AUTONOMOUS CAMERA CONTROL BY NEURAL MODELS IN ROBOTIC VISION SYSTEMS

Tyler W. Garaas, Frank Marino and Marc Pomplun
Department of Computer Science, University of Massachusetts Boston
100 Morrissey Boulevard, Boston, MA 02125-3393, U.S.A.

Keywords: Robotic Vision, Neural Modeling, Camera Control, Auto White Balance, Auto Exposure.

Abstract: Recently there has been growing interest in creating large-scale simulations of certain areas in the brain. The areas that are receiving the overwhelming focus are visual in nature, which may provide a means to compute some of the complex visual functions that have plagued AI researchers for many decades; robust object recognition, for example. Additionally, with the recent introduction of cheap computational hardware capable of computing at several teraflops, real-time robotic vision systems will likely be implemented using simplified neural models based on their slower, more realistic counterparts. This paper presents a series of small neural networks that can be integrated into a neural model of the human retina to automatically control the white-balance and exposure parameters of a standard video camera to optimize the computational processing performed by the neural model. Results of a sample implementation including a comparison with proprietary methods are presented. One strong advantage that these integrated sub-networks possess over proprietary mechanisms is that 'attention' signals could be used to selectively optimize areas of the image that are most relevant to the task at hand.

1 INTRODUCTION

Recent advances in neuroscience have allowed us unprecedented insight into how assemblies of neurons can integrate together to establish complex functions such as the deployment of visual attention (Moore & Fallah, 2004), the conscious visual perception of objects (Pascual-Leone & Walsh, 2001), or the remapping of visual items between eye movements (Melcher, 2007). This accumulation of detailed knowledge regarding the structure and function of the individual neurons and neural areas that are responsible for such functions has led a number of neuroscientists to prepare large-scale neural models in order to simulate these areas. Some of the modeled areas include the primary visual cortex (McLaughlin, Shapley, & Shelly, 2003), the middle temporal area (Simoncelli & Heeger, 1998), or an amalgamation of areas (Walther & Koch, 2006). Although the usual motivation for creating such models is to ultimately make predictions about their possible mechanisms or functional roles in biological organisms, recent advances in parallel computing – in particular, the introduction of the Cell processor as well as the graphics processing unit (GPU) – will likely direct the attention of robotics researchers toward developing comprehensive neural models for use in robotic applications.

Robotic vision systems that are based on biologically inspired neural models represent an initially promising path to finally achieving intelligent vision systems that have the power to perform the complex visual tasks that we take for granted on a daily basis. A classic example is that of object recognition, at which computer vision systems are notoriously poor performers. Humans, on the other hand, can quickly – on the order of hundreds of milliseconds – and effortlessly recognize complex objects under a variety of situations – e.g., various lightning conditions, rotations, or levels of occlusion. Various models of how this processing may occur in humans have been proposed, which have resulted in increased object recognition abilities by artificial systems (e.g., Riesenhuber & Poggio, 1999; Walther & Koch, 2006). Consequently, it is likely that subsequent iterations of these models will make their way into future robotic vision systems.

W. Garaas T., Marino F. and Pomplun M. (2009).
AUTONOMOUS CAMERA CONTROL BY NEURAL MODELS IN ROBOTIC VISION SYSTEMS.
In *Proceedings of the 6th International Conference on Informatics in Control, Automation and Robotics - Robotics and Automation*, pages 305-311
DOI: 10.5220/0002217303050311

Most neural models of visual areas operate in an idealized space (e.g., Lanyon & Denham, 2004); receiving pre-captured and manipulated images, whereas in robotic vision systems certain constraints are necessarily imposed. These constraints largely revolve around the need to process information in realistic time-frames – optimally real-time – as well as to interact directly with the physical world; likely through some form of video camera. Since these models will operate on input that cannot be known ahead of time, the system should be designed to handle a wide range of situations that may arise.

Most video cameras suitable for a robotic vision system include some ability to automatically monitor and adjust white-balance, exposure, and focus. However, in a robotic vision system that employs large-scale neural models, these automatic functions may lead to suboptimal processing conditions or even conflicts with the neural mechanisms. This paper presents a proof-of-concept method for manually controlling certain parameters in the camera to optimize the processing of a neural model of the retina, which will likely form the initial processing stage of future biologically inspired vision systems. In particular, the control of white-balance (WB) and exposure parameters are considered. Implementation details and a comparison to proprietary methods are given in the following sections.

2 SYSTEM OVERVIEW

The vision system presented here consists of a number of simple neural layers (2D layout of neurons that process image signals from nearby neruons) interconnected to form the basis for a robotic vision system; Figure 1 gives a simplified illustration of how the neurons within each layer are connected. The layers are modeled after a subset of the neurons present in the human retina. Figure 2 illustrates the individual neuron types and connections that constitute the artificial retina, which are briefly described below in order to establish the motivation for the WB and exposure sub-networks (subnets) presented hereafter. The neural model (excluding WB and exposure subnets) consists of 17 layers which total to approximately 225,000 individual neurons and 1.5 million connections. The network is designed to be executed on the GPU of a standard video card.

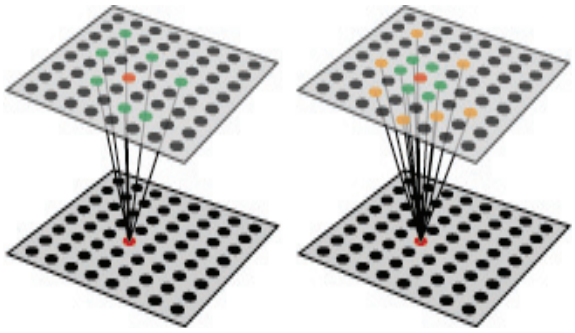

Figure 1: Simplified illustration of the connections used in the neural model and camera control subnets: (left) random connections to cells in previous layers and (right) random connections demonstrating a center-surround organization. Information in both cases flows from the upper layer to the bottom layer.

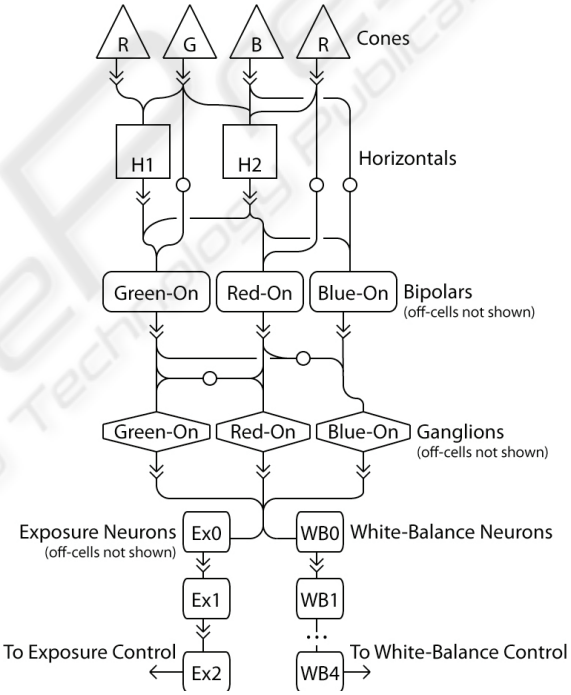

Figure 2: Connections between the various neuron types in the retina and camera control subnets. Solid arrows indicate excitatory connections while arrows with a white circle indicate inhibitory connections; ellipses between WB1 and WB4 indicate a continuation of the connection pattern directly above.

2.1 Apparatus

The network presented hereafter was simulated on two computer graphics cards (Nvidia 380 gtx) using an SLI setup. The OpenGL shading language (GLSL) was used to implement the computations for individual neurons. Video input was retrieved using

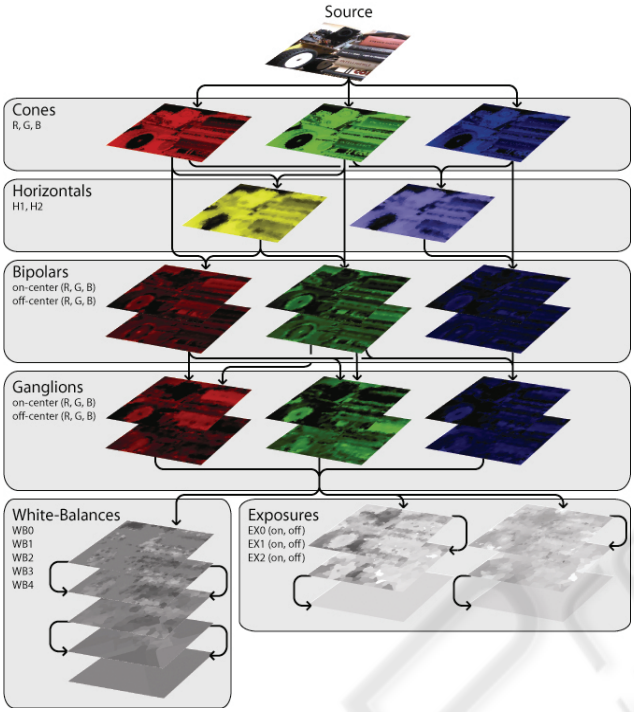

Figure 3: Activation maps and connection structure of the neural model and camera control subnets. Lighter areas represent higher activations while the colors indicate the spectral contributions to the activations.

a Cannon VCC4 video camera, and images were captured from the camera using the Belkin Hi-Speed USB 2.0 DVD Creator. Activations of the entire network can be computed very quickly: 100 iterations of computing activations for every single neuron take approximately 0.5 seconds.

2.2 Simulated Retina

The human retina is often considered a simple means for sensing light that enters the eye. On the contrary, the retina is actually a complex extension of the brain that is responsible for both reducing the amount of information transmitted to the various visual centers of the brain and converting the incoming signal into a form that is suited for higher-level processing by cortex. In the neural model presented here, we simulate the cones (R, G, B; referred to as long-, medium-, and short-wavelength cones in biological organisms, which are responsible for extracting the contributions of three primary color-components of the image), the horizontal cells (H1 & H2 cells, which essentially compute a 'blurred' version of the incoming image), on- and off-center bipolar cells (R, G, B cells, which compute an initial contrast-sensitive activation due to the antagonistic center-surround arrangement), and on- and off-center ganglion cells (R, G, B cells, which also compute a center-surround, contrast-sensitive signal that is also spectrally opposed, due to the inhibitory connections from bipolar cells). For the sake of brevity, we do not describe the specifics of individual neuron activations and connections. However, the essentials of the retinal neurons simulated here follow very closely those laid out by Dacey (2000) and Dowling (1987).

3 WHITE-BALANCE CONTROL

WB control in cameras was included so that changes in illumination could be countered to keep white areas within an image looking white. For instance, lighting that is stronger across the red spectrum of visible light will cause white areas to take on a reddish hue. Many different algorithms, such as white point estimation (Cardei, Funt & Barnard, 1999), chromaticity estimation using neural networks (Funt & Cardei, 1999), and gray world (Buchsbaum, 1980), have been proposed to control for changes in color due to the infinite spectrum of light sources. Although humans do not have the ability to directly control the color of objects as they are being received by the various early visual areas, neural mechanisms do exist to counter the effect of

illuminants on the actual *perception* of color (Brainard, 2004). This ability is aptly referred to as color constancy.

The automatic white-balance mechanism described here is a subnet of the neural model portrayed above. The basic goal is largely the same as that of previously proposed mechanisms; that is, to make white objects project white color onto the incoming image regardless of the illumination color. As such, the proprietary automatic white-balance mechanism would provide an adequate means to achieve this; however, there are a few caveats that may make a specifically designed WB control mechanism desirable. First, ganglion cells, from which the WB function will be computed, do not directly encode the primary image colors (i.e., RGB); instead, they encode a spatially and spectrally opponent signal that encodes the differences between red/green and blue/yellow signals. This property may introduce differences between an optimal white-balance parameter set by proprietary mechanisms and the optimal white-balance parameter for network computation. Second, certain biologically inspired mechanisms may take advantage of having the computation of such things implemented directly inside the network. This will be discussed in detail later.

The WB subnet introduced here is conceptually very simple. It begins by including a layer into the network (WB0) that 'extracts' areas of the image that represent candidates for white or light gray regions (technically, B/Y – R/G neutral). The candidate areas are exactly those areas in which the on-center ganglion cells have nearly the same level of activation and where the sum of the activations is greater than some threshold. A small amount of programming code is given below which gives a basic idea of how neurons' activations in layer WB0 are computed; red, green, and blue variables store the average activations of incoming red, green, and blue on-center ganglion cells, respectively; on-center activations range from 0.0 (no activation) to 1.0 (full activation).

```
float intensity = red + green + blue;
float R = red / intensity;
float G = green / intensity;
float B = blue / intensity;

if(R > 0.25 && G > 0.25 && B > 0.25 &&
 total > 1.0)
   activation = (B - R)*4.0;
else
   activation = 0.0;
```

After layer WB0 has extracted the areas that are potentially white or light gray, neurons in WB1 then compute a local maximum of the WB0 neurons to which it is connected; Figure 1 (left) illustrates the basic connection structure. Finally, layers WB2 through WB4 perform a simple averaging of the neuron activations from incoming layers; however, only neurons with non-zero activation (i.e., those representing a candidate area) will contribute to the average. The end-product of the WB subnet is a value that can be used to step the white-balance parameter either towards a more blue hue or a more red hue depending on the situation. If, for instance, the activations of red and blue ganglion cells are close to equal across the image, the step functions will be zero and the white-balance parameter will not change. However, if red ganglion cells have larger activations, in general, then the stepping function will be negative, which will cause the camera to introduce a slightly bluer hue to the image.

The WB subnet was designed to balance the activations across red and blue on-center ganglion cells. Consequently, the subjective view of the image cannot be used to assess the performance of the subnet, which is contrary to proprietary mechanisms. With that said, the WB subnet adjusts

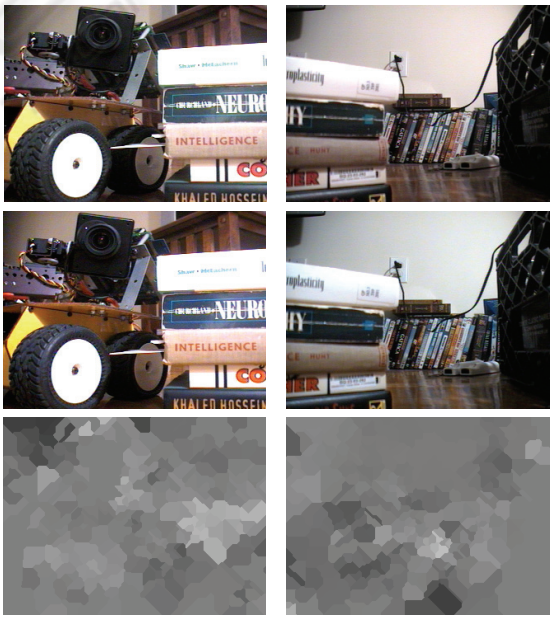

Figure 4: White-balance results: (top) adjusted image using proprietary auto-WB mechanism, (middle) adjusted image using the WB subnet, and (bottom) activation map of WB2. Lighter portions in WB2 represent candidate areas that contain greater activations of blue ganglion cells, while darker portions represent candidate areas greater activations of red ganglion cells.

the WB of the camera in much the same way as the proprietary mechanism in certain situations; see Figure 4 (left), for example. In contrast, other situations can produce deviations in WB settings between the subnet and proprietary mechanisms; see Figure 4 (right), for example. The size of WB steps should also be considered, as too large a step size will introduce over-correction and, ultimately, a ping-ponging of the WB parameter as the subnet slowly narrows in on the correct value; on the other hand, too small a step size will lead to a very slowly adjusting WB. Finally, in the current network, following a change in the white-balance parameter, it was necessary to insert a short delay before another step could be made; this was needed to allow the changes in image color due to the WB parameter change to spread through the various neural layers.

4 EXPOSURE CONTROL

One of the most remarkable properties of the human vision system is its ability to function over a strikingly large range of luminance conditions, a span of approximately 10 billion to 1 (Dowling, 1987). The human eye has essentially two ways of dealing with the variation it experiences in day-to-day luminance levels. (1) The pupil can reduce its area by a factor of approximately 16 due to changes in ambient illumination. (2) The circuitry in the retina is specially designed to handle two general lighting conditions: dim light, primarily handled by the rod-pathway in the retina; and bright light, handled by the cone-pathway in the retina.

Video cameras, on the other hand, do not have the luxury of such robust input mechanisms. Nevertheless, various methods have been developed to allow cameras to function under a rather impressive span of luminance levels – at least when all things are considered. The camera used for the present study employs two primary parameters that can be adjusted to compensate for luminance levels: iris size and gain control.

The network control of exposure is similar to that of WB in that a conceptually simple subnet progressively computes various properties of the incoming image which allows it to 'step' the relevant parameter towards optimizing some computation. The computation that is optimized in exposure control is contrast; too much light entering and the image gets 'washed-out'; too little light creates an underexposed image. This mechanism in particular will likely be very important to robotic vision systems using biologically neural models, as contrast has shown to play a particularly critical role in the neural computations that take place in the primate visual cortex (Sceniak et al., 1999).

As with the WB subnet, the functioning of the exposure subnet is conceptually very simple. Essentially, the subnet attempts to maximize the contrast of two spatially adjacent areas using the on- and off-center ganglion cells. Recall that in the neural model of the retina (and the biological retina) contrast plays a specific role for two classes of neurons, bipolar cells and ganglion cells. That is, these cells compute an activation that highlights high contrast areas of the image. Consequently, much of the work required for computing our exposure control function is already implemented.

The remaining work is performed by two independent subnets, an off-subnet and an on-subnet. Each subnet first computes a local maximum of the incoming ganglion cell activations (on-center ganglion cells will have higher activations in bright areas, especially if it is adjacent to a dark area, and vice-versa for off-center ganglion cells). This maximum is then averaged across the image to produce an exposure step-value similar in nature to the WB step-value, with one difference, the step value for the exposure control must work to control both the iris and gain of the camera, which is handled by a simple scheme: changes to the iris take precedence over changes to gain, which instead serves to fine-tune the exposure using small step-values.

Sample results of the exposure subnet are shown in Figure 5 for both bright-light conditions (left) and dim-light conditions (right). Exposure values in dim-light conditions closely follow those computed by the proprietary control mechanisms. However, significant differences can be seen in bright-light conditions. This is likely due to the goal of the exposure subnet to extract a maximum contrast signal, which may only occur in a portion of the image, as opposed to enhancing a global contrast signal. This difference is most prominent around the wheel in Figure 5 (left) where the proprietary mechanism introduces spectral highlighting on the rubber (black) portion of the wheel; whereas when under the exposure subnet's control, the gain is weaker to allow the natural blackness of the wheel to maximize the contrast between the plastic (white) and rubber (black) portions.

5 CONCLUSIONS

The current paper introduced small subnets that can be integrated into biologically inspired neural models of human visual areas to control the WB and exposure parameters on most standard video cameras. In contrast to their usual functions, WB and exposure parameters are used to optimize the actual processing that occurs in the neural model, as opposed to simply providing a clearer image. The subnets of this particular implementation are based on the activations of on-center and off-center ganglion cells from a neural model of an artificial retina.

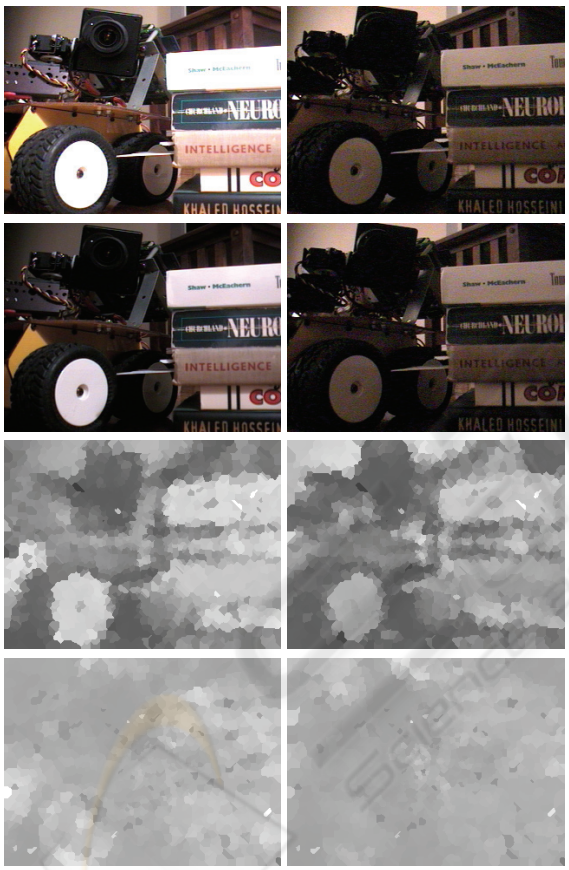

Figure 5: Exposure results: (top) adjusted image using the proprietary auto-exposure mechanism, (upper-middle) adjusted image using the exposure subnet, (bottom-middle) activation map of on-center EX0, and (bottom) activation map of off-center EX0.

Aside from customizing the parameters of the camera to optimize model computation, the subnets introduced here have other features that would make them a desirable replacement for proprietary mechanisms. One feature in particular could provide a substantial benefit, which is the ability to selectively optimize computation for areas in the image in which the neural model is 'interested'. Indeed, one of the most studied neural signals in biological organisms is that of attention, which is often implemented in artificial neural models (Lanyon & Denham, 2004). Consequently, with very little modification, the subnets presented here could be modified to selectively provide emphasis to attended areas based on incoming attention signals.

For instance, imagine a robotic vision system that is placed in a daylight setting receiving very bright light from the sun. If the robot wishes to examine a dark portion of the incoming image – say the lettering of a poster printed on a black background, proprietary mechanisms will be inadequate as they will selectively optimize the range of high pixel values – i.e., those representing bright areas. Instead, if the image is adjusted to optimize the range of low pixel values – i.e., those representing the poster, the robot may then successfully achieve its goal. Attention signals representing the robot's desire to inspect the poster would provide a perfect indicator by which to optimize the correct portion of the incoming image. Future implementations will be directed toward realizing such models.

REFERENCES

Brainard, D. H. (2004). Color constancy. In L. Chalupa & J. Werner (Eds.), *The Visual Neurosciences* (pp. 948-961): MIT Press.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of Franklin Institute*, 310, 1-26.

Cardei, V., Funt, B. & Barnard, K. (1999). White point estimation for uncalibrated images. *Proceedings of the IS&T/SID seventh color imaging conference*. (pp. 97-100). Scottsdale, AZ, USA.

Dacey, M. (2000). Parallel pathways for spectral coding in primate retina. *Annual Review of Neuroscience*, 23, 743-775.

Dowling, J. E. (1987). *The Retina: An Approachable Part of the Brain*. Cambridge, MA, USA: Belknap Press.

Funt, B. & Cardei, V. (1999). Bootstrapping color constancy. *SPIE Electronic Imaging '99*.

Lanyon, L. H. & Denham, S. L. (2004). A model of active visual search with object-based attention guiding scan paths. *Neural Networks*, 873-897.

McLaughlin, D., Shapley, R. & Shelly (2003). Large-scale modeling of the primary visual cortex: influence of cortical architecture upon neuronal response. *Journal of Physiology-Paris*, 97, 237-252.

Melcher, D. (2007). Predictive remapping of visual features precedes saccadic eye movements. *Nature Neuroscience*, 10, 903-907.

Moore, T. & Fallah, M. (2004). Microstimulation of the frontal eye field and its effects on covert spatial attention. *Journal of Neurophysiology*, 91, 152-162.

Pascual-Leone, A. & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, 292, 510-512.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019-1025.

Sceniak, M. P., Ringach, D. L., Hawken, M. J. & Shapley, R. (1999). Contrast's effect on spatial summation by macaque V1 neurons. *Nature Neuroscience*, 2, 733-739.

Simoncelli, E. P. & Heeger, D. J. (1998). A model of neuronal responses in area MT. *Vision Research*, 38, 743-761.

Walther, D. & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, 19, 1395-1407.