Perceptual Dynamic Range for In-Camera Image Processing

Praveen Cyriac
praveen.cyriac@upf.edu
David Kane
david.kane@upf.edu
Marcelo Bertalmío
marcelo.bertalmio@upf.edu

Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

Abstract

Digital cameras apply a non-linearity to the captured sensor values prior to quantisation. This process is known as perceptual linearisation and ensures that the quantisation rate is approximately proportional to human sensitivity. We propose an adaptive in-camera non-linearity that ensures that the detail and contrast visible in the processed image match closely with the perception of the original scene. The method has been developed to emulate basic properties of the human visual system including contrast normalisation and the efficient coding of natural images via adaptive processes. Our results are validated visually and also quantitatively by two image quality metrics that model human perception. The method works for still and moving images and has a very low computational complexity, accordingly it can be implemented on any digital camera. It can also be applied off-line to RAW images or high dynamic range (HDR) images. We demonstrate the performance of the algorithm using images from digital cinema, mobile phones and amateur photography.

1 Introduction

Image sensors in digital cameras capture values which are proportional to light intensity and within a range of between 3-4 orders of magnitude. These values are typically passed through a non-linearity prior to quantisation to a range of around 2 orders of magnitude. The purpose of the non-linearity is to ensure that the quantisation rate is approximately proportional to human sensitivity (see [4] and references therein). The problem of reducing the dynamic range of an image while preserving the perceived detail, is known as tone mapping (see [25] for a thorough study on the problem). Digital cameras generally do this by means of gamma correction [23], which is a very simple, image independent approach, designed to produce good results on average. Petit and Mantiuk [22] demonstrated that the simple S-shaped response curve used by many cameras only works well for a subset of images. In the literature there are a number of articles that propose adaptive, image dependant tone mapping curves, often based on psychophysical models, for example, Stevens' law is used in [20], the Naka-Rushton formula in [21, 24], the Weber-Fechner law in [2, 18]. Approaches related to the method presented in this paper include: Hwung et al. [15] who propose an

Pages 19.1-19.11
DOI: https://dx.doi.org/10.5244/C.29.19

adjustable non-linear transform of medium intensities and a linear transform of low and high intensities; Finlayson et al. [12] who find the gamma exponent that maximises the entropy of the image; Brunner et al. [6] determine upper and lower slopes of the tone curve based upon the number of pixels in the low, mid and high region of the image histogram; Mantiuk et al. [19] determine a piece-wise linear curve that minimises the contrast distortion with respect to the original image; Drago et al. [9] apply a logarithmic curve using a base that varies from 2 to 10 depending on the pixel intensity; Larson et al. [18] apply a perceptually constrained cumulative histogram. The current work is most closely related to Ferradans et al. [11] and Cyriac et al. [8]. Both approaches apply two stages; The first stage achieves some degree of histogram equalisation via the application of a point-wise non-linearity. The second stage applies contrast enhancement. Ferradans et al. [11] compute a non-linearity that combines properties of the Weber-Fechner and Naka-Rushton equations. In contrast, the approach of Cyriac et al. [8] is based on new psychophysical research showing that subjects prefer images that have a flat lightness histogram [16], where lightness is the non-linear perception of luminance. Accordingly, Cyriac et al. [8] find the gamma exponent that best flattens the lightness histogram.

This paper presents a method of in-camera image processing that ensures that images look natural and that the detail and contrast visible in the processed image closely match those that can be seen by an observer present at the original scene. Our algorithm adapts to each scene. The adaptions are based on research into natural image statistics, the efficient representation of natural images in the human visual system and contrast normalisation. We validate the model with two image quality metrics that incorporate a model of human vision. Finally, we demonstrate the performance of the model on moving and still images. The model has low computational complexity and thus can potentially operate on camera hardware. The method can also be applied to RAW and HDR images.

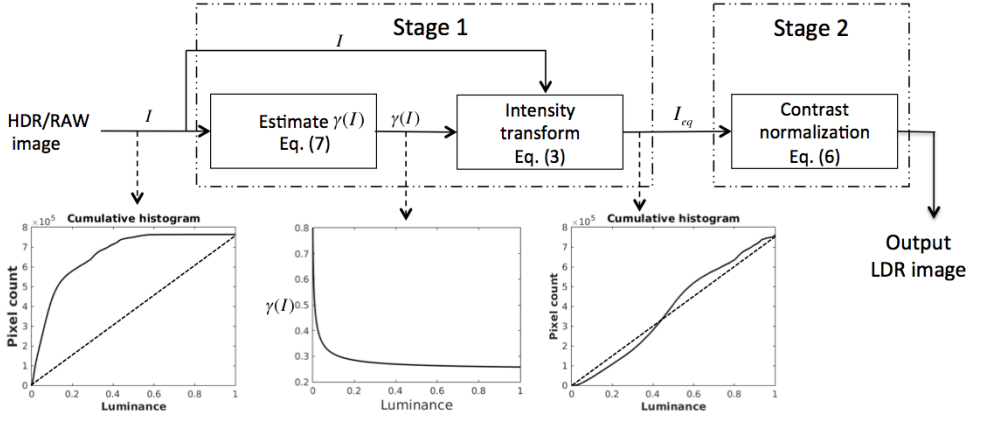

Figure 1: Block diagram of the proposed method.

2 Proposed method

The proposed method finds, for a given image, a transform which ensures that the output values are more evenly distributed over the available range. The transformations applied to the image are based on natural image statistics, psychophysical data and neurophysiological models. Figure 1 shows the block diagram of our proposed method. We have followed the same approach as Ferradans et al. in [11] and our method consists of two main stages, a global one followed by a local contrast enhancement one:

1. Using natural image statistics, a function γ is estimated using some key features of the cumulative histogram of the input intensity image. This function γ is used to perform a transform of the intensity values in a manner that complies with users' preference data as obtained through psychophysical tests [16].

2. The output of the previous stage is passed through an additional contrast normalisation procedure that replicates efficient coding behaviour that occurs both in the retina and cortical areas of the human visual system.

2.1 Natural image statistics and histogram equalisation

In the vision science community the prevailing view is that the visual system transforms the input image to ensure an efficient representation (see [20] and references therein). The human visual system has evolved so as to adapt best to the statistics of natural images. Several works on natural image statistics (e.g. [14, 26]) report that the average shape of the luminance histogram for a natural image is triangular in log-log coordinates: it increases linearly up to a peak, obtained for an image intensity value of M (related to the average of the intensity), and then decreases linearly with a slightly different slope, see Figure 2. This implies that the cumulative histogram, being simply the integral of the histogram, will also be a piece-wise linear function in log-log coordinates, increasing linearly with some slope γ_L until the intensity value M, then increasing linearly with a different slope γ_H. In our method we use this insight from the above-mentioned results on natural image statistics to estimate, for the input image, the particular values of M, γ_L and γ_H that best fit the specific histogram of the image. That is, instead of using fixed values of M, γ_L and γ_H that may

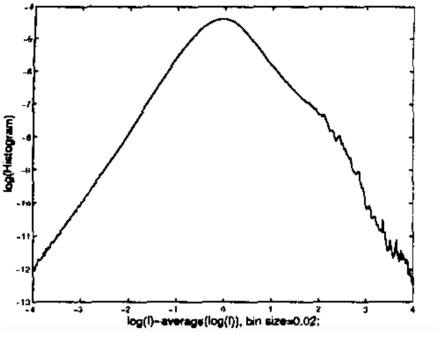

Figure 2: Average histogram of natural scenes, in log-log coordinates. From [14].

adequately represent the average statistics of natural images, we tailor these values to the particular image at hand, obtaining an image-dependent and smooth sigmoid curve $\gamma(I)$ such that $\gamma(I) \simeq \gamma_L$ for small intensities, and $\gamma(I) \simeq \gamma_H$ for large intensities. This approach is also supported by neurophysiological evidence [28] showing that the retina adapts to the light intensity distribution over the image.

Natural scenes tend to have a low-key luminance distribution. This means that low luminance values occur much more frequently than high luminance values. This is especially true for HDR images [16] such as a picture taken directly into sunlight. The results is that many images are dominated by dark, low contrast regions when presented linearly on lower dynamic range media, such as commercial and consumer displays (see Figure 3, left). This problem can be mitigated by a process called histogram equalisation that flattens the luminance distribution of an image. This technique is well established and is effective at increasing the contrast, and in turn, the detail visible in an image. Complete histogram equalisation is achieved by computing the cumulative histogram of an image and applying this as a point wise non-linearity as follows, where H is the normalised cumulative histogram, I the original normalised image and x a pixel location:

$$I_{eq}(x) = H(I(x)). \tag{1}$$

Although complete histogram equalisation is highly effective at increasing image contrast, it can lead to very sharp changes in contrast and frequently produces unnatural looking images (see Figure 3, middle). Thus some form of constrained histogram equalisation is necessary (see Figure 3, right). One approach is to apply a smooth function that approximates the cumulative histogram. In this work we apply a smooth function that is derived from the statistics of natural scenes. As we demonstrate above, for natural images the average cumulative histogram can be modelled in log-log coordinates as a piecewise linear function with two different slopes; therefore, in linear-linear coordinates the cumulative histogram has the form

$$H(I) = I^{\gamma(I)}, \tag{2}$$

where $\gamma(I)$ is the function described in the previous section. From Equations (1) and (2) we introduce the first stage of our model, which produces constrained histogram equalisation:

$$I_{eq}(x) = (I(x))^{\gamma(I(x))}. \tag{3}$$

Recent research has demonstrated that subjects display a preference for images with a flat lightness distribution [16], where lightness is the non-linear perception of world luminance values. We model the visual pipeline from capture to perception as consisting of two

Figure 3: Comparison of complete histogram equalisation and proposed method. From left to right: linearly scaled HDR image, result of complete histogram equalisation of the HDR image, result of constrained histogram equalisation (proposed first stage) applied on the HDR image.

main nonlinearities. The first is from image capture (I) to presentation on a display (I_{disp}). This non-linearity is known as the system gamma, a power law that can be modelled as the product of the camera nonlinearity (encoding gamma) and the display nonlinearity (decoding gamma):

$$I_{disp} = I^{\gamma_{sys}}, \quad \gamma_{sys} = \gamma_{enc}\gamma_{dec}. \tag{4}$$

The second nonlinear function is the psychological relationship Ψ between real world luminance (e.g. as displayed on a monitor) and perceived luminance L:

$$L = \Psi(I_{disp}). \tag{5}$$

In this work we make the simplifying assumption that the encoding gamma is the inverse of the decoding gamma, therefore the initial image I can be said to be proportional to the displayed image I_{disp}. In that case we can see the first stage of our model, Eq. (3), as an approximation to the perceptual function in Eq. (5).

2.2 Contrast normalisation

In the neuroscience literature there is abundant neurophysiological evidence (see [6, 7] and references therein) that the visual system is performing an operation called contrast normalisation, in which the contrast (the difference between light intensity and its mean value) is divided by a factor depending on the standard deviation of the light intensity. This re-scaling already occurs at the retina and optimises information transmission and coding efficiency [5, 17]. Given that contrast normalisation is a key element of the human visual system we have incorporated it to our method with the following second and final stage:

$$O(x) = \mu(x) + (I_{eq}(x) - \mu(x)) * k/\sigma, \tag{6}$$

where x is a pixel, $I_{eq}(x)$ is the value at pixel x computed at the previous stage of our method, $\mu(x)$ is the local mean of I_{eq}, k is a constant, σ is the standard deviation of I_{eq}, and $O(x)$ is the final output value of our method for pixel x.

3 Implementation

In this section we will present the implementation details. Our method consists of the two stages described by Eqs. (3) and (6). Both equations are applied separately to each of the red, green and blue colour channels, which have previously been normalised into the range $[0,1]$ and where values above the 99 percentile have been clipped.

The $\gamma(I)$ curve in Eq. (3) is defined by estimating its parameters (γ_L, γ_H and M) from the cumulative histogram of the luminance channel in log-log coordinates, as illustrated in Figure 4 and explained in what follows. The luminance channel is the Y-channel of the input image when converted into XYZ colour space. Our estimate of M will be the average of the intermediate values L_m and L_M on the horizontal axis (log luminance), which respectively correspond to the values of 1 and 90% in the vertical axis (log cumulative histogram). Next we define M^+ as the value above M at one third the difference between L_M and M.

The values of γ_L and γ_H are estimated with respect to M^+: the slope of the line from the value at M^+ to the end of the curve gives γ_H, and the slope of the line that goes from the value at M^+ to the point on the curve that is 1 unit lower than M^+ along the vertical axis

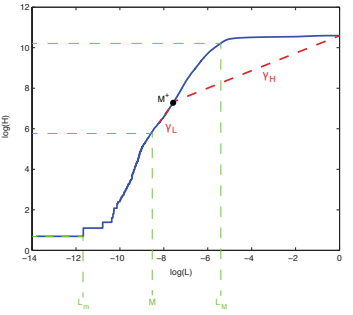

Figure 4: Example of a cumulative histogram for a single natural image (in log-log axes) and our estimated parameters γ_L, γ_H and M.

will be our estimate for γ_L. Having the three parameters γ_L, γ_H and M we can now define the function $\gamma(I)$ as follows:

$$\gamma(I) = \gamma_H + (\gamma_L - \gamma_H)(1 - \frac{I^n}{I^n + M_{lin}^n}), \tag{7}$$

where M_{lin} is simply the exponential of M (since I is in linear coordinates but M was computed on log luminance values) and n is a fixed exponent that regulates the steepness of the curve (in practice we set $n = \gamma_L$). The function $\gamma(I)$ thus defined goes from γ_L to γ_H with a smooth transition at M_{lin}, as it was argued in Section 2.1 that it should behave in order for Eq. (3) to perform histogram equalisation of natural images.

For the second stage, Eq. 6, we have to specify the value of k and the way the local mean $\mu(x)$ is computed. The value $\mu(x)$ is obtained by convolving I_{eq} with a kernel W which is a linear combination of two Gaussian kernels, one with small standard deviation ($\sigma = 5$) which is given a weight of 0.9 and the other with a larger standard deviation ($\sigma = 25$) which is given a weight of 0.1. The constant k controls the level of contrast of the image, and a value of $k = 0.33$ produces results with good contrast and a natural appearance.

For video sequences, applying the above method to each frame separately may result in flickering artefacts due to the possibility of a sudden change in the tone curve from one frame to the next. Therefore for video we adopt a two pass approach. In the first pass, we estimate γ_L, γ_H and M for each frame separately, then we apply a temporal low pass filter to these values. In the second pass, we apply our method (Eqs. (3) and (6)) with the new parameters.

As a final comment, we note that all the operations described above (computation of histogram, estimation of parameters, computation of mean and standard deviation, etc.) are of low computational complexity and therefore the proposed method is a very good candidate for real-time applications.

4 Experiments and comparisons

We note that all the results reported in this section use the same procedure of automatic parameter estimation described above. One potential application of our method is for the in-camera automated processing and compression of images. To illustrate the potential of

(a) (b) (c)

(d) (e) (f)

Figure 5: First row: original JPEG images as recorded by the camera, with the exception of image (b) which is generated by applying a S-shaped curve to the LogC data of the RAW image. Second row: results of applying our method to the corresponding RAW images. Camera models: left column, Nikon D3100 consumer photography camera; middle column, ARRI Alexa digital cinema camera[1]; right column, Nexus 5 smartphone camera.

this method we compare the regular JPEG output versus the result of applying our method to the corresponding RAW image. The results are illustrated with images from consumer, smartphone and cinema cameras and the results are shown in Figure 5. Note the natural appearance, enhanced contrast and absence of halos, spurious colours or visual artefacts of any kind.

Figure 6 shows how our method can also be used offline as a tone mapping operator, applied to HDR images. Again, no visual artefacts can be observed.

Next, we perform a quantitative evaluation of our method. We use the Fairchild dataset [10] of HDR images and compare our algorithm with six state-of-the-art tone mapping operators, using the metrics DRIM [3] and TMQI [30]. The results are shown in Table 1. The values for DRIM are global error scores, thus the lower this number, the better the method is, whereas the opposite is the case for TMQI: the values for TMQI are estimates of structural fidelity (S) and naturalness (N) of a tone-mapped image with respect to the original HDR image, with an overall quality (Q) being a weighted average of S and N, thus the higher the values, the better the method is. We can see from Table 1 that, according to DRIM, the amount of distortions produced by our approach matches that of Mantiuk et al. [19] and is significantly less than the error introduced by all other tested methods. In terms of TMQI, our method ranks as the best in terms of overall quality (Q), and matches to Mantiuk et al. [19] for structural fidelity (S). It is worth noting that our results have a very high naturalness (N) index. The evaluation by DRIM is illustrated in Figure 7. The overlaid colours on the grayscale images represent the error introduced by the tone mapping procedure: green portrays a loss of contrast, red an amplification of contrast and blue, inversions of contrast. The saturation of the colour corresponds to the error magnitude. We can see in the two examples that our method produces significantly less error in general and, in particular, a reduced loss of contrast.

Finally, in Figure 8 we compare our method with state-of-the-art tone mapping operators

Figure 6: Results of our method applied to HDR images from the Fairchild dataset (top row) [10] and to images from the ARRI dataset (bottom row)[13].

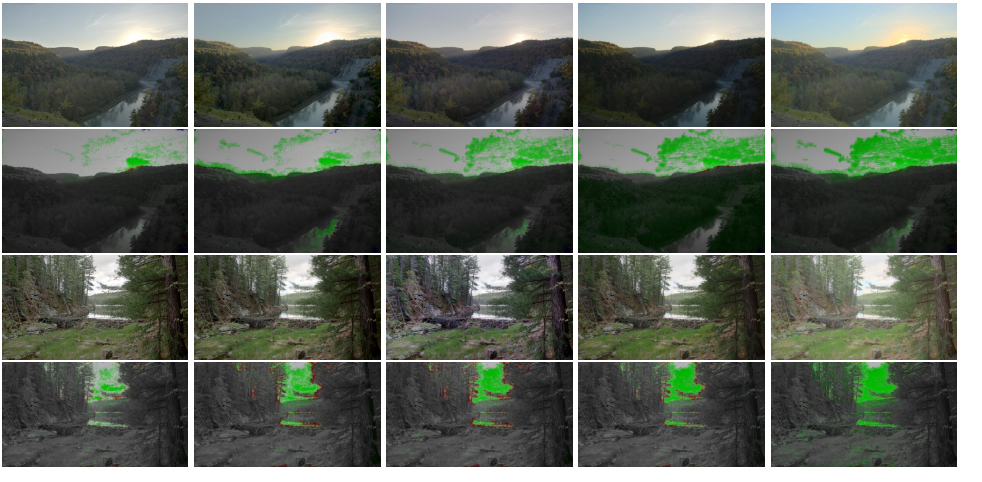

Figure 7: Comparison of our method with state-of-the-art tone mapping operators. The first and third rows show the results of various algorithms and the second and fourth rows show the corresponding error maps computed with DRIM [3]. From left to right: the proposed method, Mantiuk et al. [19], Ferradans et al. [11], Reinhard et al. [24] and Drago et al. [9].

Images	TMO	DRIM[8]	TMQI[30]		
			Q	S	N
Average of 41 images for DRIM 106 images for TMQI	Only 1st stage, Eq. (3)	0.49	0.90	0.91	0.55
	Proposed method	**0.46**	**0.91**	**0.92**	**0.58**
	Mantiuk et al. [19]	**0.46**	0.90	**0.92**	0.51
	Ferradans et al. [11]	0.50	0.89	0.89	0.50
	Drago et al. [9]	0.54	0.87	0.87	0.42
	Reinhard et al. [24]	0.51	0.87	0.88	0.44
	Singnoo et al. [27]	0.54	0.89	**0.92**	0.48
	Cyriac et al. [8]	0.48	0.89	**0.92**	0.51

Table 1: Quantitative evaluation using the Fairchild dataset [10].

Figure 8: Comparison of our method with state-of-the-art tone mapping operators on two HDR video sequences from the ARRI dataset [13]. Top row: "smith_hammering". Bottom row: "fishing_longshot". From left to right, the proposed method, Mantiuk et al. [19], Ferradans et al. [11], Reinhard et al. [24], and Drago et al. [9].

using two HDR video sequences from the ARRI dataset [13]. We can see that our algorithm produces results that are natural looking and with very little noise[1]. The video displays no visible flicker nor any sort of spatiotemporal artefacts.

5 Conclusion

We have presented a method for in-camera non-linear mapping that adapts to the particularities of each image and produces results that look natural even for challenging scenes. It is based on basic properties of the human visual system, works for still and moving images and has very low computational complexity. The experiments and comparisons show that the results obtained with our method, for a variety of camera types, are visually and quantitatively very good, without halos, flicker or spatiotemporal artefacts of any kind. We are currently improving the automated parameter selection.

6 Acknowledgment

A patent application based on the research in this article has been filed at the European patent office, Application no 15154172.9-1906. This work was supported by the European Research Council, Starting Grant ref. 306337, by the Spanish government, grant ref. TIN2012-38112, and by the Icrea Academia Award.

[1]This casts some doubts on the applicability of the TMQI metric, since the naturalness score N that we get for these images is much smaller than that of Mantiuk et al. [19]

References

[1] Stefano Andriani, Harald Brendel, Tamara Seybold, and Joseph Goldstone. Beyond the kodak image set: A new reference set of color image sequences. In *ICIP*, pages 2289–2293, 2013.

[2] Michael Ashikhmin. A tone mapping algorithm for high contrast images. In *Proceedings of the 13th Eurographics workshop on Rendering*, pages 145–156. Eurographics Association, 2002.

[3] Tunç Ozan Aydin, RafałMantiuk, Karol Myszkowski, and Hans-Peter Seidel. Dynamic range independent image quality assessment. In *ACM SIGGRAPH 2008 Papers*, SIGGRAPH '08, pages 69:1–69:10, 2008.

[4] Marcelo Bertalmío. *Image Processing for Cinema*. CRC Press, 2014.

[5] Naama Brenner, William Bialek, and Rob de Ruyter Van Steveninck. Adaptive rescaling maximizes information transmission. *Neuron*, 26(3):695–702, 2000.

[6] Ralph Brunner. Automatic tone mapping curve generation based on dynamically stretched image histogram distribution, November 20 2012. US Patent 8,314,847.

[7] Matteo Carandini and David J Heeger. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62, 2012.

[8] Praveen Cyriac, Marcelo Bertalmio, David Kane, and Javier Vazquez-Corral. A tone mapping operator based on neural and psychophysical models of visual perception. In *SPIE/IS&T 2015 Human vision and electronic imaging XX*, 2015.

[9] Frédéric Drago, Karol Myszkowski, Thomas Annen, and Norishige Chiba. Adaptive logarithmic mapping for displaying high contrast scenes. In *Computer Graphics Forum*, volume 22, pages 419–426. Wiley Online Library, 2003.

[10] Mark D Fairchild. The hdr photographic survey. In *Color and Imaging Conference*, volume 2007, pages 233–238. Society for Imaging Science and Technology, 2007.

[11] Sira Ferradans, Marcelo Bertalmio, Edoardo Provenzi, and Vicent Caselles. An analysis of visual adaptation and contrast perception for tone mapping. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(10):2002–2012, 2011.

[12] Graham Finlayson and Ruixia Xu. Gamma adjustment for maximizing information in images, September 17 2013. US Patent 8,538,145.

[13] Jan Froehlich, Stefan Grandinetti, Bernd Eberhardt, Simon Walter, Andreas Schilling, and Harald Brendel. Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays, 2014.

[14] Jinggang Huang and David Mumford. Statistics of natural images and models. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1. IEEE, 1999.

[15] Den-Jen Hwung, Jen-Chuan Wang, and Der-Song Su. Digital gamma correction system for low, medium and high intensity video signals, with linear and non-linear correction, December 5 1995. US Patent 5,473,373.

[16] David Kane and Marcelo Bertalmío. Is there a preference for linearity when viewing natural images? In *SPIE/IS&T 2015 Image Quality and System Performance XX*, 2015.

[17] David B Kastner and Stephen A Baccus. Insights from the retina into the diverse and general computations of adaptation, detection, and prediction. *Current opinion in neurobiology*, 25:63–69, 2014.

[18] Gregory Ward Larson, Holly Rushmeier, and Christine Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *Visualization and Computer Graphics, IEEE Transactions on*, 3(4):291–306, 1997.

[19] RafałMantiuk, Scott Daly, and Louis Kerofsky. Display adaptive tone mapping. *ACM Trans. Graph.*, 27(3):68:1–68:10, August 2008.

[20] Bruno A Olshausen and David J Field. Vision and the coding of natural images. *American Scientist*, 88(3):238–245, 2000.

[21] Sumanta N Pattanaik, Jack Tumblin, Hector Yee, and Donald P Greenberg. Time-dependent visual adaptation for fast realistic image display. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 47–54. ACM Press/Addison-Wesley Publishing Co., 2000.

[22] Josselin Petit and Rafał K Mantiuk. Assessment of video tone-mapping: Are cameras? s-shaped tone-curves good enough? *Journal of Visual Communication and Image Representation*, 24(7):1020–1030, 2013.

[23] Charles Poynton. *Digital video and HD: Algorithms and Interfaces*. Elsevier, 2012.

[24] E. Reinhard and K. Devlin. Dynamic range reduction inspired by photoreceptor physiology. *Visualization and Computer Graphics, IEEE Transactions on*, 11(1):13–24, Jan 2005. ISSN 1077-2626. doi: 10.1109/TVCG.2005.9.

[25] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[26] Daniel L Ruderman. The statistics of natural images. *Network: computation in neural systems*, 5(4):517–548, 1994.

[27] Jakkarin Singnoo. *A simplified HDR image processing pipeline for digital photography*. PhD thesis, University of East Anglia, 2012.

[28] Stelios M Smirnakis, Michael J Berry, David K Warland, William Bialek, and Markus Meister. Adaptation of retinal processing to image contrast and spatial scale. *Nature*, 386(6620):69–73, 1997.

[29] G. Tumblin and H. Rushmeier. Tone reproduction for computer generated images. *IEEE Computer Graphics and Applications*, 1993.

[30] H. Yeganeh and Zhou Wang. Objective quality assessment of tone-mapped images. *Image Processing, IEEE Transactions on*, 22(2):657–667, Feb 2013.