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:

1 Introduction

In daylight, when the Human Visual System (HVS) works best in terms of color vision and perception of details, the amount of the light arriving to our retinas spans many orders of magnitude, from $10^2 \frac{cd}{m^2}$ indoors to $10^9 \frac{cd}{m^2}$ outdoors with the brightest sunlight [4]. But our photoreceptor neurons in the retina, rods and cones, produce electrical outputs which span only two orders of magnitude [14] (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 [7] (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) [16].

An excellent survey of the many TM methods proposed up to 2005 can be found in [13]. Among the more recent works we would like to mention [12, 11, 15], which use a perceptual-based approach involving the Naka-Rushton equation [9]; [5], which uses the anchoring theory of visual perception; and [6], 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 [3]. 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 [3] 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 [3] and discuss its limitations. Section 3 introduces our method and explains how to overcome most of the problems encountered in [3]. 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 Review of the TSTM: a perceptually-inspired two-stage tone mapper

In this section we briefly review TSTM, the tone mapping operator described in [3]. TSTM stands for ‘Two Stage Tone Mapper’ and, as the name indicates, it is composed by two stages, each one inspired by a particular feature of the HVS: the first step implements visual adaptation while the second operates a spatially local contrast enhancement.

Before recalling the formal details, let us introduce the notation that will be used throughout the paper. Let $\mathcal{I} : \mathcal{J} \rightarrow (0, +\infty)^3$, be the radiance map representing the input HDR image, $\mathcal{J}$ being its spatial domain: $\mathcal{J} = \{1, \ldots, W\} \times \{1, \ldots, H\} \subset \mathbb{Z}^2$, 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 $\mathcal{I}$, $c \in \{R, G, B\}$, 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. $\lambda(x) = \frac{1}{3} \left[ I_R(x) + I_G(x) + I_B(x) \right]$, of the pixel $x \in \mathcal{J}$ and with $\lambda$ a generic luminance value, i.e. $\lambda \in [\lambda_{\text{min}}, \lambda_{\text{max}}] \subset (0, +\infty)$, where $\lambda_{\text{min}}$ and $\lambda_{\text{max}}$ are the extreme luminance values. In order to avoid singularities in $\lambda = 0$ we add to the whole luminance image a value of $10^{-12}$.

Let us begin with the explanation of the visual adaptation step. For the sake of simplicity, we will first explain the whole process using the luminance image and later on we will extend the method to the full color image.

Visual adaptation occurs mostly in the retina, where photoreceptors (cones and rods) respond to the incoming light and send the information towards the brain. Experiments performed in controlled conditions have shown that the phenomenological law that represents this response is the so-called Naka-Rushton equation [14]:

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

where $\mu$ is called the semi-saturation value and $r$ measures the normalized response of the photoreceptors to light. Notice that the range of $r$ is a subset of the unit real interval $[0, 1]$. In the neuroscience community the semi-saturation value has been associated with the luminance of the widest area of the scene, however, in the tone mapping community, $\mu$ has been translated as a suitable luminance average.

It has been proven [9] that the normalized response of the photoreceptors (functions $r$) is directly related to the sensation magnitude, which at a psychophysical level, is expressed by Weber-Fechner’s law:

$$\frac{\Delta \lambda}{\lambda + m} = k' \Delta r$$

where $k' > 0$ is a perceptual constant, $\Delta \lambda$ is the minimal perceivable intensity increment, $m > 0$ is related to noise in the visual mechanism and $\Delta r$ is the minimum increment in sensation magnitude.

It is easy to check ([14], [3]) that the Naka-Rushton formula expressed as in eq.(1) fails to be a valid description of the visual adaptation phase for generic real-world luminance conditions [14] because it does not follow Weber-Fechner’s contrast law. The immediate consequence is that it leads to the so-called saturation catastrophe effect: depending on the value of $\mu$, eq.(1) either saturates to 0 (i.e. black) the darkest luminance values or saturates to 1 (i.e. white) the brightest ones, thus losing many details that can be visible in the real-world scene.

As suggested in [14], this problem can be eliminated by replacing $\mu$ with an appropriate non-negative function $f_\mu(\lambda)$, i.e. considering this generalization of the Naka-Rushton equation:

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

where we have maintained the same symbol $r$ to avoid a cumbersome notation.

In [3], the authors proposed to univocally determine $f_\mu$ by imposing the function $r(\lambda)$ in eq.(3) to satisfy Weber-Fechner’s law of contrast perception. This requirement can be formalized through the following differential equation [3] obtained by rearranging eq.(2) and taking infinitesimal differences:

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

By integrating both members with respect to the variable $\lambda$, one obtains:

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

being $C$ an integration constant and $k \equiv \frac{1}{k'}$. Introducing this expression of $f_\mu(\lambda)$ in eq.(3), one finds that the expression of the generalized Naka-Rushton formula coherent with Weber-Fechner’s contrast perception is:

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

The triplet of parameters $C$, $k$ and $m$ can be determined by imposing some general conditions. As commented before the semisaturation value in the Naka-Rushton equation is usually identified with a suitable average of the scene that we translate
mathematically as \( r(\mu) = 1/2 \). Substituting in eq.(6) we can set \( C \equiv \frac{1}{k} - k \log(m + \mu) \). To obtain \( k \) and \( m \) we force the final range of \( r \) to be extended over the entire normalized interval \([0, 1]\); it can be proven [3] that this corresponds to setting

\[
k \equiv \frac{1}{\log\left(\frac{m + \lambda_{\text{max}}}{m + \lambda_{\text{min}}}\right)};
\]

\[
m \equiv \frac{\mu^2 - \lambda_{\text{max}}\lambda_{\text{min}}}{\lambda_{\text{max}} + \lambda_{\text{min}} - 2\mu}.
\]

With these parameters, the function \( f_\mu \) is well defined and non-negative within the range \([\lambda_{\text{min}}, \lambda_{\text{max}}]\) and, by substituting the explicit value of \( C \), we get the expression:

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

with \( k \) and \( m \) defined as in eqs. (7), (8), respectively.

Let us now extend the Naka-Rushton equation to a full color image. In [3], the authors commented that the Naka-Rushton implementation on color images that gives the best results in terms of color rendition is the following:

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

with the same choice of the parameters appearing in \( f_\mu \) as above. In this paper, we will follow this choice.

Let us now pass to the second stage of TSTM: enhancement of spatially local contrast. The phenomenological characteristics of the HVS have been used in [10] 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.

### 2.1 Drawbacks of TSTM

The empirical tests performed by using eq. (10) 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 \( \mu \). In [3], this value has been represented by a convex linear combination in the logarithmic domain between the arithmetic \( \mu_a \) and geometric \( \mu_g \) luminance averages:

\[ \mu(\rho) = \mu_a^\rho \mu_g^{1-\rho}, \text{ where } \rho \in [0, 1]. \]

The best results were achieved using values of \( \rho \) that vary between 0.7 and 1, depending on the particular image. The effect of varying \( \rho \) is a modification in the overall brightness of the output: the bigger the value of \( \rho \), the darker the output.

To bypass the problem related with HDR images whose range extend beyond 5 orders of magnitude, in [3] the authors proposed to redefine \( \lambda_{\text{min}} \) and \( \lambda_{\text{max}} \) in the following way: one computes the histogram of the luminance image in the log-scale and defines a window of 5 orders of magnitude; then slides the window over the log-histogram and for each position computes the number of pixels that fall inside the window; finally, one selects the position where the number of pixels inside the window is maximum. The new values \( \lambda_{\text{min}} \) and \( \lambda_{\text{max}} \) are taken as the extreme points of that window. All values above \( \lambda_{\text{max}} \) are clipped to \( \lambda_{\text{max}} \) and all values under \( \lambda_{\text{min}} \) are clipped to \( \lambda_{\text{min}} \).

However, this does not overcome the problem and all the information contained in the clipped luminance regions below the new value of \( \lambda_{\text{min}} \) and above the new value of \( \lambda_{\text{max}} \) 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 \( \mu \) can fall into a poorly populated region of the histogram, resulting in the under- or over-exposure 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.(6) over each interval and merge them in such a way that the global transformation on \( \lambda \) 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 [1]. 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 \( \bar{\mu}_j \) and standard deviations \( \sigma_j \), \( j = 1, \ldots, N \) in the log-domain\(^1\).

For the \( j \)-th Gaussian, the values \( \tilde{\mu}_j^- \equiv \bar{\mu}_j - 2\sigma_j \) and \( \tilde{\mu}_j^+ \equiv \bar{\mu}_j + 2\sigma_j \).

\(^1\) \( N \), the total number of modes, depends on the orders of magnitude of the image, we have taken \( N = \text{round} \left( \log_{10} \left( \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}} \right) \right) \).
\( \hat{\mu}_j + 2\tilde{\sigma}_j \) 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:

\[
\lambda_{min}^j = \begin{cases} 
\lambda_{min} \left( e^{\hat{\mu}_j - \tilde{\sigma}_j}, e^{\hat{\mu}_j + \tilde{\sigma}_j} \right) & \text{if } j = 1; \\
\max (e^{\hat{\mu}_j - \tilde{\sigma}_j}, e^{\hat{\mu}_{j-1} + \tilde{\sigma}_{j-1}}) & \text{if } j \in \{2, \ldots, N\}, 
\end{cases}
\]

\[
\lambda_{max}^j = \begin{cases} 
\lambda_{max} \left( e^{\hat{\mu}_j + \tilde{\sigma}_j}, e^{\hat{\mu}_j - \tilde{\sigma}_j} \right) & \text{if } j = N \\
\min (e^{\hat{\mu}_j + \tilde{\sigma}_j}, e^{\hat{\mu}_{j+1} - \tilde{\sigma}_{j+1}}) & \text{if } j \in \{1, \ldots, N-1\}, 
\end{cases}
\]

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:

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

where \( r(\lambda) \) is defined as in eq.(6), with the following parameters:

\[
k_j \equiv \frac{1}{\log (m_j + \mu_j + \lambda_{max}^j)};
\]

and

\[
m_j \equiv \max \left( 0, \frac{\mu_j^2 - \lambda_{max}^j \lambda_{min}^j}{\lambda_{max}^j \lambda_{min}^j - 2\mu_j} \right).
\]

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 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:

\[
r_G(\lambda) = \sum_{j=1}^{N} \chi(j)(C_j + h_j r_j(\lambda)),
\]

where function \( \chi(j) \) is the characteristic function of the \( j \)-th sub-interval:

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

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 \( \lambda \) 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 \([r_G(e^{\hat{\mu}_j}), r_G(e^{\tilde{\sigma}_j})] \). The length of the \( j \)-th interval in the perceptual brightness domain is then:

\[
\tilde{h}_j \equiv r_G(e^{\hat{\mu}_j}) - r_G(e^{\tilde{\sigma}_j}) = k \log \left( \frac{m + e^{\hat{\mu}_j}}{m + e^{\tilde{\sigma}_j}} \right).
\]

By normalizing, we obtain the final expression of \( h_j \):

\[
h_j = \frac{\log \left( \frac{m + e^{\hat{\mu}_j}}{m + e^{\tilde{\sigma}_j}} \right)}{\sum_{i=1}^{N} \log \left( \frac{m + e^{\hat{\mu}_i}}{m + e^{\tilde{\sigma}_i}} \right)}.
\]

where \( m \) is computed with global values as in eq. (8) with \( \mu \equiv \mu_G \). This expression guarantees that each \( j \)-th subrange is mapped into a subrange \([e^{\hat{\mu}_j}, e^{\tilde{\sigma}_j}] \subset [0, 1] \) coherently with the Weber-Fechner’s law, eq.(2). 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:

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

The extension of this method to color images is exactly the same as for the TSTM method. We use eq. (10) but we substitute \( f_\mu(\lambda) \) with:

\[
f_G(\lambda(x)) = \frac{\lambda(x)}{r_G(\lambda(x))} - \lambda(x).
\]

4 Results and comparisons

In this section we will show some results obtained by the multimodal 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
Figure 1: (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.

Figure 2: Result obtained by tonemapping with the TSTM algorithm (right) and the multi-modal TSTM algorithm (left) the synthetic image ‘Bathroom’, courtesy of Ward Larsson.

higher amount of details in areas misrepresented by $\mu(\rho)$ in the TSTM method. The effect of computing eq.(9) 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. (3). 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 $\rho$ and the proper location of the final value of $\mu$. By locating the modes automatically the user-dependent parameter is not required and the control of the over-all brightness depends mostly on the $b_j$ values. The consequence can be observed in the image ‘Cars’ in fig. (3): 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. (3), thus the Naka-Rushton functions are more precisely located.

As it was stated in section 2.1, 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. (3). 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. On the other hand, 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. 4).

Let us now discuss the results obtained by multi-modal TSTM in comparison with three methods of the state of the art: [2], [13] and [8].

From a color reproduction point of view, multimodal 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. 4, 5 and 6, respectively.

Taking into consideration over-all contrast, multimodal 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 [3], 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
Figure 3: 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, 'Cars' (second row) and 'GroveC' (third row), courtesy of Paul Debevec.
unchanged. Our proposal is to use the GMM in order to approximate 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 [3]. 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 [3] 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. 14, 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 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

August 2006.


Figure 5: Results of the ‘Memorial’ (courtesy of Paul Debevec) 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.
Figure 6: Results of the ‘Desk’ (courtesy of Industrial Light & Magic) 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.