

A Robotic Eye Controller Based on Cooperative Neural Agents

Oscar Chang, Pascual Campoy *Member, IEEE*

Carol Martínez *Student Member, IEEE* and Miguel A. Olivares-Méndez *Student Member, IEEE*

Abstract—A neural behavior initiating agent (BIA) is proposed to integrate relevant compressed image information coming from others cooperating and specialized neural agents. Using this arrangement the problem of tracking and recognizing a moving icon has been solved by partitioning it into three simpler and separated tasks. Neural modules associated to those tasks proved to be easier to train and show a good general performance. The obtained neural controller can handle spurious images and solve an acute image related task in a dynamical environment. Under prolonged dead-lock conditions the controller shows traces of genuine spontaneity. The overall performance has been tested using a pan and tilt camera platform and real images taken from several objects, showing the good tracking results discussed in the paper.

keywords : Computer vision, neural agents, behavior initiators, agents cooperation, genuine spontaneity, neural networks.

I. INTRODUCTION

Nature invented an acute mobile eye and a big spark toward intelligence flared in the evolutive night. Indeed experts believe [1] [2] that the ability to carry out survival related task in a dynamical environment, with the help of an acute vision system, is a fundamental issue in the development of true intelligence. Sharp vision depends on complementary image handling abilities such as efficient tracking and fast recognition of predator/prey situations. In any case the associated controller has to respond with a quick, conflict free behavior to images provided by the eye. In many situations the objective moves so fast that the eye correlated muscles are unable to mechanically compensate the displacement and as a result the controlling vision system has to deal with an image that travels through the associated retina. Biological vision systems have evolved diverse mechanisms to handle image movement and recognition. Modeling and application of endocrine principles been reported in [3]. Special purpose algorithms for fast pixel tracking have been studied in [4]. In [5] competitive artificial neural networks (ANN) are devoted to eye tracking in video sequences and in [6] a convolutional neural network is trained for tracking purposes.

Biological neural controllers conceals a rich repertory of behavior initiating agents which make real life neurons tick with enthusiastic self determination. At the highest level of brain development the existence of specific behavior initiating mechanisms in monkeys have been studied and modeled [7]. Recent experiments have found clear indication of initiating mechanisms in the *Drosophila* brain [8]. Efforts

to model and put to use these phenomena have been reported. In [9] fuzzy rules are used to infer in the behavior of a navigation expert system. In [10] the behavior-release mechanism of the basal ganglia was used in a robot controller. In [11] it was experimentally demonstrate that behavior initiating agents based upon ANN and working in collaboration with other image processing agents, produce highly proactive vision guided robots.

For this paper the quality of being and agent implies the capacity to perform a useful job without external intervention and satisfying the four weak conditions of [12]: Autonomy, Social ability, Reactivity and Pro-activeness. In isolated conditions some of our agents have zero reactivity or zero proactiveness and can be described as neural modules.

II. BEHAVIOR INITIATING AGENTS

In the *drosophila* brain studies [8] researches found non linearities and the possible competence of only a few neurons in the final behavior initiating mechanism, deep buried in the fly's brain. Such mechanism or agent provide the fly with genuine spontaneity -a distinctive label of living creatures-enabling the insect to get bored about tedious situations and deciding (self decision) to take radical (and possibly life saving) changes in its current line of behavior. These biological behavior initiators mechanisms also produce, in the long run, a conflict-free time sequence of behaviors, which preserves the insect from physical damage.

The first concern of this work is to define a behavior initiation mechanism which, mimics the capacity of genuine spontaneity found in flies, and use it to auto impulse robotic eye activities. The presented solution is based upon an n-flop, a robust neural network constructed with sigmoidal neurons and sharing a common excitatory input K . Being robust, it serves as foundation for other large scale optimization structures such as the TSP neural solver. The n-flop is the basic building block beyond the concept of programming with neurons [13], the term is derived from flip-flop a computer circuit that has only two stable states. n-flops have n stable states and the capacity to solved high dimensional problems [14].

In an n-flop, neurons are programmed by their inter-connections to solve the constraint that only one of the n will be active when the system is in equilibrium. To this end, the output of each neuron is connected with an inhibitory input weight (-1) to each of the other $n-1$ neurons inputs (lateral inhibition). In addition each neuron receives a common excitatory input K which, on controlled situations, tends to force all neuron outputs towards 1. A solution is

found by rising K and forcing all neurons to some near-equilibrium but unstable "high-energy" state. At this point K is set to almost zero, releasing the network to seek a low-energy stable or equilibrium state. The solution given by a non biased n-flop is an unique but unpredictable winner. A unique winner guarantees a conflict-free operation in terms of robotic actuators. In our terminology an isolated n-flop behaves as an independent agent with high proactiveness and null reactiveness, capable of producing random sequences of conflict-free states. Proper biasing can launch statistical reactiveness in this primal behavior.

The n-flop accepts elaborate data structured modulation which makes possible to solve resource allocation problems such as the TSP [13]. When properly done modulation changes the statistics of the n-flop behavior from pure random to quasi optimal. In terms of agent operations this is equivalent to go from pure proactive to reactive-proactive behavior. We use the n-flop as an integrating, modulable behavior initiating agent, modulated with compressed image information coming from others agents. In previous experiments this arrangement showed encouraging results [11].

A. Image Compressing Agents

For this paper an image compressing agent is a trained neural network with the following characteristics:

- Uses a 100x 100 pixels input retina
- Has one hidden layer
- Has a small number of output neurons (less than twenty for our current modeling)

This finite network has to be successfully trained with backpropagation to execute a task which requires image processing capabilities [15].

We will use the term image compression to denote that the high dimensional set of all possible definable images is mapped into the outputs of a classifier network with a few output neurons. In practice all high dimensional image activities are compressed into the conduct of a few individuals. In this work the main role of an image compressing agent is to be trained in object recognition or object tracking duties, in order to produce compressed image information used, thereafter, to modulate a behavior initiating agent.

III. METHOD

Our goal is to construct neural machinery capable of controlling a robotic eye that executes basic image oriented duties while maintaining genuine spontaneity. The system works independently, with no intervention from higher level layers, although due to the neural nature of the controller such intervention is feasible. For now the controller should actively search for a particular kind of moving object and do tracking and visual servoing of it for as long as possible. Is the object is lost the eyed should assume a proactive search. The system has to be able to escape from unavoidable mistakes and traps that occur in image recognition based upon trained ANN. The control activities are partitioned into three neural sub-agents:

- Behavior initiating agent
- Image centering agent
- Image recognition agent

The behavior initiating agent or BIA is based upon an n-flop and acts as the main control administrator. In non modulated conditions this agent produces a conflict-free, random eye movements by its own.

The image centering agent or ICA is a trained ANN specialized in processing the image information that conveys the relative position of an arbitrary image in the retina. Its job is to help the main controller to bring interesting images to the center of the retina.

The image recognition agent or IRA is a trained ANN specialized in recognizing a particular kind of object (icon) and discriminates it from other objects or from the background noise. Its job is to help the main controller in the recognition of proper icons.

Once all participant agents are debugged and tested structural modulation is activated so that the behavior initiating agent is modulated with information coming from image processing agents (figure 1).

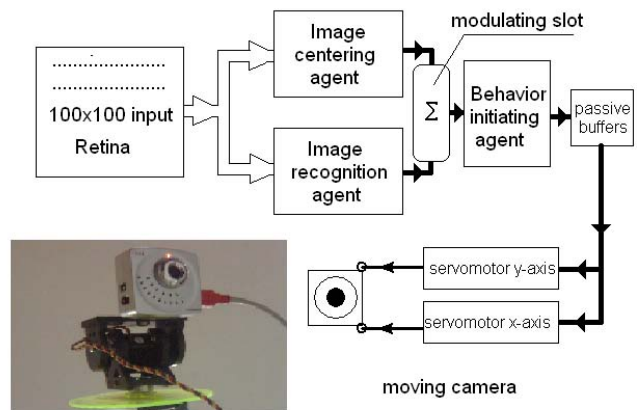


Fig. 1. Schematic of the robotic eye controller composed by three cooperating agents and a 100x100 retina, which provides collective information

A. The Behavior Initiating Agent

The image capturing system is constructed with a camera, mounted in a pan and tilt platform (figure 2). The servo motors are connected through non intelligent buffers to the output of two independent 5-flops fired in parallel. These flops acts as a behavior initiators agent (BIA) and the output (winner) of each 5-flop is used to move the camera in the x,y axis (horizontal/vertical) according to the 5x5 grid arrangement shown in figure 2. Here after a given firing the neurons x_1, y_3 have become winners. Non intelligent buffers realized plain calculations and move the servos in order to bring this virtual point to the center of the retina. When flops are fired again a new virtual point in the grid appears, buffers make calculations and servos respond accordingly. When the BIA is repeatedly fired the system respond with a random eye movement that chases an virtual moving point. End-ofrun information which come from the buffers is fed back toward

the n-flops so that in free running conditions the eye improves its random search (figure 2). Considerations about this search are given in [11].

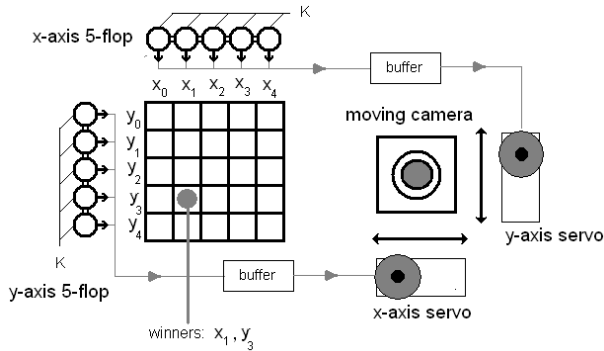


Fig. 2. Behavior initiating Agent. The output of two isolated 5-flops is used to activate two servomotors according to a pre established 5x5 grid arrangement. When flops are repeatedly fired the servomotors begins to chase and imaginary moving point.

B. The Image Centering Agent

It is well known the importance of image centering in image recognition processes [16]. Based in our experimental results ANN are much more difficult to train, by any method when the processed images move too much in the retinal area. So a training scheme where the image to be recognized stays around the center of the retina is adopted. This raises the recognition rate dramatically but also makes necessary to pull any interesting image to the retina center before it can be recognize.

Image centering is done by providing compressed image information to the previously defined BIA, which remains the main administrator of camera movement. Compressed image information comes from an ANN trained with a copy of the grid system used in by the behavior initiating agent. A figure (circle) of variable size is randomly located in a i,j grid position and the respective information is presented as targets to a trainable neural net with 10 outputs, which map the 5x5 grid positions. In the example of figure 3 a circle has been located in the x_1,y_3 position. From this the 10 required targets are generated. This net has 19 hidden and 10000 inputs neurons. Training is done with standard backpropagation in about 100.000 cycles. After training the output neurons indicate the grid position of a received image. The output becomes noisy and contradictory when complex images circulate in the retina. Figure 3 shows schematics.

C. The Image Recognition Agent

The image recognition agent is defined by the 100x100 retina input, 19 hidden and two output neurons. Its duty is to recognize a specific icon or object. For control applications we choose this icon to be a fat arrow. When training, a figure of variable size and randomly rotated, is randomly located near the center of the retina. If the figure is a fat arrow the two targets are fixed at 0.9–0.1. Otherwise targets are fixed at 0.1–0.9. A total of 100 objects, including circles and rectangles

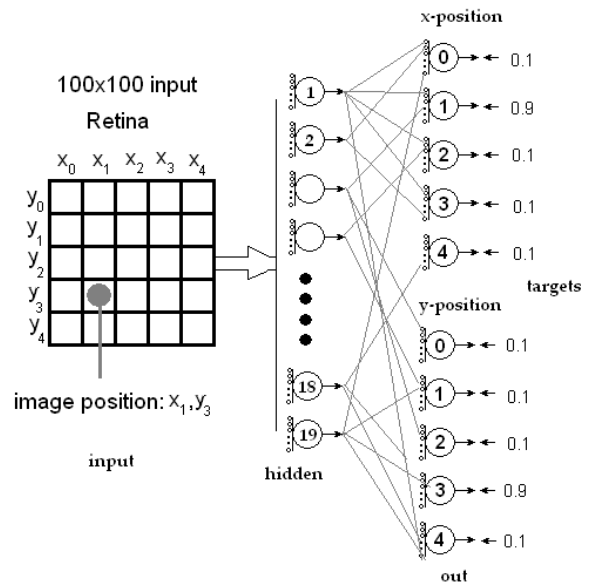


Fig. 3. The image Centering Agent ICA. An ANN is trained to indicate image location in a 5x5 grid superimposed over the 100x100 retina. For isolated, not too big images obtained information is reliable. For complex imagery, however, information becomes very noisy and contradictory.

of different size, are presented as counterexample for arrows. Training is done with standard backpropagation and it takes about 100.000 cycles to get a properly qualified individual. After training, the network successfully recognizes fat arrows located near the center of the retina in any position. It also clearly rejects other icons such as circles or rectangles of any size. Although trained as a classifier the output become noisy and falls in contradictions when complex images reach the retina. Figure 3 shows the schematic of a qualified image recognition agent.

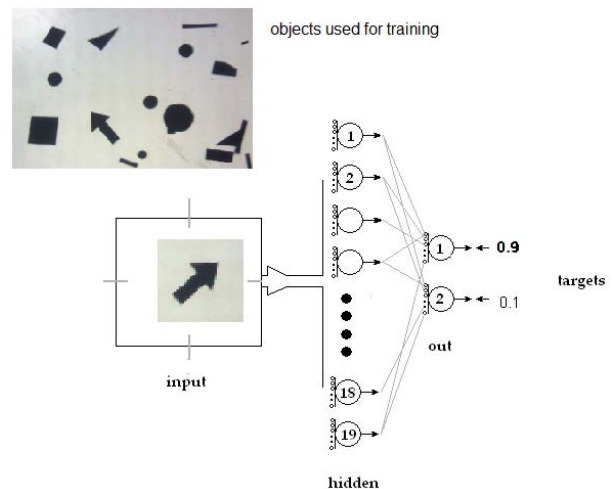


Fig. 4. The image recognition agent IRA. An ANN is trained to recognize a unique icon (arrow) against other 100 icons. For isolated images near the center of the retina the obtained information is reliable. For complex imagery, information becomes very noisy and contradictory.

D. Agents Cooperation: The Modulation Process

In an isolated n-flop the probability of neuron i to become a winner depends on its internal potential V(i), given by :

$$V(i) = \sum_{j=1}^n a_j * (-1) + K \quad \forall j \neq i \quad (1)$$

where:

- V(i) is the internal potential,
- a_j the output from neuron j
- K common excitatory input

For modulation purposes to this internal potential a modulating scalar b_k is added:

$$V(i) = \sum_{j=1}^n a_j * (-1) + K + b_k \quad (2)$$

The scalar b_k is in turn the weighted summation of signals coming from the image centering agent.

That is:

$$b_k = \sum_{j=1}^{10} o_j * w_j \quad (3)$$

where

- b_k modulating scalar
- o_j jth output of the ICA
- w_j inter agent connecting weight

Equation 3 allows that the compressed image information, coming from the ICA, modifies the statistical behavior of the BIA.

The selection of appropriate inter agents weights w_j is critical. They change the behavior of the BIA from being pure proactive and random, to being reactive-proactive or only reactive. In living creatures a good balance between these two characteristics is carefully tuned by evolution [8].

In the previous work, the selection of the inter agent weights was done with simulated evolution [11]. For this paper reverse engineering is used and the robotics eye is manually tuned until it reaches its top working capacity. Further improvement using simulated evolution might be considered.

In order to simplify the image centering modulation in equation 3, just the weights between connected alike neurons in the ICA and the BIA, are set different from zero:

$$b_i = o_j * w_i \quad \text{only for} \quad i = j \quad (4)$$

This leaves only ten weights which were experimentally set to 0.1.

Image recognizing modulation:

To complete the modulation process each inter agent weight w_j is in turn modulated (switched on-off) by the average value of the difference of the image recognizing agent in the last ten BIA firing that is, let

$$g = \frac{1}{10} \sum_{j=-10}^0 o_{1j} - o_{2j} \quad (5)$$

where

- g=average of error
- o_{1j} =output 1 of IRA
- o_{2j} =output 2 of IRA

this average is compared against an operative threshold which indicates how likely the icon analyzed in the last ten BIA firing a correct one, i.e.

$$\text{if } g > \text{threshold then close switch} \quad (6)$$

else open

The appropriate setting of inter agent weights and operative threshold causes that when the centralized icon is the correct one the BIA accepts centering information and performs real tracking. If observed icon is not correct, centering information is disconnected and BIA initiates virtual eye movements (improved random search). See figure 5.

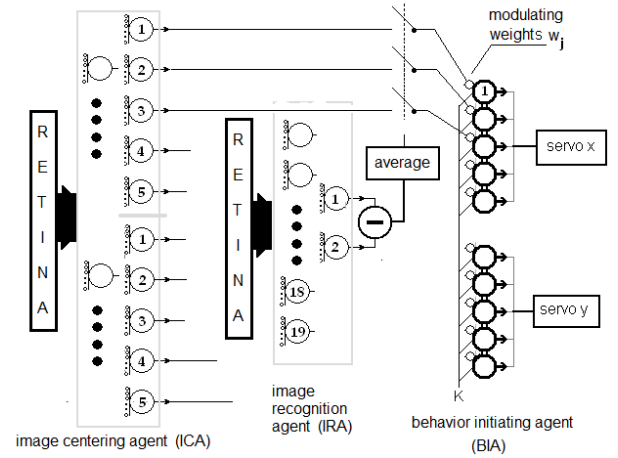


Fig. 5. Modulated Robotic eye controller. BIA accepts compressed image information from cooperating ICA. This information is enough to carry on an effective image tracking. The image recognizing agent (IRA) controls through average error if an found icon deserves to be tracked or not. When not tracking BIA performs imaginary eye movement.

IV. RESULTS

The system behaves well with simulated and real images and shows some promising capacities. For instance when the eye explores a complex background formed by superimposed icons, some spurious figures are mistaken as arrows, causing a potential deadlock. In this condition the controllers mimics a living drosophila and from time to time changes spontaneously its behavior trying to find a way out (or so it seems to an external observer). Under complex imagery lost and recuperation of icon frequently occur. In such cases the controller keeps a hard working, proactive attitude and, in our experiments, no continuous lost of the icon ever occurs.

In figure 6 a typical tracking sequence is shown. The image recognition agent has been deactivated and the controller has to track-servoing a moving circle which changes in size and have some associate noise. During the tracking the BIA

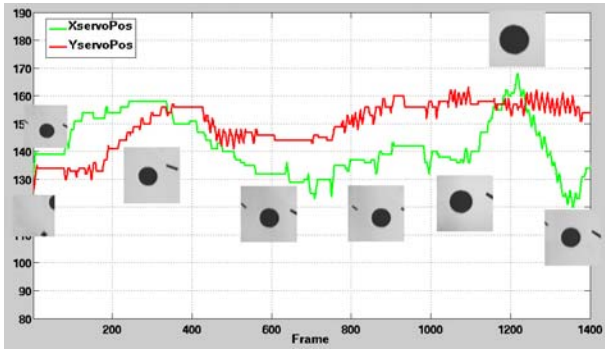


Fig. 6. The robotic eye controller realizes a tracking-servoing exercise by following a moving circle which changes in size and has some lateral noise. The two servo positions are generated by the BIA with the help of centering information coming from the ICA.

generates the x-y servos positions with the help of centering information coming from the ICA.

Figure 7 shows the x-y servo activities with the three participating agents fully activated. Icons move up to 10 cm/seg and rotate up to 60 rpm., images are processed at about 20 frames/sec. When an arrow is spotted servo behavior is modulated by centering information and tracking begins. If the arrow is lost servos change their behavior to a semi-random search which produces many spurious images.

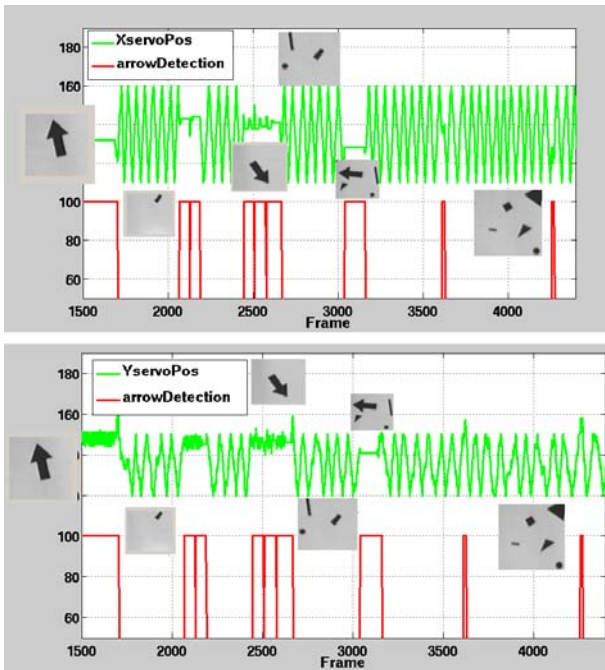


Fig. 7. Typical search-track sequence completed by the proposed neural machinery. When an arrow is spotted the BIA accomplish tracking. If the arrow is lost the BIA adopts a semi-random search. Many spurious images are found and processed during the sequence. Icons move up to 10 cm/seg and rotate up to 60 rpm.

V. CONCLUSION

The problem of tracking and recognizing a moving target has been solved by partitioning it in three simpler and

separated tasks that can be executed by three independent neural agents. Neural agents associated to those simpler tasks proved to be easier to train and show a good general performance. The proposed behavior initiating agent (BIA) provides and integrating low bandwidth channel where others specialized agents can contribute with relevant compressed information. As a result the obtained controller accomplishes an acute image related task, in a dynamical environment without requiring the intervention of higher control layers. As a disadvantage each decision taken by the controlling neural machinery remains stochastic variable so erroneous decisions will always be present. This behavior is shared with living organisms. On the other hand the BIA has the capacity to successfully keep on moving the robotic eye under very noisy conditions and for long periods. Under such circumstances spurious images are usually deadly traps for image-driven machinery. This enduring capacity appears because under appropriate modulation conditions the BIA keeps a dash of genuine spontaneity.

ACKNOWLEDGMENT

The authors would like to thank the Spanish Science and Technology Ministry for funding the experiments of this work through the R&D project CICYT DPI 2007-66156. The authors also like to thank UPM for the grant as Visitant Professor of the first author and also the Consejera de Educacin del a Comunidad de Madrid and the Fondo Social Europeo (FSE) for some of the authors PhD Scholarship.

REFERENCES

- [1] R. A. Brooks, "Intelligence without representation," 1991, no. 47, pp. 139-159. [Online]. Available: <http://citeseer.ist.psu.edu/old/brooks91intelligence.html>
- [2] H. P. Moravec, "Locomotion, Vision and Intelligence," in *Robotics Research, The First International Symposium*, M. Brady and R. Paul, Eds. Cambridge, Ma: MIT Press, 1984, pp. 215-224.
- [3] J. J. Henley and D. P. Barnes, "An artificial neuro-endocrine kinematics model for legged robot obstacle negotiation." 8th ESA Workshop on Advanced Space Technologies for Robotics and Automation, Nov 2004.
- [4] F. Dellaert and R. Collins, "Fast image-based tracking by selective pixel integration," in *ICCV 99 Workshop on Frame-Rate Vision*, September 1999.
- [5] M. G. Marcone, G. and L. L., "Eye tracking in image sequences by competitive neural networks," *Neural Processing Letters*, vol. 7, no. 3, pp. 133-138, 1998.
- [6] D. D. Lee and H. S. Seung, "A neural network based head tracking system," in *NIPS '97: Proceedings of the 1997 conference on Advances in neural information processing systems 10*. Cambridge, MA, USA: MIT Press, 1998, pp. 908-914.
- [7] L. D. Soltani, A. and W. Xiao-Jing, "Neural mechanism for stochastic behavior during a competitive game," *Sciences Direct, Neural Networks*, vol. 19, no. 8, 2006.
- [8] A. Maye, H. Ch., G. Sugihara, and B. B., "Order in spontaneous behavior," *PLoS One*, vol. 10, no. 1371, 2007. [Online]. Available: <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0000443>
- [9] G. V. Parasuraman, S. and B. Shirinzadeh, "Fuzzy decision mechanism combined with neuro-fuzzy controller for behavior based robot navigation," *Industrial Electronics Society, 2003. IECON apos;03. The 29th Annual Conference of the IEEE*, vol. 3, no. 2, 2003.
- [10] T. J. Prescott, M. F., G. K., H. M., and R. P., "A robot model of the basal ganglia: Behavior and intrinsic processing," *Sciences Direct, Neural Networks*, vol. 19, no. 1, 2006.

- [11] O. Chang, "Evolving cooperative neural agents for controlling vision guided mobile robots," *Proceedings of 2009 8th IEEE International Conference on Cybernetic Intelligent Systems, September 9TH-10Th, 2009, Birmingham University, UK*, 2009.
- [12] M. Woolridge and N. R. Jennings, "Intelligent agents: Theory and practice," *Knowledge Engineering Review*, vol. 10, no. 2, 1994.
- [13] J. B. Shackleford, "Neural data structures: programming with neurons - technical," *Hewlett-Packard Journal*, pp. 69–78, 1989.
- [14] G. Serpen, "Managing spatio-temporal complexity in Hopfield neural network simulations for large-scale static optimization," *Mathematics and Computers in Simulation*, pp. 279–293, 2004. [Online]. Available: <http://dx.doi.org/10.1016/j.matcom.2003.09.023>
- [15] D. Pomerleau, "Knowledge-based training of artificial neural networks for autonomous robot driving," in *Robot Learning*, J. Connell and S. Mahadevan, Eds., 1993, pp. 19–43.
- [16] R. J. Vogelstein, M. U., E. Culurciello, R. Etienne-Cummings, and G. Cauwenberghs, "Spatial acuity modulation of an address-event imager," in *Proceedings of the IEEE International Conference on Electronics, Circuits and Systems*. IEEE, 2004, pp. 207–210.