Brightness Perception, Dynamic Range and Noise: a Unifying Model for Adaptive Image Sensors

Vladimir Brajovic
Carnegie Mellon University
brajovic@cs.cmu.edu

Abstract

Many computer vision applications have to cope with large dynamic range and changing illumination conditions in the environment. Any attempt to deal with these conditions at the algorithmic level alone are inherently difficult because: 1) conventional image sensors cannot completely capture wide dynamic range radiances without saturation or underexposure, 2) the quantization process destroys small signal variations especially in shadows, and 3) all possible illumination conditions cannot be completely accounted for. The paper proposes a computational model for brightness perception that deals with issues of dynamic range and noise. The model can be implemented on-chip in analog domain before the signal is saturated or destroyed through quantization. The model is "unified" because a single mathematical formulation addresses the problem of shot and thermal noise, and normalizes the signal range to simultaneously 1) compress the dynamic range, 2) minimize appearance variations due to changing illumination, and 3) minimize quantization noise. The model strongly mimics brightness perception processes in early biological vision.

1 Introduction

Natural scenes routinely produce high dynamic range (HDR) radiance maps that can span more than six orders of magnitude. Today's cameras fail to capture such a wide dynamic range because their image sensors can only measure about three orders of magnitude (about 10 bits per pixel per color). Consequently, saturation and underexposures in images are common. Vision algorithms that rely on such images inevitably have difficulty performing well; in fact, failure is often accepted as "a fact of life" when the scene dynamic range exceeds the capabilities of a sensor.

Biological eyes, on the other hand, cope very well with the wide dynamic range radiance maps from nature. This is done through a series of adaptive mechanisms for brightness perception. It is unlikely that natural brains encode perceived sensations with more than 6 bits of dynamic range.[2] How do we then handle and encode the dynamic range of stimuli that can span six or more orders of magnitude? The brightness perception is at the core of this process. It forms a low dynamic range representation of the perceived sensation from a high dynamic range stimulus in such a way to preserve perceptually important

Figure 1: A scene captured by a digital camera (top). The same scene as it may be perceived by a human observer (bottom). The top example is an 8-bit display of the 10-bit original image. The bottom example is an 8-bit display produced by our proposed brightness perception model based on the 10-bit original image.

This work has been supported by the National Science Foundation under Grant No. 0102272.

[2] This speculative statement is based on the Pierre Bourguer finding in the 1700s which states that the incremental threshold in the photopic state is 1/64 of the background stimuli, a value still accepted today.

visual features. A majority of these compression mechanisms are based on lateral processing found at the retinal level [8][1], but some are taking place in the early visual cortex. Figure 1 shows an example of how a camera may see a scene versus how a human observer may perceive it.

This paper presents a model of brightness perception suitable for an on-chip implementation that can result in an adaptive image sensor that receives high dynamic range (HDR) radiance maps and outputs low dynamic range (LDR) images in which all perceptually important visual features are retained. Needless to say, the model can also be applied numerically to HDR radiance maps to produce LDR images for rendering on conventional LDR devices (e.g., displays and printers). In the remainder of the paper it will become apparent that such a compression of the dynamic range largely removes appearance variations caused by harsh illumination conditions and reduces the visual feature degradation due to signal quantization. The model can also be successfully applied to minimize thermal and shot noise from the images. As such our model unifies the three issues: brightness perception, dynamic range and noise (thermal, shot and quantization).

In the remainder of the paper, we will first briefly review classical brightness perception models and prior approaches to HDR radiance map capture. Then we will introduce the mathematical model for image denoising, followed by an explanation of how it is used for dynamic range compression. We will conclude by illustrating how the model can be solved in an analog electronic network.

1.1 Brightness Perception as Tone Mapping Operator

The brightness perception in humans, extensively modeled by computational neuroscientists [12][14][29], determines how world radiance maps are perceived. Experimental studies reveal the existence of two distinct subsystems, one for contour extraction and another that assigns surface properties (e.g., lightness, brightness, texture) to regions bound by those contours. The center-surround processing in the retina gives rise to contour extraction at contrast discontinuities in the input image. At the higher level, regions are formed from local border contrast information by means of "spreading mechanisms" or "filling-in" [14]. This two-step model preserves discontinuities somewhat because the contrast discontinuities are detected first. Nonetheless, the "halo" artifacts (a.k.a. "inverse gradient") are exhibited around the boundaries due to the center-surround processing involved in the boundary computation. These halos actually model what is called "Mach bands" in human brightness perception. However, the halos are objectionable when contained in an image observed by a human.

The brightness perception relates HDR input stimuli into LDR percepts. As such, the brightness perception could be thought of as a tone-mapping operator. In computer graphics, the general problem of rendering HDR on low dynamic range devices is called *tone mapping* [38]. Tone mapping techniques are classified into global and local methods. Global methods specify one mapping curve that applies equally to all pixels. Local methods provide a space-varying tone mapping curve that takes into account the local content of the image. Excellent reviews on tone mapping techniques can be found in Refs. [10][30][9]. Let us review some closely related methods.

The main contributor to the creation of a HDR radiance map is the illumination field. In the presence of shadows the illumination field can induce a dynamic range of 100,000:1 or more. In the absence of wide illumination variations, the reflectance of the scene has a relatively low dynamic range, perhaps 100:1. Therefore, finding the reflectance is one way to compress the dynamic range of a HDR image.

In its most simplistic form, an image $I(x, y)$ is regarded as a product [16][17]:

$$I(x, y) = R(x, y)L(x, y)$$

where $R(x,y)$ is the reflectance and $L(x,y)$ is the illuminance at each point (x,y). Computing the reflectance and the illuminance fields from real images is in general, an ill-posed problem. Therefore, various assumptions and simplifications about L, R, or both are proposed in order to attempt to solve the problem. A common assumption is that L varies slowly while R can change abruptly. For example, homomorphic filtering uses this assumption to extract R by high-pass filtering the logarithm of the image [35]. Horn [16] assumes that L is smooth and that R is piece-wise constant. Then taking the Laplacian of the image's logarithm removes slowly varying L while marking discontinuities caused by the changes in R.

Of course in most natural images, the assumptions used in these examples are violated. For example, shadow boundaries on a sunny day will create abrupt changes in L. Under such conditions the homomorphic filtering would create a "halo" artifact (i.e., inverse gradient) in the recovered reflectance at the shadow boundary. The Horn's method would interpret the abrupt change in L as the change in R.

There are numerous other methods that attempt to compute the "local gain" in order to compress HDR images I to a manageable range and produce R. Estimating the local gain is analogous to estimating the illumination field since the local gain equals $1/L$.

Land's "Retinex" theory [21] estimates the reflectance R as the ratio of the image $I(x,y)$ and its low pass version that serves as the estimate for $L(x,y)$. The "halo" effects are produced at large discontinuities in $I(x,y)$. Jobson, et al [20] extended the Retinex algorithm by combining several low-pass copies of $I(x,y)$ using different cut-off frequencies for each low-pass filter. Since this combination retains some moderately high spatial frequencies, the estimate of

L(x,y) can better describe abrupt changes in *L*. This helps reduce halos, but does not eliminate them entirely.

In order to eliminate the notorious halo effect, Tumblin and Turk [39] introduced the *low curvature image simplifier* (LCIS) hierarchical decomposition of an image. Each component in this hierarchy is computed by solving partial differential equation inspired by anisotropic diffusion [28]. At each hierarchical level, the method segments the image into smooth (low-curvature) regions while stopping at sharp discontinuities. The hierarchy describes progressively smoother components of *I(x,y)*. *L* is then mostly described by smooth components (accounting for abrupt discontinuities) while *R* is described with components containing a greater spatial detail. Tumblin and Turk attenuate the smooth components and reassemble the image to create a low-contrast version of the original while compensating for the wide changes in the illumination field. This method drastically reduces the dynamic range, but tends to overemphasize fine details. The algorithm is computationally intensive and requires the selection of at least 8 different parameters.

Durand and Dorsey [9] use bilateral filtering to account for illumination field discontinuities. They call this estimate the "base level". Using bilateral filters [36] is a way to perform discontinuity-preserving smoothing. The pixels in the Gaussian filter kernel are ignored if the intensity of the underlying pixel is too far away from the intensity of the pixel in the center of the kernel. Similarly, Pattanaik and Yee [27] perform discontinuity-preserving filtering by only smoothing over pixels that are within the prescribed intensity range of the central pixel. All of these anisotropic smoothing methods for the estimation of *L* respect well-defined discontinuities, thus eliminating notorious halos. However, the small discontinuities in the deep shadows may be smoothed over, even though their relative strength may be as significant as the relative strength of the discontinuities in the bright image regions. In addition, performing anisotropic smoothing in a spatial domain is somewhat cumbersome. The bilateral filter has been suggested [9] as an alternative to anisotropic diffusion model [28] in order to overcome difficulties with convergence times and stopping criteria.

Fattal, et al. [10] takes a somewhat different, yet intuitive and elegant approach. To compress wide dynamic range images while preserving details, Fattal suggests amplifying the small gradients and attenuating the large gradients. Then, by integrating (i.e., solving an elliptic PDE), the resultant image whose gradients best approximate desired gradients (in the least square variational sense) is found. The halos are eliminated because there is never a division with a smoothed version of the original image. Fatall's method however, may have a slight problem with enhancing details in shadows. Major features of objects in shadows create gradients that are as small in the image as the minor details are in the brightly illuminated regions. To "pull up" the major object features from the shadows, the shadow gradients need to be amplified much more than small detail gradients in the bright regions. Since there is only a single mapping curve used for mapping the gradients, this cannot be achieved.

1.2 Capturing HDR Radiance Maps

All tone-mapping operators assume that the HDR radiance map is available. To capture high dynamic range radiance maps, many approaches have been proposed. Ref. [25] provides a good review of various techniques. We will only be illustrating the most common ones:

Multiple exposures: The easiest method is to use several images collected at different exposures [7]. This approach works well and is possible with conventional cameras, but is restricted to largely static scenes or requires optically aligned multiple cameras for simultaneous multiple exposure capture. Another approach is to create multiple exposures by introducing fixed optical attenuation (or exposure control) at pixel level [26]. A third method for acquiring multiple exposures of a point in a scene is through mosaicing, which uses a moving camera whose field of view is selectively attenuated [32].

Optical Contrast Masking: Another approach to capturing the HDR radiance map, is to spatially control the amount of light entering the camera lens by either using micro mirrors [6], or by using spatial light modulators [25]. These systems are somewhat complex since they require additional components and precise optical alignment. Also, the spatial resolution and image quality may be affected by the components that are imposed in the optical path. These approaches are reminiscent of contrast masking – a tone-mapping operator used in traditional photography where a negative contrast mask (a transparency) is first made by blurring the original photo. Then the contrast mask is superimposed in the optical path to spatially modulate the amount of light that projects the film negative onto the photographic paper.

Adaptive Imaging Sensors: Considerable effort has been made toward designing image sensors that will directly capture HDR radiance map. Various on-chip techniques have been proposed to capture wide dynamic range images [3][4][23][33][42]. Generally these methods create intelligent pixels that try to measure wide dynamic range at pixel level without regard for the local distribution of the radiance map impinging onto the sensitive surface. These sensors develop various on-chip signal representations that can encode wide dynamic range.

Neuromorphic Imaging Sensors (Silicon Retinas): The biological retina detects contrast boundaries (i.e., edges), among other image processing functions it performs. It has served as an inspiration for a class of imaging sensors loosely called silicon retinas, the first stage of many neuromorphic circuits[22][24]. One of the main issues being addressed by the silicon retinas is the robustness they provide to illumination and dynamic range

of sensed radiance maps. However Mead's Silicon Retina, the seminal work in this area, essentially implements Land's Retinex algorithm [21], and since Retinex suffers from halos, the silicon retina exhibits the same problem. The majority of silicon retinas built so far are contrast sensitive retinas, therefore, they implement some form of edge detection (i.e., high pass filtering) on the received image [34][2][40].

2 The Proposed Model

Our algorithm is motivated by a widely accepted assumption about human vision: human vision responds to local changes in contrast rather than to global stimuli levels. Having these assumptions in mind, our goal is to find the estimate of *L(x,y)* such that when it divides *I(x,y)* it produces *R(x,y)* in which local contrast is appropriately enhanced. In this view *R(x,y)* takes the place of the perceived sensation, while *I(x,y)* takes the place of the input stimulus. Then, *L(x,y)* is the "perception gain" that maps inputs sensation into the perceived stimulus, that is:

$$I(x,y)\frac{1}{L(x,y)} = R(x,y) \qquad (1)$$

With this biological analogy, *R* is *mostly* the reflectance of the scene and *L* is *mostly* the illumination field, but they may not be "correctly" separated in a strict physical sense. After all, humans perceive reflectance details in shadows as well as in bright regions, but they are also cognizant of the presence of shadows. A small part of the illumination field mixes with the scene reflectance and finds its way into our perceived sensation. This is important because shadows are used for interpretation of the scene. From this point on, we may refer to *R* and *L* as reflectance and illuminance respectively, but they are to be understood as the *perceived sensation* and the *perception gain* respectively. Durand and Dorsey [9] call these two quantities "detail" and "base" respectively.

Once the perception gain is computed (see below), the perceived sensation *R* is found according to Equation (1). To create a photo realistic display luminance I_d, the highlights are reintroduced by blending the perceived sensation *R* with the attenuated illumination *L* :

$$I_d(x,y) = \mathrm{S}\{R(x,y) + aL(x,y)\}, \quad 0 < a < 1 \qquad (2)$$

where *a* is the attenuation constant and **S** is a linear scaling operator that scales the result to the full signal range.

Figures 2 and 3 illustrate this process. The model first computes *L* from the original image, the details of which will be described in a moment, then this estimate of *L* divides *I* to produce *R*. Finally, *R* is blended with *L* to produce a photorealistic rendition with a compressed dynamic range.

Using Figure 2 and Figure 3, we can make several observations about our method. First, the details in the shadows are appropriately enhanced and brought into the comparable dynamic range with highlights for rendering on the LDR devices. Second, the perceived gain *L* respects discontinuities essential in eliminating halo artifacts. It is important to keep in mind that the above graphs are 1D scan lines from a 2D result. Both the pixels in the scan line direction as well as in the orthogonal direction played a role in determining how to smooth or not to smooth over various discontinuities. Finally, the absolute global ordering of intensities in the original does not hold in the result; however just like in human vision, the local relationships are properly maintained to deliver a visually appealing and photo-realistic result.

Figure 2: The original image I is decomposed into a) perceived sensation *R* and perception gain *L*. The highlights are introduced by blending

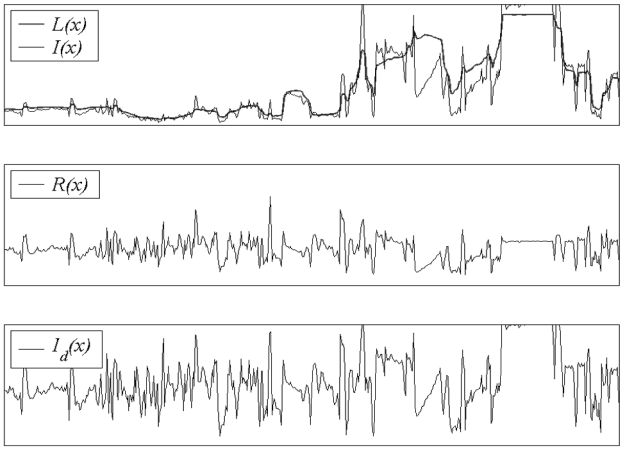

Figure 3: Horizontal intensity line profiles through the middle of images shown in Figure 2. The top graph shows the original image (thin line) and the computed perception gain (thick line). The middle graph shows the perceived sensation calculated by Equation (1). The bottom graph shows the output image formed by Equation (2).

2.1 Computing Perception Gain

The computation of the perception gain of *L* is at the core of our model. We seek to calculate an *L* that generally preserves large discontinuities in order to eliminate halos. We also seek a method that is simple

from both a numerical and a circuit implementation point of view. Finally, we seek a method that treats small discontinuities in the shadows as important as large discontinuities in the bright regions. Both describe equally important visual features—they are only different in the HDR radiance map because of the difference in illumination levels.

Discontinuity-preserving smoothing is analogous to image denoising; both methods seek to smooth the image without destroying underlying visual features. Therefore, we shall introduce the computation of the perception gain *L* by first considering the denoising aspects of the model. Then we will follow by showing how the same model applies to our goals for the dynamic range compassion.

2.2 Shot Noise and Thermal Noise

The most significant random noise sources in image sensors include shot noise and thermal noise[3]. Another point of concern, depending on the architecture of the sensor, is fixed pattern noise. The quantization noise is commonly excluded from consideration during the image sensor design, but it must be considered at the system level. Common ways of reducing thermal noise include improving the purity of the process, designing the pixel to minimize leakage currents, and cooling the entire sensor.

Our idea for denoising is simple, we will allow pixel signals to mix in smooth areas of the impinging radiance map while preventing mixing across discontinuities. In a sense, we have self-programmable pixel bins that are driven by the distribution of the radiance map. This is a relatively simple idea. What is more challenging is determining the necessary amount of signal mixing while still having a method that is realizable in a high-resolution image sensor chip.

We derive our "chip-friendly" denoising model as follows. Pixel measurement v_i is a random variable given by:

$$v_i = u_i + n_i^{th} + n_i^{sh} \tag{3}$$

where u_i is the true underlying pixel signal, n_i^{th} and n_i^{sh} are random thermal and shot noise respectively, and i is the spatial pixel index.[4] The thermal noise is considered normally distributed $n_i^{th} = N(0, \sigma_{th}^2)$. The shot noise follows Poisson distribution with an average arrival rate of u_i, which can be approximated with continuous normal distribution $n_i^{sh} = N(0, u_i)$.[5] Following Bayesian

[3] Common noise sources in image sensors include: shot noise, thermal noise, reset noise (if clocked), dark current noise (thermal and shot), fixed pattern noise, clocking noise and amplifier noise.

[4] We consider one-dimensional signal to simplify notational complications. The formulae directly extends to 2D.

[5] The mean of the Poisson distribution is u_i, but that is excluded from the noise distribution because it is captured in Equation (3) as the true underlying signal.

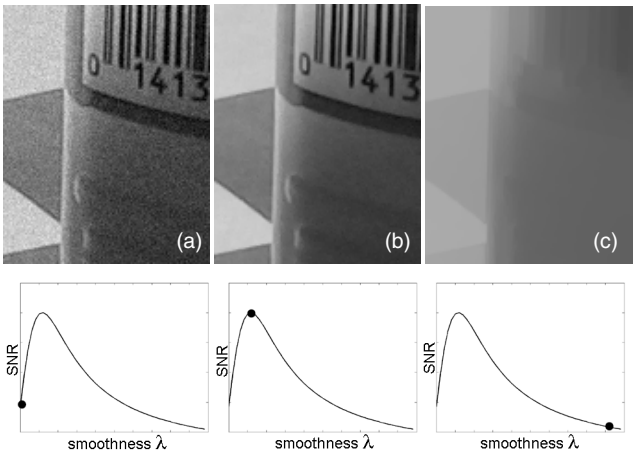

Figure 4: Result of denoising with our model applied to a window taken from Figure 1(top): Starting with a noisy image (a), the SNR is improved by increasing the smoothness parameter λ b). If the smoothness parameter λ is further increased, the SNR begins to degrade because the underlying structure of the signal begins to be destroyed; nonetheless, the major discontinuities are still preserved.

formulation, we seek to recover signal *u(x)* that maximizes the conditional probability distribution:

$$P(u \mid v) \propto P(v \mid u) P(u) \tag{4}$$

where *P(v|u)* is the probability distribution of observing the measurements *v,* given that the true underlying radiance map *u* impinges on the detectors. *P(u)* is our prior probability distribution describing the underlying radiance map as a Markov random field (MRF). Equation (4) is further written as:

$$P(u \mid v) \propto \prod_{i \in D} \exp\left(-\frac{(v_i - u_i)^2}{2(\sigma_{th}^2 + u_i)} \right) \exp\left(-\frac{|\nabla u_i|}{2\sigma_{RMF}^2} \right) \tag{5}$$

where *D* is the image domain and normalizing constants for probability distributions are omitted. The first term is the likelihood term that says that the measurement data are normally distributed around the true radiance map *u* with $(\sigma_{th}^2 + u)$ variance. The second term is an exponential distribution prior for the total variation smoothness over the RMF first-neighbor cliques. This prior imposes the total variation constrain over radiance map field *u* [31]. The total variation prior is chosen because it preserves discontinuities [31][5] and, as we will see in a moment, it allows for a convenient electronic analog network implementation. The total variation prior derives motivation from robust statistics [19] as to exclude gross outliers (large discontinuities in our case) from the estimation process. The optimal *u* is found by minimizing:

$$\sum_{i \in \mathbf{D}} \frac{(v_i - u_i)^2}{2(\sigma_{th}^2 + u_i)} + \frac{|\nabla u_i|}{2\sigma_{RMF}^2} \tag{6}$$

After some manipulation and introduction of more convenient notation $\lambda = 1/\sigma_{RMF}^2$ and $\alpha = \sigma_{th}^2$ our goal is to find *u(x)* that minimizes:

$$J(u) = \sum_{i \in D} (v_i - u_i)^2 + \lambda (u_i + \alpha) |\nabla u_i| \qquad (7)$$

Compared to classical total variation denoising, our model is peculiar in that it contains the multiplication of the total variation prior with the signal estimate *u*. It is worth pointing out that other robust statistics prior could be used in lieu of the total variation. The key is that in our model the signal *u*, or (*u*+α) multiplies such a prior.

Equation (7) can be numerically minimized using the gradient descend method. Figure 4 shows the results of denoising with our model. We see that with appropriate selection of the smoothness parameter λ, the model improves shot and thermal noise. If the parameter λ is further increased, the model produces a result that can be used as the estimate for the perception gain *L* for dynamic range compression (see Section 2.3 below).

2.3 Dynamic Range Compression

Recalling the example shown in Figure 4, it is clear that our model can compute the smooth version of the original image to serve as the estimate for the perception gain *L*. Since the model preserves discontinuities, the perceptual halos are not present in the perceived sensation R. The fact that our model essentially uses Weber-Fechner contrast to preserve discontinuities has an additional intuitive justification when used for the recovery of *R*. Namely, the small discontinuities in shadows are not necessarily "small". They only appear small in the radiance map because they are attenuated by the absence of strong illumination. Weber-Fechner contrast is a relative measure of contrast so that the small changes in shadows can "compete" with large changes in bright regions and not be smoothed over. This feature is absent from approaches proposed in the past.

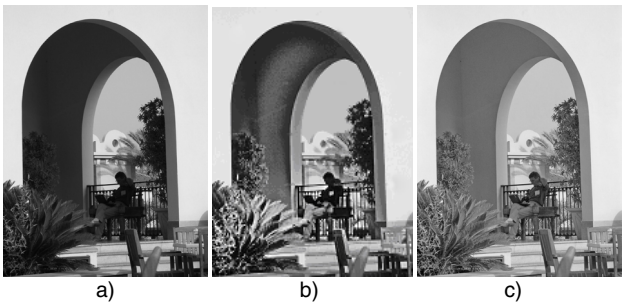

Figure 5: Comparison of Multi Scale Retinex with Color Restoration [20] to our proposed method. (a) Original image, (b) Retinex, and (c) our method. Large halo can be observed in (b) on the shadowed wall.

Recently the Multi Scale Retinex algorithm has been proposed as an improvement that minimizes halos [20]. Figure 5 compares our dynamic range comparison method to the Multi Scale Retinex. It is obvious that unlike Retinex, our method does not generate objectionable halos.

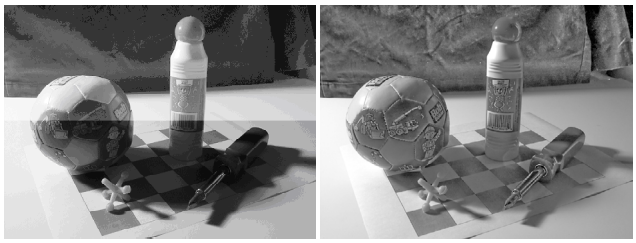

Figure 6: Quantization noise example: (a) An original 10-bit image is quantized to 8 bits (same as the example in Figure 1). Then the signal in the top half of the image is amplified to reveal the signal quality in dark fabric drape in the background. We see that due to quantization to 8 bits the signal details are degraded exhibiting the significant posetrization in the image. (b) When our algorithm normalizes the 10-bit radiance map before the quantization to 8 bits, the details of the folds in the background drape are much better preserved.

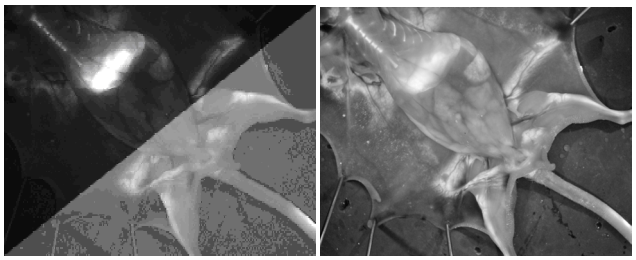

Figure 7: Another quantization noise example: a) 12-bit radiance map of a laboratory animal injected with a fluorescent dye is quantized to 8-bits for display. The important spatial variations in the liver (the brightest blob) as well as all the details in the shadows (amplified in the lower righ corner) are lost due to quantization. If the 12-bit radiance map is first normalized with our algorithm and then quantized to 8-bits for display, all perceptually important details survive the quantization process.

2.4 Quantization Noise

Quantization noise is caused when an analog signal is quantized into a discrete quantization level. Small signal variations that fall within a quantization bin are lost thus causing the quantization noise. It is too late to combat the quantization noise after the quantization is done. From the quantization signal-to-noise ratio (SNR_q) point of view, it is very important that the signal amplitude be carefully matched to the full-scale amplitude of the quantizer (i.e., A/D converter). The quantization noise is inevitable. Traditionally, when the quantization noise is analyzed, there is an implicit assumption that the signal variations are well matched to the input range of an A/D converter. For images, this assumption is largely untrue. By referring to the top graph of Figure 3, we see that the variations of the signal in the shadows are so small that they will be quantized with only a few quantization levels. This means that this part of the signal will undergo severe degradation due to the quantization and have very bad SNR_q. On the other hand signals in the middle and bottom graph of Figure 3 exhibit signal variations that are better matched to

the A/D's full range. As such, the important spatial information contained in the image will span many quantization levels and will be quantized with much better SNR_q. Figure 6 illustrates this point for the dark regions.

The quantization noise is not only visible in the dark region, but is generally present throughout the image. Figure 7 illustrates a case where the signal details describing small transitions in the bright regions are lost due to quantization noise. Our method again normalizes the dynamic range to preserve those perceptually important details.

Increasing the number of bits for representing the intensity pixels can reduce quantization noise. Indirect methods for filtering some of the quantization noise could involve averaging of multiple exposures [13] as to increase the bit resolution. Our method excels in retaining visual details within limited bit resolution of the quantizer (e.g., 6-8 bits).

2.5 Illumination Invariance

Since our method takes large signal variations out of the raw HDR radiance maps while retaining important visual details, it can be used to remove wide illumination variations from images. Figure 8 shows a face imaged under different illumination directions. The perceptual sensation R computed by our method largely removes these variations. When a vision algorithm interprets such preprocessed images, the algorithm may be more robust to illumination variations (see Figure 9).

Figure 8: A face taken with three different illuminations: original images (left), and perceived sensation R computed by our method (right).

Figure 9: An example of simple image processing in the presence of illumination-induced variations: Edge detection on original images of Figure 8 (left), and edge detection on the preprocessed images (right). The illumination variations are reduced making image processing algorithms operate more reliably regardless of illumination conditions.

3 Analog Network Solver

Equation (7) can be rewritten as:

$$J(u) = \sum_{i \in D} (v_i - u_i)^2 + \lambda \frac{1}{R_{hi}} |\nabla u_i|^2 \qquad (8)$$

$$R_{h_i} = \frac{|\nabla u_i|}{(u_i + \alpha)} \qquad (9)$$

The L_1-norm in the total variation prior in Equation (7) is rearranged and replaced by the quadratic in Equation (8), while taking the rearrangement into account in *Rh*. of Equation (9). Equation (8) is the expression for the power dissipated by the resistive grid shown in Figure 10. The first term is the power dissipated on the vertical resistors Rv, while the second term is the power dissipated on the horizontal resistors *Rh*. The voltage sources v_i represent the pixel measurements. Since the network will settle to a minimum entropy state, which is formally stated by Maxwell's minimum heat theorem, the nodal voltages u_i assume values that minimize Equation (8). Therefore, the network of Figure 10 finds our denoised image u_i.

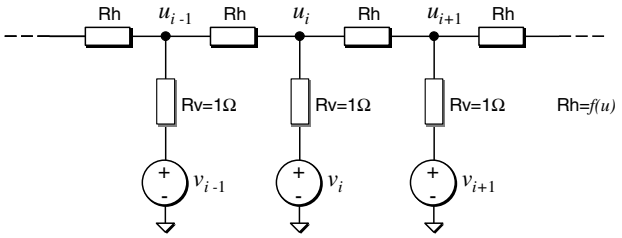

Figure 10: Resistive grid that minimizes energy in Equation (8), therefore finding the denoised image u from noise inputs v. The horizontal resistors are controlled with the local Weber-Fechner contrast, thus preserving the discontinuities in the image.

It is interesting to note that the values of the horizontal resistors *Rh* given in Equation (9) follow local Weber-Fechner contrast established in experimental psychology more than a century ago [41]. Weber's-Fechner's contrast is given as $C_w = |\nabla u| / u$. An alternative form that also models the contrast perception at isotopic state (e.g., very low background illuminations) is $C_w = |\nabla u| / (u + \alpha)$ where the constant α, usually small, represents a baseline level of activity that must be surpassed. This is exactly the expression for the horizontal resistors in our denoising network that determines which discontinuities should be smoothed over, and which should not.

The inability of uniform resistive grids to properly account for discontinuities has been recognized for some time. In place of horizontal resistors, so-called resistive fuses have been proposed [15]. The idea is to "break" neighbor-to-neighbor connections if the difference among neighbors exceeds a predetermined threshold. Breaking local links has been suggested for stochastic image restoration [11]. Unlike our resistive grid, the resistive grids proposed for image segmentation and processing thus far treat small discontinuities in shadows the same as small discontinuities in the bright regions. According to our earlier argument this may not be the best thing to do in the presence of wide illumination variations.

4 Conclusion

We have introduced a simple computational model that unifies image denoising, dynamic range compression and brightness perception. The model can be solved on a relatively simple analog network, promising to result in an adaptive image sensor that can handle wide dynamic range scenes.

References

[1] Kolb, Helga, "How the Retina Works", *American Scientist*, 91(1), January-February 2003.

[2] K. A. Boahen and A. Andreou, "A contrast-sensitive retina with reciprocal synapses," in J. E. Moody, ed., Advances in Neural Information Processing Systems 4, San Mateo, CA: Morgan Kaufman, 1991.

[3] Brajovic, V., Miyagawa, R., and Kanade, T., "Temporal Photoreception for Adaptive Dynamic Range Image Sensing and Encoding," *Neural Networks*, 11(7-8), 1149-1158.

[4] Brajovic, V. and T. Kanade, "A VLSI Sorting Image Sensor: Global Massively Parallel Intensity-to-Time Processing for Low-Latency, Adaptive Vision," IEEE Transactions on Robotics and Automation, 15(1), February, 1999, pp. 67-75.

[5] T. F. Chan and J. Shen. "On the role of the BV image model in image restoration," UCLA Department of Mathematics CAM Report 02-14 available at: www.math.ucla.edu/~imagers. To appear in AMS Contemporary Mathematics, 2002.

[6] Christensen, Marc P. et al., "Active-eyes: an adaptive pixel-by-pixel image-segmentation sensor architecture for high-dynamic-range hyperspectral imaging," Applied Optics-IP, Volume 41, Issue 29, 6093-6103, October 2002.

[7] Debevec, P. E., and Malik, J. Recovering high dynamic range radiance maps from photographs. In Proceedings of SIGGRAPH 97, ACM SIGGRAPH Addison Wesley, Los Angeles, California, Computer Graphics Proceedings, Annual Conference Series, 369–378, 1997

[8] Dowling, J.E., *The Retina: An Approachable Part of the Brain*, Harward University Press, 1987.

[9] Durand, F., and Dorsey, J. 2002 Fast Bilateral Filtering for the Display of High-Dynamic-Range Images. ACM Transactions on Graphics, 21(3) (Proc. ACM SIGGRAPH 2002).

[10] Fattal, R., Lischinski, D., AND Werman, M. 2002, Gradient domain high dynamic range compression. ACM Transactions on Graphics, 21(3) (Proc. ACM SIGGRAPH 2002).

[11] Geman, S. and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," PAMI-6, November 1984, pp.721-7

[12] Grossberg, S., "Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena," *Perception & Psychophysics*, 43, 241-277, 1988.

[13] M. D. Grossberg and S. K. Nayar, "High Dynamic Range from Multiple Images: Which Exposures to Combine?" In Proc. ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV), Nice, France, October 2003.

[14] Hansen, T. and Neumann, H. "Neural mechanisms for Representing Surface and Contour Features," *Emergent Neural Computational Architectures*, LNAI 2036, S. Wermter et al. (Eds.), 139-153. 2001

[15] J. G. Harris, C. Koch, E. Staats, and J. Luo, "Analog hardware for detecting discontinuities in early vision," *Int. J. Comput. Vis.*, vol. 4, pp. 211–223, 1990.

[16] Horn, B.K.P. 1974 "Determining Lightness from an Image," *Computer Graphics and Image Processing*, 3(1). 277-299.

[17] Horn, B. K. P. 1986. *Robot Vision*, MIT Press,.

[18] Horn, B.K.P. 1988 Parallel Networks for Machine Vision. *AI Memo. 1071*, AI Lab, MIT.

[19] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.

[20] Jobson, D.J., Rahman, Z., and Woodell, G.A., "A multi-scale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Tran. On Images Processing*, 6(7). 965-976, 1997

[21] Land, E.H., and McCann, J.J. "Lightness and Retinex theory," *Journal of the Optical Society of America*, 61(1), 1-11, 1971

[22] Mead, C. *Analog VLSI and Neural Systems*, Reading, MA: Addison-Wesley, 1989

[23] Mead, C. Neuromorphic Electronic Systems. *Proc. IEEE*, 78, 1629-1636, 1990

[24] C. A. Mead and M. A. Mahowald, "A silicon model of early visual processing," *Neural Networks*, vol. 1, pp.91-97, 1988.

[25] Nayar, S.K. and Bronzoi, V., "Adaptive Dynamic Range Imaging: Optical Control of Pixel Exposures Over Space and Time," *Proc of the 9th IEEE Inter. Conf. on Computer Vision (ICCV 2003)*.

[26] S. K. Nayar and T. Mitsunaga, "High Dynamic Range Imaging: Spatially Varying Pixel Exposures," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, June 2000.

[27] Pattanaik, S. N. and Yee, H., "Adaptive Gain Control for High Dynamic Range Image Display," *Proceedings of Spring Conference in Computer Graphics (SCCG2002)*, April 24-27, 2002, Budmerice, Slovak Republic.

[28] Perona, P., and Malik, J., "Scale-space and edge detection using anisotropic diffusion," *IEEE Tran. Pattern Analysis and Machine Intelligence*, 12(7), 629-639, 1990

[29] Pessoa, L. Mingolla, E. and Neumann, H.. "A Contrast- and Luminance-driven Multiscale Network Model of Brightness Perception," *Vision Research*, 35(15), 1995

[30] Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J. "Photographic Tone Reproduction for Digital Images," ACM *Transactions on Graphics*, 21(3), (Proceedings of SIGGRAPH 2002).

[31] L. Rudin, S. Osher, and C. Fatemi, "Nonlinear total variation based noise removal algorithm", *Physica, 60 D* (1992), pp. 259--268.

[32] Y. Y. Schechner and S. K. Nayar "Generalized Mosaicing: High Dynamic Range in a Wide Field of View," International Journal of Computer Vision, Vol 53, No. 3, July 2003

[33] Schanz, M., Nitta, C., Bubmann, A., Hosticka, B.J., and Wertheimer, R.K., "A High-Dynamic-Range CMOS Image Sensor for Automotive Applications," *IEEE Journal of Solid-State Circuits*, Vol. 35, No. 7, 932-938, 2000

[34] Shi, B.E. and Choi, T. "A Michelson Contrast Sensitive Silicon Retina," 8th International Conference on Neural Information Processing, ICONIP2001, Shanghai, China, November 14-18, 2001.

[35] Stocham, J.T.G., "Image Processing in the context of visual model," *Proceedings of the IEEE*, vol. 60, 1972, pp 828-842,

[36] Tomasi, C., and Manduchi, R. 1998. Bilateral filtering for gray and color images. In Proc. *IEEE Int. Conf. on Computer Vision*, 836–846.

[37] Trottenberg, U., Oosterlee, C.W., and Schuller A. 2001. *Multigrid*, San Diego, Calif., Academic Press.

[38] Tumblin, J., AND Ruschmeier, H. 1993. Tone reproduction for realistic images. IEEE Comp. Graphics & Applications 13, 6, 42–48.

[39] Tumblin, J., and Turk, G. 1999, LCIS: A boundary hierarchy for detail-preserving contrast reduction. In Proc ACM SIGGRAPH 99, A Rockwood, Ed., 83-90.

[40] P. Venier, "A contrast sensitive silicon retina based on conductance modulation in a diffusion network," Proc. 6th Intl. Conf. on Microelectronics for Neural Networks, Evolutionary and Fuzzy Systems, Dresden, Germany, pp. 145- 148, Sep. 1997.

[41] Wandel, B.A. 1995. Foundations of Vision. Sunderland MA: Sinauer.

[42] Yang, D., Gamal, A. E., Fowler, B., and Tian, H., 1999 A 640x512 CMOS image sensor with ultrawide dynamic range floating-point pixel-level ADC. IEEE Journal of Solid State Circuits 34, 12 (Dec.), 1821–1834.