A Perceptually Based Model of Visual Adaptation

Armin Samii
James O'Brien

Acknowledgement

\#DGE1106400.

A Perceptually Based Model of Visual Adaptation
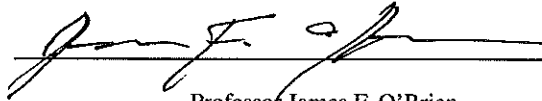
by K. Armin Samii

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:

Committee:
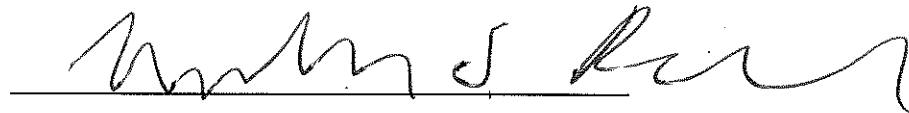
Professor James F. O'Brien

Research Advisor

May 14, 2014

(Date)

* * * * * * *

Professor Marty Banks

Second Reader

_5-16-14_

(Date)

Abstract

In this paper we describe a perceptual model that accounts for the time-varying changes to perceived color and brightness that occur due to time-varying adaptation and the transition between cone- and rod-mediated vision. Given a multispectral, high-dynamic-range image of a scene and the viewer's current adaptive state, this model produces a low-dynamic-range image that, when viewed photopically, creates a perception similar to what would have been experienced by the viewer in the original scene. When applied to a video sequence, our model produces temporally coherent output which models the viewer's adaptation state. A calibrated four-color-channel video camera is used to obtain video-rate data for scenes with moving objects. We describe a new demosaicing algorithm for working with this type of camera, and how the four color channels can be use to estimate independent rod and cone responses.

Contents

(a) *Mesopic viewing just after sunset.*   **(b)** *After dusk without any sunlight.*   **(c)** *Colored lights illuminate the scene.*

Figure 1: *A timelapse sequence showing the same scene in different mesopic illumination conditions*

1 Introduction

In dimly lit scenes, the human visual system perceives the world with a color shift known as the Purkinjie effect. It is caused by a nonlinear mixing of the signals from the rod photoreceptors, which are active at low light levels, into the neural pathways of the cone photoreceptors, which are active at higher light levels.

Kirk and O'Brien [2011] modeled this effect by estimating the rod and cone responses to a scene. Their method samples the visible light spectrum using multiple images, a time-consuming process which only works for static scenes and assumes a fully-adapted viewer. In this paper, we estimate the photoreptor responses using a single image from a camera with four color channels. This process allows us to accurately model the Purkinjie effect for video sequences.

We then model the time-course of adaptation in order to capture perceptual effects due to adaptation and changing illumination levels. While adaptation happens at many different scales both temporally and spatially [Webster 2011; Kohn 2007; Rieke and Rudd 2009], we only explicitly model the coarse temporal adaptation which occurs early in the visual pipeline. Other adaptation mechanisms are implicitly handled using any high dynamic compression algorithm; roughly speaking, these algorithms aim to maximize visibility within a specified dynamic range.

Ideally, the spectral responses of each color channel of the camera would exactly correspond to those of the long, medium, and short cones, and rods (LMSR). Since no such a camera is commercially available, we use a camera with four linearly independent channels: red, green, blue, panchromatic (RGBP)[1]. The panchromatic channel only has an IR-cut filter and is sensitive to light across the entire visible spectrum. We use the RGBP data to approximate the LMSR responses and the resulting perceptual response.

To compress the resulting high dynamic range image to the display's dynamic range, we account for how the perceived brightness will change over time. Our final result is a low dynamic range RGB image that, when viewed photopically on a monitor, evokes a perceptual response which matches the original scene.

---

[1]Throughout this paper, we will use 'R' for rods in LMSR, and red in RGBP. The usage should be clear from context.
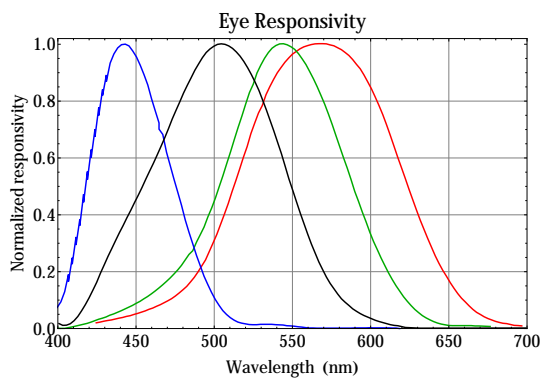
(a) *Normalized responsivity of the long (red), medium (green), and short (green) wavelength cones, and the rods (black).*
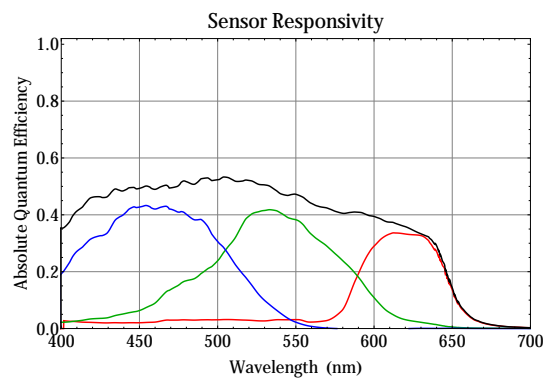
(b) *Absolute responsivities of the red, green, blue, and panchromatic channels of the Kodak TRUESENSE Sparse CFA sensor. These curves are after the infrared-cut filter, and thus differ from the manufacturer-provided curves.*

Figure 2: *In Section 3.2, we show how to approximate the rod and cone responses from the sensor data.*

2   Related Work

Spectral imaging: A standard RGB camera can nearly meet the Luther-Ives condition, which states that each channel must be a linear combination of cone photoreceptor responses to create a color transformation [Seymour 1997]. When taking into account the fourth channel in the human visual system, the rods, a standard RGB camera will lead to large errors in the approximation of the four color channel responses. More sensor data is necessary for a good approximation.

Our work builds on that of Kirk and O'Brien [2011]. Their method finely samples and reconstructs the spectral distribution of energy in a scene, which requires taking many images successive images through a set of filter. The amount of manual labor needed to acquire a spectral image prevents video capture.

There are several options available for acquiring spectral images with a single capture, though often at the cost losing spatial resolution [Du et al. 2009; Cao et al. 2011]. Cao et al. [2011] compensate for the lost spatial resolution by combining the data with a high-resolution non-spectral RGB camera. The KaleidoCamera uses optical filters similar to Kirk and O'Brien [2011], but splits the sensor into nine regions, each encoding the full image captured with a different filter [Manakov et al. 2013]. This allows the same data to be captured with a single exposure at lower resolution. While all of these options are viable solutions for our problem, we have decided to use a single, four-color-channel camera for three reasons: (1) the panchromatic channel is more sensitive to light than the RGB color channels, making it well-suited for capturing low-light scenes; (2) it is simple to implement in most current camera architectures and complements current HDR imaging technologies; (3) our color transformation meets the Luther-Ives condition with low error, so we do not need the more accurate spectral response estimations that these other cameras provide.

There are many other options for capturing HDR videos that are more accurate than our simple method of alternating exposure times. For example, Tocci et al. [2011] simultaneously capture three images at different exposures using three sensors with different sensitivities. Nayar and Branzoi [2003] optically adjust the amount of light that hits the sensor, aiming for bright but unsaturated pixel sites across the sensor, which maximizes the signal-to-noise ratio. Our method can easily complement these architectures by replacing each of their sensors with a four-channel sensor.

High Dynamic Range Compression: Several High Dynamic Range (HDR) compression algorithms also take into account psychophysical effects. Kuang et al. [2007] present a method which accounts for the lack of perceived color in low-light conditions. Durand and Dorsey [2000] discuss time-dependent chromatic adaptation, but only in terms of discounting-the-illuminant (i.e. neural white balancing). Irawan et al. [2005] combine a psychophysical model of visual adaptation [Pattanaik et al. 2000] with a specific global HDR compression algorithm [Larson et al. 1997] to handle large dynamic ranges in HDR video streams. Our method can handle arbitrary HDR compression algorithms, not just the histogram adjustment of Larson et al. [1997].

Time Course of Light and Dark Adaptation: Ledda et al. [2004] modeled visibility using psychophysical data of the time-course of adaptation. They use a model of local adaptation to perform high dynamic range compression. Their method does not model the Purkinjie effect.

Pattanaik's algorithm for time-dependent adaptation used the Hunt 94 color appearance model [Pattanaik et al. 2000; Hunt 1995], focusing on Hunt's brightness model. Their method does not model the Purkinjie effect. In earlier work, they model psychophysically-based spatial vision effects such as acuity, noise, brightness, contrast, and the opponent color space [Pattanaik et al. 1998]; our model complements these effects.

Demosaicing RGBP images To estimate the LMSR responses, we use a commercially-available camera sensor which has four color channels, red, green, blue, and "panchromatic" (see Figure 3 and Section 3.2). The so-called panchromatic pixels do not have a color filter over them, only an infrared-blocking filter. As a result, they respond to light across the visible light spectrum and are therefore more sensitive in low-light conditions (see Figure 2b). RGBP demosaicing algorithms have also been proposed by the engineers of the Kodak RGBP sensor [Kumar et al. 2009] and in an algorithm for image deblurring [Wang et al. 2012]. Wang *et al.* rely on reducing the image to standard Bayer Demosaicing; in doing so, they throw away half of the available color resolution. Kodak's algorithm bilinearly interpolates the colors to all pixels without using a naïve Bayer demosaicing algorithm, and retains full color resolution. However, it does not properly interpolate across edges, resulting in color fringes (see Figure 4).
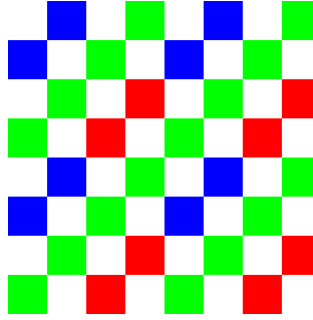
Figure 3: *The mosaic pattern of the Kodak sensor we used.*

3 Estimating Rod and Cone Responses

Using a camera sensor with four color channels, we construct a four-channel RGBP image (Section 3.1). We then linearly combine these channels to form a four-channel LMSR image (Section 3.2). The resulting LMSR response is more accurate than if we were to use a standard RGB image, and it provides linearly independent estimates of the rod and cone responses.

3.1 Four-Channel Image Demosaicing

We use an Imperx Bobcat IGV-B1320T camera[2] with a Kodak TRUESENSE Sparse Color Filter Patter RGBP sensor[3]. The resolution of the camera is 1320x736. The mosaic pattern is shown in Figure 3. Half of the pixels are panchromatic: they are sensitive to the full visible light spectrum. Their response curve (see Figure 2) is linearly independent from RGB.

Previous demosaicing algorithms do not properly interpolate colors near edges, creating color fringing artifacts in high-frequency regions. See Figure 4 for a comparison between our method and the method of Kumar *et al.* [2009]. Our algorithm is very similar to Kumar et al. [2009], but avoids the color fringes.

In our demosaicing algorithm, we first construct the high-resolution panchromatic channel and low-resolution color channels independently. We then restore high frequencies in the color channels using the signal from the panchromatic channel. Let $m_s(\mathbf{I}))$ be the masking function which selects pixel sites $s \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}, \mathrm{P}\}$ from the mosaiced image $\mathbf{I}$, replacing unselected pixel sites with zeros. Let $b(m_s(\mathbf{I}))$ be the bilinear interpolation function which produces a fullsize image using the selected pixels in $m_s(\mathbf{I})$. The demosaicing algorithm consists of the following steps:

1. Construct $\mathbf{P}_{\mathrm{highres}} = b(m_{\mathrm{P}}(\mathbf{I})))$, the full resolution panchromatic image.

---

[2]http://www.imperx.com/bobcat/igv-b1320

[3]http://www.truesenseimaging.com/technologies/truesense-sparse-color-filter-pattern

(a) *An image demosaiced using the algorithm of Kumar* et al. *[2009].*

(b) *The same image demosaiced using our algorithm.*

Figure 4: *A comparison of our demosaicing algorithm to the algorithm described in Kumar* et al. *[2009]. There is significant color fringing in high frequency regions. Our result avoids the color fringes. While it seems Kumar* et al. *'s method is sharper, the black color fringes are hallucinated and should not be present in the image.*

2. Construct $\mathbf{C}_{\text{lowres}} = b(m_c(\mathbf{I}))$, $c \in \{\text{R}, \text{G}, \text{B}\}$, the three low resolution color images.

3. Denoise $\mathbf{P}_{\text{highres}}$ and $\mathbf{C}_{\text{lowres}}$.

4. Construct $\mathbf{P}_{c,\text{lowres}} = b(m_c(\mathbf{P}_{\text{highres}}))$, the low resolution panchromatic image constructed using only the pixels with color c.

5. Assuming $\mathbf{P}_{c,highres}$ contains the high frequency information that is missing from $\mathbf{C}_{lowres}$, construct

$$\mathbf{C}_{\text{highres}} = \left(\frac{\mathbf{P}_{\text{highres}}}{\mathbf{P}_{\text{lowres}}} \right) \cdot \mathbf{C}_{\text{lowres}}.$$

Note that we denoise each color channel independently. This is especially important with RGBP images, since a single noisy pixel influences $4 \times 4 = 16$ pixels, whereas with a standard Bayer mosaic pattern, it only influences $2 \times 2 = 4$ pixels. Our denoising function is a median filter with a radius of 3 pixels, which we selected because the sensor we use tends to have salt-and-pepper noise. Other sensors may benefit from more intelligent denoising algorithms.

3.2 Converting RGBP to LMSR

Given an RGBP pixel value, we want to estimate the response of the long, medium, and short cones (LMS) and the rods (R). The first step is to find a linear mapping from the camera's color space to the eye's color space. Specifically, we want to find the matrix $\mathbf{A}$ such that $\mathbf{A}\mathbf{b}_s = \mathbf{b}_e$, where $\mathbf{b}_s$ is the sensor's RGBP vector and $\mathbf{b}_e$ is the eye's LMSR vector.

The RGBP vector obtained from the camera is the integral of the scene's reflectance spectra under the RGBP spectral responsivity curves. The LMSR vector is similarly obtained in the retina. Figure 2a shows the normalized LMSR sensitivity curves, and Figure 2b shows the absolute RGBP sensitivity curves.

Let $l(\lambda)$ be the radiance of a scene point at wavelength λ. Let $r_s(\lambda)$ be the sensitivity of the sensor at color $s \in$
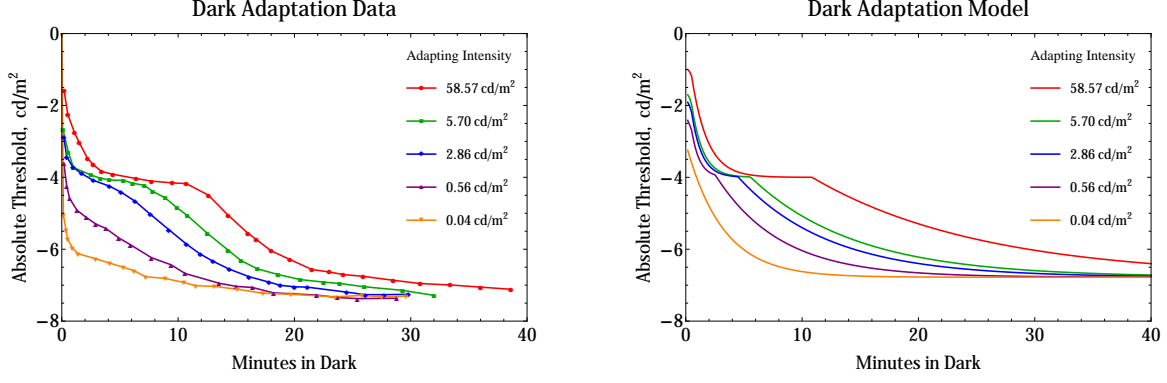
(a) *Data from observer S.H. in Table 1 of Hecht* et al. *[1937].*

(b) *Our model of the data.*

Figure 5: *Time-course of dark adaptation. The minimum perceptible threshold is plotted. The initial segment of the curve is the cone adaptation, and the later segment is the rod adaptation. Our model has the same characteristics of the psychophysical data: the rate of exponential convergence and the point at which rods become more sensitive than cones.*

$\{R, G, B, P\}$ to wavelength λ, and $r_e(\lambda)$ be the sensitivity of the eye at receptor $e \in \{L, M, S, R\}$. A single point $\mathbf{b}_s$ in the resulting image is

$$\mathbf{b}_s = \int_\lambda r_s(\lambda)l(\lambda)d\lambda. \tag{1}$$

The eye's response to the same point is similarly derived:

$$\mathbf{b}_e = \int_\lambda r_e(\lambda)l(\lambda)d\lambda. \tag{2}$$

To estimate $\mathbf{b}_e$, we discretize:

$$\mathbf{b}_s \approx \sum_{\lambda=400\text{nm}}^{700\text{nm}} r_s(\lambda)l(\lambda) \tag{3}$$

$$\mathbf{b}_e \approx \sum_{\lambda=400\text{nm}}^{700\text{nm}} r_e(\lambda)l(\lambda). \tag{4}$$

Which samples the frequency spectrum at 1nm intervals between 400nm and 700nm (the visible range for both the eye and the sensor). We turn this sum into a matrix-vector multiplication by stacking each of the n elements of the sum and combining the channels into a single matrix:

$$\mathbf{b}_s = \mathbf{r}_s\mathbf{l} \tag{5}$$

$$\mathbf{b}_e = \mathbf{r}_e\mathbf{l}. \tag{6}$$

8

where $\mathbf{b}_s$ and $\mathbf{b}_e$ are 4×1 vectors, $\mathbf{r}_s$ and $\mathbf{r}_e$ are $4 \times n$ matrices, and $\mathbf{l}$ is an $n \times 1$ vector. We can now compute the 4×4 projection $\mathbf{A}$ of RGBP onto LMSR:

$$\mathbf{A} = \mathbf{r}_e \mathbf{r}_s^{-\mathsf{P}} \tag{7}$$

$$\mathbf{b}_e = \mathbf{A} \mathbf{b}_s. \tag{8}$$

Where the symbol $^{-\mathsf{P}}$ denotes the psuedoinverse.

Using the data shown in Figures 2a and 2b, we obtain the matrix given in Appendix 8.

3.3   Constructing Radiance Maps with Physical Units

All pixel values must have physically accurate units so that any choice of shutter speeds, apertures, and analog gains (ISO) for the same scene will ideally result in the same values. We combine multiple exposures of a scene into a single radiance map using the method of Debevec and Malik [1997]. This method computes the scene radiances by inverting the camera response function to illuminance on each channel of the sensor. When obtaining video, we program the camera to sequentially cycle exposure settings.

We precompute the camera's response function with high precision using twelve exposures of a standard color calibration chart. We then compute the proper scale factor using a scene of known illuminance (calibrated using a photometer). These values are constant for all images taken with the same camera setup.

A fully saturated pixel site would cause the conversion to LMSR to produce incorrect results. The preferred solution is to combine enough exposures to cover the entire dynamic range of the scene. If this requires prohibitively many exposures, it is necessary to store the list of fully saturated pixels until the end of the processing pipeline, and ensure they are fully saturated in the final image.

We construct radiance maps after demosaicing the raw image into an RGBP image.

4 Low-Light Tone Mapping

The rods are active between approximately 10^{-6} and 10^2cd/m^2; the cones are active between approximately 10^{-4} and 10^8cd/m^2 [Hood and Finkelstein 1986]. In *mesopic* viewing conditions, where both the rods and cones are active, the final image is a sum of the *photopic* contribution from the cones and the *scotopic* contribution from the rods.

Given the LMSR response $\mathbf{b}_e$ computed in Equation 8, we compute the display RGB values which will produce the same LMS response in photopic viewing conditions. For the cones, this is a direct mapping from $\mathbf{b}_e$ to the display RGB. For the rods, we must also account for the Purkinjie effect.

The 3×4 matrix $\mathbf{M}$ which maps the neural LMSR responses onto the display RGB has elements

$$m_{ij} = \int_\lambda d_i(\lambda) r_j(\lambda) d\lambda, \tag{9}$$

where $r_j(\lambda)$ is the spectral response of photoreceptor j, and d_i is the spectral emission of the display channel i. This matrix for an Apple Cinema HD Display is provided in Appendix 8.

The scotopic contribution to the final RGB image uses the Purkinjie effect function described by Kirk *et al.* [2011], with their model derived from Cao *et al.* [2008]. We summarize this function in Appendix 8.

The photopic and scotopic contributions are thus:

$$c_{\text{cones}} = \mathbf{M} \cdot \mathbf{b}_e \tag{10}$$

$$c_{\text{rods}} = \mathbf{M} \cdot \text{Purkinjie}(\mathbf{b}_e). \tag{11}$$

These quantities will be scaled light and dark adaptation and combined based on to produce the final image.

5   Light and Dark Adaptation

The rods and cones adapt independently to changing illumination conditions [Hecht et al. 1937]. We model this adaptation by computing the minimum perceptible illuminance level l at each frame in a video sequence. The maximum perceptible illuminance level is $l \cdot \rho$, for a constant dynamic range interval ρ.

Dark adaptation takes over half an hour to complete fully (see Figure 5a), and has the following properties [Hecht et al. 1937]:

1. Cones complete adaptation in about 5 minutes.

2. Rods complete adaptation in about 40 minutes.

3. Dimmer illumination levels result in faster adaptation rates.

When adapting to a low-light mesopic scene, the scene will first become dimly visible with photopic colors in the first few minutes (Property 1), then the adaptation of the rods will cause the Purkinjie effect to take place over the next 40 minutes (Property 2).

Light adaptation happens much more rapidly than dark adaptation, reaching full adaptation in about a minute [Adelson 1982].

5.1   Fitting the Psychophysical Data

The psychophysical dark adaptation data from Hecht et al. [1937] (shown in Figure 5a) provides the minimum visibility threshold over time. In Hecht's experiments, the human observer was first shown a stimulus at the adapting intensity, a, for several minutes. After fully adapting to this intensity, Hecht measured the dimmest perceptible stimulus over time.

For a given adapting intensity, the rate of adaptation is well-approximated by an exponential convergence to the threshold. The exponential convergence rate is a function of the adapting intensity. We model the adaptation of the rods and the cones separately by estimating the rate of adaptation r as a function of the adapting intensity a.

The five curves shown in Figure 5a correspond to five different adapting intensities a. We fit the following function for both the rod and cone segments independently, for a total of ten curves:

$$s_p(a_t, t) = 1 - \exp(-t \cdot r_{p,a_t}) \tag{12}$$

Where the variables are:

$p \in \{\text{cones,rods}\}$,

t: the time since the start of adaptation

a_t: the adapting intensity of the scene in Candelas per meter squared (cd/m^2),

$s_p(a_t, t) \in [0, 1]$: the relative state of adaptation at time t, normalized such that $s = 1$ indicates full adaptation

r_{p,a_t}: the exponential parameter of segment p for adapting intensity a_t.

We model the adaptation rate as a function of the adapting intensity using power function of the form αx^β. Changing the discrete variable r_{p,a_t} into a continuous function $r_p(a_t)$,

$$r_p(a_t) = \alpha_p a_t^{\beta_p} \tag{13}$$

solving for α_p and β_p.

Combining Equations 12 and 13:

$$\begin{aligned} s_p(a_t, t) &= 1 - \exp(-t \cdot r_p(a_t)) \\ &= 1 - \exp(-t \cdot \alpha_p a_t^{\beta_p}) \end{aligned} \tag{14}$$

We fit the model using the following parameters:

$$\alpha_{\text{cones}} = 0.015203$$

$$\beta_{\text{cones}} = -0.13432$$

$$\alpha_{\text{rods}} = 0.0017289$$

$$\beta_{\text{rods}} = -0.20612$$

with time in units of seconds, and l in log-units of cd/m^2.

The adaptation rate in the psychophysical experiments was determined by the single adapting intensity, a_0. In a video sequence with changing illumination levels, it does not make sense to only use the adapting intensity of the first frame. We therefore set the adapting intensity to be proportional to the current minimum threshold:

$$a_t = l_t \cdot \rho/2. \tag{15}$$

The adaptation curves of our final model are shown in Figure 5b. Note the similarity between our data and the psychophysical data in Figure 5a. Notably, there are two distinct segments of each curve: a cone adaptation, then rod adaptation.

Light adaptation occurs much more rapidly than dark adaptation. There is less psychophysical data available on the time course of light adaptation, so we assume a fixed rate of adaptation r_p independent of the adapting intensity a_t. Following the model of Pattanaik *et al.* [2000] and Hood and Finkelstein [1986], the light adaptation constants are:

$$\begin{aligned} r_{\text{cones}} &= .0091 \\ r_{\text{rods}} &= .0025. \end{aligned} \tag{16}$$

5.2 Applying the Model

Given $s_p(t)$, the relative adaptation state at time t, we must find $l_p(f_t, t)$, the absolute adaptation state for the frame f_t at time t. Let $l_p(f_t, \infty)$ be the minimum perceptible illumination threshold when fully adapted to the scene f_t. We call this the *goal* adaptation threshold. We choose the $l_p(f_t, \infty)$ which maximizes the number of pixels within the visible illumination interval:

$$\text{argmax}_{l_p(\infty)} \sum_{v \in f_t} \mathbb{1}\{l_p(\infty) \le v \le l_p(\infty) \cdot \rho\} \tag{17}$$

for scene illumination levels v, where $\mathbb{1}$ is the indicator function. We maximize Equation 17 by uniformly sampling the scene illumination values in log units, and find the minima within .01 cd/m^2.

We then clamp the l_p tuple to their working range (see Table 5.1, [Boff et al. 1986]):

$$10^{-6} \le l_{\text{rods}} \le 10^2/\rho \tag{18}$$

$$10^{-4} \le l_{\text{cones}} \le 10^8/\rho. \tag{19}$$

At any given instant, $\rho \approx 100$, but rapidly changing pupil radii (between 2mm and 8mm) gives an increased dynamic range by a factor of 16 [Dowling 1987]. Since we do not have information about where the eye is looking or what the current pupil size is, we use $\rho = 100 \cdot 16 = 1600$.

To discretize Equation 12, at each timestep we take a step towards the goal threshold using the discrete exponential filter:

$$l_p(f_t, t) = (1 - Tr_t) \cdot l_p(f_{t-1}, t-1)$$
$$+ (Tr_t) \cdot l_p(f_t, \infty) \tag{20}$$

where T is the length of the timestep in seconds and t is the discrete frame index.

In summary, at each timestep, we compute four values:

1. a_t, the starting adapting intensity (Equation 15),

2. r_p, the adaptation rates (Equations 13 and 16),

3. $l_p(f_t, \infty)$, the goal adaptation thresholds (Equation 17), and

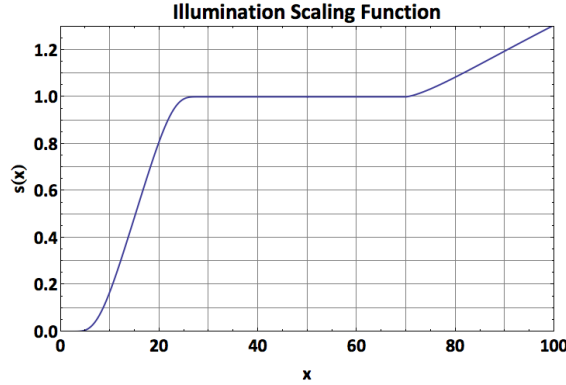4. $l_p(f_t, t)$, the current minimum thresholds (Equation 20)

Figure 6: *Our scaling function: it smoothly drops to 0 at the minima, diverges beyond the maxima, and remains constant between the minimum and maximum threshold. For visualization purposes, we use the parameters $l = 30$, $\rho = 2.3$, $\alpha_0 = 100$, $\alpha_1 = 1.1$. To generate the shown images, we use $\rho = 1600$ and $\alpha_1 = 1.2$ instead.*

6   Range Compression

Given the visible illumination interval $[l_p, l_p \cdot \rho]$, we need a function to render pixels below l_p as fully black, and above $l_p \cdot \rho$ as fully white. Other methods have used the S-shaped Naka-Rushton equation for this purpose [Pattanaik et al. 2000; Irawan et al. 2005] based on psychophysical data first introduced by Naka and Rushton [1966]. While the Naka-Rushton function is psychophysically correct, it will not preserve constant brightness within the visible interval. This is important because many HDR methods that compress the dynamic range in log scale [Tumblin and Rushmeier 1993; Larson et al. 1997; Fattal and Lischinski 2002; Durand and Dorsey 2002] based on the assumption that humans perceive brightness in roughly a logarithmic scale. This assumption will not hold if the brightnesses are first passed through the Naka-Rushton equation.

We therefore deviate from the psychophysical literature and build a heuristic function. We want our function to have the following properties:

1. **Continuity:** Having only C^0 continuity produces artifacts, so the function requires C^1 continuity or smoother.

2. **Linearity:** As justified above, the function should preserve linearity for computations in log space to remain accurate.

3. **Minimum converges to zero:** The function should reach zero below the minimum threshold to match the data [Hecht et al. 1937]. The Naka-Rushton equation approaches but never reaches zero, which is undesirable because many HDR compression algorithms will attempt to place all nonzero values in the display's dynamic range, rather than render them as fully black.

14

4. **Maximum diverges to infinity for cones:** The function should diverge towards infinity above the maximum threshold for cones. It will later be clamped to the display dynamic range. This will allow even very dim luminances to be rendered as fully white, which is desirable when these luminances are above the threshold $l \cdot \rho$.

5. **Smoothness parameter:** While not necessary, we would like to be able to control the smoothness of the attenuation and amplification to find a value which avoids noticeable discontinuities in the image.

We therefore use the piecewise function visualized in Figure 6. At the minima, we use a Planck-Taper Window [Tu 2008] which converges to zero with C^∞ continuity:

$$t_{0,p} = l_p/\alpha_0$$

$$t_{1,p} = l_p$$

$$Y_p = \frac{t_{1,p} - t_{0,p}}{(x - t_{1,p})} + \frac{t_{1,p} - t_{0,p}}{(x - t_{0,p})}$$

$$s^-(x, l_p) = \begin{cases} 0 & x \leq t_{0,p} \\ 1/(\exp(Y_p) + 1) & t_{0,p} < x < t_{1,p} \\ 1 & x \geq t_{1,p} \end{cases} \tag{21}$$

where α_0 is the minimum-threshold smoothing constant.

Rod saturation results in a decrease of the rod contribution to the final image [Stockman and Sharpe 2006], which is modeled in the Purkinjie effect function described in Appendix 8. For the cone contribution, illumination levels above the maximum threshold are amplified with a polynomial:

$$t_{2,p} = l_p \cdot \rho$$

$$Z_p = t_{2,p} - (1/\alpha_1)^{\alpha_1}$$

$$s^+(x, l_p) = \begin{cases} 1 & x < t_{2,p} \\ ((x^{\alpha_1} - Z_p)^{\alpha_1} + Z_p)/x & x \geq t_{2,p} \end{cases} \tag{22}$$

$$\tag{23}$$

where α_1 is the maximum-threshold smoothing constant. The complete scaling function for the cones is

$$s(\cdot, l_p) = s^+(\cdot, l_p) \cdot s^-(\cdot, l_p). \tag{24}$$

We set $\alpha_1 = 100$ and $\alpha_2 = 1.2$ for all examples shown.

6.1 Applying the Scaling Function to Static Images

We apply the amplification function $s^+(\cdot)$ *after* HDR compression because HDR compression algorithms do not gracefully handle diverging values. We apply the attenuation function $s^-(\cdot)$ *before* HDR compression because it must
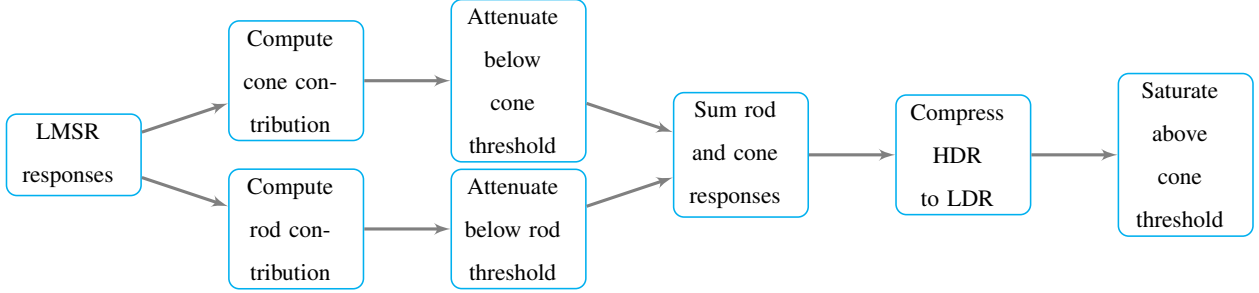
Figure 7: *Our pipeline for computing the correctly-scaled image given the initial LMSR responses and the minimum perceptible threshold for the rods and cones.*

be applied to the rod and cone contributions independently. After HDR compression, the image no longer has physical units so we apply $s(\cdot)^+$ based on retained values from c_{cones}. The final computation is performed as follows, with a flowchart of this computation shown in Figure 7: where the function hdrToLdr is any suitable high dynamic range

---

Algorithm 1 Computing the Brightness

1: **function** $b(c, l, \mathrm{hdrToLdr})$

2:     $\mathrm{hdr}_{\mathrm{cones}} \leftarrow c_{\mathrm{cones}} \cdot s^-(c_{\mathrm{cones}}, l_{\mathrm{cones}})$

3:     $\mathrm{hdr}_{\mathrm{rods}} \leftarrow c_{\mathrm{rods}} \cdot s^-(c_{\mathrm{rods}}, l_{\mathrm{rods}})$

4:     $\mathrm{hdr} \leftarrow \mathrm{whitebalance}(\mathrm{hdr}_{\mathrm{cones}}) + \mathrm{hdr}_{\mathrm{rods}}$

5:     $\mathrm{ldr} \leftarrow \mathrm{hdrToLdr}(\mathrm{hdr})$

6:     $\mathrm{result} = \mathrm{ldr} \cdot s^+(c_{\mathrm{cones}}, l_{\mathrm{cone}})$

7:     **return** result

8: **end function**

---

compression algorithm. We have found the Durand and Dorsey [2002] dynamic range compression algorithm works well for our examples. We then compress values in the mesopic and scotopic range as done in Kirk *et al.* , then clamp the resulting image to the display's dynamic range.

6.2   Applying the Scaling Function to Video Sequences

While the algorithm in Section 6.1 works for single images, it produces two types of artifacts when applied per-frame to a video with changing values for l_p:

1. **Intensity**: regions which should remain at constant visibility will become dimmer as the contrast increases due to more of the scene becoming visible, and

16

2. **Color**: as the rod contribution adaptively increases in the mesopic range, the hue may change without a corresponding change in brightness.

To prevent these artifacts, we compute the ratio κ of the current brightness at time t to the final brightness after full adaptation. We then set the RGB image at time t to be κ times the fully-adapted RGB image. We compute κ using the ratio between the brightness at the current and final adaptation state, without any HDR compression:

$$\kappa = b(c, l(t), \cdot)/b(c, l(\infty), \cdot). \tag{25}$$

where the function b is computed in Algorithm 1. The final image at time t is then

$$f_t = \kappa \cdot b(c, l(\infty), \text{hdrToLdr}), \tag{26}$$

where κ will be equal to 1 when the viewer is fully adapted.

7   Results and Discussion

We demonstrate the results of our method for dark adaptation, light adaptation, timelapse videos, and real-time videos.

Dark Adaptation: Figure 8 simulates what an observer would see after stepping into a candlelit room after being adapted to bright sunlight. The cones have reached near-full adaptation after five minutes, and the rods after forty minutes. Between five and forty minutes, the increase in perceived brightness are from the color shifted rod contribution. The supplementary video contains the full adaptation sequence, as well as simulated photopic-only and scotopic-only responses for comparison.

Light Adaptation: Figure 9 shows the much quicker rate of light adaptation. This example simulates a viewer stepping outside after being adapted to a dark room. Notice that the sky is initially overexposed, whereas the balcony is visible. Over time, the sky becomes increasingly visible, whereas the balcony does not change brightnesses since it remains in the visible illumination interval. See the supplementary video for the full adaptation sequence.

Photopic to Mesopic Scene: Figure 1 shows frames from a two-hour timelapse from just before sunset to only moonlight and artificial light. The time-course of adaptation corresponds with the changing illumination levels, and therefore has little effect on the final result. See the supplementary video for the full timelapse.

Additional Photographs: Figure 12 shows comparisons between images with and without our perceptual model. The photopic images were generated by transforming the camera's RGB to the display's RGB, then compressing the resulting high dynamic range image. We did not tune any parameters or do any additional processing to obtain these results.

Real-time Videos: Figures 10 and 11 show several frames from a real-time, high dynamic range videos. These examples simulate moonlight conditions by using lamps filtered to emit light at 4100K, the color temperature of moonlight. In Figure 10, the candlelight would be brighter relative to the rest of the scene in actual moonlight; since we simulate moonlight with a bright lamp, this additional visual cue may break the illusion of dim lights.
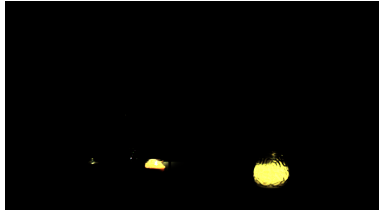
**(a)** *Upon entering the dark room, only the brightest parts of the scene are visible.*

(b) *After five minutes, cone adaptation is nearly complete, and the brightest regions are above the visible threshold.*
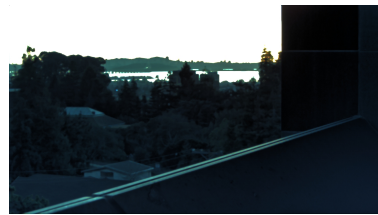
(c) *After forty minutes, rod adaptation is nearly complete. Notice that more letters are visible on both charts.*
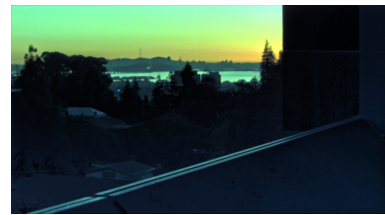
Figure 8: *The time-course of dark adaptation: simulation of entering a dark room after being adapted to bright sunlight. Compare the legibility of the letters between the middle frame and the right frame.*
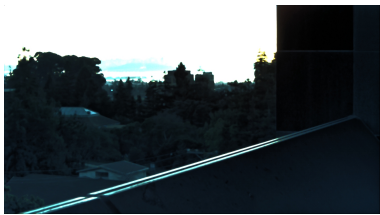
(a) *After stepping into bright light*

(b) *After 30 seconds of adaptation*

(c) *After 90 seconds of adaptation*

Figure 9: *The time course of light adaptation: entering a balcony after being adapted to a dimly lit room. Over time, the saturated regions in the background become visible, whereas the foreground remains nearly constant.*
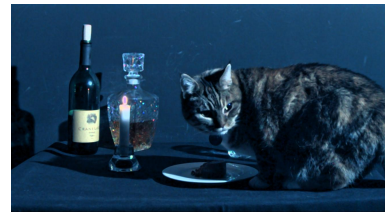

Figure 10: *Frames extracted from a video clip with simulated moonlight (color temperature of 4100K).*

Figure 11: *Frames extracted from a video clip with both simulated moonlight (color temperature of 4100K) and artificial ambient light (color temperature of 3200K).*

7.1 Discussion

Our results demonstrate the application of our methods to several scenes where low-light perception and adaption effects are significant. They also demonstrate the application of our methods to video, while demonstrating good temporal coherence.

Some of the adaptation effects we have modeled occur over long time intervals, up to forty minutes. Modeling these effects at realtime rates would likely not be useful. However, being able to model these effects and predict their impact on a viewer's perception is useful. Accelerated application of adaptation may also be useful to convey the rapid advance of time or to generate artistic effects.

Figure 12: *(Left) Images generated without modeling the Purkinjie effect or the perceptible illumination interval. (Right) Images generated with our method. We did not perform any color correction or additional processing other than using the algorithm described in this paper.*

8 Acknowledgements

Appendix 1: The Purkinjie Effect Model

The Purkinjie effect is caused by signals from the rods mixing with those from the cones early in the visual pipeline [Cao et al. 2008]. This mixing results in signals from the retina being perceived with an effect that is roughly a blue color shift. This color shift occurs in the opponent color space. Our goal is to find $\hat{q}$, the photopic LMS response which will be most perceptually similar to the scotopic response of the input LMSR signal q. We use the model described in Kirk *et al.* [2011], which we summarize here. To keep the constants the same as in Kirk *et al.* , we convert our LMSR signal from cd/m^2 (which we use in the rest of this paper) to units of Trolands.

First, the cone signals are regulated by the rod signal:

$$g_c = (1 + .33(q_c + \kappa_c q_{rod}))^{-.5}/m_c \tag{27}$$

where q_c is the LMSR response for photoreceptor c, $\kappa_{\text{long}} = \kappa_{\text{medium}} = .33$, $\kappa_{\text{short}} = .5$, and m_c is the maximum value of the cone fundamentals: $m_{\text{long}} = .637$, $m_{\text{medium}} = .392$, $m_{\text{short}} = 1.606$.

The matrix which transforms an LMS signal into the corresponding value in the oppenent color space is:

$$\mathbf{M} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & -1 & 1 \\ 1 & 1 & 0 \end{pmatrix}. \tag{28}$$

Expressing the color shift in the opponent color space, we get the three rows:

$$o_1 = x k_1 (\rho_1 g_{\text{med}} - \rho_2 g_{\text{long}}) q_{\text{rod}} \tag{29}$$

$$o_2 = y(\rho_3 g_{\text{short}} - k_1 \rho_4 (\alpha g_{\text{long}} + (1 - \alpha) g_{\text{med}})) q_{\text{rod}} \tag{30}$$

$$o_3 = z(\alpha g_{\text{long}} + (1 - \alpha) g_{\text{med}}) q_{\text{rod}} \tag{31}$$

where the parameters are: $x = y = 15$, $z = 4$, $\rho = [1.111, .939, .4, .6]$, and $\alpha = .619$. Bringing o out of the opponent color space, we obtain our final photopic LMS value:

$$\hat{q} = \mathbf{M}^{-1} o. \tag{32}$$

For a more complete discussion of the parameter choices and computation, refer to Cao *et al.* [2008] and Kirk *et al.* [2011].

Appendix 2: Transformation Matrices

The matrix $\mathbf{A}$ which converts RGBP values to LMSR is

$$\mathbf{A} = \begin{pmatrix} -0.1592 & 0.5772 & -2.2128 & 1.7319 \\ -0.5385 & 1.7578 & -0.9488 & 0.6684 \\ -1.0384 & -1.5815 & 0.6083 & 1.1282 \\ 2.2845 & 4.1602 & 3.4431 & -2.5556 \end{pmatrix}.$$

The matrix $\mathbf{M}$ which converts LMSR values to RGB is

$$\mathbf{M} = \begin{pmatrix} 0.65216 & -0.75068 & -0.03467 & 0.19764 \\ -0.01482 & 0.36133 & 0.00046 & 0.09840 \\ 0.06301 & -0.19355 & 0.35516 & 0.31689 \end{pmatrix}.$$

Bibliography

ADELSON, E. H. 1982. Saturation and adaptation in the rod system. *Vision Research 22*, 10, 1299–1312.

BOFF, K. R., KAUFMAN, L., AND THOMAS, J. P., Eds. 1986. *Handbook of Perception and Human Performance, Vol. 1: Sensory Processes and Perception*. John Wiley & Sons.

CAO, D., POKORNY, J., SMITH, V. C., AND ZELE, A. J. 2008. Rod contributions to color perception: linear with rod contrast. *Vision research 48*, 26, 2586–2592.

CAO, X., TONG, X., DAI, Q., AND LIN, S. 2011. High resolution multispectral video capture with a hybrid camera system. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, IEEE, 297–304.

DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *Computer Graphics and Interactive Techniques*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, SIGGRAPH '97, 369–378.

DOWLING, J. E. 1987. *The Retina: An Approachable Part of the Brain*. Harvard University Press.

DU, H., TONG, X., CAO, X., AND LIN, S. 2009. A prism-based system for multispectral video acquisition. In *Computer Vision, 2009 IEEE 12th International Conference on*, IEEE, 175–182.

DURAND, F., AND DORSEY, J. 2000. Interactive tone mapping. Springer Verlag. Held in Brno, Czech Republic.

DURAND, F., AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics*.

FATTAL, R., AND LISCHINSKI, D. 2002. Gradient domain high dynamic range compression. *ACM Transactions on Graphics*.

HECHT, B. Y. S., HAIG, C., AND CHASE, A. M. 1937. The Influence of Light Adaptation on Subsequent Dark Adaptation of the Eye.

HOOD, D., AND FINKELSTEIN, M. 1986. *Handbook of Perception and Human Performance, Vol. 1: Sensory Processes and Perception*. John Wiley & Sons, ch. Chapter 5, Sensitivity to Light.

HUNT, R. W. G. 1995. *The reproduction of colour, 5th edition*. Fountain Press.

IRAWAN, P., FERWERDA, J. A., AND MARSCHNER, S. R. 2005. Perceptually based tone mapping of high dynamic range image streams. In *Proceedings of the Sixteenth Eurographics conference on Rendering Techniques*, Eurographics Association, 231–242.

KIRK, A. G., AND O'BRIEN, J. F. 2011. Perceptually based tone mapping for low-light conditions. *ACM Transactions on Graphics 30*, 4 (July), 1.

KOHN, A. 2007. Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of neurophysiology 97*, 5, 3155–3164.

KUANG, J., JOHNSON, G. M., AND FAIRCHILD, M. D. 2007. icam06: A refined image appearance model for hdr image rendering. *Journal of Visual Communication and Image Representation 18*, 5, 406–414.

KUMAR, M., MORALES, E. O., ADAMS, J. E., AND HAO, W. 2009. New Digital Camera Sensor Architecture for Low Light Imaging. *ICIP 42*, 7 (June), 77A.

LARSON, G. W., RUSHMEIER, H., AND PIATKO, C. 1997. A visibility matching tone reproduction operator for high dynamic range scenes. *Visualization and Computer Graphics, IEEE Transactions on 3*, 4, 291–306.

LEDDA, P., SANTOS, L. P., AND CHALMERS, A. 2004. A Local Model of Eye Adaptation for High Dynamic Range Images. *AFRIGRAPH 1*, 212, 151–161.

MANAKOV, A., RESTREPO, J. F., KLEHM, O., HEGEDÜS, R., EISEMANN, E., SEIDEL, H.-P., AND IHRKE, I. 2013. A reconfigurable camera add-on for high dynamic range, multi-spectral, polarization, and light-field imaging. *ACM Trans. Graph. (Proc. SIGGRAPH 2013) 32*, 4 (July), 47:1–47:14.

NAKA, K.-I., AND RUSHTON, W. A. 1966. S-potentials from luminosity units in the retina of fish (cyprinidae). *The Journal of physiology 185*, 3, 587–599.

NAYAR, S., AND BRANZOI, V. 2003. Adaptive Dynamic Range Imaging: Optical Control of Pixel Exposures over Space and Time. In *IEEE International Conference on Computer Vision (ICCV)*, vol. 2, 1168–1175.

PATTANAIK, S. N., FERWERDA, J. A., FAIRCHILD, M. D., AND GREENBERG, D. P. 1998. A multiscale model of adaptation and spatial vision for realistic image display. *Proceedings of the 25th annual conference on Computer graphics and interactive techniques - SIGGRAPH '98*, 287–298.

PATTANAIK, S. N., TUMBLIN, J., YEE, H., AND GREENBERG, D. P. 2000. Time-dependent visual adaptation for fast realistic image display. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., 47–54.

RIEKE, F., AND RUDD, M. E. 2009. The challenges natural images pose for visual adaptation. *Neuron 64*, 5, 605–616.

SEYMOUR, J. C. 1997. Why do color transforms work? In *Electronic Imaging'97*, International Society for Optics and Photonics, 156–164.

STOCKMAN, A., AND SHARPE, L. T. 2006. Into the twilight zone: the complexities of mesopic vision and luminous efficiency. *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists) 26*, 3 (May), 225–39.

TOCCI, M. D., KISER, C., TOCCI, N., AND SEN, P. 2011. A versatile hdr video production system. In *ACM SIGGRAPH 2011 Papers*, ACM, New York, NY, USA, SIGGRAPH '11, 41:1–41:10.

TU, L. W. 2008. Bump functions and partitions of unity. *An Introduction to Manifolds*, 127–134.

TUMBLIN, J., AND RUSHMEIER, H. 1993. Tone reproduction for realistic images. *Computer Graphics and Applications, IEEE 13*, 6, 42–48.

WANG, S., HOU, T., BORDER, J., QIN, H., AND MILLER, R. 2012. High-quality Image Deblurring with Panchromatic Pixels. *ACM Transactions on Graphics (TOG) 1*, 212.

WEBSTER, M. A. 2011. Adaptation and visual coding. *Journal of vision 11*, 5, 3.