

A multi-modal approach to perceptual tone mapping

Sira Ferradans, Marcelo Bertalmío, Edoardo Provenzi, Vicent Caselles

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

e-mails: {sira.ferradans,marcelo.bertalmio,edoardo.provenzi,vicent.caselles}@upf.edu.

Tel. +34 93 542 2937. Fax. +34 93 542 2569.

Abstract

We present an improvement of TSTM, a recently proposed tone mapping operator for High Dynamic Range (HDR) images, based on a multi-modal analysis. One of the key features of TSTM is a suitable implementation of the Naka-Rushton equation that mimics the visual adaptation performed by the human visual system coherently with Weber-Fechner's law of contrast perception. In the present paper we use the Gaussian Mixture Model (GMM) in order to detect the modes of the log-scale luminance histogram of a given HDR image and then we use the information provided by GMM to properly devise a Naka-Rushton equation for each mode. Finally, we properly select the parameters in order to merge those equations into a continuous function. Tests and comparisons to show how this new method is capable of improving the performances of TSTM are provided and commented, as well as comparisons with state of the art methods.

Keywords: High Dynamic Range Images, Tone Mapping, Naka Rushton formula, Weber-Fechner Contrast, Gaussian Mixture Model

Digital Peer Publishing Licence

Any party may pass on this Work by electronic means and make it available for download under the terms and conditions of the current version of the Digital Peer Publishing Licence (DPPL). The text of the licence may be accessed and retrieved via Internet at
<http://www.dipp.nrw.de/>.

First presented at the Conference on Visual Media Production 2009, extended and revised for JVRB

1 Introduction

In daylight the Human Visual System (HVS) works best in terms of color vision and perception of details. The amount of light arriving to our retinas can span many orders of magnitude, from the scotopic lower bound $10^2 \frac{cd}{m^2}$ indoors to the glare upper bound $10^9 \frac{cd}{m^2}$ outdoors with the brightest sunlight [FPSG96] but our photoreceptor neurons in the retina, rods and cones, produce electrical outputs which span only two orders of magnitude [SEC84] (pag.326). Therefore, the HVS cannot operate over the entire range of physical radiances simultaneously. Rather, it adapts to an average intensity and handles a smaller magnitude interval, through a process called *visual adaptation*. In photography and film production we are faced with the same situation: most cameras (both photo and video cameras) take Low Dynamic Range (LDR) pictures, spanning only two orders of magnitude, so some sort of adaptation mechanism is required. In film production, the equivalent of the visual adaptation of the HVS is achieved by flooding the scene with more light, as the great director Sydney Lumet so clearly explains in [Lum95] (page 83):

If you've ever passed a movie company shooting on the streets, you may have seen an enormous lamp pouring its light onto an actor's face. We call it an arc or a brute, and it gives off the equivalent of 12,000 watts. Your reaction has probably been: What's the matter with these people? The sun's shining brightly and they're adding that big light so that the actor is practically squinting. Well, film is limited in many ways. It's a chemical process, and one of its limitations is the amount of contrast it can take. It can adjust to a lot of light or a little bit of light. But it can't take a lot of light and a little bit of light in the same frame. It's a poorer version of your own eyesight. I'm sure you've seen a

person standing against a window with a bright, sunny day outside. The person becomes silhouetted against the sky. We can't make out his features. Those arc lamps correct the "balance" between the light on the actor's face and the bright sky. If we didn't use them, his face would go completely black.

Therefore, the use of movie cameras capable of capturing High Dynamic Range (HDR) images would greatly simplify the process of shooting outdoors: less artificial lights to transport and set-up, less time spent, less hassle for the actors. These sort of cameras are becoming more popular, but still we are faced with the problem that most *displays* are LDR, so a HDR to LDR conversion must be performed in order to screen the picture. This HDR to LDR conversion, if it is performed trying to emulate as much as possible the contrast and color sensation of the real-world scene, i.e. achieve an image that looks natural (as it is our case, as opposed to trying to maximize the visible details even if the resulting image appears artificial), is called 'Tone Mapping' (TM) or 'Tone Reproduction' (TR) [LRP97].

An excellent survey of the many TM methods proposed up to 2005 can be found in [RWPD05]. Among the more recent works we would like to mention [RSSF02, RD05, TAMS08], which use a perceptual-based approach involving the Naka-Rushton equation [NCY79]; [KMS05], which uses the anchoring theory of visual perception [GKB⁺99]; and [LFUS06], where the authors propose an interactive method that allows the user to create better subjective results.

In this paper we propose a perceptual-based approach for TM which is an extension of the method introduced in [FPBC09b]. Given that the goal is to obtain LDR pictures that appear natural, it seems reasonable to try to mimic basic features of the HVS: in our case, we are trying to emulate *visual adaptation* and *spatially local contrast enhancement*. Our contribution is to propose a method for TM which compares well in terms of image quality with the state-of-the-art, is able to deal with images where the luminance histogram has modes which are far apart, and is fast. It is an improvement of the method introduced in [FPBC09b] and which was not capable of dealing well with multi-modal histogram images. It is presented for still images, but in the final section we suggest how it could be extended to motion pictures.

This paper is organized as follows. In section 2 we review the technique proposed in [FPBC09b, FPBC09a] and discuss its limitations. Section 3 intro-

duces our method and explains how to overcome most of the problems encountered in [FPBC09b, FPBC09a]. Section 4 presents some results of our algorithm as well as comparisons with state-of-the-art methods. Finally, section 5 presents some conclusions and possibilities for future research.

2 Combining the Naka-Rushton equation and the Weber-Fechner contrast

In this section we will present the fundamental concepts about *visual adaptation* and *contrast perception*. These concepts will allow us to discuss a modification of the Naka-Rushton equation that is vital for the construction of our tone mapping operator. These conclusions were already presented in [FPBC09b, FPBC09a]. We will also present the advantages and drawbacks of the method and, in next section, we will propose a new improvement based on Gaussian Mixture Models (GMM).

Let us begin 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/m² (scotopic limit) to 10^6 cd/m² (glare limit) [Pra07]. The automatic process that allows the HVS to operate over such a huge range is called visual adaptation [SEC84].

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 ([SEC84] (pag.326)) to four ([KS08] (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 [SEC84]:

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

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, ΔV_{\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 the first psychophysical experiments using hand held weight, discovering that the perception of weight followed a ratio law. Regarding visual perception, these experiments were developed 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 a phenomenological approach was applied 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*:

$$\frac{\Delta\mathcal{I}}{\mathcal{I}} = k, \quad (2)$$

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:

$$\frac{\Delta\mathcal{I}}{\mathcal{I} + m} = k, \quad (3)$$

where $m > 0$ is a quantity often interpreted as internal noise in the visual mechanism, e.g. quoting [CW03] (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 $\mathcal{I} \gg m$ the slope of $\log(\text{JND})$ as a function of $\log(\mathcal{I})$ is 1; but when $\mathcal{I} \ll m$ 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 [WS82] (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:

$$\frac{\Delta\mathcal{I}}{\mathcal{I} + m} = k\Delta s, \quad (4)$$

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

We would now like to underline the identification between the electric response of visual neurons in the retina and the perceived brightness function 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 ([PTYG00, RD05, TAMS08]) 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

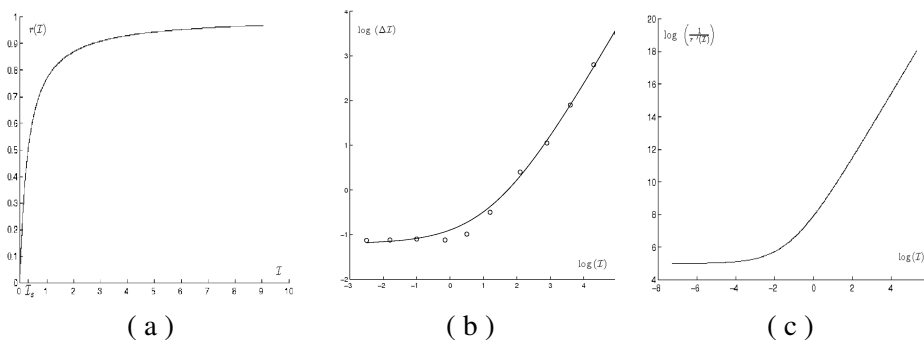


Figure 1: **(a)** Graph of the Naka-Rushton equation $r(I)$ vs. I , with $I_s = 0.1$. **(b)** Increment threshold versus intensity in log-log scale. The dots represent experimental data taken from [Shapley and Enroth-Cugell 1984](pag. 291). The curve that interpolates the data was obtained using a function 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.

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 $r(I)$ must be modified to reproduce the correct (Weber-Fechner) perception of contrast.

Given that we can identify $s(I)$ and $r(I)$, let us rewrite eq.(4) by substituting function s with function r and taking again infinitesimal differences:

$$\frac{dr}{dI} = \frac{1}{k(I + m)}. \quad (6)$$

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

$$\frac{dr}{dI} = \frac{I_s}{(I + I_s)^2}. \quad (7)$$

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 I_s is modified [SEC84] from a constant to a function of I .

As remarked in [SEC84], we can argue the same conclusion by analyzing the behavior of the Naka-Rushton equation. In fact, if we set a constant value for I_s , the function will map to 1 *all* light levels I significantly bigger than 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 I_s with a suitable function, $f_{I_s}(I)$, we will avoid this problem. We will refer to the modified Naka-Rushton equation as:

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

2.3 Applying perceptual laws to the digital world

In this section we will apply the concepts already presented to digital images. For the sake of simplicity, we will first consider the luminance image and then extend the method to the full color image. We use the simplest, equally weighted, luminance since the results that we are going to present are practically invariant with respect to the many luminance definitions available in the literature.

Let us introduce the notation that will be used throughout the paper. Let $\vec{I} : \mathcal{J} \rightarrow (0, +\infty)^3$, be the radiance map representing the input HDR image, \mathcal{J} being its spatial domain: $\mathcal{J} = \{1, \dots, W\} \times \{1, \dots, H\} \subset \mathbb{Z}^2$, 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 $x = (x_1, x_2) \in \mathcal{J}$ 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. Finally, we denote with $\lambda(x)$ the *luminance*, i.e. $\lambda(x) = \frac{1}{3} [I_R(x) + I_G(x) + I_B(x)]$, of the pixel $x \in \mathcal{J}$ and with λ a generic luminance value, i.e. $\lambda \in$

$[\lambda_{\min}, \lambda_{\max}] \subset (0, +\infty)$, where λ_{\min} and λ_{\max} are the extreme luminance values. In order to avoid singularities in $\lambda = 0$ we add to the whole luminance image a value of 10^{-12} .

The translation of the equation (8) to the digital world is:

$$r(\lambda) = \frac{\lambda}{\lambda + f_{\mu}(\lambda)}, \quad (9)$$

where we have maintained the same symbol r to avoid a cumbersome notation. Note that we are assuming that the semi-saturation constant to be translated as the value μ , that we will leave unspecified for now.

In [FPBC09b, FPBC09a], the authors proposed to univocally determine f_{μ} by imposing the function $r(\lambda)$ in eq.(9) to satisfy Weber-Fechner's law of contrast perception. This requirement can be formalized through the following differential equation [FPBC09b, FPBC09a]:

$$\frac{d}{d\lambda} \left(\frac{\lambda}{\lambda + f_{\mu}(\lambda)} \right) = \frac{1}{k(m + \lambda)}, \quad \forall \lambda \in [\lambda_{\min}, \lambda_{\max}], \quad (10)$$

By integrating both members with respect to the variable λ , one obtains:

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

being C an integration constant. Introducing this expression of $f_{\mu}(\lambda)$ in eq.(9), one finds that the expression of the generalized Naka-Rushton formula coherent with Weber-Fechner's contrast perception is:

$$r_{C,k,m}(\lambda) = C + k \log(m + \lambda). \quad (12)$$

The triplet of parameters C , k and m can be determined by imposing some general conditions that were discussed in [FPBC09b, FPBC09a]. Their analytical expressions are:

$$C \equiv \frac{1}{2} - k \log(m + \mu); \quad (13)$$

$$k \equiv \frac{1}{\log \left(\frac{m + \lambda_{\max}}{m + \lambda_{\min}} \right)}; \quad (14)$$

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

With these parameters, the function f_{μ} is well defined and non-negative within the range $[\lambda_{\min}, \lambda_{\max}]$

and, by substituting the explicit value of C , we get the expression:

$$f_{\mu}(\lambda) = \frac{\lambda}{\frac{1}{2} + k \log \left(\frac{m + \lambda}{m + \mu} \right)} - \lambda, \quad (16)$$

with k and m defined as in eqs. (14), (15), respectively.

Let us now extend the Naka-Rushton equation to a full color image. In [FPBC09b], the authors commented that the Naka-Rushton implementation on color images that gives the best results in terms of color rendition is the following:

$$r(I_c(x)) = \frac{I_c(x)}{I_c(x) + f_{\mu}(\lambda(x))} \quad x \in \mathcal{I}, \quad c \in \{R, G, B\}, \quad (17)$$

with the same choice of the parameters appearing in f_{μ} as above. In this paper, we will follow this choice. Note that we are applying independently the function $r(I_c(x))$ to each R, G, B color channel.

This transformation constitutes the first step of the Two Stage Tone Mapper (TSTM). Let us now review the second stage of TSTM: enhancement of spatially local contrast. The phenomenological characteristics of the HVS have been used in [PAPBC09] to build a variational energy functional $E(I)$ whose minimization gives rise to an explicit algorithm able to perform a *balance* between two opposite mechanisms: one provides a spatially local contrast enhancement and the other controls that the intensity value dispersion does not depart too much from the input image. This step, apart from improving detail visibility, permits to partially discard the presence of a possible color cast.

2.4 Drawbacks of TSTM

The empirical tests performed by using eq. (17) have shown that TSTM performs very well on HDR images whose range is up to 5 orders of magnitude and whose histogram is not sharply multi-modal. Moreover, its output results strongly depend on the choice of μ . In [FPBC09b], this value has been represented by a convex linear combination in the logarithmic domain between the arithmetic μ_a and geometric μ_g luminance 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 0.7 and 1, depending on the particular image. The effect of varying ρ is a modification in the overall brightness of the output: the bigger the value of ρ , the darker the output.

To bypass the problem related with HDR images whose range extend beyond 5 orders of magnitude, in

[FPBC09b] the authors proposed to reduce the image range. However, this does not overcome the problem and all the information contained in the clipped luminance regions below the new value of λ_{\min} and above the new value of λ_{\max} are lost.

Finally, the presence of sharply separated ‘modes’ in the histogram can result in an incorrect rendition of contrast because, in that situation, the value of μ can fall into a poorly populated region of the histogram, resulting in the under- or over-exposition of some image areas.

3 An improvement of TSTM: the multi-modal approach

In order to overcome the problems related to the possible presence of sharp modes in the HDR image histogram, we propose here an extension of the first step of TSTM based on a multi-modal approach.

The main idea of our method is to divide the whole luminance range into smaller intervals, apply eq.(12) over each interval and merge them in such a way that *the global transformation on λ results continuous*. The main advantage of this approach is that, while preserving Weber-Fechner’s contrast globally in the image, it tone-maps correctly the details in all the areas of the histogram. Moreover, outliers or luminance values far in the histogram will not influence the current intervals.

Once again we will introduce the method for the luminance plane and then extend it to a color image. We will start by explaining how to divide the luminance range into intervals, then how to process each of them and, finally, we will show how to link the different intervals in order to process the complete luminance range.

We propose to divide the log-luminance histogram in intervals using the Gaussian Mixture Model (GMM) method [Bis06]. Understanding GMM as a density estimator, if we compute GMM over the histogram we obtain where the modes are located. In the present paper, we have chosen to compute GMM over the histogram in the log scale because we obtain more robust results.

The result of the GMM is a set of N Gaussians defined by their mean values $\tilde{\mu}_j$ and standard deviations $\tilde{\sigma}_j$, $j = 1, \dots, N$ in the log-domain¹.

¹ N , the total number of modes, depends on the orders of magnitude of the image, we have taken $N = \text{round} \left(\log_{10} \left(\frac{\lambda_{\max}}{\lambda_{\min}} \right) \right)$.

For the j -th Gaussian, the values $\tilde{\mu}_j^- \equiv \tilde{\mu}_j - 2\tilde{\sigma}_j$ and $\tilde{\mu}_j^+ \equiv \tilde{\mu}_j + 2\tilde{\sigma}_j$ can be considered as the extrema of its area of influence, since the area under the Gaussian between these extrema is approximately 95,4% of the total area.

We notice that, on one side, the support of the N Gaussians may not cover the entire log-luminance range of the image because of isolated pixels; on the other side, Gaussians may overlap, so, in order to overcome these problems, we define the limits of the N sub-intervals in the linear luminance range as follows:

$$\lambda_{\min}^j = \begin{cases} \lambda_{\min} & \text{if } j = 1; \\ \max \left(e^{\frac{\tilde{\mu}_{j-1}^+ + \tilde{\mu}_j^-}{2}}, e^{\tilde{\mu}_{j-1}^+} \right) & \text{if } j \in \{2, \dots, N\}. \end{cases}$$

$$\lambda_{\max}^j = \begin{cases} \lambda_{\max} & \text{if } j = N \\ \min \left(e^{\frac{\tilde{\mu}_j^+ + \tilde{\mu}_{j+1}^-}{2}}, e^{\tilde{\mu}_{j+1}^-} \right) & \text{if } j \in \{1, \dots, N-1\}, \end{cases}$$

note that we are forcing the intervals to be within the range defined by the contiguous Gaussian means.

Let us now define the j -th interval as $[\lambda_{\min}^j, \lambda_{\max}^j]$ and the normalized Naka-Rushton formula over it as:

$$r_j(\lambda) = \frac{r(\lambda) - r(\lambda_{\min}^j)}{r(\lambda_{\max}^j) - r(\lambda_{\min}^j)}, \quad \lambda \in [\lambda_{\min}^j, \lambda_{\max}^j], \quad (18)$$

where $r(\lambda)$ is defined as in eq.(12), with the following parameters:

$$k_j \equiv \frac{1}{\log \left(\frac{m_j + \lambda_{\max}^j}{m_j + \lambda_{\min}^j} \right)}; \quad (19)$$

and

$$m_j \equiv \max \left(0, \frac{\mu_j^2 - \lambda_{\max}^j \lambda_{\min}^j}{\lambda_{\max}^j + \lambda_{\min}^j - 2\mu_j} \right). \quad (20)$$

Thus, instead of having only one m and k for the whole image, we have one for each interval.

Note that we are forcing m_j to be positive. A value of m_j less than zero may correspond to a convex r_j , a condition that we want to avoid because the HVS response to light stimuli [SEC84] is described by a concave function.

Now that we have defined the normalized Naka-Rushton formula over each interval, we will show how to link all of them and construct a continuous function over the entire range that we express as:

$$r_G(\lambda) = \sum_{j=1}^N \chi_j(\lambda) (C_j + h_j r_j(\lambda)), \quad (21)$$

where function $\chi_j(\lambda)$ is the characteristic function of the j -th sub-interval:

$$\chi_j(\lambda) = \begin{cases} 1 & \text{if } \lambda \in [\lambda_{\min}^j, \lambda_{\max}^j]; \\ 0 & \text{otherwise.} \end{cases}$$

The values h_j and C_j allow us to stick together the different Naka-Rushton equations and set their output within the range $[0, 1]$. Let us start by the scaling factor h_j . This value defines the height of the j -th Naka-Rushton within the final range $r_G(\lambda_{\max}) - r_G(\lambda_{\min})$. In the λ domain the j -th Naka-Rushton formula will be applied over the luminance values in the the j -th Gaussian domain, thus, in the perceptual brightness domain, these values will be mapped to $[r_G(e^{\tilde{\mu}_j^-}), r_G(e^{\tilde{\mu}_j^+})]$. The length of the j -th interval in the perceptual brightness domain is then:

$$\tilde{h}_j \equiv r_G(e^{\tilde{\mu}_j^+}) - r_G(e^{\tilde{\mu}_j^-}) = k \log \left(\frac{m + e^{\tilde{\mu}_j^+}}{m + e^{\tilde{\mu}_j^-}} \right). \quad (22)$$

By normalizing, we obtain the final expression of h_j :

$$h_j = \frac{\log \left(\frac{m + e^{\tilde{\mu}_j^+}}{m + e^{\tilde{\mu}_j^-}} \right)}{\sum_{i=1}^N \log \left(\frac{m + e^{\tilde{\mu}_i^+}}{m + e^{\tilde{\mu}_i^-}} \right)} \quad (23)$$

where m is computed with global values as in eq. (15) with $\mu \equiv \mu_a$. This expression guarantees that each j -th subrange is mapped into a subrange $r_j(e^{\tilde{\mu}_j^-}, e^{\tilde{\mu}_j^+}) \subset [0, 1]$ coherently with the Weber-Fechner's law, eq.(4). Note that we are computing the h_j values over the extrema of the Gaussians instead of the extrema of the intervals. Some HDR images present outliers produced by numerical errors while creating the HDR image, i.e. small amounts of pixels with values far away from the main mass of the histogram. If the amount of outliers is small, the GMM algorithm does not model their group of values, thus, they will not be taken into account in the computation of h_j . This fact is important, given that we define h_j as a distance between the extrema, and therefore, a single outlier can modify the final value of the h_j . Therefore, the main advantage of taking the extrema of the Gaussians is that the effect of the outliers is minimized.

Finally, we can stick together all the scaled Naka-Rushton functions by defining C_j as:

$$C_j = \begin{cases} 0 & \text{if } j = 1; \\ \sum_{k=1}^{j-1} h_k & \text{if } j = 2, \dots, N. \end{cases}$$

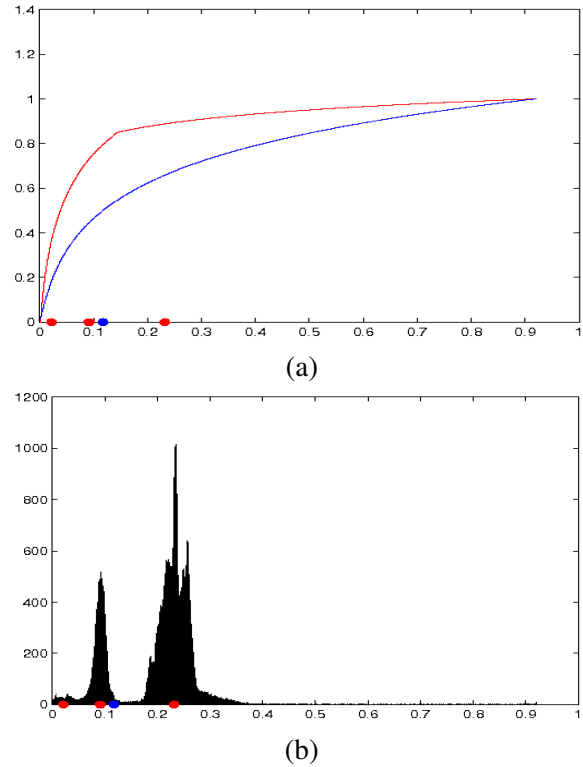


Figure 2: (a) In red we have the multimodal Naka-Rushton function with its three $\mu(j)$ as red circles, in blue the Naka-Rushton obtained with TSTM and its $\mu(\rho)$ value. (b) Histogram of the 'Cars' image with the modes obtained with GMM in red and the $\mu(\rho)$ with $\rho = 0.5$ for the TSTM algorithm in blue. Note how the modes represent better the mass of the histogram.

The extension of this method to color images is exactly the same as for the TSTM method. We use eq. (17) but we substitute $f_\mu(\lambda)$ with:

$$f_G(\lambda(x)) = \frac{\lambda(x)}{r_G(\lambda(x))} - \lambda(x). \quad (24)$$

4 Results and comparisons

Comparing Tone Mapping results is a very difficult issue given the perceptual nature of the problem. Here we propose a perceptual approach, that is, the subjective taste of the user, which has being the standard for the last years in the Tone Mapping community. Although we would like to point out that some interesting work are being produced in this topic: Drago et al. [DMAC03], Yoshida et al. [YBMS05], Kuang et al. [KYL⁺07], Ledda et al. [LCTS05], Cadík et al. [CWNA08] all ran psychological experiments and obtained a set of the best TMO following the users'



Figure 3: Result obtained by tonemapping with the TSTM algorithm (right) and the multi-modal TSTM algorithm (left) the synthetic image 'Bathroom', courtesy of Ward Larsson.

answers. Recently, Aydin et al.[AMMS08] have proposed a measure that is able to output a numerical error by comparing the TM image and the original HDR, allowing a more objective and reproducible way of judging the quality of a TMO.

In this paper we leave the numerical evaluation of our TMO for further research. We will show some results obtained by the multi-modal TSTM method and compare them, both to TSTM and to some methods of the state of the art in Tone Mapping.

Let us start by comparing TSTM and multi-modal TSTM. Multi-modal TSTM was first introduced in order to obtain higher amount of details in areas misrepresented by $\mu(\rho)$ in the TSTM method. The effect of computing eq.(16) over sub-intervals of luminance range is a greater accuracy in the representation of all image areas, as can be seen in the first row of fig. (4). Note how multi-modal TSTM renders well the stained-glass window without diminishing the overall contrast of the image.

The overall brightness of the TSTM output images depended on two factors: the user dependent parameter ρ and the proper location of the final value of μ . By locating the modes automatically the user-dependent parameter is not required and the control of the overall brightness depends mostly on the h_j values. The consequence can be observed in the image 'Cars' in fig.(4): while TSTM produces good results contrast-wise, the overall brightness of the image is low for a midday scene. The reason of this improvement is that the modes obtained with GMM represent better the mass of the histogram (see the histogram in fig. (4), thus the Naka-Rushton functions are more precisely

located.

As it was stated in section 2.4, when the value of $\mu(\rho)$ misrepresents the mass of the histogram, TSTM produces output images with unbalanced contrast between different areas of the histogram, i.e. while some areas present a great amount of details others tend to be flat. An example can be seen in third row of fig. (4). The TSTM result gives a high amount of details in the brighter areas (the sky) leaving the darker areas too bright to reproduce the perception of a shadowy scene. However, multi-modal TSTM balances better the contrast in all the image, obtaining more contrast on the darker areas.

On the other hand, multi-modal TSTM seems to give unnatural results for synthetic HDR images. The reason could be that the method is assuming features of a natural scene that may not be fulfilled by the synthetic image (see fig. 3).

Let us now discuss the results obtained by multi-modal TSTM in comparison with three methods of the state of the art: [DD02], [RWP05] and [MMS06].

From a color reproduction point of view, multi-modal TSTM produces natural colors in dark and bright areas without over- or under-saturating tones, as can be seen in the sky of the 'Office' image or in the stained glass of the 'Memorial' and 'Desk' images, see fig. 5, 6 and 7, respectively.

Taking into consideration over-all contrast, multi-modal TSTM reproduces details in bright and dark areas while maintaining over-all contrast in the image (see the dark areas of 'Memorial' and 'Desk'). Therefore, to the authors opinion multimodal TSTM compares well to the state of the art.

5 Conclusions and perspectives

We have proposed a multimodal extension of TSTM [FPBC09b, FPBC09a], a recent tone mapping operator for HDR images inspired by two sequential stages of the HVS: visual adaptation and local contrast enhancement. The first step is implemented through a suitable modification of the classical Naka-Rushton equation which combines range compression and global rendition of contrast following Weber-Fechner's law. The second step is performed through a variational algorithm for contrast enhancement which is also able to reduce the effect of a possible color cast. The multimodal extension proposed in this paper only affects the first step, while the second is kept unchanged. Our proposal is to use the GMM in order to approximate

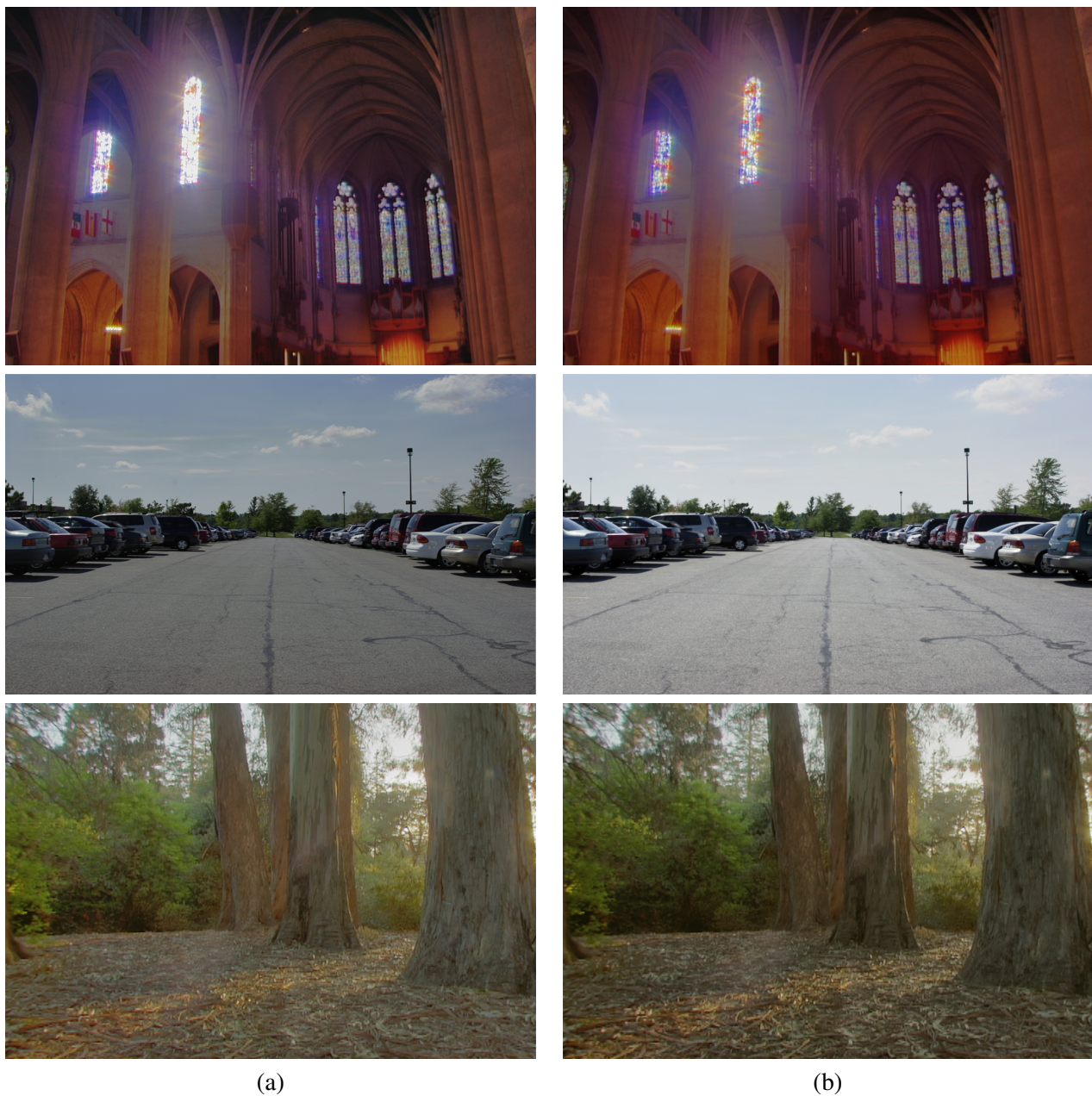


Figure 4: Results obtained by tonemapping with (a) the TSTM algorithm and (b) the multi-modal TSTM algorithm of the images: 'Nave' (first row) courtesy of Paul Debevec[DM97], 'Cars'(second row) and 'GroveC' (third row), courtesy of Paul Debevec[DM97].

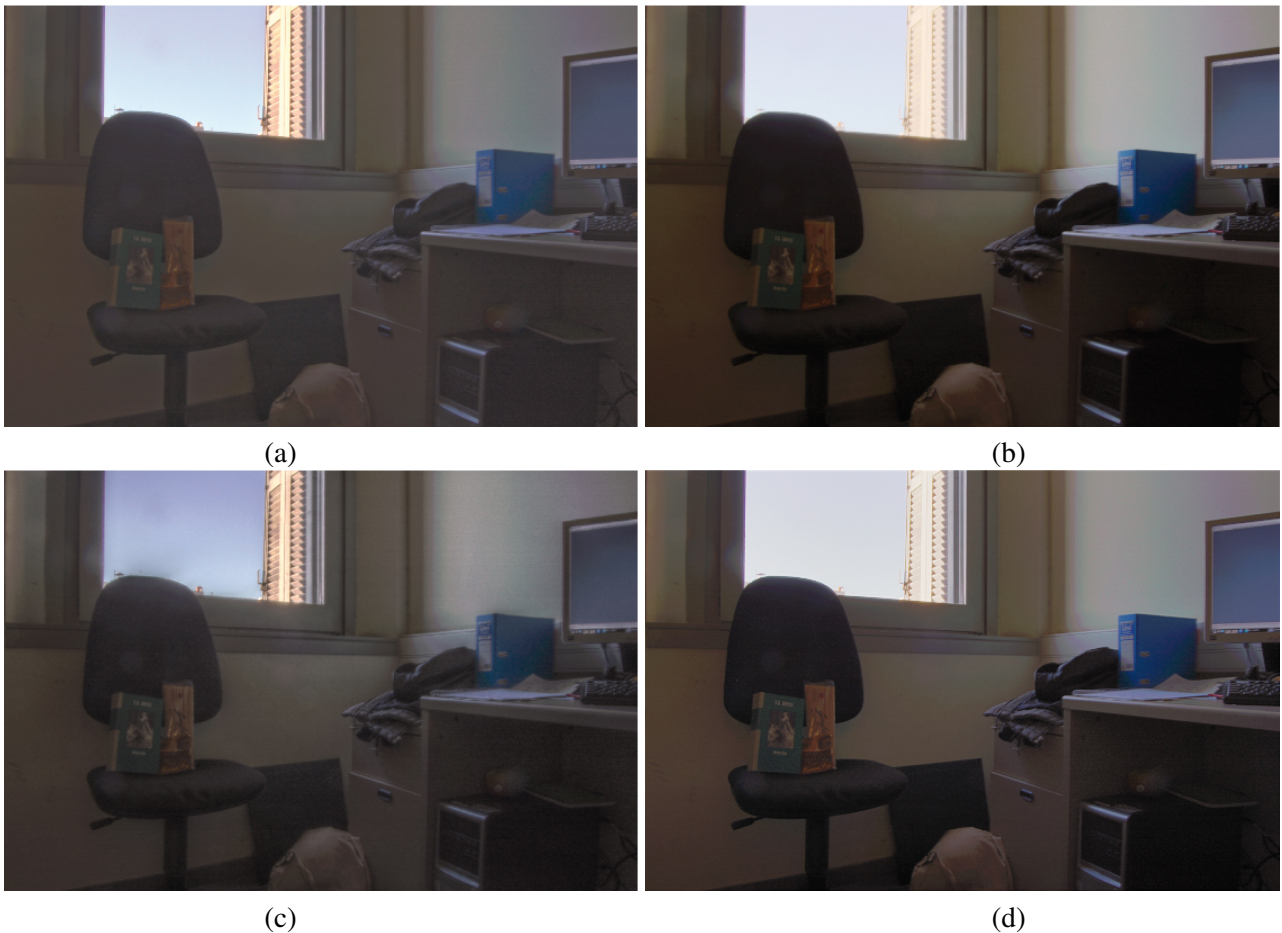


Figure 5: Results of ‘Office’ produced by the methods based on the papers of (a) Durand et al. (b) Reinhard et al. (c) Mantiuk et al. (d) multi-modal TSTM.

the modes of the logarithmic histogram through a collection of Gaussians and use this information in order to implement, for each mode, a Naka-Rushton function that better distributes the tone values with respect to a global transformation. Finally, all these restricted Naka-Rushton functions are merged together.

Our tests have shown that this multimodal extension indeed corresponds to a better rendition of contrast with respect to the global Naka-Rushton transformation proposed in [FPBC09b]. Besides this improvement, with the new method we do not need to restrict the range of HDR images to 5 orders of magnitude anymore, while in [FPBC09b] this was essential in order to avoid a weak rendition of global contrast.

We are currently working on the extension to motion pictures of the technique presented in this paper. As an initial approach we would like to use several correlative frames, corresponding to the same shot, to build the function in eq. (21), and then apply our TM operator, using this function, to all the frames in the

shot. It is expected that if there are no sudden and abrupt changes of luminance in the sequence, the output will not show mapping artifacts.

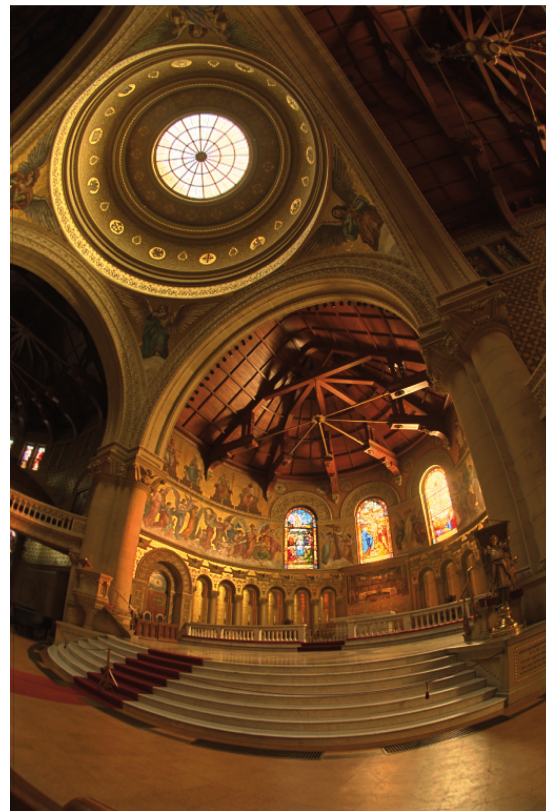
As a theoretical drawback of our model, we point out that the way in which we build domain and codomain of the Naka-Rushton functions corresponding to each Gaussian is not perfectly coherent in the case of overlapping Gaussians. As a future work, it would be interesting to find a smooth way to overcome this problem.

Acknowledgements

The authors would like to thank L. Sánchez for his photograph, 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 PNP GC project, reference MTM2006-14836. V. Caselles also wants to acknowledge “ICREA Acadèmia” prize for excellence in



(a)



(b)



(c)



(d)

Figure 6: Results of the ‘Memorial’ (courtesy of Paul Debevec[DM97]) produced by the methods based on the papers of (a) Durand et al. (b) Reinhard et al. (c) Mantiuk et al. (d) multi-modal TSTM.

research founded by the Generalitat de Catalunya. E. Provenzi acknowledges the Ramón y Cajal fellowship by Ministerio de Ciencia y Tecnología de España.

Disclaimer

Copyright (c) 2004, Industrial Light & Magic, a division of Lucasfilm Entertainment Company Ltd. Portions contributed and copyright held by others as indicated. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- Neither the name of Industrial Light Magic nor the names of any other contributors to this software may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

References

- [AMMS08] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H. P. Seidel, *Dynamic range independent image quality assessment*, ACM SIGGRAPH 2008 papers (Los Angeles, California), 2008, pp. 1–10.
- [Bis06] Christopher M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, 2006, ISBN 0387310738.
- [CW03] L. M. Chalupa and J. S. Werner (eds.), *The visual neuroscience*, MIT Press, 2003, ISBN 0-262-03308-9.
- [CWNA08] M. Cadík, M. Wimmer, L. Neumann, and A. Artusi, *Evaluation of HDR tone mapping methods using essential perceptual attributes*, *Computers & Graphics* **32** (2008), no. 3, 330–349, ISSN 0097-8493.
- [DD02] F. Durand and J. Dorsey, *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, 2002, pp. 257–266, ISBN 1-58113-521-1.
- [DM97] Paul E. Debevec and Jitendra Malik, *Recovering high dynamic range radiance maps from photographs*, Proceedings of the 24th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co, 1997, pp. 369–378, ISBN 0-89791-896-7.
- [DMAC03] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, *Adaptive logarithmic mapping for displaying high contrast scenes*,

Citation

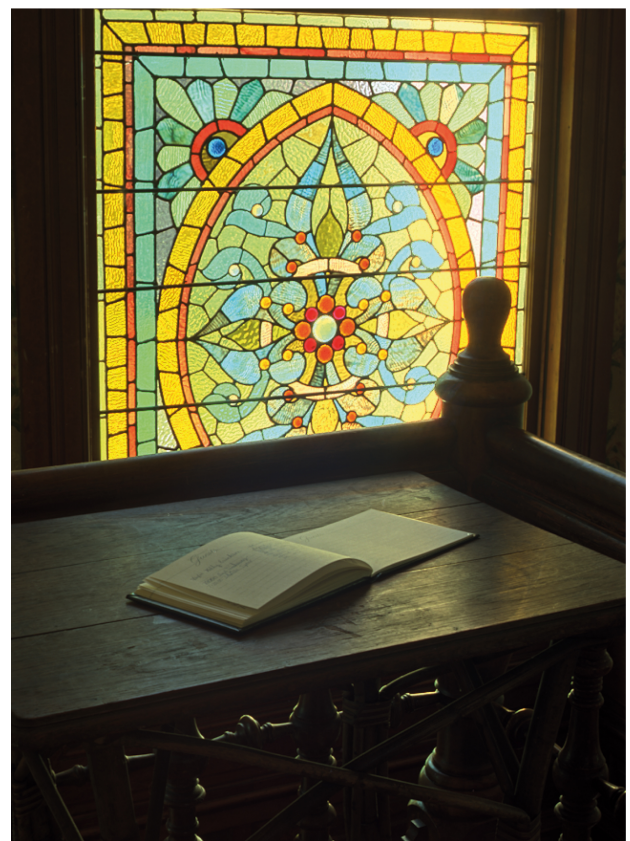
Sira Ferradans, Marcelo Bertalmio, Edoardo Provenzi, and Vicent Caselles, *A multi-modal approach to perceptual tone mapping*, *Journal of Virtual Reality and Broadcasting* 9(2012), no. 7, December 2012, urn:nbn:de:0009-6-35145, ISSN 1860-2037.

- Computer Graphics Forum **22** (2003), 419–426, ISSN 1467-8659.
- [FPBC09a] S. Ferradans, E. Provenzi, M. Bertalmío, and V. Caselles, *An analysis of visual adaptation and contrast perception for a tone mapping operator*, IEEE Transactions on Pattern Analysis and Machine Intelligence **33** (2009), no. 10, 2002–2012, ISSN 0162-8828.
- [FPBC09b] S. Ferradans, E. Provenzi, M. Bertalmío, and V. Caselles, *TSTM: A two-stage tone mapper combining visual adaptation and local contrast enhancement*, IMA Preprint, <http://www.ima.umn.edu/preprints/may2009/may2009.html> (2009).
- [FPSG96] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg, *A Model of Visual Adaptation for Realistic Image Synthesis*, Proceedings of SIGGRAPH 96, Computer Graphics Proceedings, Addison Wesley, 1996, pp. 249–258, ISBN 0-89791-746-4.
- [GKB⁺99] A. Gilchrist, C. Kossyfidis, F. Bonato, T. Agostini, J. Cataliotti, X. Li, B. Spehar, V. Annan, and E. Economou, *An anchoring theory of lightness perception*, Psychol Rev **106** (1999), no. 4, 795–834, ISSN 0033-295X.
- [KMS05] G. Krawczyk, K. Myszkowski, and H. P. Seidel, *Lightness Perception in Tone Reproduction for High Dynamic Range Images*, Computer Graphics Forum **24** (2005), no. 3, 635–645, ISSN 1467-8659.
- [KS08] J. Keener and J. Sneyd, *Mathematical Physiology*, Springer, 2008, ISBN 978-0387094199.
- [KYL⁺07] J. Kuang, H. Yamaguchi, C. Liu, G. M. Johnson, and M. D. Fairchild, *Evaluating HDR rendering algorithms*, ACM Trans. Appl. Percept. **4** (2007), no. 2, 9, ISSN 1544-3558.
- [LCTS05] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen, *Evaluation of Tone Mapping Operators using a high dynamic range display*, Proceedings ACM Transactions on Graphics **24** (2005), no. 3, 640–648, ISSN 0730-0301.
- [LFUS06] Dani Lischinski, Zeev Farbman, Matt Uyttendaele, and Richard Szeliski, *Interactive local adjustment of tonal values*, SIGGRAPH '06: ACM SIGGRAPH 2006 Papers (Boston, Massachusetts), ACM, 2006, pp. 646–653, ISBN 1-59593-364-6.
- [LRP97] G. Ward Larson, H. Rushmeier, and C. Piatko, *A visibility matching tone reproduction operator for high dynamic range scenes*, IEEE Transactions on Visualization and Computer Graphics **3** (1997), no. 4, 291–306, ISSN 1077-2626.
- [Lum95] S. Lumet, *Making movies*, Alfred A. Knopf, 1995, ISBN 0-679-43709-6.
- [MMS06] R. Mantiuk, K. Myszkowski, and H. P. Seidel, *A Perceptual Framework for Contrast Processing of High Dynamic Range Images*, ACM Transactions on Applied Perception (TAP) **3** (2006), no. 3, 286–308, ISSN 1544-3558.
- [NCY79] K. I. Naka, R. Y. Chan, and S. Yasui, *Adaptation in catfish retina*, Journal of Neurophysiology **42** (1979), no. 2, 441–454, ISSN 0022-3077.
- [PAPBC09] R. Palma-Amestoy, E. Provenzi, M. Bertalmío, and V. Caselles, *A perceptually inspired variational framework for color enhancement*, IEEE Transactions on Pattern Analysis and Machine Intelligence **31** (2009), no. 3, 458–474, ISSN 0162-8828.
- [Pra07] W. K. Pratt, *Digital Image Processing: PIKS Scientific inside*, 4. ed., newly updated and rev. ed. ed., J. Wiley & Sons, 2007, ISBN 0-471-76777-8.
- [PTYG00] S. N. Pattanaik, J. Tumblin, H. Yee, and D. P. Greenberg, *Time-Dependent Visual Adaptation For Fast Realistic Image Display*, Proceedings of SIGGRAPH, 2000, pp. 47–54, ISBN 1-58113-208-5.

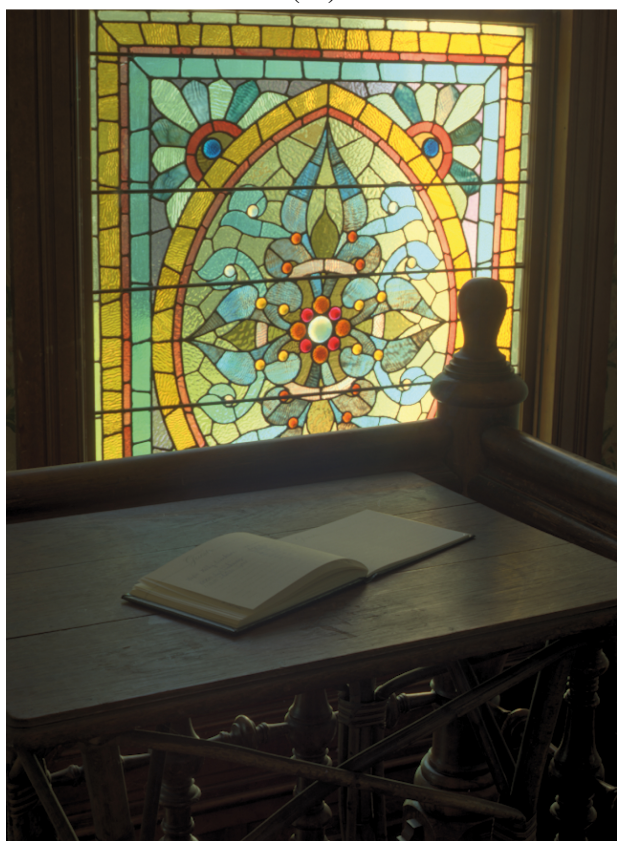
- [RD05] E. Reinhard and K. Devlin, *Dynamic Range Reduction Inspired by Photoreceptor Physiology*, IEEE Trans. on Visualization and Computer Graphics **11** (2005), no. 1, 13–24, ISSN 1077-2626.
- [RSSF02] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, *Photographic tone reproduction for digital images*, ACM Trans. Graph. **21** (2002), no. 3, 267–276, ISSN 0730-0301.
- [RWPD05] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High Dynamic Range Imaging, Acquisition, Display, And Image-Based Lighting*, Morgan Kaufmann Ed., 2005, ISBN 0-12-585263-0.
- [SEC84] R. Shapley and C. Enroth-Cugell, *Visual adaptation and retinal gain controls*, Progress in Retinal Research **3** (1984), 263–346, ISSN 0278-4327.
- [TAMS08] D. Tamburrino, D. Alleysson, L. Meylan, and S. Süsstrunk, *Digital camera workflow for high dynamic range images using a model of retinal processing*, Proceedings SPIE, vol. 6817, 2008, LCAV.
- [WS82] G. Wyszecky and W. S. Stiles, *Color science: Concepts and methods, quantitative data and formulae*, John Wiley & Sons, 1982, ISBN 0-471-02106-7.
- [YBMS05] A. Yoshida, V. Blanz, K. Myszkowski, and H. P. Seidel, *Perceptual evaluation of tone mapping operators with real-world scenes*, Human Vision and Electronic Imaging X, SPIE, 2005, pp. 192–203, ISBN 9780819456397.



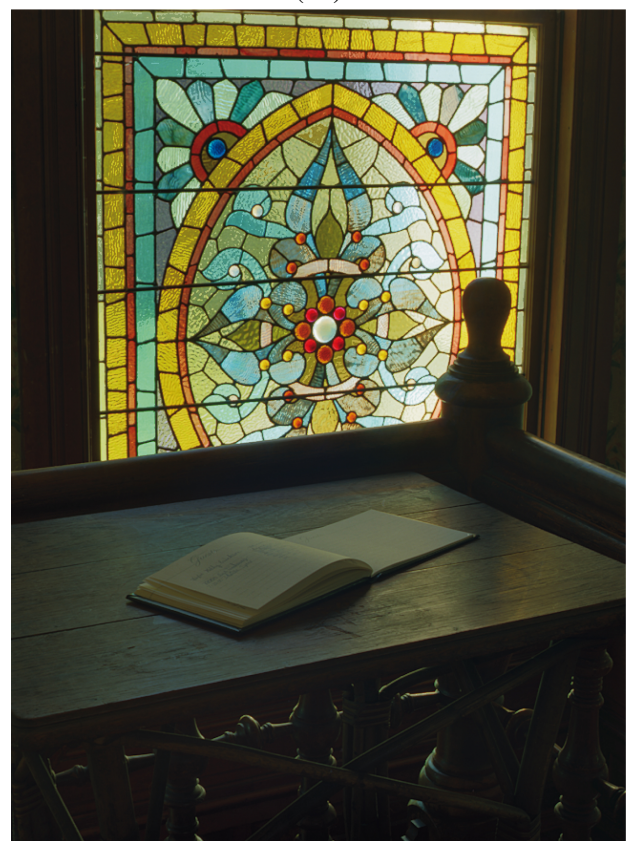
(a)



(b)



(c)



(d)

Figure 7: Results of the ‘Desk’ (courtesy of Industrial Light & Magic, see Disclaimer) produced by the methods based on the papers of (a) Durand et al. (b) Reinhard et al. (c) Mantiuk et al. (d) multi-modal TSTM.