Title
Perceptually Based Tone Mapping for Low-Light Conditions

Permalink
https://escholarship.org/uc/item/9sn0x756

Authors
Kirk, Adam G
O'Brien, and James F

Publication Date
2011-07-01

Supplemental Material
https://escholarship.org/uc/item/9sn0x756#supplemental

Peer reviewed

Perceptually Based Tone Mapping for Low-Light Conditions

Adam G. Kirk   James F. O'Brien

University of California, Berkeley

Images copyright Adam Kirk and James O'Brien.

Figure 1: *Left: A high dynamic range (HDR) image showing UC Berkeley's South Hall captured at night shown without perceptual tone mapping. Center: Perceptual tone mapping for low-light conditions. Right: Perceptual tone mapping for low-light conditions with scene intensities scaled to one-eighth that of the center image.*

Abstract

In this paper we present a perceptually based algorithm for modeling the color shift that occurs for human viewers in low-light scenes. Known as the Purkinje effect, this color shift occurs as the eye transitions from photopic, cone-mediated vision in well-lit scenes to scotopic, rod-mediated vision in dark scenes. At intermediate light levels vision is mesopic with both the rods and cones active. Although the rods have a spectral response distinct from the cones, they still share the same neural pathways. As light levels decrease and the rods become increasingly active they cause a perceived shift in color. We model this process so that we can compute perceived colors for mesopic and scotopic scenes from spectral image data. We also describe how the effect can be approximated from standard high dynamic range RGB images. Once we have determined rod and cone responses, we map them to RGB values that can be displayed on a standard monitor to elicit the intended color perception when viewed photopically. Our method focuses on computing the color shift associated with low-light conditions and leverages current HDR techniques to control the image's dynamic range. We include results generated from both spectral and RGB input images.

Keywords: High dynamic range imaging, tone mapping, human perception, scotopic mesopic photopic vision, Purkinje effect, day-for-night processing.

CR Categories: I.4.3 [Image Processing]: Enhancement—Tone Mapping; I.4.8 [Image Processing]: Scene Analysis—Color.

Links: ◆DL 🔗PDF 🌐WEBSITE AND DATA

1 Introduction

Reproducing the perception of low-light scenes presents challenges due to changes in how the human visual system responds at different light levels. In well-lit scenes, the eye behaves photopically with light perception mediated by the short, medium, and long cone cells. The three types of cone cells have distinct spectral response functions and they allow perception of a three-dimensional color space. In near-dark scenes, the eye functions scotopically, with only the rod cells active. The rod cells have a spectral response function that is distinct from the cones, and when only the rods are active color discrimination is dominated by a single perceptual axis, leading to primarily monochromatic vision. In between the photopic and scotopic regimens are low-light scenes, such as the one shown in Figure 1, where the eye functions mesopically. In mesopic vision all four types of receptors are active and contribute to color perception.

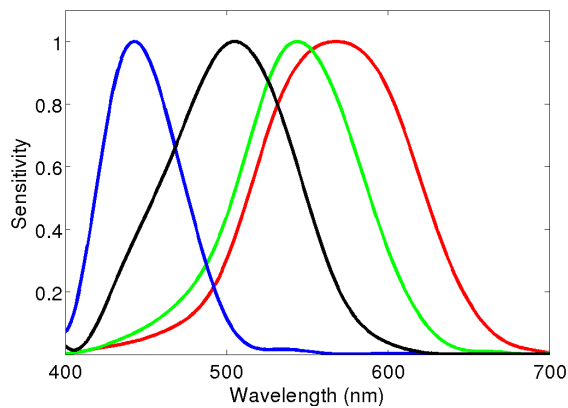Contact email: {akirk, job}@eecs.berkeley.edu

Figure 2: *Plot of the normalized cone and rod spectral sensitivities. Long cone sensitivities are shown in red, medium cone sensitivites in green, short cone sensitivites in blue, and rod sensitivites in black [Bur, 1951; Stockman and Sharpe, 2000].*

Each of these four types of photoreceptor has a spectral response that is linearly independent from the others. (See Figure 2.) This independence implies that the eye measures a four-dimensional colorspace. However, there are only three distinct types of signaling pathways to the brain from the retina. These pathways are dominated by the cones in photopic vision and by the rods in scotopic vision, but in mesopic vision the signals from the rods and cones are combined. Further, the relative weighting in this combination varies across the retina as local illumination varies within the mesopic regime. Therefore, although the eye senses in four-dimensions, communication of color to the brain still occurs within a three-dimensional colorspace. The compression from four to three dimensions is nonlinear and depends locally on the intensity of the scene. Perceptually, this combination of signals manifests as a shift in hue that is commonly known as the Purkinje effect.

Unfortunately, the Purkinje effect is not replicated by current cameras or displays. Because humans photopically perceive a three-dimensional colorspace, commercially available camera and display technologies also operate in three dimensions, for example using three color sensors or three color phosphors. Color space standards, such as ISO RGB or CIE XYZ, explicitly assume that three dimensions are sufficient to model human color perception. Monitors and printers that include more than three primaries typically do so to expand the gamut within this three-dimensional space, not to add extra dimensions of color. These systems assume photopic viewing and cameras/displays have no way to capture/stimulate rod response independent from cones. As a result, low-light images captured by standard cameras either look underexposed or they look very similar to well lit images. Color variations that do occur between photographs of dimly and brightly lit scenes are due to different colored illuminants, and do not account for the perceptual color changes due to the Purkinje effect.

Color shifts associated with mesopic vision are an important lighting cue. Accordingly, photographers and cinematographers have developed ad hoc techniques to mimic color effects of vision under low-light conditions. Known as day-for-night techniques, they include adjusting the color balance or shooting with colored filters. Other approaches filter the color of environmental lighting [Malkiewicz, 1992]. Perceptually motivated models of scotopic and mesopic vision also play an important role in designing spaces and making lighting choices, for both aesthetic and safety reasons. One such example is highway design, where achromatic simula-

tions of mesopic vision are used to measure the effects of lighting conditions on driving performance [Bullough and Rea, 2000].

We present a computational model of the changes to perceived color that occur for low-light scenes, and a perceptually based tone mapping algorithm that generates standard RGB images that appear closer to what the standard viewer would experience in low-light conditions. This algorithm uses only a single point of control, uniform scaling of scene luminance, to determine a color change corresponding to the Purkinje shift. We use the term exposure to refer to this scaling as the apparent effect of uniform doubling/halving of scene luminance would correspond to doubling/halving the length of exposure when taking a photograph. Our algorithm captures low-light perceptual effects by approximating the psychophysical process through which color is perceived. It computes a standard observer's rod and cone receptor responses to a scene, and determines how those responses would be combined to produce the output signals from the eye. Our method then computes the photopic RGB display image that most closely elicits the corresponding cone response.

In order to determine how the human visual system responds to a given scene, our tone mapping algorithm ideally works on spectral input images. While standard images approximate the spectral density function at a pixel using only three channels, a spectral image approximates the continuous distribution of energy at every pixel using a higher dimensional representation. This representation allows the cone and rod responses to be computed directly by integrating their response functions against the pixel's spectral density function.

While spectral images are useful, they are also difficult to acquire. We review several existing acquisition methods and describe our own approach which uses multiple images taken through a series of colored filters. In cases when only a standard RGB image is available, we reconstruct a plausible spectral distribution that would produce the given RGB values. Although this preprocessing allows us to apply our algorithm to RGB input images and produces reasonable results, working with true spectral data is preferred as it captures effects that are lost by premature transition into a three-dimensional color space. In particular, photopic metamers may appear distinct when viewed mesopically or scotopically due to independent rod contribution. Premature projection into a three-dimensional color space precludes estimating a rod response independent from cone responses.

We include results for several images to demonstrate the effectiveness of our algorithms. We also provide supplemental data including measured filter transmission curves and camera sensor responses that may be of use to others working on similar problems.

2 Background

The human eye perceives light using four types of photoreceptor: long, medium, and short cone cells, and rod cells. These types of receptor each detect a specific distribution of wavelengths, and their sensitivity functions are linearly independent. The spectral sensitivities are shown in Figure 2. In addition to spectral sensitivity, the receptors are also characterized by a range of intensities over which they are active. Cones operate in photopic, or well-lit, conditions and rods operate under dark, or scotopic, conditions. Photopic and scotopic illumination levels overlap in a mesopic range, where both rods and cones contribute to color perception.

While we have four types of photoreceptor, we only perceive a three-dimensional colorspace. This limitation exists because the rod cells use the same pathways as the cones for conveying information to the brain. This communication of color to the brain is be-
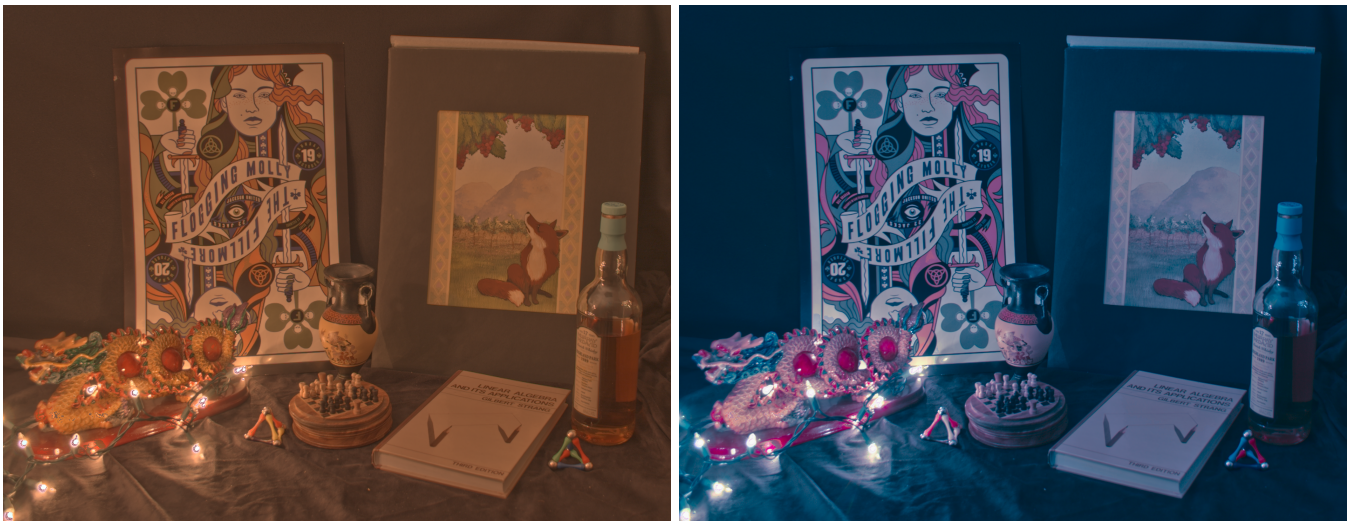
Images copyright Adam Kirk and James O'Brien. Painting untitled by Susan Kim.

Figure 3: *Left: HDR still life scene with no tone mapping. Right: The image has been tone mapped for low-light conditions.*

lieved to be characterized by the opponent color model [Buck et al., 2000]. In scotopic vision, subjects do not perceive gray scale, but instead report the perception of monochromatic scenes. In mesopic vision, subjects experience three-dimensional color perception, albeit with a reduced gamut and nonlinear color shift that varies with intensity [Shin et al., 2004b].

The basic biological mechanism for color perception can be described by the opponent color model. This model states that there are three perception channels. The first is a combination of the long and medium cone cells, which corresponds approximately to the red versus green quality of color. The second is a combination of all three cone cells, which generally describes the blue versus yellow qualities. The final channel is another combination of the long and medium cone cells, and gives an approximation to overall intensity. Experiments show the rod cells also convey information to the brain using all three of these channels. This particular biological configuration explains how the rod cells can create the sensation of color, and it also gives a sense of why the Purkinje phenomenon is difficult to capture with a simple model [Cao et al., 2008].

Our method models the responses of the cones and rods to light in a scene and then uses them to compute perceived color. These responses can be computed directly from a spectral image, which gives an approximate distribution of energy in the visible wavelengths for each pixel. There are several existing methods for acquiring spectral images. Burns and Berns [1996] use a set of band pass filters. They sampled the scene under these filters and solved for the image spectra. Mohan and colleagues [2008] built a system to dynamically select the spectra best associated with their device's input channels. While they get limited data in a single photograph, such a system could be used to construct a spectral image with multiple images. The benefit of their system is that the spectral bands used in reconstruction can be selected dynamically. Schechner and Nayar [2002] show that it is possible to spectrally image a scene using video taken through a spatially varying filter. By panning across the scene, they capture each scene point under the entire range of their filter. They then construct a spectral image using correspondence. Park and colleagues [2007] showed that one can reconstruct the reflectance spectra of objects in a scene by imaging the scene under known multiplexed lighting conditions. This process requires multiple images and a calibrated light source with a controller.

Yasuma and colleagues [2008] simulated a camera designed to capture more than three unique frequency bands. They proposed a

modification to the sensor's Bayer pattern to introduce additional filters at a lower resolution. With such a system, a traditional RGB image can still be reconstructed, and at the same time one can obtain lower-resolution spectral information in a single picture. This approach has the strong advantage of being applicable to scenes that contain moving objects and where the lighting cannot be controlled. The availability of commercial imaging devices that could easily acquire spectral images could potentially have many applications including our low-light tone mapping method.

Our work uses spectral images to perform perceptually based tone mapping. However, others have addressed the issue using regular RGB image data. Tumblin and Rushmeier originally introduced the problem [1993]. Ferwerda and colleagues [1996] model several characteristics of the visual system, including: threshold visibility, color appearance, acuity, and time sensitivity. While this work addresses color change due to decreases in light level, at scotopic levels their model produces a greyscale image, which is not what viewers report. Durand and Dorsey [2000] address this problem by blending a particular color of blue into scenes as light is decreased. Khan and Pattanaik [2004] use a similar approach to Durand and Dorsey. In their work, they hypothesize that the blue shift in moonlit scenes is caused by rod receptors contributing color as if they were short cones. They estimate a cone response from the initial scene and, as the light is reduced, blend an attenuated rod response into the short cone response. Finally, they convert to RGB space. While both Khan and Pattanaik and Durand and Dorsey introduce a color shift, blending a linear amount of a single blue color is inconsistent with psychophysical data.

Pattanaik and colleagues [1998] examined a variety of important perceptual effects related to high-dynamic-range scenes, including effects that occur at low-light levels. They also explicitly consider rod contribution to perception. Our approach to modeling the Purkinje effect goes beyond what they present with a more comprehensive model of the color effects. Kuang and colleagues [2007] presented the iCAM06 model for rendering HDR images. Their framework accounts for rod contribution in low-light conditions, but with a less detailed model than ours that operates directly in RGB space. Both of these methods examined a host of related issues, such as increased noise and loss of spatial acuity at low light levels, that would complement our work, which focuses only on color effects.

In addition to these automated methods, it is possible for an artist to tone map an image for low-light conditions using standard image

Images copyright Adam Kirk and James O'Brien.

Figure 4: *Left: An HDR night scene with no tone mapping featuring UC Berkeley's Sather Tower. Right: The image has been tone mapped for low-light conditions. Artifacts on the clock face occur because the hands moved during image acquisition.*

Images copyright Adam Kirk and James O'Brien.

Figure 5: *Left: The result of a standard grayscale image conversion. Right: Our measure of the mesopic value of a pixel, where lighter indicate a more photopic pixel and darker indicates a more scotopic pixel. Both images have been scaled to fit the range* [0 255] *to highlight differences.*

editing tools. Photo editing tools in the hands of a skilled artist can create stylized effects that can be quite attractive and convey the general impression of low light. However, manual tone mapping requires the artist to make arbitrary decisions on how the colors should be modified and is generally not perceptually correct.

In the vision science literature, researchers have done experiments to characterize the Purkinje effect and proposed theories to explain the underlying mechanism. Shin and colleagues [2004a] have proposed a partial model based on data from user studies. Their data-driven model is based in LMS space. They collected data from several subjects on how color perception shifts at a set of illumination levels that are evenly separated perceptually. They then built a multilinear model based on LMS values as well as illumination level. Cao and colleagues [2008] also fit a model based on user studies, however they use a more detailed model of the biology. As a result, their model is effective at all levels of illumination. Our tone mapping algorithm is based on the Cao model.

3 Methods

Given the spectral distribution of energy for the light at each pixel in an image of a scene, our goal is to compute the perceptual response of the human visual system and then determine a standard RGB image that will elicit a similar response when viewed photopically on a standard monitor. With this system a user can capture a spectral image and use a single exposure parameter to adjust the perceived brightness of the scene. The resulting image colors will take into account the Purkinje effect so that photopic, mesopic, and scotopic vision are modeled correctly. Increasing the exposure value will move the image towards photopic colors, and decreasing the value will move the image towards scotopic colors.

3.1 Acquiring Spectral Images

The first step in this process is to acquire a spectral image of the scene. Our image acquisition method is not the major contribution of this paper and, as mention previously, a wide variety of other methods for spectral image acquisition have appeared in the literature. We build spectral images using multiple RGB images, each taken through a different color filter. Because we know the sensitivity functions of the camera sensors and transmission functions of the filters, we can solve for the unknown spectral distribution of the light entering the camera at each pixel. This method is analogous to

the method of Park and colleagues [2007]. In their work, they image a scene under a series of differently colored lights and solve for surface reflectance. Mathematically the methods are nearly identical, but our method solves for the light entering the camera rather than surface properties and does not require controlled illumination.

Let p_c be the value recorded by a camera for some pixel p in color channel $c \in \{\text{red}, \text{green}, \text{blue}\}$. If the camera sensor's sensitivity function is $\mathcal{S}_c(\lambda)$, then we can write

$$p_c = \int_\Omega \mathcal{S}_c(\lambda)\, \mathcal{R}(\lambda)\, \mathrm{d}\lambda \tag{1}$$

where $\mathcal{R}(\lambda)$ is the spectral distribution of light incident at pixel p, and $\Omega = [400\text{nm } 700\text{nm}]$.

We wish to solve for $\mathcal{R}(\lambda)$, but the three color channel limit of commercial cameras is insufficient for uniquely resolving the function [Kohonen et al., 2006]. Ideally we could add additional channels to the camera to further sample the spectra, but because doing so is infeasible we instead take multiple images with each image taken through a different transmissive filter. If $\mathcal{F}_i(\lambda)$ is the transmission function for one of these filters then the pixel value in an image taken through that filter is given by

$$p_{c,i} = \int_\Omega \mathcal{S}_c(\lambda)\, \mathcal{F}_i(\lambda)\, \mathcal{R}(\lambda)\, \mathrm{d}\lambda \ . \tag{2}$$

To determine the spectral image we need to compute the spectra $\mathcal{R}(\lambda)$ that best predicts all the observed values $p_{c,i}$.

We represent the spectral functions using cubic B-spline basis functions so that, for example, we have

$$\mathcal{R}(\lambda) = \mathcal{B}(\lambda)^\mathsf{T} \boldsymbol{r} \tag{3}$$

where $\mathcal{B}(\lambda)$ is a vector of uniform cubic B-spline basis functions over the interval Ω, $\boldsymbol{r}$ is the vector of control points for $\mathcal{R}(\lambda)$, and $\cdot^\mathsf{T}$ denotes transpose. We use 10 knot points uniformly distributed over Ω. This number was selected as sufficient based on spectrometer measurements from a collection of Munsell color squares. Modern illuminants, such as fluorescent or LED lights, tend to have sharply peaked spectral distributions that would not be well resolved using a smooth B-spline basis. However, we will integrate the distributions against the smooth rod and cone sensitivity functions, so a

smoothed representation of sharp peaks will not adversely impact our results.

With this discretization we can rewrite (2) as

$$p_{c,i} = \int_\Omega \mathcal{B}(\lambda)^\mathsf{T} \boldsymbol{s}_c \; \mathcal{B}(\lambda)^\mathsf{T} \boldsymbol{f}_i \; \mathcal{B}(\lambda)^\mathsf{T} \boldsymbol{r} \; \mathrm{d}\lambda \; . \tag{4}$$

All quantities except $\boldsymbol{r}$ are known, so we move $\boldsymbol{r}$ out of the integral and rewrite as

$$p_{c,i} = \left(\int_\Omega \mathcal{B}(\lambda)^\mathsf{T} \boldsymbol{s}_c \; \mathcal{B}(\lambda)^\mathsf{T} \boldsymbol{f}_i \; \mathcal{B}(\lambda)^\mathsf{T} \; \mathrm{d}\lambda \right) \boldsymbol{r}$$

$$p_{c,i} = \boldsymbol{v}_{c,i}^\mathsf{T} \boldsymbol{r} \tag{5}$$

where $\boldsymbol{v}_{c,i}^\mathsf{T}$ is equal to the expression in parentheses.

We collect all of these equations for each channel and all n filters together as

$$V = \begin{bmatrix} \boldsymbol{v}_{red,1}^\mathsf{T} \\ \boldsymbol{v}_{green,1}^\mathsf{T} \\ \boldsymbol{v}_{blue,1}^\mathsf{T} \\ \boldsymbol{v}_{red,2}^\mathsf{T} \\ \vdots \\ \boldsymbol{v}_{blue,n}^\mathsf{T} \end{bmatrix} \quad \boldsymbol{p} = \begin{bmatrix} p_{red,1} \\ p_{green,1} \\ p_{blue,1} \\ p_{red,2} \\ \vdots \\ p_{blue,n} \end{bmatrix} \; . \tag{6}$$

We then solve for $\boldsymbol{r}$ in a least-squares sense while constraining the solution to be positive using

$$\min_{\boldsymbol{r}} \quad \|\boldsymbol{V}\boldsymbol{r} - \boldsymbol{p}\|_2^2$$

$$s.t. \quad \mathcal{B}(\lambda)^\mathsf{T} \boldsymbol{r} \geq 0 \; . \tag{7}$$

To capture the spectra in a scene, we use high dynamic range (HDR) input images [Debevec and Malik, 1997]. If a sensor becomes saturated it becomes impossible to determine exactly how much energy was incident at that pixel. In our examples, we captured HDR images under each transmissive filter, registered the HDR images to each other, and normalized them all to a one second exposure.

All images were captured using a Canon EOS 300D in RAW format and without gamma correction so that the data was in a linear energy space. The camera's spectral sensitivity was determined experimentally by imaging a known broad-spectrum light source through a set of known transmissive filters. The light source was a Koehler Illuminant with a SoLux Q50MR16/CG/47/36 4700 Kelvin Halogen Bulb driven by a regulated DC power source. The filters used for calibration were a set of Roscolux polycarbonate filters.

The spectral image data were captured for static scenes using Cokin P-series filters, specifically filters P001, P002, P003, P004, P005, P006, P020, P036, P047, P050, and P231. The goal in selecting these filters was to find a linearly independent set that sufficiently spanned the visible wavelengths. We determined the spectral transmission of each filter as a percentage of incident light using an Edmund Optics Visible Wavelength USB CCD Spectrometer, model BRC111A-USB-VIS, which has a wavelength resolution of 1 nm. A sample input series with an inset plot of percentage of light transmitted per wavelength by the filter can be seen in the left column of Figure 11.

With a three channel camera sensor and eleven filters we have 33 independent measurements of the $\mathcal{R}(\lambda)$ which is somewhat excessive for estimating the four values of the rod and cone responses at each pixel. However, these spectral images have uses in other applications. The camera and filter calibration data and the raw image data is included in the supplemental data for this paper.

Images copyright Adam Kirk and James O'Brien.

Figure 6: *Top: An HDR image of the Fremont Troll with no tone mapping. Bottom: The image has been tone mapped for low-light conditions.*

3.2 Tone Mapping for Low-Light Conditions

Once we have acquired spectral image data, the first step of our tone mapping procedure is to determine the cone and rod response to the acquired data. To convert the spectral image to LMSR space, we integrate each pixel's spectral distribution against the standard rod and cone response curves [Bur, 1951; Stockman and Sharpe, 2000]. If a pixel has spectrum $\mathcal{R}(\lambda)$ and the normalized response curve of the eye for a given receptor is $\mathcal{E}_j(\lambda)$, where $j \in \{\text{Long, Medium, Short, Rod}\}$, then the normalized four-dimensional LMSR response $\boldsymbol{q}$ can be calculated as

$$q_n = \int_\Omega \mathcal{E}_n(\lambda) \, \mathcal{R}(\lambda) \, \mathrm{d}\lambda \tag{8}$$

We compute this integral using the same B-spline representation described in the previous section.

At this stage, the image represents the rod and cone responses that would result from the scene being viewed by a standard observer. We use this information to determine the Purkinje Shift, which we model as an offset to the long, medium, and short channels such that the result is the perceptually closest photopic signal to the mesopic signal [Cao et al., 2008]. More specifically, let $\hat{\boldsymbol{q}}$ be a three-dimensional long, medium, and short cone response. We

Images copyright Adam Kirk and James O'Brien.

Figure 7: *Top: An HDR image of the Seattle skyline as seen from Kerry Park at night with no tone mapping. Bottom: The image is tone mapped for low-light conditions. Some differences are due to lights being switched on and off in the scene during acquisition.*

show how to compute Δq such that

$$\hat{q} = [q_{\text{Long}} \ q_{\text{Medium}} \ q_{\text{Short}}]^{\mathsf{T}} + \Delta q \qquad (9)$$

where $\hat{q}$ will be the three-dimensional cone response most perceptually similar to the four-dimensional cone and rod response q.

As discussed previously, rods and cones share neural pathways. The signals passing through those pathways are a combination of rod and cone receptors, subject to sensitivity regulation [Miyahara et al., 1993]. The regulated signals g are

$$g_{\text{Long}} = 1/(1 + 0.33(q_{\text{Long}} + \kappa_1 q_{\text{Rod}}))^{0.5}$$
$$g_{\text{Medium}} = 1/(1 + 0.33(q_{\text{Medium}} + \kappa_1 q_{\text{Rod}}))^{0.5} \qquad (10)$$
$$g_{\text{Short}} = 1/(1 + 0.33(q_{\text{Short}} + \kappa_2 q_{\text{Rod}}))^{0.5}$$

where κ_1 and κ_2 determine the rod contribution to the neural signals [Cao et al., 2008]. These mixing terms depend on the mesopic viewing conditions. They are near zero during photopic vision, and reach an approximate value of $\kappa_1 = 0.25$ and $\kappa_2 = 0.4$ during fully adapted scotopic vision [Cao et al., 2008]. For our results we assume full adaptation and use the above values. We found that modulating κ_1 and κ_2 according to standard mesopic adaptation produced a system that was too sensitive to exposure and made it difficult to obtain a desired result. We could construct a system with a similar response but higher κ_1 and κ_2 fidelity by giving the user a non-linear exposure control.

As discussed earlier, our tone mapping procedure is based on rod contribution causing a shift in the opponent color model space. The opponent color model expresses color as three different combinations of the cone sensors. For a pixel in LMS space, $\hat{q}$, the corre-

sponding pixel in opponent color space o is computed as:

$$o_{\text{Red/Green}} = \hat{q}_{\text{Medium}} - \hat{q}_{\text{Long}}$$
$$o_{\text{Blue/Yellow}} = \hat{q}_{\text{Short}} - (\hat{q}_{\text{Long}} + \hat{q}_{\text{Medium}}) \qquad (11)$$
$$o_{\text{Luminance}} = \hat{q}_{\text{Long}} + \hat{q}_{\text{Medium}}$$

where the first component of o roughly measures depth of red versus depth of green, the second channel measures depth of blue versus depth of yellowness, and the third channel measures luminance. Note that this transformation can be expressed as an invertible matrix such that $o = A\hat{q}$.

To determine the amount of color shift, each term in the opponent color model has a weighted amount of the rod response added. The weight is a function of the sensitivity regulated receptor response g. Specifically, the color shift in the opponent color model Δo is:

$$\Delta o_{\text{Red/Green}} = x \kappa_1 \left(\rho_1 \frac{g_{\text{Medium}}}{m_{\text{max}}} - \rho_2 \frac{g_{\text{Long}}}{l_{\text{max}}} \right) q_{\text{Rod}}$$

$$\Delta o_{\text{Blue/Yellow}} = y \left(\rho_3 \frac{g_{\text{Short}}}{s_{\text{max}}} - \rho_4 \left(\alpha \frac{g_{\text{Long}}}{l_{\text{max}}} + (1 - \alpha) \frac{g_{\text{Medium}}}{m_{\text{max}}} \right) \right) q_{\text{Rod}}$$

$$\Delta o_{\text{Luminance}} = z \left(\alpha \frac{g_{\text{Long}}}{l_{\text{max}}} + (1 - \alpha) \frac{g_{\text{Medium}}}{m_{\text{max}}} \right) q_{\text{Rod}}$$

$$(12)$$

Where $l_{\text{max}} = 0.637$, $m_{\text{max}} = .392$, and $s_{\text{max}} = 1.606$ are the maximum values of the cone fundamentals [Stockman and Sharpe, 2000], and psychophysical data [Cao et al., 2008] is used to fit the ρ and α as: $\rho_1 = 1.111$, $\rho_2 = 0.939$, $\rho_3 = .4$, $\rho_4 = 0.15$, and $\alpha = .619$. The above leaves free parameters x, y, and z. These parameters control the scale of the shift for each of the channels in the opponent color model. Psychophysical data suggests that x and

y are the same, but z is independent [Cao et al., 2008]. For all the examples in this paper, we used $x = y = 15$ and $z = 5$. Returning to Equation 9, we have

$$\hat{q} = [q_{\text{Long}} \, q_{\text{Medium}} \, q_{\text{Short}}]^{\mathsf{T}} + A^{-1} \Delta o \qquad (13)$$

giving the photopic response $\hat{q}$ most perceptually similar to the original mesopic response.

The final step is to convert from LMS space into RGB space for display. At this stage, we know the desired excitation of cones in the retina and we seek to find the display settings that produce that excitation. This process is essentially a color matching problem where we need to know the mapping from the display's primaries, $\mathcal{M}_{\text{Red,Green,Blue}}(\lambda)$, to viewer stimulus. These spectra are device dependent and one can map values for one display onto another using standard color matching functions. Further, cone responses depend on adaptation to the ambient viewing conditions [Mantiuk et al., 2009]. Our results were computed using the measured spectral output of an Apple Cinema HD Display with the default display profile, and we assume viewing in a dark room. Common experience shows that viewing RGB images calibrated for one display on different display typically produces reasonable results, but applications demanding precision should use color matching functions appropriate to the actually display device, display driver color profile, and viewing conditions.

Let M be the matrix that maps the display's spectral emission onto the cone fundamentals

$$M = \begin{bmatrix} m_{\text{Long,Red}} & m_{\text{Long,Green}} & m_{\text{Long,Blue}} \\ m_{\text{Medium,Red}} & m_{\text{Medium,Green}} & m_{\text{Medium,Blue}} \\ m_{\text{Short,Red}} & m_{\text{Short,Green}} & m_{\text{Short,Blue}} \end{bmatrix}, \qquad (14)$$

with the components of M given by $m_{x,y} = \int_{\Omega} \mathcal{E}_x(\lambda) \, \mathcal{M}_y(\lambda) \, d\lambda$. We solve for the final RGB settings p^* that maps to our desired photopic response $\hat{q}$ with

$$\min_{p^*} \quad \|M p^* - \hat{q}\|_2^2$$
$$s.t. \quad p^* \geq 0 \qquad (15)$$

When solving for a given pixel, we first try the unconstrained solution, $p^* = M^{-1}\hat{q}$. In most cases, the result of the unconstrained problem satisfies the constraint. In the event that it doesn't, a solution is found by solving a small quadratic programming problem.

The above optimization converts pixels from LMS space into RGB space, but the values still need to be compressed to the appropriate dynamic range. For photopic images, the target dynamic range would be the full range of the display. However, low-light scenes are generally perceived as dim because illumination that falls in the mesopic or scotopic range doesn't approach the saturation level of cone receptors. As a result, images tone mapped for low light levels look odd when displayed on a bright monitor using the device's full range. We account for this effect by mapping to a reduced dynamic range. This method is similar to standard cinema approaches which use neutral density filters or shortened exposures in filming what should appear to be night scenes [Malkiewicz, 1992].

Our range reduction technique starts with standard HDR compression. We use the HDR compression algorithm of Durand and Dorsey [2002] implemented using the bilateral filter approach developed by Paris and Durand [2006]. From here, our operator further compresses pixels in the low dynamic range (LDR) image based on their mesopic levels. Our measure of mesopic level is based on rod attenuation in the rod-mediated opponent color model. Consider $\Delta o_{\text{Luminance}}$ in Equation 12. Rod contribution to $\Delta o_{\text{Luminance}}$ is at a minimum during photopic vision, which corresponds to the

Images copyright Adam Kirk and James O'Brien.

Figure 8: *Left Column: Spectral image tone mapped for low-light conditions. Right Column: Non-spectral HDR image of the same scene mapped into LMSR space before applying our tone mapping procedure. For both columns, exposure decreases from top to bottom. This scene was not used in training our mapping from RGB space to LMSR space.*

rod multiplier on $\Delta o_{\text{Luminance}}$ evaluating to zero. If we pull out this scalar factor as

$$w = \left(\alpha \frac{g_{\text{Long}}}{l_{\text{max}}} + (1 - \alpha) \frac{g_{\text{Medium}}}{m_{\text{max}}} \right), \qquad (16)$$

then a pixel is fully photopic when $w = 0$. The larger w grows, the further into the scotopic range we can consider the pixel and the more we reduce the dynamic range. By examining the average value of w across an image for a variety of exposures, we can pick a constant β for which we consider the pixel fully scotopic. In our examples, $\beta = 1$. Finally, we choose a parameter $\gamma \in [0\ 1]$ which is the maximum ratio by which we reduce the range of a scene. In our examples, γ ranges between 0.25 and 0.5. Given these parameters, the maximum value x for a pixel with mesopic factor w is $x = \max(1 - \frac{w}{\beta}(1 - \gamma), \gamma) * v$, where v is the maximum value of the display's dynamic range, normally 255. A pixel with mesopic factor w is then linearly mapped into the range $[0\ x]$. See Figure 5 for a comparison of the mesopic factor to a standard grayscale image.

4 Approximation for Non-Spectral Images

Spectral images contain information necessary to compute rod and cone responses for the tone mapping procedure we have described, however they also require special effort to acquire which may not always be feasible. For cases where spectral information isn't available, we describe a data-driven alternative that, when given an RGB value, infers a plausible LMSR value. This procedure requires lofting from a three-dimensional space to a four-dimensional one, and

Original image from the OpenEXR dataset, copyright Industrial Light and Magic.

Figure 9: *Left: A non-spectral HDR image. Right: The result of our tone mapping applied to the image mapped into LMSR space.*

in the process some extra information must be generated. In particular, while the standard RGB primaries allow a good estimate of cone responses, they don't provide enough information for an independent estimate of rod response.

Other researchers have reported that the reflectance spectra for many common scenes can be fairly well approximated using a small number of bases [Maloney and Wandell, 1987]. This result implies that for many scenes the rod response is strongly correlated with the cones. For plain RGB images we take advantage of this correlation by using a linear mapping from RGB space to LMSR space. For the examples in this paper shown in Figures 8, 9, and 10 we build the mapping using the spectra we acquired for other images. Alternatively, one could build the mapping using existing material databases [Kohonen et al., 2006].

To build the mapping we solved for the matrix H that best satisfies $Q = HP$, where Q is a matrix of LMSR values from observed spectra and P is a matrix of the corresponding RGB values. For our data, this fit had high residual indicating that rod responses generally are linearly independent from the cones. However, it is not clear how sensitive viewers would be to the types of errors introduced, particularly if they have never viewed the scene firsthand under scotopic or mesopic conditions. Furthermore, mesopic vision is characterized by a reduced gamut which appears to make many errors less noticeable. We found the results to be generally satisfactory for low light levels, but artifacts appear in brighter scenes. Luckily we have the original RGB values as a photopic reference. Based on this reasoning, we propose the following pipeline for non-spectral images. First, generate LMSR data as described above. Next, tone map the LMSR data to produce an initial mesopic image. Finally, for each pixel blend between the source image and the mesopic image based on w to produce the final result.

5 Results and Discussion

We present several scenes for evaluation. None of the images in this paper have been subjected to any photo-processing other than the methods described, plus cropping, rotating, and downsampling the final images for publishing. The raw input images and resulting spectral images are included in the supplemental data for this paper. We also include LMSR images and standard RGB HDR images.

Figure 11 shows a visual comparison of the steps in our method. The left column shows our input sources. Each image in the left column is an HDR image under a different transmissive filter. The second through fifth columns demonstrate different aspects of our methods as exposure varies. Each row in the second through fifth columns has the same exposure. The second column shows the result of our range reduction technique without performing tone mapping. Note that even in the darkest image it is still possible to distinguish color. The third column shows how our mesopic factor changes with exposure. The fourth column shows the result of our tone mapping procedure mapped to the full range of the dis-

play. These images clearly demonstrate distinctive hue shifts. The final results of our full algorithm are shown in the fifth column. In the fourth and fifth column, note that similar reds in the apple and mango map to drastically different colors in the shortest exposure.

The majority of the images in this paper demonstrate our tone mapping on spectral images. In Figure 3, we show how our tone mapping process varies with proximity to lighting. In the original image, the reddish-brown of the Scotch whisky in the bottle furthest from the light source deepens considerably, while the reds in the dragon remain bright. Figure 4 shows an outdoor scene. Neighboring pixels can have extreme lighting discontinuities, and the contrast between the trees and the tower shows our method respects these discontinuities. The well-lit tree between the tower and the building as well as the white walls of the building are good examples of how different colors shift in the transition from photopic to scotopic light conditions. Figure 6 shows an outdoor statue captured at night. The tone mapped version is not only closer to the viewer's experience in near dark, but also demonstrates a useful artistic effect. Figure 7 shows a night skyline. The original HDR image contains a large amount of light scattering. The tone mapped version demonstrates a reduction in the perceptual effect of this light. Figure 1 shows another night scene. The original image contains little contrast between the doorway and the walls of the building, while there is a strong change in hue in the tone mapped image. Comparing the hue shift for the red brick to the hue change for the bushes and trees demonstrates the difference in relative hue change for different colors. The red brick underneath the street light contains another example of hue change as a function of proximity to light sources.

Finally, we demonstrate our tone mapping process using non-spectral images as input. In Figure 8 we compare results from tone mapping spectral data and plain RGB data. Note that while spectral data has a smoother transition as the exposure decreases, the RGB data results in a similar scotopic image and still demonstrates hue change based on proximity to light sources. Figures 9 and 10 compare standard HDR images with our tone mapped versions. Note that bright areas in the HDR images and tone mapped images are similar, while the differences in the shadows are more significant.

The human visual system has many adaptive mechanisms, and our work focuses on just one, namely the interplay between rods and cones in the early stages of vision. Subsequent to the processing that occurs in the eye there are other adaptive mechanisms, such as color constancy, that are not incorporated into our perceptual mapping. Furthermore, there are additional scotopic and mesopic effects that we chose not to model. These include well-studied effects such as the loss of acuity and increased noise [Ferwerda et al., 1996]. If desired, these effects could be combined with our perceptual tone mapping method.

One somewhat arbitrary aspect of our method is that the range of intensities in an image can be scaled however one desires. For example, one could adjust the exposure parameter so that a bright, sunlit scene was rendered with scotopic treatment. Generally doing so will not produce the impression of a nighttime scene without additional processing. Among other issues, the color of the illumination from the sun would be inconsistent with a night scene as would the detail and color in the sky. A less extreme example appears in Figure 1. The lamp in the center of the scene and other light sources provided a fair amount of illumination, and the middle image is most consistent with our firsthand impression of the scene. The right image takes some artistic license and corresponds to a nonexistent location with much dimmer lighting.

The methods we have described provide a perceptually motivated model of how human viewers perceive low-light scenes. We believe

Original image "Foggy Night" copyright Jack Tumblin, Northwestern University.

Figure 10: *Left: A non-spectral HDR image. Right: The result of our procedure applied to the image mapped into LMSR space.*

that they will have uses in applications where one wishes to predict actual viewer experiences, convey the correct impression of low-light scene, or generate artistic effects.

Acknowledgements

References

BUCK, S. L., KNIGHT, R. F., AND BECHTOLD, J. 2000. Opponent-color models and the influence of rod signals on the loci of unique hues* 1. *Vision research 40*, 24, 3333–3344.

BULLOUGH, J. D., AND REA, M. S. 2000. Simulated driving performance and peripheral detection at mesopic and low photopic light levels. *Lighting Research and Technology 32*, 4, 194.

BUREAU CENTRAL DE LA CIE. 1951. CIE Proceedings (1951), vol. 3.

BURNS, P. D., AND BERNS, R. S. 1996. Analysis of multispectral image capture. In *Proc. Fourth Color Imaging Conference: Color Science, Systems, and Applications*, 19–22.

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.

DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *Proceedings SIGGRAPH 97*, 369–378.

DURAND, F., AND DORSEY, J. 2000. Interactive tone mapping. In *Proceedings of the Eurographics Workshop on Rendering Techniques*, 219–230.

DURAND, F., AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. In *Proceedings SIGGRAPH 2002*, 257–266.

FERWERDA, J. A., PATTANAIK, S. N., SHIRLEY, P., AND GREENBERG, D. P. 1996. A model of visual adaptation for realistic image synthesis. In *Proceedings SIGGRAPH 96*, 249–258.

KHAN, S. M., AND PATTANAIK, S. N. 2004. Modeling blue shift in moonlit scenes by rod cone interaction. *Journal of Vision 4*, 8.

KOHONEN, O., PARKKINEN, J., AND JÄÄSKELÄINEN, T. 2006. Databases for spectral color science. *Color Research and Application 31*, 5 (Oct.), 381–390.

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*, 406–414.

MALKIEWICZ, K. 1992. *Film Lighting: Talks with Hollywood's Cinematographers and Gaffers*. Simon and Schuster.

MALONEY, L. T., AND WANDELL, B. A. 1987. Readings in computer vision: issues, problems, principles, and paradigms. ch. Color constancy: a method for recovering surface spectral reflectance, 293–297.

MANTIUK, R., REMPEL, A. G., AND HEIDRICH, W. 2009. Display considerations for night and low-illumination viewing. In *Applied Perception in Graphics and Visualization*, 53–58.

MIYAHARA, E., SMITH, V. C., AND POKORNY, J. 1993. How surrounds affect chromaticity discrimination. *Journal of the Optical Society of America A 10*, 4, 545–553.

MOHAN, A., RASKAR, R., AND TUMBLIN, J. 2008. Agile spectrum imaging: Programmable wavelength modulation for cameras and projectors. *Computer Graphics Forum 27*, 2, 709–717.

PARIS, S., AND DURAND, F. 2006. A fast approximation of the bilateral filter using a signal processing approach. In *European Conference on Computer Vision*.

PARK, J., LEE, M., GROSSBERG, M. D., AND NAYAR, S. K. 2007. Multispectral Imaging Using Multiplexed Illumination. In *IEEE International Conference on Computer Vision (ICCV)*.

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. In *Proceedings SIGGRAPH 98*, 287–298.

SCHECHNER, Y. Y., AND NAYAR, S. K. 2002. Generalized Mosaicing: Wide Field of View Multispectral Imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 10 (Oct.), 1334–1348.

SHIN, J. C., MATSUKI, N., YAGUCHI, H., AND SHIOIRI, S. 2004. A color appearance model applicable in mesopic vision. *Optical Review 11*, 4, 272–278.

SHIN, J. C., YAGUCHI, H., AND SHIOIRI, S. 2004. Change of color appearance in photopic, mesopic and scotopic vision. *Optical Review 11*, 4, 265–271.

STOCKMAN, A., AND SHARPE, L. T. 2000. The spectral sensitivities of the middle-and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research 40*, 13, 1711–1737.

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

YASUMA, F., MITSUNAGA, T., ISO, D., AND NAYAR, S. K. 2008. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. Tech. rep., Nov.
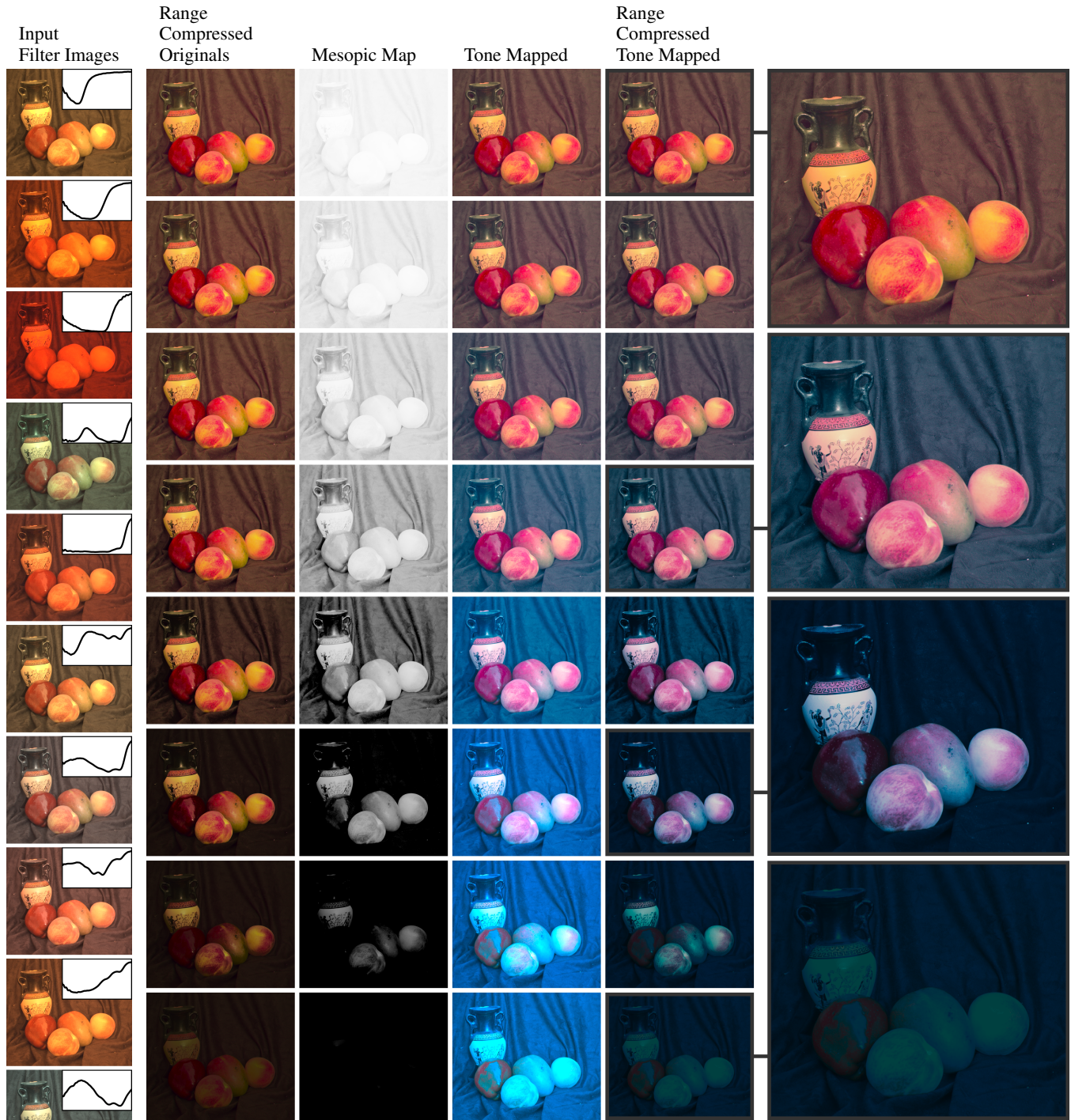
Images copyright Adam Kirk and James O'Brien.

Figure 11: *The first column on the left shows our input images. Each of these images is taken under a different filter, and the transmission properties of the filter are shown inset as a percentage of light transmitted versus wavelength from 400 nm to 700 nm. All rows in the second through fifth columns are the same exposure, which decreases from top to bottom. The second column shows the original image with our reduced range technique to simulate darkness. Note the lack of color shift. The third column shows an image representing the mesopic blend, where light indicates more photopic and dark indicates more scotopic pixels. The fourth column shows the original scene with a color shift, but the scene is mapped to the full dynamic range. The fifth column shows our range reduction technique applied to the fourth column. The images on the far right are higher resolution views of selected images in the fifth column.*