Surveillance Applications of Biologically-Inspired Smart Cameras

Kosta Haltis[1], Lee Andersson[1], Matthew Sorell[1], and Russell Brinkworth[2]

[1] School of Electrical and Electronic Engineering,
[2] School of Molecular and Biomedical Science
University of Adelaide SA 5005, Australia
russell.brinkworth@adelaide.edu.au

Abstract. Biological vision systems are capable of discerning detail and detecting motion in a wide range of highly variable lighting conditions. We describe the real-time implementation of a biological vision model using a high dynamic range video camera and a General Purpose Graphics Processing Unit (GPGPU) and demonstrate the effectiveness of the implementation in two surveillance applications: dynamic equalization of contrast for improved recognition of scene detail; and the use of biologically-inspired motion processing for the detection of small or distant moving objects in a complex scene.

Keywords: Surveillance, digital video processing, biological vision, motion detection, image enhancement.

1 Introduction

Flying insects have extraordinary visual capabilities that allow them to navigate in cluttered environments without collisions, perform spectacular aerobatic manoeuvres [1] and discriminate the motion of visual targets camouflaged within complex background textures (*visual clutter*). These are all challenging tasks for artificial vision systems that have attracted substantial attention from scientists and engineers.

One such challenge is discerning detail in high dynamic range images, given that typically only 8-bit (256 level) luminance is usually available in digital imaging. This presents itself as a limitation on the information content, by which we mean, specifically, the distinguishability of objects within an image, available for capture from a scene under difficult lighting conditions. High dynamic range (HDR) imaging processes allow capture over a much larger luminance range, however such images are not supported by the majority of image display and storage media. While there are conventional engineering solutions, such as gamma correction and tone mapping, we consider a biologically-inspired approach based on photoreceptor cells, whose purpose is precisely luminance range compression [2]. A model for this process has been developed, and allows the operation to be carried out on HDR images in software.

The photoreceptor model is well suited to parallel computation but is computationally expensive in a conventional serial computing environment. The model has therefore been implemented on a General Purpose computing Programmable Graphic

M. Sorell (Ed.): e-Forensics 2009, LNICST 8, pp. 65 – 76, 2009.

Unit (GPGPU). This allows the model to be applied to HDR images and the resulting images to be displayed in real time.

A user study has been performed to explore the performance of the photoreceptor HDR compression technique against typical linear scaling of the luminance range and subsequent gamma adjustment for display. This study concluded that the photoreceptor model is consistently able to retain equal or greater information content of scene through the biologically inspired range compression technique, particularly under complex lighting conditions.

This system has the potential for application within military and consumer imaging systems including surveillance, target detection, security monitoring, and face and text recognition software.

2 Biologically-Inspired Vision Model

Many conventional camera systems tend to have poor performance when capturing images of scenes with complex lighting conditions. They tend to either struggle to capture details in the darkest or brightest parts of the scene, and occasionally in both.

Fig. 1. Images captured with a conventional camera with different exposure settings and global gain adjustments

The reason for this is that these conventional cameras generally use a relatively simple global adjustment to the luminance on each frame as a whole, in an attempt to improve the final image output quality. However, in many cases the resulting image quality is poor. Conventional cameras usually can only capture an 8-bit range of luminance levels (256 possible values) despite the huge range of possible luminance levels that natural lighting conditions provide (in the order of 10^8 possible values) [3]. This is evident in Figure 1. Overall, the processing that is performed on each image is approximately linear across all the pixels within the image.

2.1 Spatial and Temporal Image Processing

Spatial image processing is a form of processing performed on each pixel based on surrounding pixels of the same image. For HDR images, some spatial processing techniques that are used include tone mapping and gamma correction.

"Tone mapping" is a technique that is used to map a set of colours to another set. This is useful when attempting to perform dynamic range compression [4].

"Gamma Correction" is a nonlinear operation used to increase the dynamic range of the luminance of pixels within an image [5]. A simple format of a gamma correction encoding is power encoding, which takes the form shown in Equation 1. This function is generally used with a gamma value 0.45-0.50 for gamma correction prior to displaying images on a computer monitor.

$$V_{out} = V_{in}^{\gamma} \tag{1}$$

The results of post captured image stitching with spatial processing applied on a couple of image can be seen in Figure 2.

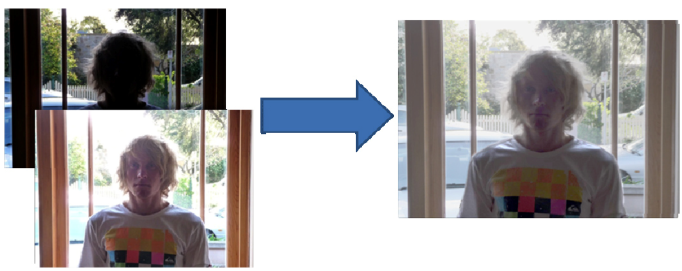

Fig. 2. High Dynamic Range image created by stitching spatially processed images

In terms of image quality, spatial image processing performs very well for images with both moving and stationary scenes. However, it requires largely iterative, and sometimes complex, calculations that depend on all surrounding pixels within the frame. This would make it difficult to perform real-time image processing on video footage using these techniques, especially if the frame resolution is large.

Temporal processing is the action of processing a pixel's new data value based on its previous states in time (that is, previous frames of a video sequence). Pure temporal image processing requires no knowledge of the surrounding pixels, and utilises the temporal characteristics for moving images for the processing.

The issue with temporal systems is that if there is no movement of an object within an image, the object will slowly fade to become less distinguishable from surrounding objects. Hence, the only way to constantly see detail in stationary objects is to continuously move the camera. This can be seen in the images in Figure 3.

The middle image is a snapshot in time after the camera had been stationary for several seconds. The right image is a snapshot in time during slow movement of the camera. It can be seen here that while the camera is moving, the markings on the cards and the card edges are much more defined.

In a system that performs spatial processing, the markings and card edges would be very visible, even in the stationary images. The reason that we can see stationary objects with our eyes is because our eyes are constantly jittering in fast but small amplitude motion known as *saccades*.

Fig. 3. Image Comparison between Raw (left), Temporal processing on stationary images (middle), Temporal processing on moving images (right). The left image has saturated luminance and the detail has faded and in the centre image. Constant small-scale motion in the right image refreshes the detail.

2.2 Photoreceptor Model

While biological vision systems vary in complexity and capability, our interest is in insect vision which incorporates temporal, but not spatial, processing in the photoreceptors [6, 7]. Insect motion vision can be modelled in three primary stages - Photoreceptors, Lamina Monopolar Cells, and Motion Processing.

Photoreceptors receive light through the optical system of the eye, applying non-linear dynamic range compression to achieve a high dynamic range of optical luminance. Photoreceptors are equivalent to pixels and act on a pixel-by-pixel basis and are discussed in more detail below.

Lamina Monopolar Cells in flies remove redundancy in both space and time in an optimal way based on the local light level [8]. Processing steps include variable and relaxed high-pass filtering in both space and time depending on light levels [9], signal amplification and a saturating non-linearity. The end result, as seen in Figure 4, highlights edges and areas of relative movement.

Motion Processing is used to calculate the motion of every pixel relative to the camera. This is modelled by the so-called Reichardt Correlator [10], which compares a pixel with a delayed signal from other surrounding pixels. This is done via a non-linear multiplicative interaction between the two channels. This motion calculation

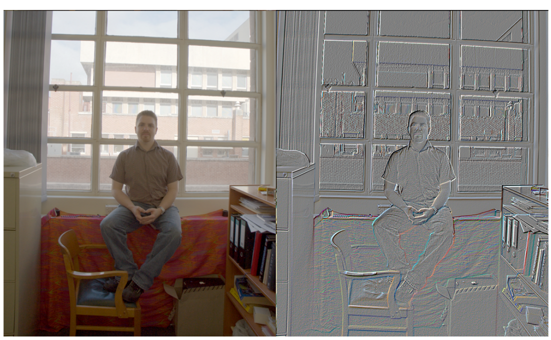

Fig. 4. Lamina Monopolar Cells in flies remove redundancy in both space and time, highlighting edges and areas of relative movement. The image on the left is a single frame after simulated photoreceptor processing; on the right is the output of the simulated LMC stage in which motion detail is retained.

relative to the visual system can then be used to perform velocity estimation of object within the field of view.

Photoreceptors

The photoreceptors provide the first stage of processing for the Biological Visual system. They are responsible for Dynamic Range Reduction of the input images. The biological eye is able to see a much larger dynamic range than can be linearly encoded by the photoreceptors [11].

Biological Visual Systems use complex Non-Linear Dynamic Range Reduction at the pixel level in an attempt to improve the final image output. This system dynamically adjusts the dark and bright areas of the images independently through temporal pixel-wise operations.

The dark pixels are brightened and the bright pixels are darkened to equalize the luminance throughout the whole image, hence reducing the dynamic range of the image.

A model for the Biological Photoreceptors, which can be seen in Figure 5, was proposed by Mah et al in [11] as an extended version of the model by Van Hateren and Snippe [12]. This photoreceptor model has been shown to closely resemble the system response of the fly's visual system and has good results on improving the image output. The model can be broken down into 4 individual stages:

- Stage 1 is a Low Pass Filter (LPF) with variable gain and corner frequency, to model varying adaptation speeds in different lighting conditions, acting to increase the information captured over a wide range of light intensities by reducing gain as intensity increases.
- Stage 2 is a Non-linear Divisive feedback via LPF, providing rapid short term adaptation of the photoreceptor response to rapid large variations in light intensity through logarithmic compression of the input.
- Stage 3 is an Exponential Divisive feedback loop via a LPF. This stage is responsible for shifting the operating range of the model. The LPF in this stage provides slow adaptation, to provide the longer term adaptation of the system to variations in light intensities.
- Stage 4 is Naka-Rushton transformation, where a constant is added to the input and the result used in a divisive feed forward operation. This stage provides a final global gain control for further amplification to the darker parts of the image.

The result of the full photoreceptor system is to compress the image from high dynamic range to a low dynamic range, by way of independent pixel gain control. This compression system also enhances the useful information capture in the process.

The full system effectively provides an approximate form of high pass filtering (HPF). Hence, the system is effective in scenes of temporal variations such as local or object movement, or through the use of saccadic like local motion of the image capture device.

The effect is that areas of change are emphasised with greater detail, while temporally stationary objects, such as background regions, fade to a common gain value as the variable gain of the system reaches steady state.

This temporal system has useful advantages over spatial based processing (which occurs primarily in the laminar monopolar cells) since the photoreceptor makes use of frame history to further improve images. Some improvements include reducing noise and filling in data that may be missing in any one frame of the image stream. An example of the output of the photoreceptor model can be seen in Figure 6.

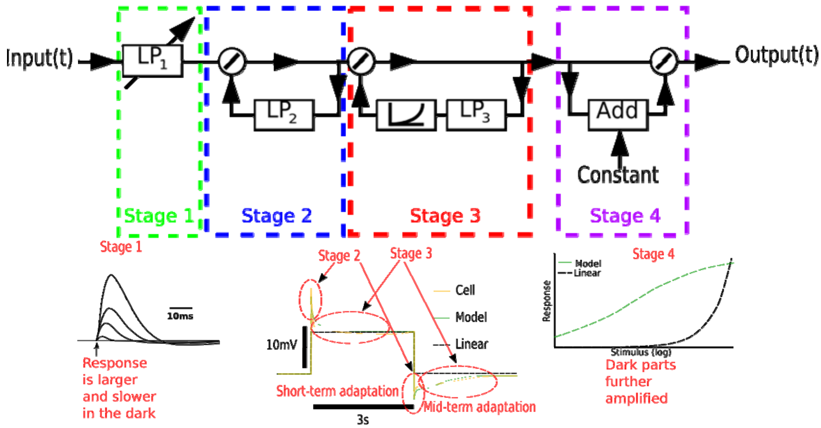

Fig. 5. Photoreceptor model block diagram and typical step responses of the model [3]

Fig. 6. Raw image (left) and improved detail after photoreceptor processing (right)

3 Implementation

The key opportunity for implementation of the proposed biologically-inspired vision system was the recognition that pixel-wise image processing could be computed in parallel using a parallel processing system designed for such graphics applications. The objective was to achieve real-time processing at the full image resolution (640 by 480 pixels) at full frame rate (25 or 30 full frames per second).

3.1 High Dynamic Range Camera

HDR images can be created using a technique called "image stitching" or through Analogue-to-Digital Conversion (ADC) techniques. State-of-the-art High Dynamic Range cameras incorporate an ADC up to 14 bits, providing a dynamic range of 16000:1 (84dB).

Image Stitching involves using multiple images of ideally the same frame at different exposure settings to generate a higher dynamic range for all pixels within the image. This technique uses these multiple images to generate HDR layers. Image stitching may be performed on within software after transferring the image layers, or in some cases within the camera's firmware.

The camera used in this work was the Basler A601f-HDR. This camera performs the image stitching on-chip post capture and delivers the 16-bit HDR image by firewire (IEEE1394) for software capture and processing.

3.2 General Purpose Graphics Processing Unit

A General Purpose computing Graphics Processing Unit (GPGPU) is a multi-processing core unit which is designed for computing many parallel, pipelined and often identical processes on very large segmented data sets. As the photoreceptor model fits this description precisely, it is an obvious candidate for GPGPU implementation, especially as software-only implementation on a dual-core Pentium processor only achieved frame rates in the order of 10 frames per second after significant code optimisation.

An NVIDIA GeForce 8800 GT processor was used for all GPU related processing requirements in our work, although any compatible GPGPU could be substituted. These processing cards are now quite cheap and are widely used for graphics-intensive applications such as video game play.

The GPGPU supports 14 multiprocessor blocks, each with 8 cores, representing a total of 112 processing cores. Each core is clocked at 600MHz, has local access to

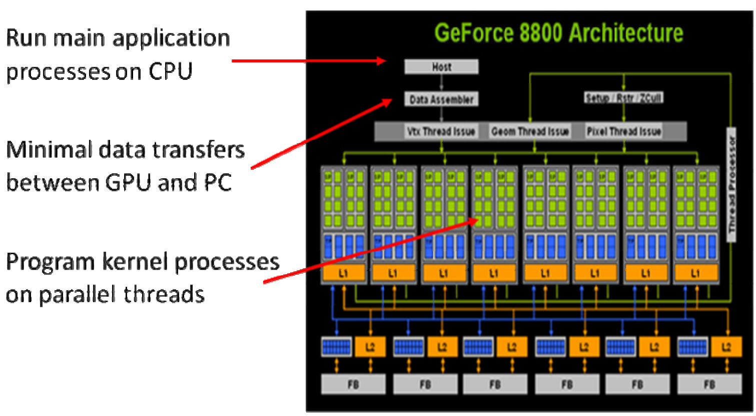

Fig. 7. The photoreceptor model was implemented on an NVIDIA GeForce 8800 GT General Purpose Graphics Processing Unit (adapted from [13])

cache memory and registers and global access to 512MB of memory. The global memory is accessible by the host CPU and has a transfer bandwidth of 57.6Gbit/s, more than sufficient to hold temporary frame data for real-time processing. The GPGPU operates as a co-processor, allowing parallel CPU management of data processing while the GPGPU implements the photoreceptor model in parallel. The architecture of the GPGPU implementation is shown in Figure 7.

3.3 Processing Bottlenecks

There are two threads running on the main CPU, in this case an Intel C2Duo dual-core processor at 2.8GHz. One thread handles image acquisition from the camera or hard drive, requiring 21.5ms. The second thread handles the complete image processing, including image scaling, gamma correction and display (19.5ms), and the photoreceptor model.

Implementing the photoreceptor model on the CPU requires 75ms, resulting in a frame period of 94.5ms, approximately 10 frames per second. Using the GPGPU requires just 8.5ms per frame, reducing the frame processing time to 28ms with a frame rate of 36 frames per second. It is therefore possible to process pre-recorded video faster than real time, and also meet the real-time requirements for direct processing from the camera. Further optimization, by scheduling the GPGPU to process in parallel with the CPU thread, could potentially increase the processing speed to 50 frames per second.

4 Applications and Performance Evaluation

Although there are many metrics of the efficacy of a surveillance camera, system or network, the ability to recognize objects and detail, and the ability to automatically detect motion are of particular interest and relevance in the case of surveillance applications of the described visual processing system. We therefore conducted an objective test to demonstrate the effectiveness of the photoreceptor-based processing for detail recognition, and a demonstration of the full model for motion detection of small objects in the field of view.

The former demonstration has direct applications in such areas as number plate pre-processing for automatic or manual character recognition and the identification of other objects under complex lighting conditions; the latter demonstrates the ability to pick out moving objects such as aircraft in the sky or people walking across a cluttered background.

4.1 Object and Detail Recognition

A test was conceived to compare the effectiveness of an 8-bit representation of an unprocessed video stream (except for global brightness control) with the photoreceptor based processing model for the application of detail recognition through the use of playing cards. Under complex lighting conditions of bright and dark areas with various levels of reflection, it is often difficult to determine the suit and the number of objects on a card, making cards a suitable basis for testing.

Five image streams that were captured using these playing cards. Each contained a different setup in terms of lighting conditions, the types of cards used, and the way in which they were displayed. The true configuration of the experiment was noted separately for comparison purposes.

Both the conventionally processed and the photoreceptor-processed image streams were saved as 8-bit JPG images and compiled into 10 individual tests (5 for each processing method). These images were then shown to a variety of people on different days, where they were asked to write down the card numbers and suits in the order that they were displayed in the images. These results were quantified to provide a measure of the efficacy of the processing for detail recognition.

Of the five sets, two resulted in 100% recognition under both unprocessed and processed conditions, and so were discarded. The fifth test included cards which were too far away to clearly identify the faces even under optimal lighting conditions due to the low resolution of the card features, and so are also not considered here. Of the two tests with meaningful results, Test 1 consisted of 20 tests and Test 2 of 40 tests, namely the identification of the correct suit and number of 20 cards as in Figure 8.

As the images were available as a sequence of frames, subjects were permitted to work back and forth through the frame set to find the best possible representation of each card. It can be seen from the test results in Figure 9 that the amount of useful information that was obtained by the test participants varies depending on the lighting situation. If users were unable to identify the card they were to leave the answer blank.

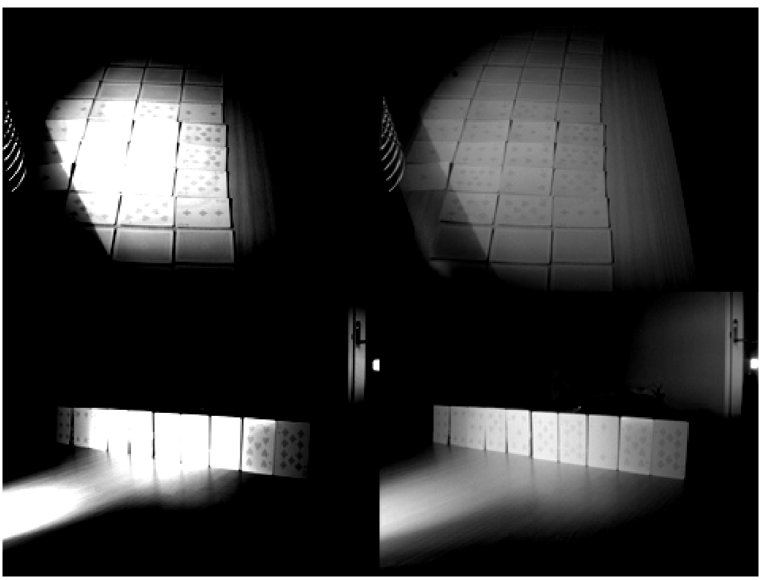

Fig. 8. Test configurations for visual comparison. Test 2 is above, Test 1 below. To the left are the unprocessed frames; to the right are frames after photoreceptor processing. Note that if global gain control were used, either the central cards would be in saturated luminance as shown, or the end cards would be too dark.

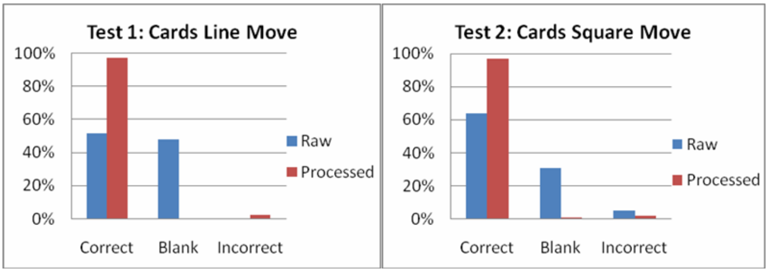

Fig. 9. Quality performance test results for both subjective recognition tests. In the raw images the no-response (blank) rate is high but drops to near zero after processing. The correct response rate after processing exceeds 95% (that is one or two errors) in each case.

Tests 1 and 2 consisted of a bright beam on the cards, causing luminance saturation on the faces of the cards. It appears as if the bio-inspired processing model has performed very well with a dramatic improvement in the number of correct readings. We conclude that under the appropriate resolution conditions, photoreceptor processing can lead to improved visibility of object artefacts in surveillance vision.

4.2 Motion Detection

Security surveillance relies on seeing as much information as possible within the cameras' fields of view. These camera systems would reduce issues with seeing details in shadowed or excessively bright areas within the fields of view at various times of the day; especially if the area under surveillance is exposed to direct sunlight.

Security surveillance could also benefit significantly from the additional stages of the Biological Vision Model (the LMC and Motion detection stages). Figure 10 shows

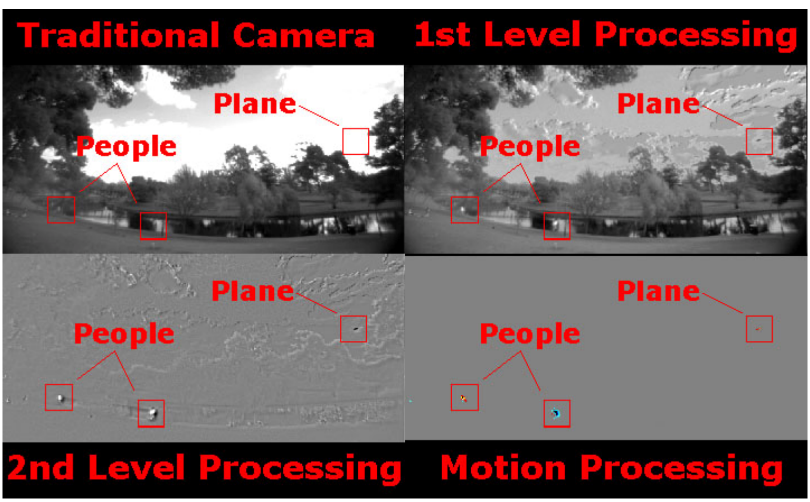

Fig. 10. Direct comparison of motion processing from each stage of a Bio-Inspired Camera System

a side-by-side comparison of the outputs of the Bio-inspired camera system prototypes at each of the Biological vision model. This image demonstrates how small objects relative to a scene can be monitored with ease. After photoreceptor luminance normalization (top right), it is possible to see objects both in the sky and on the ground simultaneously. The bottom left image shows highlighting of edges of objects with some movement within a scene. The bottom right corner shows motion detection within the scene. The intensity of these pixels represents the speed of the moving objects relative to the image capture device.

Second (LMC) and third (Motion Detection) level processing would still be useful on images retrieved from standard camera systems, however, for best results a HDR camera with the First level processing (Photoreceptor stage) will be needed as it is impossible to recover information lost when a linear global gain-controlled system is used. The complete system will ensure that useful information is maximised and the most accurate target detection and tracking will be possible. In the example images below, if the first level processing was skipped, the objects may not have been detected due to them travelling in either the darkest or brightest areas of the image. These areas are where the conventional camera system is more likely to fail in this application.

5 Future Work and Conclusions

We have demonstrated that it is possible to implement the photoreceptor stage of the biologically-inspired vision processing system using inexpensive off-the-shelf consumer graphics processing hardware, and provided some details of the surveillance applications of the proposed approach.

There are several areas of future work. High dynamic range cameras are not currently used in surveillance applications due to their relatively high cost, although there are emerging double-photodiode pixel based imaging sensors being developed according to Yamada [14] which would see a reduction in cost due to economies of scale. There is sufficient processing capability in the GPGPU in the developed system for implementation of Stage 2 and Stage 3 processing within the GPGPU environment, completing the implementation using commercially-available hardware.

Acknowledgements

This work was supported by a research grant from the Australian Research Council (LP0667744).

References

1. Land, M.F., Collett, T.S.: Chasing behavior of houseflies. J. Comp. Physiology A 89, 525–538 (1974)
2. Laughlin, S.B., Weckström, M.: Fast and slow photoreceptors — a comparative study of the functional diversity of coding and conductances in the Diptera. J. Comp. Physiology A 172, 593–609 (1993)

3. Brinkworth, R.S.A., Mah, E.L., O'Carroll, D.C.: Bioinspired Pixel-Wise Adaptive Imaging. In: Proc. SPIE, vol. 6414, pp. 641416-1–641416-7 (2007)
4. Ashikhmin, M.: A tone mapping algorithm for high contrast images. In: Proc. 13th Eurographics workshop on Rendering, pp. 145–156 (2002)
5. Sato, K.: Image-Processing Algorithms. In: Nakamura, J. (ed.) Image Sensors and Signal Processing for Digital Still Cameras, pp. 223–254. Taylor & Francis Group, Boca Raton (2006)
6. Payne, R., Howard, J.: Response of an insect photoreceptor: a simple log-normal model. Nature 290, 415–416 (1981)
7. Matić, T., Laughlin, S.B.: Changes in the intensity-response function of an insect's photoreceptors due to light adaptation. J. Comp. Physiology A 145, 169–177 (1981)
8. van Hateren, J.H.: A theory of maximizing sensory information. Biological Cybernetics 68, 68–70 (1992)
9. Juusola, M., Uusitalo, R.O., Weckstrom, M.: Transfer of graded potentials at the photoreceptor interneuron synapse. J. General Physiology 105(1), 117–148 (1995)
10. Hassenstein, B., Reichardt, W.: Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. Z. Naturforsch 11b, 513–524 (1956)
11. Mah, E.-L., Brinkworth, R., O'Carroll, D.: Implementation of an elaborated neuromorphic model of a biological photoreceptor. Biological Cybernetics 98, 357–369 (2008)
12. van Hateren, H., Snippe, H.P.: Information Theoretical Evaluation of Parametric Models of Gain Control in Blowfly Photoreceptor Cells. Vision Research 41, 1851–1865 (2001)
13. NVIDIA Corp.: NVIDIA GeForce 8800 GPU Architecture Overview, http://www.nvidia.com/object/IO_37100.html
14. Yamada, T.: CCD Image Sensors. In: Nakamura, J. (ed.) Image Sensors and Signal Processing for Digital Still Cameras, pp. 95–142. CRC Press, Boca Raton (2006)