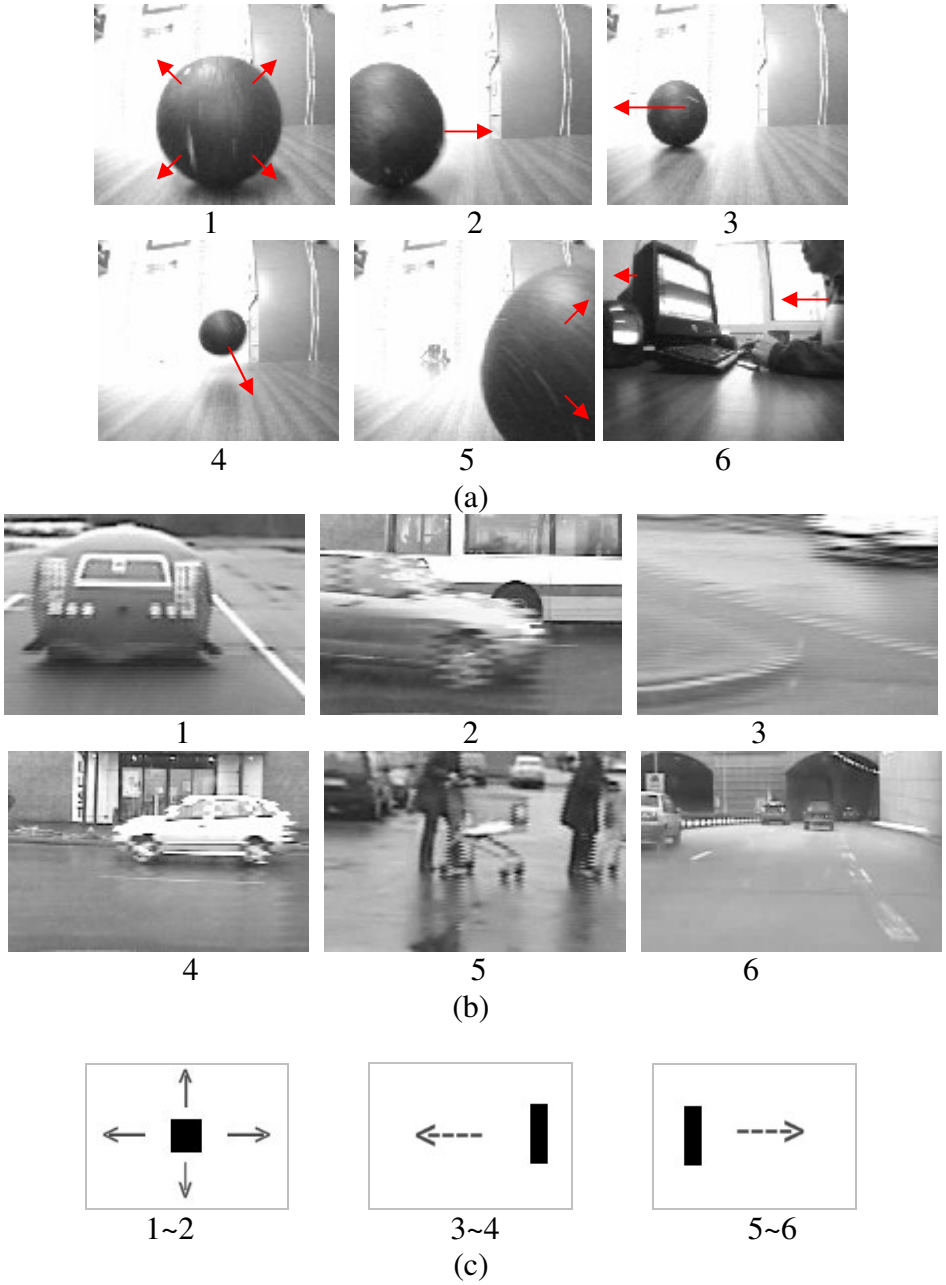


*Figure 4. Sample images from all **test sequences** that were used for testing the agents. For each of the two evolution environments and the computer generated visual stimuli "environment" six corresponding sequences were used. The number under each sample image is its corresponding video sequence number. **(a),** video sequences in robotic laboratory. 1: looming ball at about 0.4m/s; 2: translating ball nearby, taking 16 frames to move across the view field; 3: fast translating ball from right to left, taking 17 frames to move across; 4: bouncing ball from left to right; 5: looming ball near missing at about 0.45m/s; 6: robot turning at about 50°/s. Ball size: 95cm in diameter. The arrows are added for schematically indicating the visual motion direction. **(b),** driving scenes. 1: colliding with a balloon car; 2: fast translating car nearby; 3: turning; 4: translating car; 5: translating figures with trolleys; 6: getting into and out of tunnel. **(c),** computer generated visual stimuli with a 100x80 pixels field of view. 1 and 2: black square (10 pixels times 10 pixels originally) expands to 60x60 pixels with different expanding rates, 2 pixels per frame and 10 pixels per frame respectively; 3 and 4: black bar (15 pixels times 40 pixels) with different translating speeds, 4 pixels per frame and 8 pixels per frame to left respectively; 5 and 6: black bar (15 pixels times 40 pixels) with different translating speeds, 4 pixels per frame and 8 pixels per frame to right respectively. The arrows are added for schematically indicating visual stimuli moving directions.*

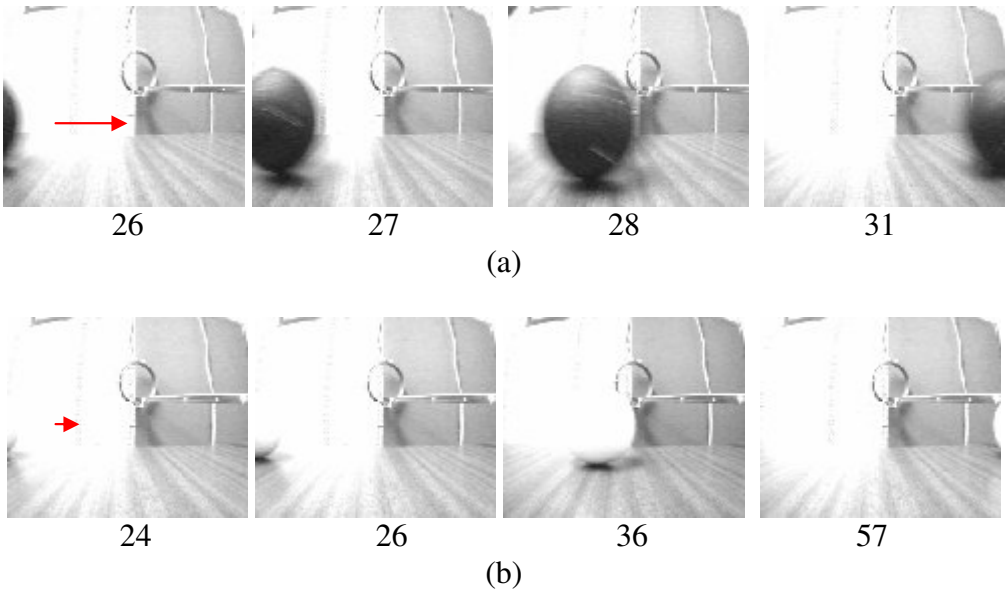
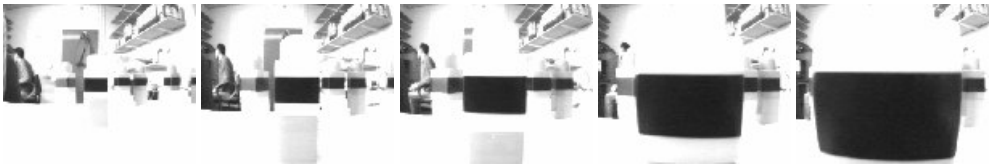


Figure 5.1. Samples of the extreme test sequences from the robotic laboratory for testing the capability of the direction selective neurons and the evolved agents. The numbers under the sampled images are the frame numbers. (a), video sequence 1: black translating ball at close range, taking 6 frames to move across the view field, i.e., from frame no.26 to frame no.31; equivalent angular velocity is 250˚/s. (b), video sequence 2: white translating ball against a white background, taking 34 frames to move across the view field, i.e., from frame no.24 to frame no.57; equivalent angular velocity 44˚/s. The black ball and the white ball were 95cm in diameter. The arrows are added to give an indication of the overall motion direction.



1 30 40 50 55

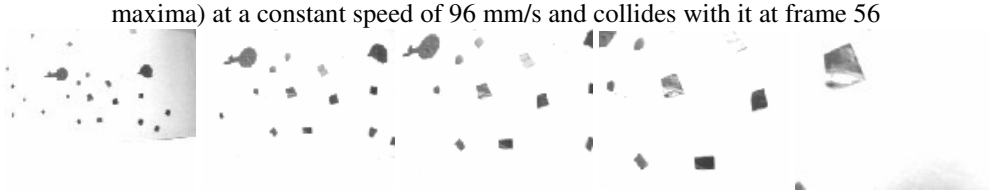
(a) The robot approaches a plastic block 95 mm in height 31 mm in width at a constant speed of 40 mm/s and collides with it at frame 55

1 30 40 50 58

(b). The robot approaches a textured ball 95 mm in diameter at a constant speed of 96 mm/s and collides with it at frame 58

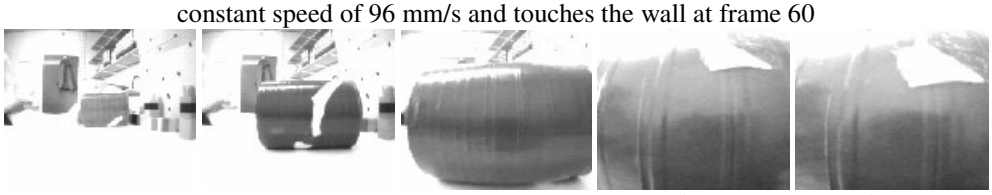
1 30 40 50 56

(c). The robot approaches a 3D irregular shaped block 95 mm in height 100 mm in width (both are maxima) at a constant speed of 96 mm/s and collides with it at frame 56



1 30 40 50 60

(d). The robot approaches a cardboard wall with irregular spots 5 mm to 30 mm in diameter at a constant speed of 96 mm/s and touches the wall at frame 60



1 30 40 50 54

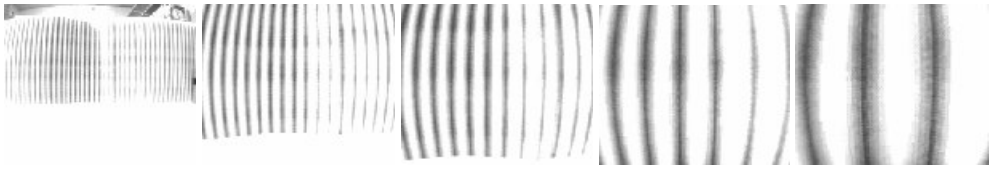
(e). A blue cylinder 130 mm in length and 95 mm in diameter with yellow marks rolls towards the robot and touches the robot at frame 54; the robot views the oncoming cylinder at a fixed position without movement.

1 30 35 40 41

(f). The robot approaches a small black ball 45 mm in diameter at speed 144 mm/s and touches the ball at frame 41

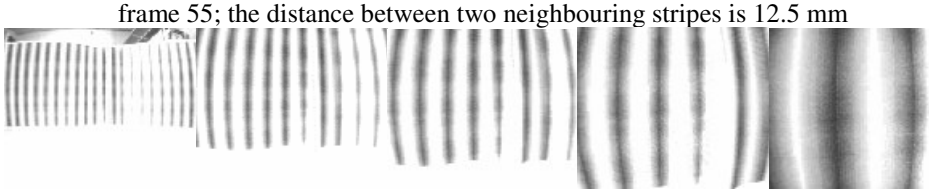
1 10 20 30 35

(g). The robot approaches a large black ball 110 mm in diameter at a constant speed of 144 mm/s and touches the ball at frame 35



1 30 40 50 55

(h). The robot approaches a paper wall with thin stripes at speed 80 mm/s and touches the paper at frame 55; the distance between two neighbouring stripes is 12.5 mm

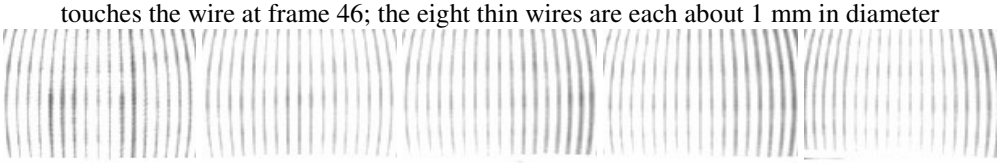


1 30 40 50 59

(i). The robot approaches a paper wall with wider stripes at speed 80mm/s and touches the paper at frame 59; the distance between two neighbouring stripes is 25 mm

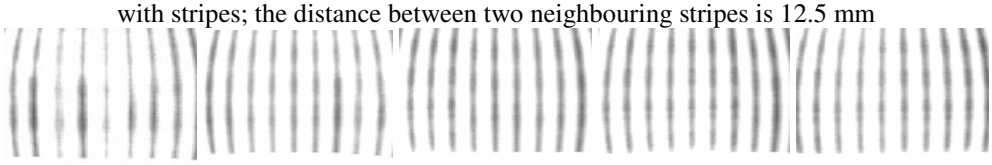
1 20 30 43 46

(j). The robot approaches a piece of RS323 wire consisting of eight thin wires at speed 128 mm/s and touches the wire at frame 46; the eight thin wires are each about 1 mm in diameter



1 20 30 40 50

(k). The robot turns counter clockwise at 63 degree/s at the centre of a cylinder 350 mm in diameter with stripes; the distance between two neighbouring stripes is 12.5 mm



1 20 30 40 50

(l). The robot turns counter clockwise at 63 degree/s at the centre of a cylinder 350 mm in diameter with stripes; the distance between two neighbouring stripes is 25 mm

Figure 5.2. Samples of the diverse test sequences from the robotic laboratory for testing the capability of the evolved agents. These collisions are caused by either the robot approaching the objects or by the object approaching the robot. The frame number is indicated underneath each image.

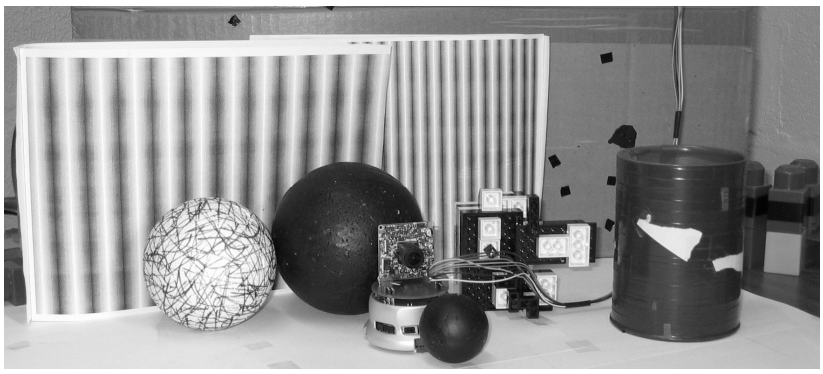


Figure 5.3. The objects and the Khepera robot used in the test sequences in the previous figure.

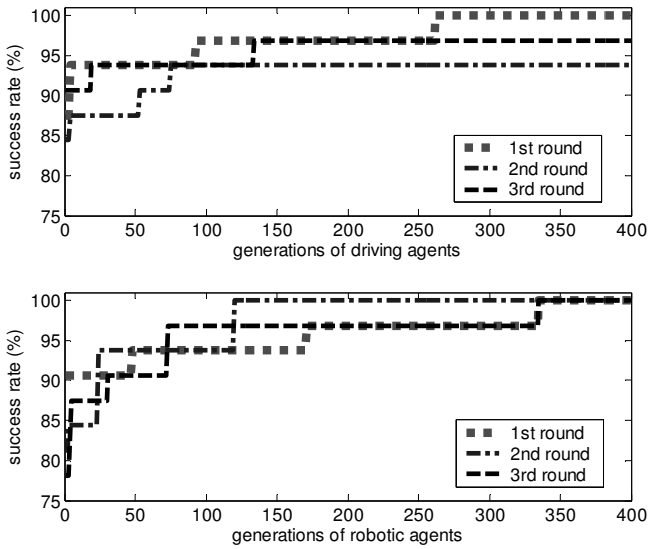


Figure 6.1. Best agents' fitness (success rate %) versus generation in the three rounds of evolution in the driving and robotic environments respectively. For each generation in each round of evolution, only one best performed agent was chosen as the best agent.

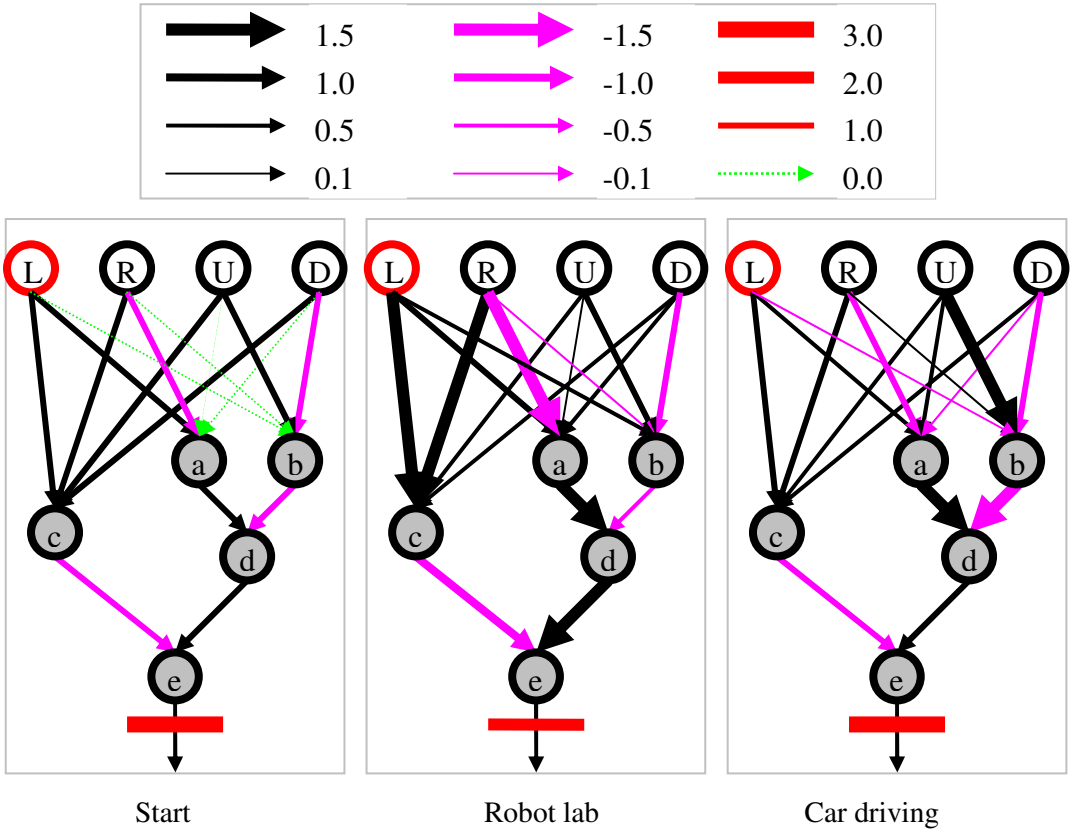
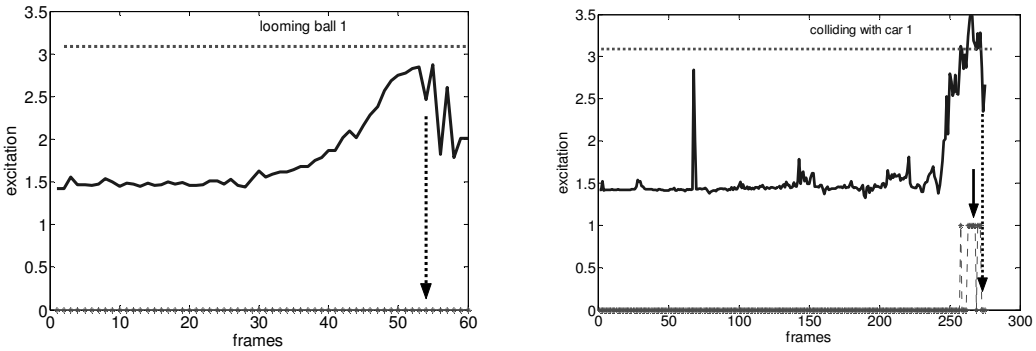
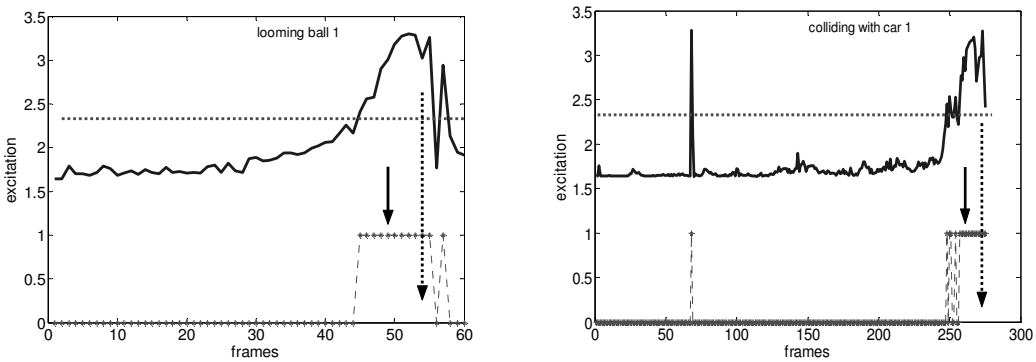


Figure 6.2. Schematic illustration of the connections between cells in start guess agent before evolution and the connections between cells in the best agents after 400 generations of evolution in a robotic laboratory and driving environment respectively. In the pictures, the strength of connections (details in table 2) are simplified into four groups,(1.5, 1.0, 0.5 and 0.1) and each group indicated by an arrow, whose thickness corresponds with the connection strength: thick arrows indicate strong connections. The connections in dotted line mean zero, in black mean positive contribution and in gray mean negative. The threshold of the e cell is also indicated by bars whose thickness reflects their relative levels, thick bars reflect high threshold values.

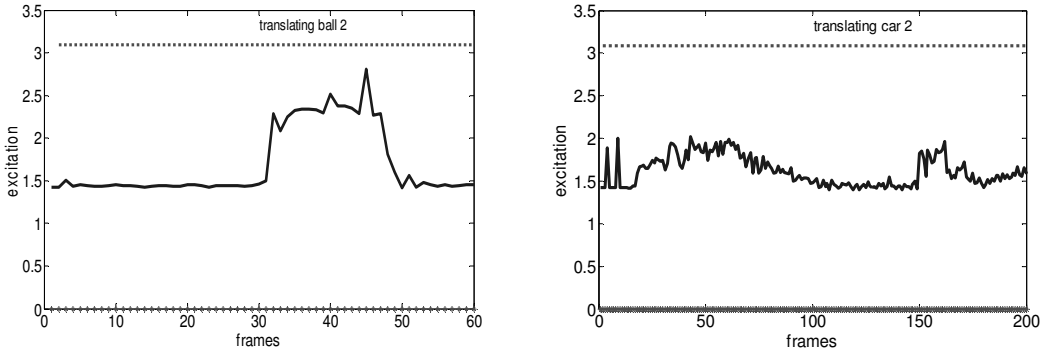


(a). processed by the best agent evolved in driving

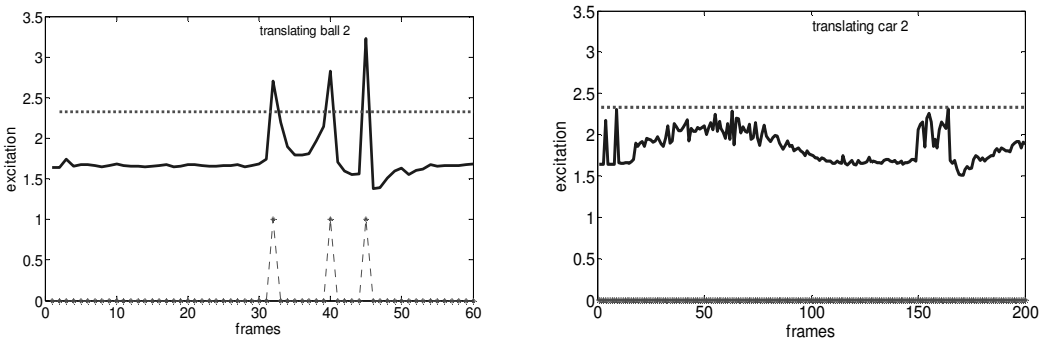


(b). processed by the best agent evolved in robotic laboratory

Figure 7. The excitation levels of the two best agents when processing the collision events in the different training environments, i.e. looming ball (Figure 4 (a), video sequence 1) and the car collision scenes (Figure 4 (b), video sequence 1). For the looming ball, the collision is actually going to be happening at frame 54 as indicated by dotted arrows; the best agent evolved in the robotic laboratory recognised the imminent collision at frame 49 as indicated by a solid arrow; no collisions were detected by the driving environment evolved best agent. For the collision with the car, the collision is actually going to occur at frame 273 as indicated by dotted arrows; the best agent evolved in the robotic lab recognised the imminent collision at frame 261 as indicated by a solid arrow; the best agent evolved in the driving environment recognised the imminent collision at frame 267 as indicated by a solid arrow. The thresholds are indicated with dotted lines. A collision was detected when five successive spikes occurred, as indicated by five elevated asterisks in the bottom trace.

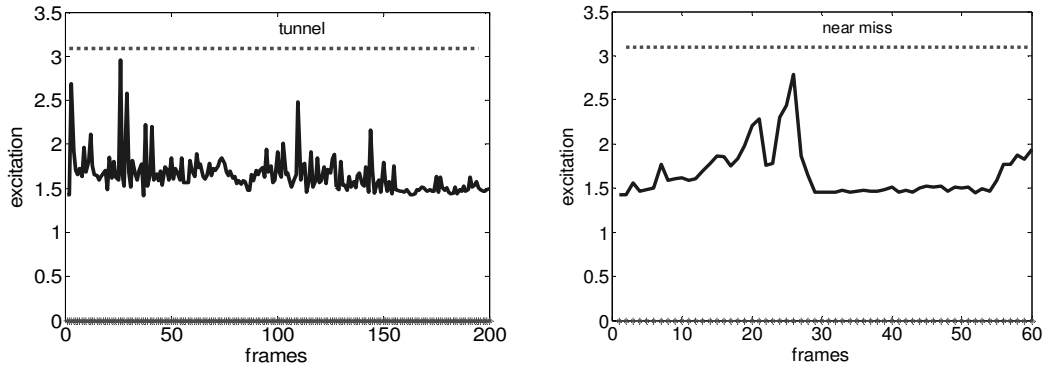


(a). processed by the best agent evolved in driving

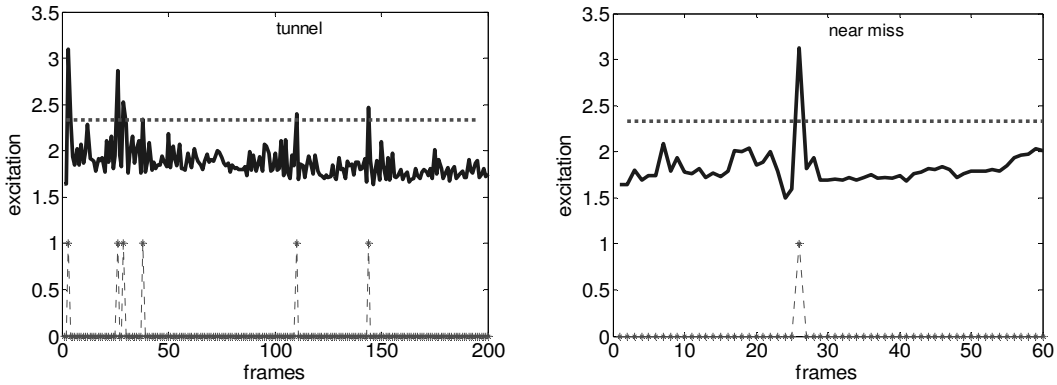


(b). processed by the best agent evolved in robotic laboratory

Figure 8. The excitation level of the two best agents when processing translating events in different environments, i.e., a translating ball (Figure 4 (a), video sequence 2) and a translating car (Figure 4 (b), video sequence 2). The thresholds are indicated with dotted lines. The spikes are indicated with elevated asterisks in the bottom trace. Both of the two best agents made correct decisions for the translating ball though patterns of the response excitation are quite different. The patterns of the response excitation of the two best agents for the translating car are quite similar when challenged with the translating car.

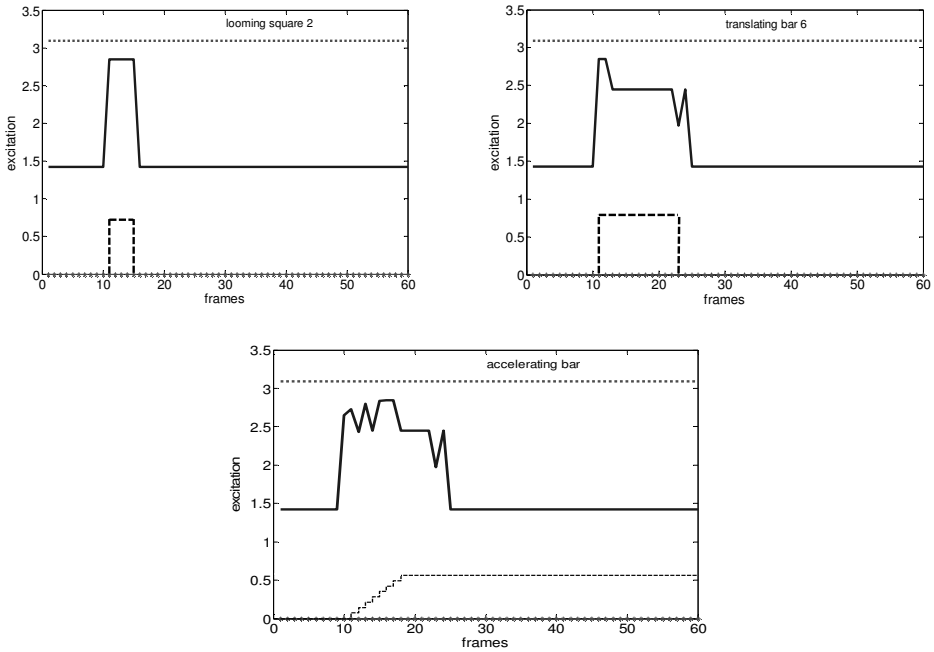


(a). processed by the best agent evolved in driving

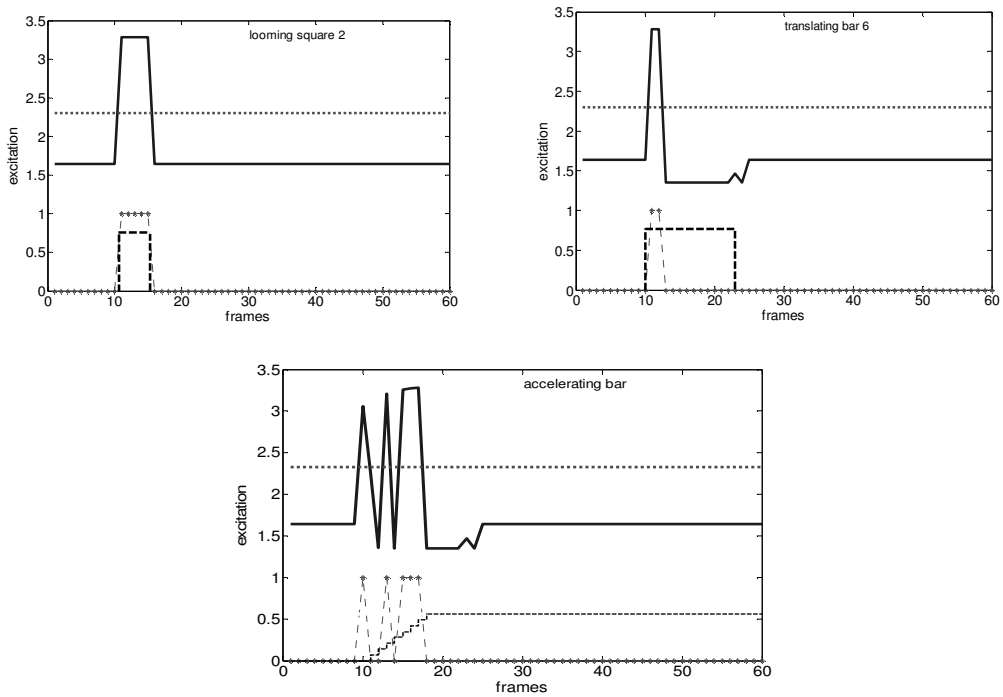


(b). processed by the best agent evolved in robotic laboratory

Figure 9. The excitation level of the two best agents when processing other looming but not colliding events in different environments, i.e., a looming ball near missing (Figure 4 (a), video sequence 5) and driving into a tunnel (Figure 4 (b), video sequence 6). The thresholds are indicated with dotted lines. The spikes are indicated with elevated asterisks in the bottom trace. Both of the two best agents responded to the scenes of driving into the tunnel in similar excitation pattern, however one generated sparse spikes and the other did not produce any spike since they have different thresholds. For the near missing ball, both of the two best agents responded with sharp impulse-like excitation and made the correct decision.

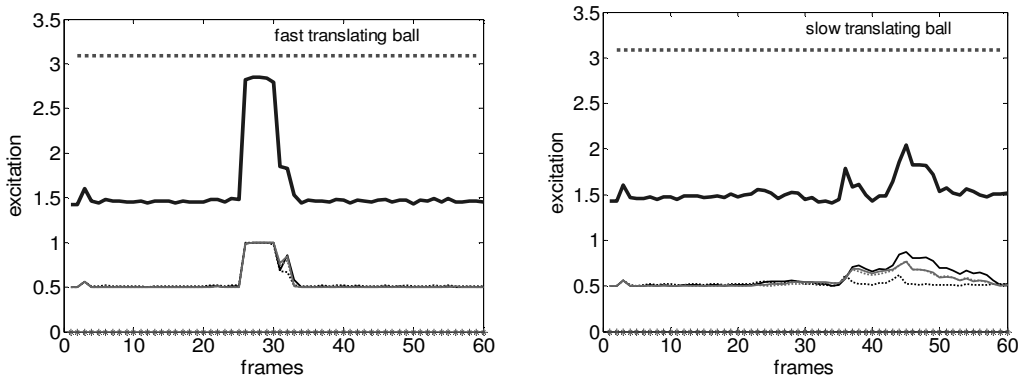


(a). processed by the best agent evolved in driving

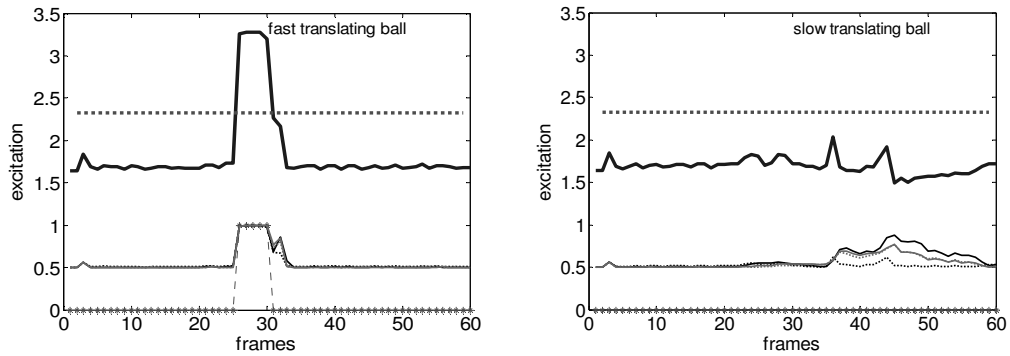


(b). processed by the best agent evolved in robotic laboratory

Figure 10. The excitation level of the two best agents when processing computer generated stimuli: a computer generated looming square (Figure 4 (c), video sequence 2), a translating bar (Figure 4 (c), video sequence 6) and an accelerating bar to right (the same translating bar but accelerated from 0 pixel per frame to 8 pixel per frame with 1 pixel step increase per frame). The thresholds are indicated with dotted lines. The spikes are indicated with elevated asterisks. The schematic speed profiles of the stimuli are indicated with dashed lines. Both of the two best agents made correct decisions for the translating bar though patterns of the response excitation are quite different. The patterns of the response excitation of the two best agents for the looming square are quite similar when challenged with the looming square. However, only the best agent evolved in the robotic laboratory made a correct decision. The best agents evolved in the robotic laboratory responds to the accelerating bar with several spikes.

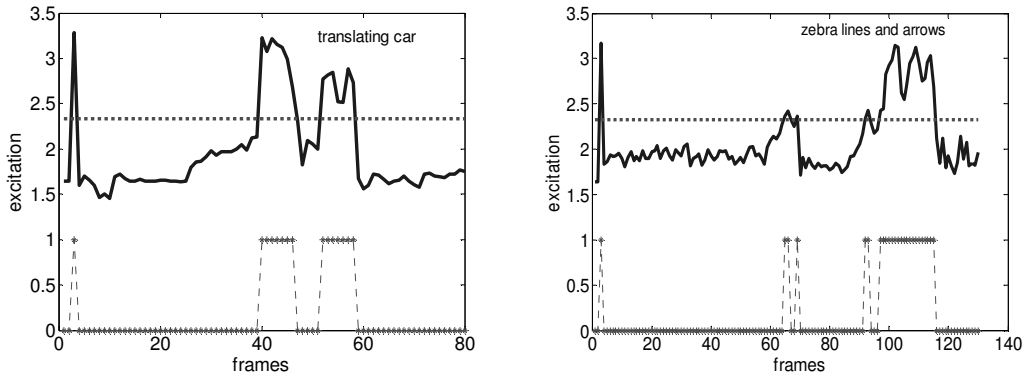


(a). processed by agent evolved in driving

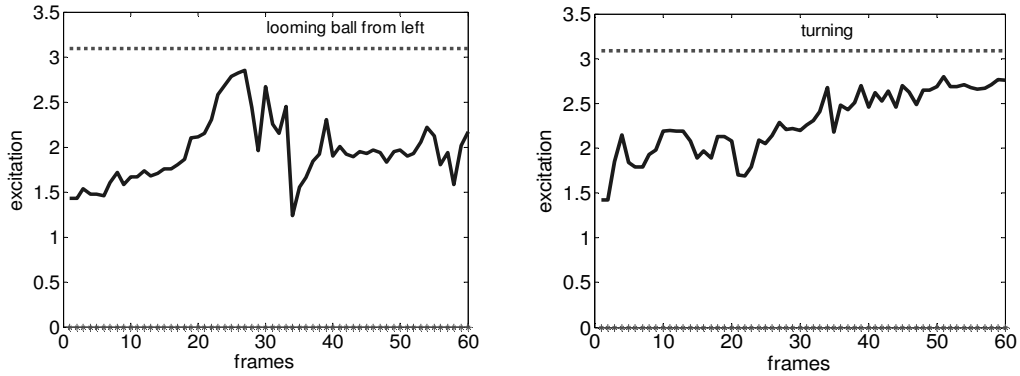


(b). processed by agent evolved in robotic laboratory

Figure 11. The excitation level of the four direction selective neurons and the two best agents in processing the two extraordinary visual scenes: a fast translating black ball (Figure 5.1, video sequence (a)) and a slow translating white ball on a white background (Figure 5.1, video sequence (b)). The thresholds are indicated with dotted lines. The spikes are indicated with elevated asterisks. The excitations from the four direction selective neurons are indicated with solid and dashed lines fluctuating between (0.5~1.0). The responses of the four direction selective neurons to the fast translating ball are without much difference. It is beyond the capabilities of the best agents in detecting the visual motion patterns in this case. The current asymmetrical lateral inhibition only spreads to 8 cells away. Therefore, the system can only cope with visual motion slower than an angular velocity of 120°/s (with current input images and camera setting: 60° field of view angle, 100 pixel in horizontal and 25 frames per seconds). In the case of white ball on a white background, the four direction selective neurons did not respond to the moving white ball until at frame no.36 when contrast became available.

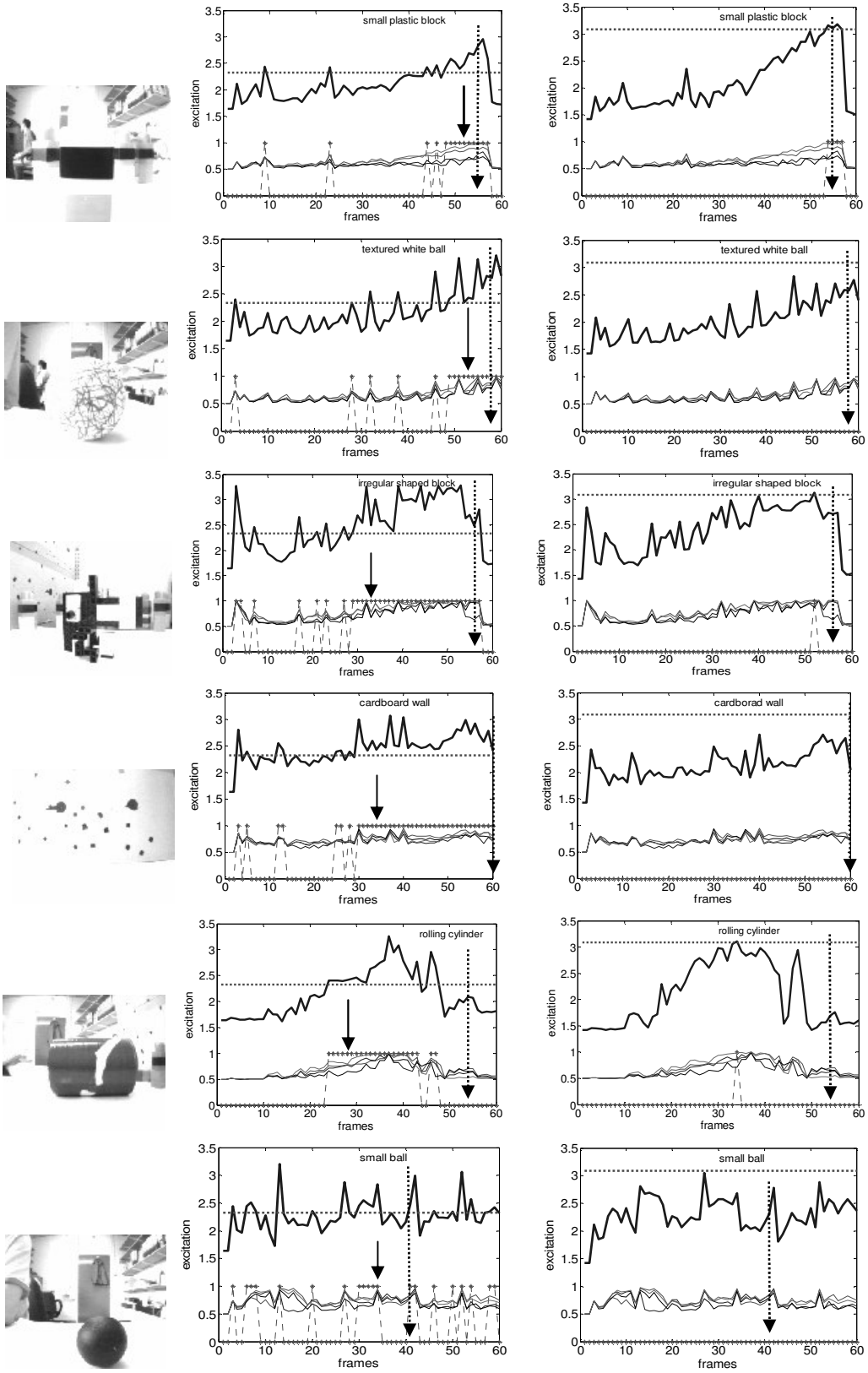


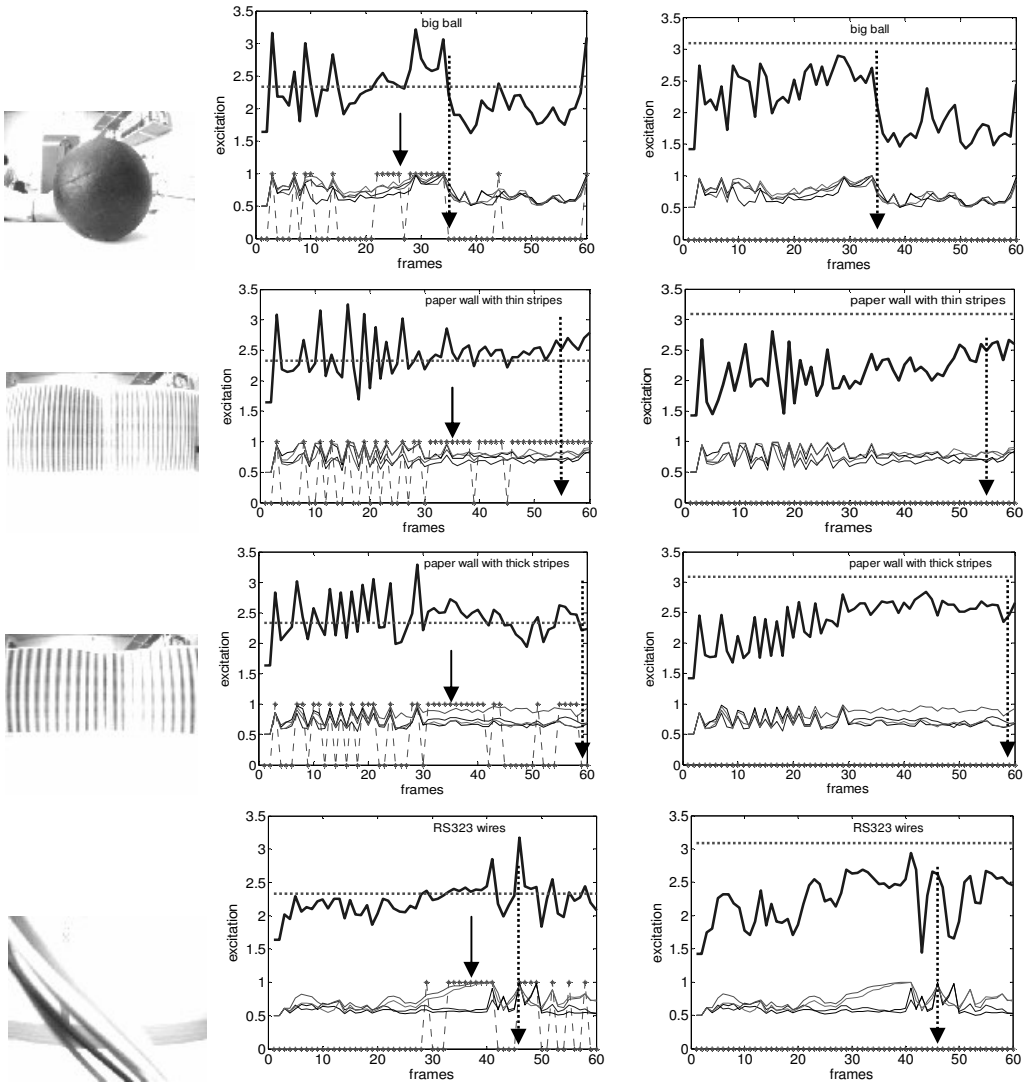
(a). processed by the best agent evolved in robotic laboratory



(b). processed by the best agent evolved in driving

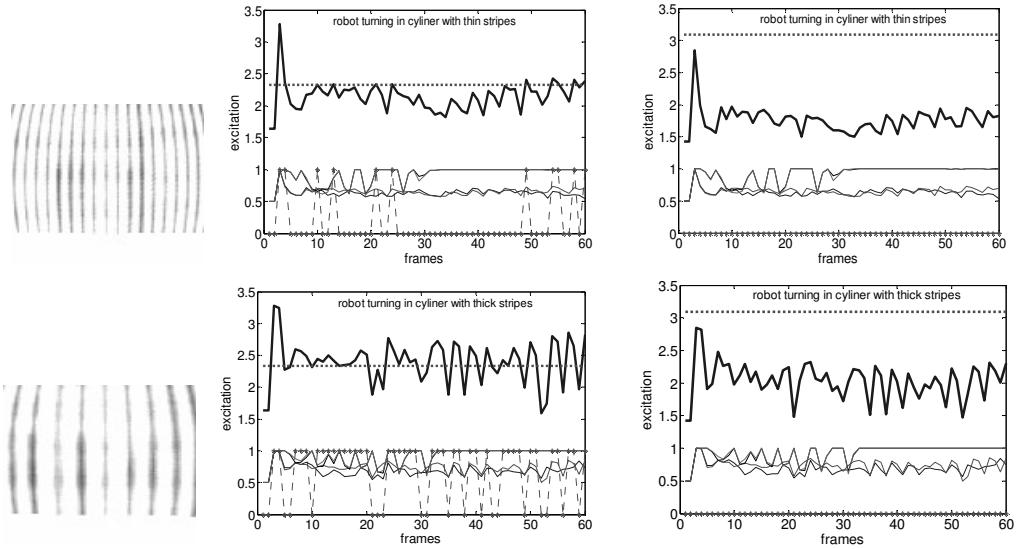
Figure 12. The excitation level of the two best agents when processing the visual scenes in swapped environments. (a), fast translating car (Figure 3 video sequence 2) and zebra lines/arrows (Figure 3 video sequence 8) processed by the best agent evolved in robotic laboratory. (b), looming ball (Figure 2 video sequence 20) and robot turning (Figure 2 video sequence 10) processed by the best agent evolved in driving. The thresholds are indicated with dotted lines. The spikes are indicated with elevated asterisks. The best agent evolved in the robotic laboratory failed in both of the road scenes: translating car and the zebra lines/arrows. It responded to these two scenes with high excitation and continues spikes. The best agent evolved in driving failed in detecting the looming ball. It failed in all the collision scenes in the test. Since it also responded to turning with high excitation ((b), left), it may send out false alarm if its threshold is lowered to fit to the looming ball ((b), right).


small plastic block
textured white ball
irregular shaped block
cardboard wall
cardborad wall
rolling cylinder
small ball
excitation
frames



(a). sample images (b). by the best robotic agent (c). by the best driving agent

Figure 13.1 The excitation level, spikes of the two best agents and the excitation level of the four direction selective neurons when processing the ***collision visual scenes*** *listed in Figure 5.2. From top to bottom, the robotic agent detected different colliding objects: plastic block at frame 52, textured white ball at frame 53, irregular shaped block at frame 33, big cardboard wall at frame 34, rolling cylinder at frame 28, small ball at frame 34, big ball at frame 26, thin striped paper wall at frame 35, thick striped paper wall at frame 35 and RS323 wires at frame 37- all with at least several frames in advance. The driving agent did not detect these colliding objects but it responded to these collisions with higher excitation level. The frames at which the robot touched the objects were indicated with dashed arrows; the frames at which the collisions were predicted were indicated with short solid arrows. The excitations from the four direction selective neurons are indicated with thin solid lines fluctuating between (0.5~1.0). (a), the sample images in the video clips. (b), the outputs of the best robotic agent when processing the visual scenes. (c), the outputs of the best driving agent when processing the visual senses.*



(a). sample images (b). by the best robotic agent (c). by the best driving agent

Figure 13.2 The excitation level, spikes of the two best agents and the excitation level of the four direction selective neurons when processing ***the turning scenes*** *listed in Figure 5.2. From top to bottom, the robotic agent did not signal a false alarm when turning within the cylinder with thin stripes; however, a false alarm was given when turning within a cylinder with thick stripes. The driving agent did not respond to these turning scenes. The excitations from the four direction selective neurons are indicated with solid lines between (0.5~1.0). (a), the sample images when the robot was turning within cylinders. (b), the outputs of the best robotic agent when processing the visual scenes. (c), the outputs of the best driving agent when processing the visual senses.*

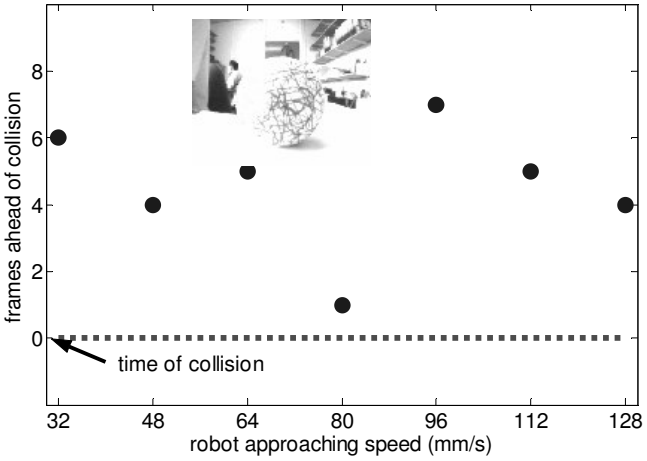
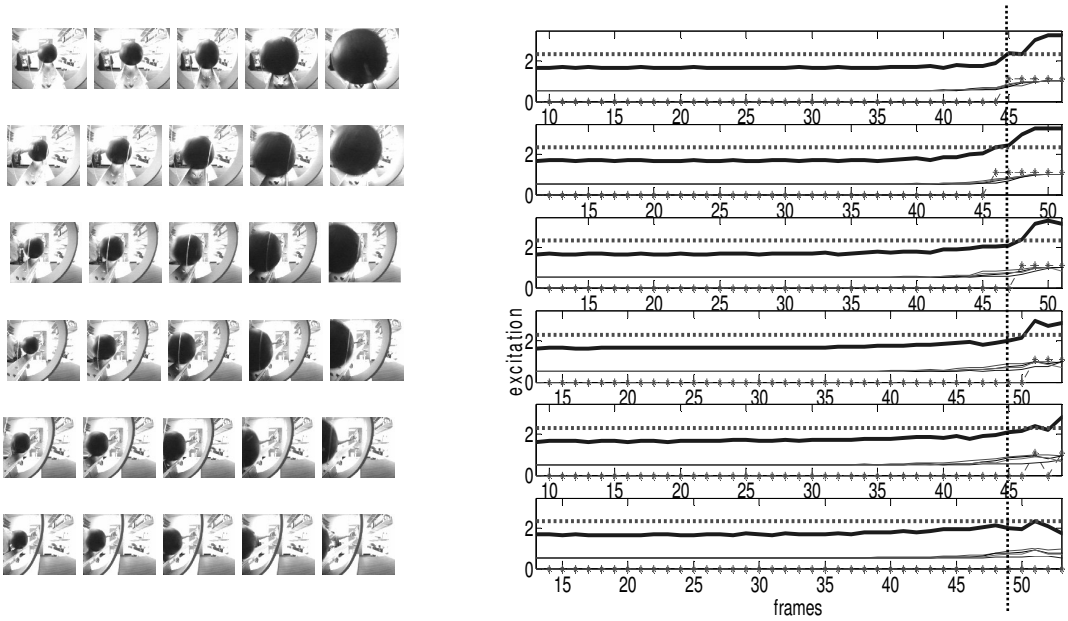


Figure 14. At different approach speeds, the evolved best robotic agent detected imminent collision 1~7 frames before it happened. The y-axis is the number of frames ahead of collision when it was detected. The exact frame numbers ahead of collision are indicated by filled circles. The ball used in the experiment is the same textured ball as described in Figure 5.2, Figure 5.3 and Figure 13. Time of collision for each case is the time when the robot touches the ball and is aligned to zero on the y-axis.



(a). the 5 successive frames (b). the response of the robotic agent

Figure 15. At different approach angles, the evolved best robotic agent responded to collision and non-collision scenes correctly. The ball used in the experiment is black 65mm in diameter. The ball was delivered on a 1000mm long 70mm high track and gained speed due to gravity. Two pieces of transparent monofilament at the lower end of the track stopped the ball from contacting the robot. The distance between the robot's body centre and the high end of the track was maintained at 1075mm in each recording. The centre of the image was fixed to the high end of the track and calibrated during the recording with a mini TV. In the 1st and 2nd video clips, the ball would hit the robot if the monofilament was not present; in the 3rd clips, the ball would graze one side of the robot; in the others, the ball is not on a collision course with the robot. From top to bottom, the deflection angle of the approaching **ball became bigger each time**. *(a), the last five successive frames before/when the ball touched the strings are shown. A line marks the time at which the image of the approaching object moved out of the view of the robot (b), the corresponding responses from the best robotic agent; excitation is shown relative to the last frame at which the ball contacted the monofilament.*

*Table 1. The fitnesses (**success rates**) of the evolved best agents per generation, over 400 generations*

Generation	Best agent evolved in driving environment	Best agent evolved in robotic laboratory
(*Start guess*)	(*41%*)	(*59%*)
1	88%	91%
25	94%	91%
100	97%	94%
400	100%	100%

Table 2. The weights of the connections between cells of the best robotic agent and the diving agent after 400 generations

Connection	Start	Rob. lab	Driving	Connection	Start	Rob. lab	Driving
L-a	1	1.1080	0.5225	a-d	1	1.3915	1.4013
R-a	-1	-1.4892	-1.0015	b-d	-1	-0.6945	-1.4951
U-a	0	0.0582	0.3573	L-c	1	1.4687	1.1403
D-a	0	0.4961	-0.0963	R-c	1	1.4140	1.2038
L-b	0	0.4277	-0.0367	U-c	1	0.5225	0.5958
R-b	0	-0.2615	0.1872	D-c	1	0.7356	0.5518
U-b	1	1.0132	1.4765	d-e	1	1.4208	1.2146
D-b	-1	-1.0005	-0.8226	c-e	-1	-0.8324	-1.1276
				e threshold	3.0	2.3268	3.0873

*Table 3. The performance of the two best agents (evolved in robotic laboratory and evolved in driving environment respectively) challenged with the three groups of **test** sequences*

	Best agent evolved in driving environment			Best agent evolved in robotic laboratory		
Test sequences	Driving scenes	Robotic lab.	Computer generated	Driving scenes	Robotic lab.	Computer generated
1	√	×	×	√	√	√
2	√	√	×	√	√	√
3	√	√	√	√	√	√
4	√	√	√	√	√	√
5	√	√	√	√	√	√
6	√	√	√	√	√	√

Note: No.1 sequence is the collision event in the three testing groups; for the computer generated test sequences, No.2 is also a collision event. √ indicates detected correctly, i.e., detected collision for collision events and did not signal collision for non-collision events; × failure in an event.

Table 4. The performance (fitnesses or success rates) of the evolved best agents in ***swapped (i.e., non-native)*** *environment*

best in generation	evolved in driving environment, tested in robotic laboratory (Figure 2)	evolved in robotic laboratory, tested in driving environment (Figure 3)
(*Start guess*)	(*59%*)	(*41%*)
25	41%	47%
100	50%	88%
400	50%	94%