TSTM: A TWO-STAGE TONE MAPPER COMBINING VISUAL ADAPTATION AND LOCAL CONTRAST ENHANCEMENT

By

Sira Ferradans

Edoardo Provenzi

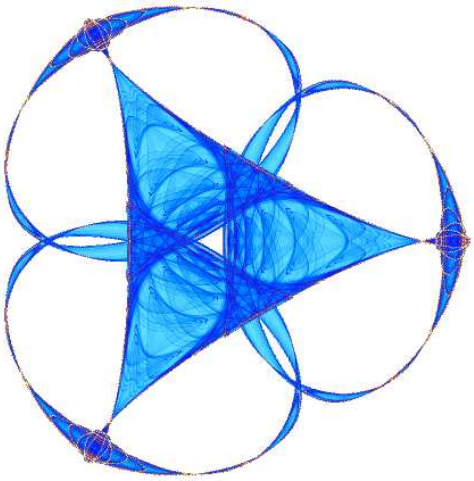
Marcelo Bertalmío

and

Vicent Caselles

IMA Preprint Series # 2253

(May 2009)



INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

UNIVERSITY OF MINNESOTA

400 Lind Hall

207 Church Street S.E.

Minneapolis, Minnesota 55455–0436

Phone: 612-624-6066 Fax: 612-626-7370

URL: http://www.ima.umn.edu

TSTM: A two-stage tone mapper combining visual adaptation and local contrast enhancement

Sira Ferradans, Edoardo Provenzi, Marcelo Bertalmío and Vicent Caselles *

Departamento de Tecnologías de la Información y las Comunicaciones, Universitat Pompeu Fabra,
Tànger 122-140, 08018, Barcelona, Spain.

Abstract

We present TSTM, a Two-Stage Tone Mapping model for high dynamic range (HDR) images inspired by two basic human visual features: visual adaptation and local contrast enhancement. The first property is usually implemented with the so-called Naka-Rushton equation, a rational transformation which permits to properly compress the range of the input HDR image. We provide a fundamental extension of this formula, emphasizing the role played by contrast perception. The second stage consists in a variational method able to perform local contrast enhancement consistent with human vision phenomenology. TSTM has computational complexity $\mathcal{O}(N \log N)$, being N the number of input pixels, and it is free from halos and artifacts. Tests and comparisons with other tone mapping algorithms are also provided and commented.

CR Categories: I.4.3 [Image Processing]: Enhancement—; I.4.8 [Scene Analysis]: Range Data— [I.2.10]: Vision and Scene Understanding—Intensity, color, photometry and thresholding

Keywords: High dynamic range images, tone mapping, Naka Rushton equation, visual adaptation, Weber-Fechner's contrast, variational contrast enhancement.

1 Introduction and State of the Art

High Dynamic Range images constitute a powerful tool to store real-world radiance values and they have become more and more accessible during the last fifteen years. In photography, the great popularity of these images is motivated by the possibility to easily create a HDR image by combining multiple photos of a static scene taken with different time exposures [Debevec and Malik 1997]. In computer graphics, the use of HDR images has provided a much realistic synthesis of artificial scenarios. Since the real-world radiance can span many orders of magnitude, a further 'compression step' is required to properly visualize the information contained in the HDR images through displays or printers. The compression can be performed on different basis: one might want to maximize detail visibility (in a particular zone or in the whole image) or one might want to emulate as much as possible *the contrast and color sensation* of the real-world scene, i.e. achieve an image that looks natural. When the aim is the latter, the range compression procedure is called 'Tone Mapping' (TM) or 'Tone Reproduction' (TR) [Ward et al. 1997].

An intrinsic difficulty in the TM problem is that it is very complicated to compare the real scene sensation with the tone mapped HDR image. In fact, the image size and the ambient conditions are, in general, very different. Up to this date, there is no universally accepted measure to quantify tone mapping fidelity with respect to the real-world sensation, so TM results are usually judged according to subjective taste. To bypass this fundamental problem, it is usually assumed that a TM operator which mimics the ability of the Human Visual System (HVS) to compress radiance would return Low Dynamic Range (LDR) images close to our very perception.

The present paper is in line with this assumption; however, we must prevent the reader that *we do not intend to give a complete or exact representation neither of the HVS, nor of some of its parts*. The reason is that the knowledge about human vision is still too vague to allow building a precise HVS model. As a consequence, we will provide a simplified interpretation of the HVS behavior by considering only two stages of human perception: visual adaptation and local contrast enhancement. Correspondingly, the model that we are going to present in this paper will implement these two HVS features independently. We are aware that we are not considering in our model processes as important to vision as time adaptation, different absorption response of the three types of cones, the influence of other cells such as amacrines, bipolars, or horizontals, the interaction between cones and rods, etc.

The visual adaptation phase occurs mostly in the retina, where photoreceptors (cones and rods) strongly compress the light signal range, and it is described by the Naka-Rushton equation [Shapley and Enroth-Cugell 1984]. We will underline a basic drawback of this formula that limits its applicability and propose a modified Naka-Rushton equation consistent with contrast perception experiments and therefore more suitable for tone mapping.

Besides visual adaptation, the HVS has spatially local features. One of the most important among them is local contrast enhancement, i.e. the HVS ability to magnify local details, a phenomenon clearly exhibited e.g. by the Mach Band effect and the simultaneous contrast effect [Gonzales and Woods 2002]. The consequence of enhancing local contrast is a color correction which permits to strongly reduce a possible color cast and to improve detail visibility. This color correction property of the HVS has been referredf to in the literature as "chromatic adaptation", i.e. the HVS not only adapts to lightness but also to the over-all color of the scene [Wyszecky and Stiles 1982]. We will reproduce these properties in the second step of our model by using a variational algorithm [Palma-Amestoy et al. 2009] that is in accordance to human vision phenomenology.

Computationally speaking, the original HDR image is the input for the visual adaptation stage of our algorithm whose output in real time is an image that shows a perceptually adequate contrast without halos, artifacts or noise. This image is still HDR (it has not been quantized) and is the input for the contrast enhancement stage. This second step does not introduce noise while enhancing details and has computational complexity $\mathcal{O}(N \log N)$, being N the number of input pixels. In the authors' opinion, the final results compare well to the state of the art which is briefly summarized in the next subsection.

1.1 State of the art in TM techniques

Many tone mapping operators have been proposed in the literature; for a thorough review and analysis of the state of the art we refer

*e-mail: sira.ferradans@upf.edu, edoardo.provenzi@upf.edu, marcelo.bertalmio@upf.edu, vicent.caselles@upf.edu. Tel. +34 93 542 2937. Fax. +34 93 542 2569.

to the excellent book [Reinhard et al. 2005]. Here we just want to give a very brief overview of the different schools of techniques that have been proposed so far.

The first works on TM used Stevens' law [Stevens and Stevens 1963] to achieve range compression [Tumblin and Rushmeier 1993; Chiu et al. 1993; Ward 1994]. A rational function very close to Naka-Rushton's formula was used for the first time in [Schlick 1994] showing improvements with respect to Stevens' law. More sophisticated vision models were taken into account in [Ferwerda et al. 1996] and [Ward et al. 1997]. A modified version of the Retinex model of color vision was optimized and applied as a tone mapping operator in [Jobson et al. 1997]. The idea to employ the Naka-Rushton equation to perform TM was presented for the first time in [Pattanaik et al. 2000] and, more recently, in [Reinhard 2005; Tamburrino et al. 2008]. Finally, the anchoring theory of visual perception was used in [Krawczyk et al. 2005].

In the literature there are also radiance compression techniques whose aim is not perceptual fidelity, but detail maximization. The common idea is to shrink large gradient variations while preserving small fluctuations, which correspond to fine details. The way in which this compression is performed constitutes the main difference among these models: in [Tumblin and Turk 1999] the authors used a hierarchical method based on a Partial Differential Equation (PDE) inspired by anisotropic diffusion, later improved in [Durand and Dorsey 2002] with the use of bilateral filtering. Instead, in [Ashikhmin 2002], [Fattal et al. 2002] and [Mantiuk et al. 2006], spatially varying compression factors were used to implement suitable manipulations of the gradient field in order to achieve the tone mapped image. Finally, in [Lischinski et al. 2006] the authors propose an interactive method that allows the user to create better subjective results.

2 Visual adaptation and the Naka-Rushton formula

Let us begin this section by recalling how the retina responds to light stimuli. The range of radiances over which the HVS can operate is very large: from 10^{-6}cd/m^2 (scotopic limit) to 10^6 cd/m^2 (glare limit) [Pratt 2007]. The automatic process that allows the HVS to operate over such a huge range is called visual adaptation [Shapley and Enroth-Cugell 1984].

It is important to stress that the HVS cannot operate over its entire range simultaneously. Rather, it adapts to an average intensity and handles a smaller magnitude interval. There is no complete agreement in the literature about the precise value of this range, which can vary from two ([Shapley and Enroth-Cugell 1984] (pag.326)) to four ([Keener and Sneyd 2008] (pag.670)) orders of magnitude.

Neuroscience experiments show that visual adaptation occurs mainly in the retina. The experiments to measure this behavior were performed using very simple, non-natural images: on a uniform background were superimposed brief pulses of light with intensity $\mathcal{I}$. When a photoreceptor absorbs $\mathcal{I}$, the electric potential of its membrane changes accordingly to the empirical law known as the Naka-Rushton equation [Shapley and Enroth-Cugell 1984]:

$$r(\mathcal{I}) \equiv \frac{\Delta V(\mathcal{I})}{\Delta V_{\max}} = \frac{\mathcal{I}}{\mathcal{I}+\mathcal{I}_s}, \tag{1}$$

where $\mathcal{I}_s$ is the light level at which the photoreceptor response is half maximal, called *semisaturation level* and which is usually associated with the level of adaptation. The change of electric potential $\Delta V(\mathcal{I})$ is the photoreceptor's physiological response to $\mathcal{I}$, which generates an electric current that propagates towards the brain. Finally, $\Delta V_{\max}$ is the highest difference of potential that can be generated. The graph of the function $r(\mathcal{I})$ is depicted in Fig. 1 (a).

Let us notice that, since $\mathcal{I}$ and $\mathcal{I}_s$ are light levels, they are positive and therefore the right hand side of Naka-Rushton's formula (1) belongs to $[0, 1]$, independently of the range of the light stimuli.

The Naka-Rushton equation describes the behavior of the HVS at any specific adaptation level.

2.1 Brightness perception: Weber-Fechner's contrast and Naka-Rushton's equation

In the mid-nineteenth century, the German physician E.H. Weber conducted psychophysical experiments with a similar set-up (flashes of light on a uniform background) as the ones described above, but instead of measuring the electric response inside the retina he applied a phenomenological approach by asking the subject when the difference between the background light $\mathcal{I}$ and the superimposed light $\mathcal{I} + \Delta\mathcal{I}$ was noticeable. The minimum difference $\Delta\mathcal{I}$ which the subject is able to perceive is called the Just Noticeable Difference, JND. Weber found out that the ratio between the JND and the background intensity $\mathcal{I}$ is constant for a wide range of values of $\mathcal{I}$, which is expressed in what is known as *Weber's Law*:

$$\frac{\Delta\mathcal{I}}{\mathcal{I}} = k, \tag{2}$$

where $k > 0$ is a perceptual constant called *Weber fraction*.

In log-log units, the relationship between $\Delta\mathcal{I}$ and $\mathcal{I}$ is linear, with slope 1: $\log(\Delta\mathcal{I}) = \log(\mathcal{I}) + \log(k)$. However, Weber's Law does not hold for low intensity values, where the slope tends to zero instead of one. To account for this, Weber's colleague G. Fechner introduced the concept of 'dark light', thus modifying Weber's Law as follows:

$$\frac{\Delta\mathcal{I}}{\mathcal{I}+m} = k, \tag{3}$$

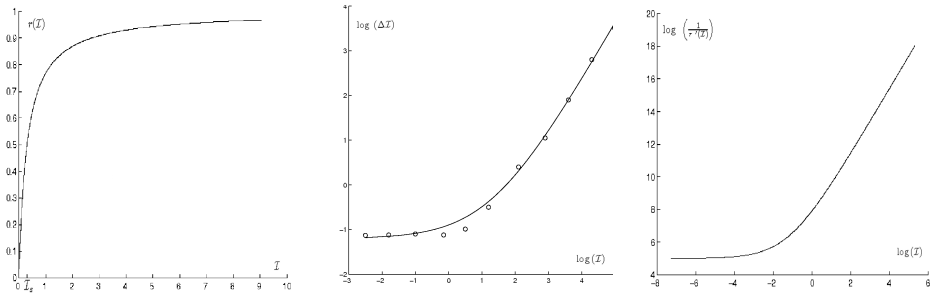
where $m > 0$ is a quantity often interpreted as internal noise in the visual mechanism, e.g. quoting [Chalupa and Werner 2003] (pag. 859) "an intrinsic activity [...] within the receptor systems that combines with the excitation produced by the background to raise the threshold". This last equation is commonly called Weber-Fechner's Law. Now, when $\mathcal{I} >> m$ the slope of $\log(\text{JND})$ as a function of $\log(\mathcal{I})$ is 1; but when $\mathcal{I} << m$ the slope is close to zero and the curve matches the experimental data also at low intensity values. See Fig. 1(b).

We now follow the approach in [Wyszecky and Stiles 1982] (pag. 490) and postulate that the JND can be used as a 'sensation magnitude' because it corresponds to (minimum) equal increments of sensation along the whole photopic range. This allows us to rewrite Weber-Fechner's Law in the following way:

$$\frac{\Delta\mathcal{I}}{\mathcal{I}+m} = k\Delta s, \tag{4}$$

where Δs is the increment in the sensation-magnitude function $s(\mathcal{I})$, also called "perceived brightness".

Naka has shown in [Naka et al. 1979] that the function $r(\mathcal{I})$ previously discussed that measures the electric response of visual neurons in the retina, represents sensation and can be identified with the perceived brightness function $s(\mathcal{I})$ introduced above. We would



(a) (b) (c)

Figure 1: ***(a)*** *Graph of the Naka-Rushton equation $r(\mathcal{I})$ vs. $\mathcal{I}$, with $\mathcal{I}_s = 0.1$.* ***(b)*** *Increment threshold versus intensity in log-log scale. The dots represent experimental data taken from [Shapley and Enroth-Cugell 1984](pag. 291). The curve that interpolates the data was obtained using a function with the same structure as in eq.(3).* ***(c)*** *Inverse of the derivative of the Naka-Rushton function in log-log scale in arbitrary units.*

now like to underline this identification by presenting the following qualitative argument. With easy manipulations of eq.(4) and by taking infinitesimal differences, we can write:

$$\log\left(\frac{d\mathcal{I}}{ds}\right) \equiv \log\left(\frac{1}{s'(\mathcal{I})}\right) = \log(k) + \log(\mathcal{I}+m). \quad (5)$$

On the other hand, if we plot the graph of $\log\left(\frac{1}{r'(\mathcal{I})}\right)$, where $r(\mathcal{I})$ is defined in eq.(1), we find the curve represented in Fig. 1(c). It can be noticed that there is a very good *qualitative match* between the curve related to the Weber-Fechner behavior and the one related to the Naka-Rushton equation.

So, from now on, we will identify the output of the Naka-Rushton equation $r(\mathcal{I})$ with the perceived brightness described by the Weber-Fechner law $s(\mathcal{I})$.

Even though this idea has not been explicitly stated yet in the tone mapping literature, all the TM works that use the Naka-Rushton equation ([Pattanaik et al. 2000; Reinhard 2005; Tamburrino et al. 2008]) are implicitly using the just described assumption.

2.2 Incompatibility between Naka-Rushton's formula and Weber-Fechner's contrast

As we have just discussed, the Weber-Fechner law and the Naka-Rushton equation refer to the same natural process: brightness perception. But they describe different aspects of the problem: on one hand we have the Weber-Fechner law that defines the perception of contrast, and on the other hand we have the Naka-Rushton equation that describes the process of adapting to the average light value of the scene and properly compressing the whole radiance range into $[0, 1]$. In order to combine these two descriptions, we are now going to show that the Naka-Rushton function $r(\mathcal{I})$ must be modified to reproduce the correct (Weber-Fechner) perception of contrast.

Given that we can identify $s(\mathcal{I})$ and $r(\mathcal{I})$, let us re-write eq.(4) by substituting function s with function r and taking again infinitesimal differences:

$$\frac{dr}{d\mathcal{I}} = \frac{k'}{\mathcal{I}+m}, \quad (6)$$

where $k' = 1/k$. This equation gives us the condition that the Naka-Rushton function $r(\mathcal{I})$ must satisfy in order to reproduce the correct perceived contrast. However, let us notice that $r(\mathcal{I})$ is expressed by formula (1), thus performing the derivative we have:

$$\frac{dr}{d\mathcal{I}} = \frac{\mathcal{I}_s}{(\mathcal{I}+\mathcal{I}_s)^2}. \quad (7)$$

Comparing eq.(6) and eq.(7) we can see that the right hand sides do not coincide. This implies that the Naka-Rushton equation does not follow the Weber-Fechner law unless $\mathcal{I}_s$ is modified [Shapley and Enroth-Cugell 1984] from a constant to a function of $\mathcal{I}$.

As remarked in [Shapley and Enroth-Cugell 1984], we can argue the same conclusion by analyzing the behavior of the Naka-Rushton equation. In fact, if we set a constant value for $\mathcal{I}_s$, the function will map to 1 *all* light levels $\mathcal{I}$ significatively bigger than $\mathcal{I}_s$, an effect called 'saturation catastrophe'.

We will show in the next section, after introducing a proper nomenclature for HDR images, that if we substitute $\mathcal{I}_s$ with a suitable function, $f_{\mathcal{I}_s}(\mathcal{I})$, we will avoid this problem. We will refer to the modified Naka-Rushton equation as:

$$r(\mathcal{I}) = \frac{\mathcal{I}}{\mathcal{I}+f_{\mathcal{I}_s}(\mathcal{I})}. \quad (8)$$

Remark: The idea of controlling the tone map using the Weber-Fechner contrast was first introduced by Ward et al. in [Ward et al. 1997]. Their TM operator is based on a modified histogram equalization where they limit the slope of the TM function so that the contrast in the final image is never *higher* than the Weber-Fechner contrast. In our case, instead, we will modify the Naka-Rushton equation so that the contrast will *exactly* match the Weber-Fechner law.

3 The modified Naka-Rushton formula in the framework of HDR images

Let us now introduce the notation that we will use throughout the paper. Let $\vec{I} : \mathfrak{I} \rightarrow (0, +\infty)^3$, be the radiance map representing the input HDR image, $\mathfrak{I}$ being its spatial domain: $\mathfrak{I} = \{1, \ldots, W\} \times \{1, \ldots, H\} \subset \mathbb{Z}^2$, where $W, H \geq 1$ are integers corresponding to the image width and height, respectively. We denote with I_c the generic value of the scalar chromatic components of $\vec{I}$, $c \in \{R, G, B\}$, with $x = (x_1, x_2) \in \mathfrak{I}$ the spatial position of an arbitrary pixel in the image, and with $I_c(x)$ the intensity value in the pixel x of the c channel. To avoid singularities when operating on $\vec{I}$ with ratios or logarithms, we add a small

positive constant: 10^{-12}. Finally, we denote with $\lambda(x)$ the *luminance*[1] of the pixel $x \in \mathfrak{I}$ and with λ a generic luminance value, i.e. $\lambda \in [\lambda_{\min}, \lambda_{\max}] \subset (0, +\infty)$, where $\lambda_{\min}$ and $\lambda_{\max}$ are the extreme luminance values.

In order to apply the Naka-Rushton formula to a HDR image, we must correctly identify which features correspond to the variables appearing in eq.(8).

For the sake of simplicity, let us begin by considering only the luminance image; later on we will extend our analysis by considering the full color image. Notice that, since the HDR image represents the radiance map of a scene, it is natural to let λ play the role of $\mathcal{I}$. Moreover, it makes sense, for a *natural scene*, to identify the semisaturation level as the average luminance, which we denote by μ. In the literature there is no agreement about which expression of μ must be used: arithmetic average μ_a, geometric average μ_g, median μ_{med}, or combinations of them [Reinhard et al. 2005]. Furthermore, neurophysiological data involving complex scenes do not give a precise indication about which value of μ is most suitable. For these reasons, we choose to let unspecified, for the moment, the formal expression of μ. We will discuss later the effects of different choices of μ.

3.1 Modification of the Naka-Rushton equation to comply with the Weber-Fechner Law

Let us now concentrate on the modified Naka-Rushton formula defined in eq.(8). Using the notation just introduced, eq.(8) can be written as:

$$r(\lambda) = \frac{\lambda}{\lambda + f_\mu(\lambda)}. \tag{9}$$

As we explained in last section, the only way to suitably combine the Naka-Rushton formula eq.(9) and the Weber-Fechner contrast law eq.(6) is by selecting the function $f_\mu(\lambda)$ as the solution of the following differential equation:

$$\frac{d}{d\lambda}\left(\frac{\lambda}{\lambda + f_\mu(\lambda)}\right) = \frac{k'}{m+\lambda}, \qquad \forall \lambda \in [\lambda_{\min}, \lambda_{\max}]. \tag{10}$$

Remark: This equation implies that r is strictly increasing for every λ, and so it preserves the order of the level lines of the original HDR image.

By integrating both members with respect to the variable λ, we obtain:

$$f_\mu(\lambda) = \frac{\lambda}{C + k' \log(m+\lambda)} - \lambda, \qquad \forall \lambda \in [\lambda_{\min}, \lambda_{\max}], \tag{11}$$

being C the integration constant.

The function $f_\mu(\lambda)$ depends on three parameters: C, k' and m. We are now going to make a suitable selection of these parameters. Let us fix C once and for all by imposing the constraint: $f_\mu(\mu) = \mu$. As commented before, we consider μ as the analogue of the semi-saturation level, hence this constraint implies that the Naka-Rushton response in μ will be 1/2 [2]. The value of C that satisfies this constraint is $C = \frac{1}{2} - k' \log(m+\mu)$, which implies that the function f_μ can be written as:

$$f_\mu(\lambda) = \frac{\lambda}{\frac{1}{2} + k' \log\left(\frac{m+\lambda}{m+\mu}\right)} - \lambda. \tag{12}$$

If we introduce this expression in the Naka-Rushton equation we find that

$$r(\lambda) = \frac{1}{2} + k' \log\left(\frac{m+\lambda}{m+\mu}\right), \tag{13}$$

which shows that $r(\lambda) \lessgtr 1/2$ whenever $\lambda \lessgtr \mu$, respectively.

Let us set the final range of r such that it is extended over the entire normalized interval $[0,1]$, passing through $1/2$ when $\lambda \equiv \mu$. We have $r(\lambda_{\max}) - r(\lambda_{\min}) = 1 = k' \log\left(\frac{m+\lambda_{\max}}{m+\lambda_{\min}}\right)$, so that

$$k' \equiv \frac{1}{\log\left(\frac{m+\lambda_{\max}}{m+\lambda_{\min}}\right)}. \tag{14}$$

Introducing this expression of k' in eq.(13), after straightforward calculations we find that $r(\lambda_{\min}) = 0$ and $r(\lambda_{\max}) = 1$ if and only if

$$m \equiv \frac{\mu^2 - \lambda_{\max}\lambda_{\min}}{\lambda_{\max} + \lambda_{\min} - 2\mu}. \tag{15}$$

We observe that, with the selection of the parameters C, k', m just described, the function $f_\mu(\lambda)$ exists, is positive for all $\lambda \in [\lambda_{\min}, \lambda_{\max}]$, and it is dimensionally correct since k' is adimensional and m has luminance dimensions.

In the next section we are going to show and comment the results of the application of this modified Naka-Rushton equation on HDR images.

3.2 Tests of the first stage of TSTM: empirical decisions and results

In last section we presented a modification of the Naka-Rushton equation that will allow us to tone map HDR images following the Weber-Fechner perceptual contrast. Let us now empirically define some variables that were left undetermined.

HDR images can span as many orders of magnitude as a natural scene but, as we already stated, the HVS can only perceive simultaneously up to four orders of magnitude. Taking into account this principle we want to apply eq.(9) only over the visible range and saturate all values outside of it. Therefore, we redefine $\lambda_{\min}$ and $\lambda_{\max}$ in eq.(14) and eq.(15) in the following way: we compute the histogram of the luminance image in the log scale and define a window of V orders of magnitude. We slide the window over the log-histogram and for each position we compute the number of pixels that fall into the window. Finally, we select the position where the number of pixels inside the window is maximum. The values $\lambda_{\min}$ and $\lambda_{\max}$ are taken as the end points of this window. All values

[1] We have verified that the dependence of the techniques proposed in the paper on the many luminance definitions available in the literature is practically inappreciable, thus we have computed the luminance as: $\lambda(x) = \frac{1}{3}\left[I_R(x) + I_G(x) + I_B(x)\right], \forall x \in \mathfrak{I}$.

[2] Actually, the 'flash over background' sort of experiments mentioned earlier show that the value of $r(\mu)$ depends on the magnitude range in which μ lies ([Shapley and Enroth-Cugell 1984] pag. 275). We could take this into account in our formulation, but since most HDR images are not calibrated (their data are in arbitrary units, not in cd/m^2) we usually lack the information to introduce this modification

(a) (b) (c)

Figure 2: *Tone mapping of the HDR image 'Nave' image (courtesy of Paul Debevec) with different values of V:* ***(a)*** $V = 2$ ***(b)*** $V = 3$ ***(c)*** $V = 5$.

above $\lambda_{\max}$ are set to $\lambda_{\max}$ and all values under $\lambda_{\min}$ are set to $\lambda_{\min}$. The code is attached in Appendix A.

Given that very few HDR images are calibrated, i.e. their values do not really represent cd/m^2, we cannot use the perceptual range defined by neuroscience experiments, so we must choose V empirically. In Fig. 2 we show the effect of changing the value of V. For all images in our test set, setting V to five gave good results, therefore, all the tests that we will present in this paper will correspond to this fixed value of V.

Another parameter that had to be empirically defined was μ. As we already said, there is no complete agreement in the definition of a proper value of the average light level. Therefore, we propose an empirical function, a convex linear combination in the logarithmic domain between the arithmetic (μ_a) and geometric (μ_g) averages: $\mu(\rho) = \mu_a^{\rho}\mu_g^{1-\rho}$, where $\rho \in [0, 1]$. The best results were achieved using values of ρ that vary between 0.5 and 1. The effect of varying ρ is a modification in the overall brightness of the final image: the bigger the value of ρ the darker the image. Since the choice of ρ is a matter of preference of the user (who, in some cases may want a darker or a lighter result) it is left as the only parameter of our algorithm.

Now that all the variables in eq.(12) have been explained, we can discuss how to extend the computation to the full color image. We have, at least, three different options.

The first one consists in the application of the popular formula (see e.g. [Fattal et al. 2002]):

$$r(I_c(x)) := \left(\frac{I_c(x)}{\lambda(x)}\right)^{\gamma} r(\lambda(x)), \qquad \forall c \in \{R, G, B\}, \quad (16)$$

where $\gamma \in (0, 1)$ is a parameter that controls the saturation of the final image: the bigger γ is, the greater the saturation. The major drawback of this method is that the optimal value of γ varies with the image and, to the best of our knowledge, there is no formula to set it automatically. More importantly, for some images, it is impossible to avoid over-saturation of bright areas or under-saturation of dark zones, see Fig. 3(a).

The second method consists in applying the modified Naka-Rushton formula eq.(13) separately to the three chromatic channels $c \in \{R, G, B\}$. The resulting images are too unsaturated and unnatural colors appear as can be seen in Fig. 3 (b).

Note that the two options just described correspond to a logarithmic transformation of the same type as eq.(13).

Finally, following the proposal of Reinhard et al. ([Reinhard 2005]) we extend the computation of the visual adaptation stage from the luminance to the full color image as follows:

$$r(I_c(x)) := \frac{I_c(x)}{I_c(x) + f_\mu(\lambda(x))}, \quad (17)$$

where $f_\mu(\lambda(x))$ is the function defined in eq.(12), with parameters k' and m determined by eq.(14) and eq.(15), respectively. We would like the reader to note that applying the Naka-Rushton formula as in eq.(17) does not correspond to a pure logarithmic transformation.

Remark: A common assumption in many color processing algorithms is that color channels are correlated, in the sense that they share the same geometry, which also coincides with the geometry of the intensity image [Kimmel 1999]. In other words, it is reasonable to assume that color channels and intensity image share the same family of level lines. Now, with this assumption, if two pixels x, y are in the same isointensity line for the channel I_c, then they will also be in the same isointensity line of $r(I_c)$. This may partially explain the good geometry preservation properties of the mapping produced by eq.(17), without halos or artifacts of any kind.

As it can be seen in Fig.(2)(c) and Fig.(3)(c), the first stage of TSTM reproduces correctly the overall contrast without over- and under-saturating colors.

The computation just described is point-wise, i.e. does not depend on the surround of each pixel. For this reason, the corresponding algorithm has minimal computational cost (see code in Appendix A) and can be run in real time.

4 Local contrast enhancement through variational techniques

The first stage of TSTM only takes into account the process of adapting to the lightness of the scene. However, as already stated, the perception of brightness, phenomenologically speaking, involves an enhacement of local contrast and a chromatic adaptation. The first one is related to the HVS ability to magnify the perception of local details. Human color constancy or chromatic adaptation is the ability to *partially* discard the color of the illumination of a scene, an essential feature of the HVS which makes it very robust in color detection, even under strong light changes [Land and McCann 1971].

These phenomenological characteristics of the HVS have been used in [Palma-Amestoy et al. 2009] to build a variational energy functional whose minimization gives rise to an explicit algorithm able to perform a perceptually inspired color correction. This algorithm is applied to the three chromatic channels independently so, to avoid

(a) (b) (c)

Figure 3: *Tone mapping of the HDR image 'Bausch lot' with the modified Naka-Rushton equation applied to:* ***(a)*** *the luminance using eq.(16) with* $\gamma = 0.8$*,* ***(b)*** *every color channel independently, and* ***(c)*** *combining luminance and color with eq.(17).*

a redundant notation, we will denote with $I(x)$ the generic intensity value of a pixel x in any color channel. The connection of this method with the Retinex theory of color vision [Land and McCann 1971] was established in [Bertalmío et al. 2009].

This perceptually inspired color correction energy functional presents a *balance* between two opposite mechanisms: one that provides a local contrast enhancement and the other that controls the intensity value dispersion. Mathematically, we can express this idea by writing the energy functional as: $E_w(I) = C_w(I) + D(I)$, where $C_w(I)$ and $D(I)$ are the contrast and dispersion terms, respectively, and $w : \mathfrak{I} \times \mathfrak{I} \to (0, +\infty)$ is a normalized spatial kernel.

A fundamental property to be fulfilled by the contrast term $C_w(I)$ in order to implement color constancy is the invariance with respect to overall changes of illumination. The simplest contrast term respecting Weber-Fechner's behavior and invariant with respect to global illuminant changes can be written as follows

$$C_w(I) = \sum_{x\in\mathfrak{I}} \sum_{y\in\mathfrak{I}} w(x,y)\, \frac{\min(I(x), I(y))}{\max(I(x), I(y))}. \qquad (18)$$

The dispersion term must avoid an excessive departure from 1/2, i.e. the center of the dynamic range in normalized units, and from the input intensity values. It can be proven that, in order to maintain dimensional coherence, an entropic dispersion term must be used [Palma-Amestoy et al. 2009]:

$$\begin{aligned} D_{\alpha,\beta}(I) &= \alpha \sum_{x\in\mathfrak{I}} \left(\frac{1}{2} \log \frac{1}{2I(x)} - \left(\frac{1}{2} - I(x) \right) \right) \\ &+ \beta \sum_{x\in\mathfrak{I}} \left(I_0(x) \log \frac{I_0(x)}{I(x)} - (I_0(x) - I(x)) \right), \end{aligned}$$

where $\alpha, \beta > 0$ represent the strength of the attachment to $1/2$ and to $I_0(x)$ (the input intensity value), respectively. In our implementation, we have set $\alpha = 0.5$ and $\beta = 1$. Given the statistical interpretation of entropy, we can say that minimizing $D_{\alpha,\beta}(I)$ amounts to minimizing the disorder of intensity levels around $1/2$ and the input data $I_0(x)$. Thus, $D_{\alpha,\beta}(I)$ accomplishes the required tasks.

Let us set $E_{w,\alpha,\beta}(I) \equiv C_w(I) + D_{\alpha,\beta}(I)$. Note that the minimum of $E_{w,\alpha,\beta}(I)$ satisfies $\delta E_{w,\alpha,\beta}(I) = 0$. To search for the minimum we use a semi-implicit discrete gradient descent strategy with respect to $\log I$ (which corresponds to using the relative entropy as a metric). The continuous gradient descent equation is $\partial_t \log I = -\delta E_{w,\alpha,\beta}(I)$, being t the evolution parameter. Since $\partial_t \log I = \frac{1}{I}\partial_t I$, we have $\partial_t I = -I\delta E_{w,\alpha,\beta}(I)$. This last equation can be discretized as follows: choosing a finite evolution step $\Delta t > 0$ and setting $I_k(x) \equiv I_{k\Delta t}(x)$, $k \in \mathbf{N}$, we can write

$$I_{k+1}(x) = \frac{I_k(x) + \Delta t \left(\frac{\alpha}{2} + \beta I_0(x) + \frac{1}{2} R_{I_k}(x) \right)}{1 + \Delta t(\alpha + \beta)}, \qquad (19)$$

where the term $R_{I_k}(x) \equiv -2I_k(x)\delta C_w(I_k)(x)$ performs contrast enhancement and can be explicitly written as $R_{I_k}(x) = \sum_{y\in\mathfrak{I}} w(x,y)\, c(I_k(x), I_k(y))$, being

$$\begin{aligned} c(I_k(x), I_k(y)) &= \frac{I_k(y)}{I_k(x)} \operatorname{sign}^+(I_k(x) - I_k(y)) - \\ &- \frac{I_k(x)}{I_k(y)} \operatorname{sign}^-(I_k(x) - I_k(y)), \end{aligned}$$

where, for every $t \in \mathbb{R}$, $\operatorname{sign}^+(t) := \frac{1}{2}(1 + \operatorname{sign}(t))$, and $\operatorname{sign}^-(t) := \frac{1}{2}(1 - \operatorname{sign}(t))$. In practice, the signum functions that appear in the previous formula are too singular to be used without producing artifacts and for this reason we used sigmoidal versions of $\operatorname{sign}^+(t)$ and $\operatorname{sign}^-(t)$ built by replacing sign with:

$$\operatorname{sign}_\sigma(t) = \frac{\arctan(\sigma t)}{\max_{t\in\mathbb{R}}\{|\arctan(\sigma t)|\}}, \qquad (20)$$

where $\sigma > 1$ defines the *slope* of the sigmoid. Our tests have shown that a value of σ that gives satisfactory results for all images in our test set is $\sigma = 20$.

Remark. The full computational complexity of the algorithm is $\mathcal{O}((WH)^2)$; however, as proven in [Bertalmío et al. 2007; Palma-Amestoy et al. 2009], using a suitable polynomial approximation of the function $c(I_k(x), I_k(y))$, the complexity can be reduced to $\mathcal{O}((WH)\log(WH))$. In fact, the computation of $R_{I_k}(x)$ can be reduced to a sum of convolutions between powers of I_k and the kernel w, namely $R^{(n)}_{I_k}(x) = \sum_{j=0}^{n} a_j(I_k(x))(w * (I_k)^j)(x)$, that can be performed through the Fast Fourier Transform (FFT). $a_j(I_k(x))$ are n real coefficients that vary with x, in our implementation $n \equiv 9$.

As commented in [Palma-Amestoy et al. 2009], this method can over-enhance noise when applied to LDR images due to quantization. However, here the input of the variational algorithm is the output of the first stage of TSTM, which is a floating point image within the range $[0, 1]$. Because it has not been quantized, this input image is still HDR and therefore,the results of the variational method are not affected by noise and provide a higher amount of detail.

(a) (b) (c) (d)

Figure 4: *Details of the 'Cars' and 'Still life' (taken from OpenEXR source) images obtained by **(a)**,**(c)** first stage and **(b)**,**(d)** second stage of TSTM.*

5 Results and comparison with the state of the art

As we have already explained, the first stage of TSTM is responsible for the range compression together with the global representation of the Weber-Fechner contrast. The effect of the second stage is to remove a possible color cast and to enhance local details.

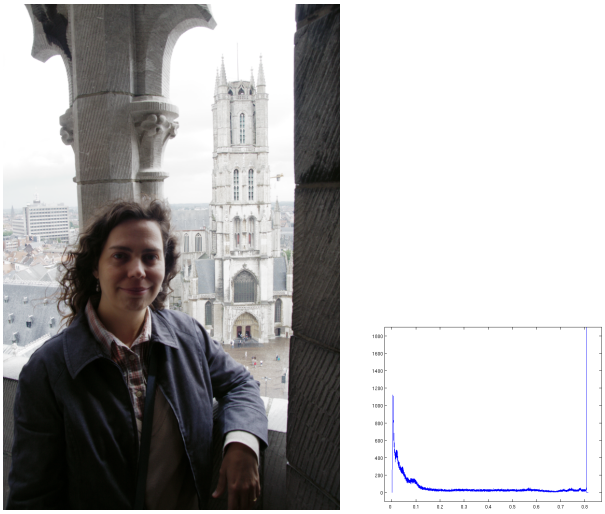
For images that do not present fine details or strong color cast, the first stage of TSTM gives satisfactory results by its own. We remark this fact with the detail presented in Fig. 4(a) and Fig. 4(b) where we can see that the action of the variational algorithm is almost inappreciable.

However, images with strong color cast or highly textured regions can be greatly improved by the action of the second TSTM step. A good example of color cast removal can be seen in Fig. 4(c) and Fig. 4(d). The output of the first stage 4(c) conserves the original color cast, while the final result 4(d) is free from color cast and has sharper details. The effect of color cast removal is particularly visible on the white napkin. The effect of local detail enhancement is particularly visible in the image 'Trees' (Fig. 5).

(a) (b)

Figure 5: *TSTM results obtained applying **(a)** only the first stage, **(b)** the complete TSTM algorithm. The effect of local detail enhancement is particularly visible in the bushes.*

We stress that our method assumes that the value of μ in eq.(12) is representative of the histogram mass. In some cases however (e.g for images with bimodal histogram) μ can fall in a poorly populated region, resulting in under- or over-exposing some image areas, as it can be seen in Fig. 6(a). Note that the background in this image is too bright, while the girl is underexposed.



(a) (b)

Figure 6: ***(a)** TSTM result obtained applying only the first stage to an image with a bimodal luminance histogram, shown in **(b)**. See text for more information.*

Although a thorough comparison is out of the scope of this paper, we would like to give an indication of the quality of our results by presenting a comparison with the output of three other tone mapping methods which are State of the Art: [Durand and Dorsey 2002], [Reinhard et al. 2005] and [Mantiuk et al. 2006]. The results presented here were produced by the code provided by the authors in the first two cases, and by the software pfstmo (http://www.mpi-inf.mpg.de/resources/tmo/) in the case of [Mantiuk et al. 2006].

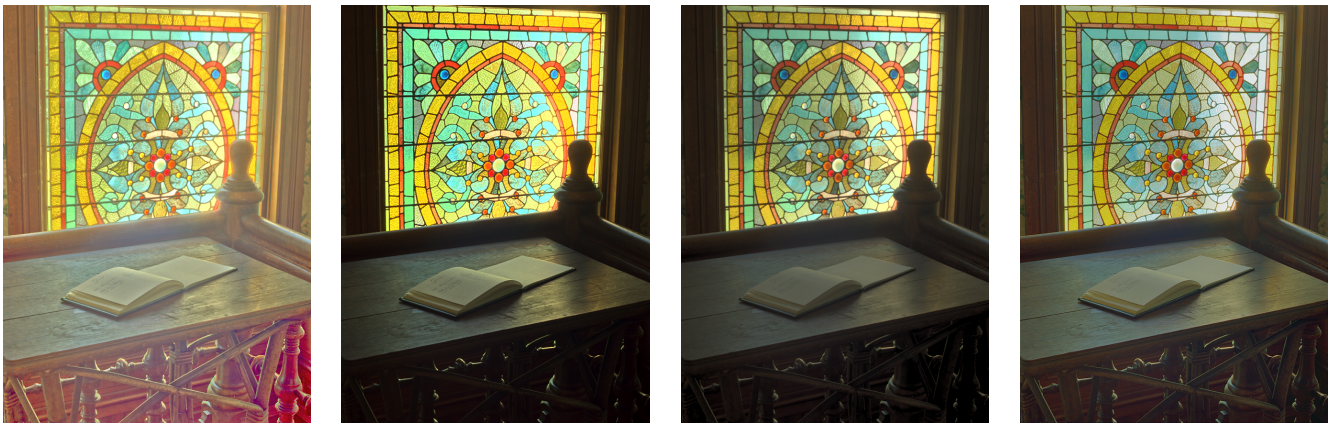
In the authors' opinion, TSTM produces reasonable results in terms of color reproduction and overall contrast visibility. It does not produce spurious colors nor over-saturation as can be checked in the sky of Fig. 7, in Fig. 8 and in the glass window area of Fig. 9. Taking into consideration contrast, TSTM does not over-enhance or diminish contrast as can be seen in Fig. 7 (see the area under the table) and in Fig. 8. After these considerations, we think that TSTM compares well with the state of the art.

(a) (b) (c) (d)

Figure 7: *Results of 'Office' produced by the methods based on the papers of* ***(a)*** *Durand et al.* ***(b)*** *Reinhard et al.* ***(c)*** *Mantiuk et al.* ***(d)*** *TSTM with parameter set to* $\rho = 0.7$.

(a) (b) (c) (d)

Figure 8: *Results of the 'Memorial' produced by the methods based on the papers of* ***(a)*** *Durand et al.* ***(b)*** *Reinhard et al.* ***(c)*** *Mantiuk et al.* ***(d)*** *TSTM with parameter set to* $\rho = 1$.



(a) (b) (c) (d)

Figure 9: *Results of the 'Desk' produced by the methods based on the papers of* ***(a)*** *Durand et al.* ***(b)*** *Reinhard et al.* ***(c)*** *Mantiuk et al.* ***(d)*** *TSTM with parameter set to* $\rho = 0.5$.

6 Conclusions and perspectives

We have presented TSTM, a tone mapping operator for HDR images inspired by two basic stages of the human vision: visual adaptation and local contrast enhancement.

We have implemented the first stage through a modified Naka-Rushton's equation, proposing a new analytical expression that allows to compress the range of the HDR image into $[0, 1]$ while reproducing the perceptual contrast of the HVS. Because this step does not significantly alter the ordering of the level lines, it is free from the typical halos and artifacts produced by spatially local methods. The results in this first stage are obtained in real-time, which suggests the potential of this method in applications to HDR cinema, as well.

After this first step, we have a floating point image with compressed range that is used as input for a variational algorithm. This stage is able to perform local contrast enhancement coherently with human perception [Palma-Amestoy et al. 2009]. The main properties of this variational algorithm are: strong color cast reduction and magnification of local details.

This second stage is spatially variant and its computational cost is $\mathcal{O}(N \log N)$, being N the number of image pixels. Though it cannot provide real-time performances, it can still be used for offline computations (it takes a few seconds for a TV-resolution image on a 3GHz 1GbRAM PC).

Finally, TSTM has only one user-dependent parameter which controls the overall brightness of the final results and, therefore, is very intuitive to set.

We leave as future work the suitable treatment of special cases such as bimodal-histogram images, where the average intensity value does not represent the highest density area of the histogram.

A further open problem refers to images with calibrated data. Since very few HDR images are calibrated, there is no exact match between the parameters of this method and the results obtained by neuroscience experiments. It would be interesting to explore a possible refinement of the new Naka-Rushton equation presented in this paper when the values of the HDR images correspond to real-world cd/m^2.

A MATLAB code of the first stage of TSTM

In this appendix we would like to provide the MATLAB code of the first stage of TSTM. This stage implements the formula expressed by eq.(17) with the specification of the function f as eq.(11) and parameters k' and m given by eq. (14) and eq.(15), respectively.

```
function output = ToneMapping(Image, rho)
  %Matlab code of the first stage of TSTM.

  L = 0.3*(Image(:,:,1)+Image(:,:,2)+Image(:,:,3));
  L = L+10^(-12);

  %Compute \lambda_{min} and \lambda_{max}
  [m_L, M_L] = getMin_and_Max(log10(L),5);

  m_L = max(m_L, min(min(L)));
  M_L = min(M_L, max(max(L)));

  L(find(L < m_L)) = m_L;
  L(find(L > M_L)) = M_L;

  %Compute the semisaturation value \mu
  mu_a = mean(mean(L));
  mu_g = 10.^mean(mean(log10(L)));
  mu = (mu_a.^rho)*(mu_g.^(1-rho));

  m = (mu*mu - M_L*m_L) /( M_L+m_L-2*mu );
  k = 1/( log(m+M_L)-log(m+m_L));

 output= zeros(size(Image));
 for i=1:size(L,1)
  for j=1:size(L,2)
    f=L(i,j)./(0.5+k*log((m+L(i,j))/(m+mu)))-L(i,j);

    for c=1:3
      output(i,j,c)=Image(i,j,c)./(Image(i,j,c)+f);
    end
   end
  end
```

In the following code, we provide the implementation of the methods described in section 3.2 to obtain values of $\lambda_{\min}$ and $\lambda_{\max}$ for the equations (14) and (15).

```
function [m_L, M_L] = getMin_and_Max(log_L,V)
  pix = size(log_L,1)*size(log_L,2);
  bins = 100000;

  m_log_L = min(min(log_L));
  M_log_L = max(max(log_L));

  range = M_log_L - m_log_L;

  %Compute the accumulated histogram
  H = cumsum(hist(reshape(log_L,pix,1), bins));

  vec_pix= zeros(bins,1);
  for b=2:bins-1
    I_h = min(b*range/bins + m_log_L + V, M_log_L);
    bin_M_L = floor((I_h - m_log_L)*bins/range);
    vec_pix(b) = (H(bin_M_L)-H(b)) / pix;

  end

  bin_m=find(vec_pix==max(max(vec_pix)),1);

  m_L = 10.^((bin_m-1) * range/bins + m_log_L);
  M_L = 10.^((bin_m-1) * range/bins + m_log_L + V);
```

Acknowledgments

The authors would like to thank P. Arias and L. Garrido for the cameras, A. Ledberg for providing useful references, L. Sánchez for his photographs, and F. Durand, E. Reinhard, and G. Ward for providing the implementations of the methods in the state of the art. M. Bertalmío and V. Caselles acknowledge partial support by PNPGC project, reference MTM2006-14836. V. Caselles also wants to acknowledge "ICREA Acadèmia" prize for excellence in research founded by the Generalitat de Catalunya. E. Provenzi acknowledges the Ramón y Cajal fellowship by Ministerio de Ciencia y Tecnología de España.

References

ASHIKHMIN, M. 2002. A tone mapping algorithm for high contrast images. In *Eurographics Workshop on Rendering*, P. Debevec and S. Gibson Eds., 1–11.

BERTALMÍO, M., CASELLES, V., PROVENZI, E., AND RIZZI, A. 2007. Perceptual color correction through variational techniques. *IEEE Trans. on Image Processing 16*, 1058–1072.

BERTALMÍO, M., CASELLES, V., AND PROVENZI, E. 2009. Issues about the retinex theory and contrast enhancement. *to appear in International Journal of Computer Vision.*

CHALUPA, L., AND WERNER, J. E. 2003. *The visual neuroscience*. MIT Press.

CHIU, K., HERF, M., SHIRLEY, P., SWAMY, S., WANG, C., AND ZIMMERMAN, K. 1993. Spatially nonuniform scaling functions for high contrast images. In *Proceedings of Graphics Interface 93*, Morgan Kaufmann, 245–253.

DEBEVEC, P., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *Proc. of the 24th annual conf. on Computer graphics*.

DURAND, F., AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. SIGGRAPH 2002, Proceedings of the 29th annual conference on Computer graphics and interactive techniques, 257–266.

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. In *ACM Trans. Graphics*, vol. 21 (3), 249–256.

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

GONZALES, R., AND WOODS, R. 2002. *Digital image processing*. Prentice Hall.

JOBSON, D., RAHMAN, Z., AND WOODELL, G. 1997. A multiscale Retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on image processing 6*, 7 (july), 965–976.

KEENER, J., AND SNEYD, J. 2008. *Mathematical Physiology*. Springer.

KIMMEL, R. 1999. Demosaicing: image reconstruction from color ccd samples. *Image Processing, IEEE Transactions on 8*, 9 (Sep), 1221–1228.

KRAWCZYK, G., MYSZKOWSKI, K., AND SEIDEL, H., 2005. Lightness perception in tone reproduction for high dynamic range images.

LAND, E., AND MCCANN, J. 1971. Lightness and Retinex theory. *Journal of the Optical Society of America 61*, 1 (Jan.), 1–11.

LISCHINSKI, D., FARBMAN, Z., UYTTENDAELE, M., AND SZELISKI, R. 2006. Interactive local adjustment of tonal values. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, ACM, New York, NY, USA, 646–653.

MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H. 2006. A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception (TAP) 3 (3)*, 286–308.

NAKA, K., CHAN, R. Y., AND S., Y. 1979. Adaptation in catfish retina. *Journal of Neurophysiology 42*, 2, 441–454.

PALMA-AMESTOY, R., PROVENZI, E., BERTALMÍO, M., AND CASELLES, V. 2009. A perceptually inspired variational framework for color enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence 31*, 3, 458–474.

PATTANAIK, S., TUMBLIN, J., YEE, H., AND GREENBERG, D. 2000. Time-dependent visual adaptation for fast realistic image display. In *Proceedings of SIGGRAPH*, 47–54.

PRATT, W. 2007. *Digital Image Processing*. J. Wiley & Sons.

REINHARD, E., WARD, G., PATTANAIK, S., AND DEBEVEC, P. 2005. *High Dynamic Range Imaging, Acquisition, Display, And Image-Based Lighting*. Morgan Kaufmann Ed.

REINHARD, E. DEVLIN, K. 2005. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Trans. on Visualization and Computer Graphics 11(1)*, 13–24.

SCHLICK, C. 1994. Quantization techniques for visualization of high dynamic range pictures. In *Proceedings of the 5th Eurographics Workshop on Rendering Workshop*, Springer Verlag, 7–20.

SHAPLEY, R., AND ENROTH-CUGELL, C. 1984. *Visual adaptation and retinal gain controls*, vol. 3. 263–346.

STEVENS, S., AND STEVENS, J. 1963. Brightness function: Effect of adaptation. *Jour. of the Opt. Soc. of America 53*, 3, 375–385.

TAMBURRINO, D., ALLEYSSON, D., MEYLAN, L., AND SUSSTRUNK, S. 2008. Digital camera workflow for high dynamic range images using a model of retinal processing. In *Proceedings SPIE*, vol. 6817.

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

TUMBLIN, J., AND TURK, G. 1999. Lcis: A boundary hierarchy for detail-preserving contrast reduction. In *SIGGRAPH: Conference Proceedings*, 83–90.

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

WARD, G. 1994. *A contrast-based scalefactor for luminance display*. Academic Press Professional, Inc., 415–421.

WYSZECKY, G., AND STILES, W. S. 1982. *Color science: Concepts and methods, quantitative data and formulas*. John Wiley & Sons.