SMPTE Meeting Presentation

Automatic, viewing-condition dependent contrast grading based on perceptual models

Praveen Cyriac

Universitat Pompeu Fabra, Barcelona, Spain, praveen.cyriac@upf.edu

David Kane

Universitat Pompeu Fabra, Barcelona, Spain, david.kane@upf.edu

Marcelo Bertalmío

Universitat Pompeu Fabra, Barcelona, Spain, marcelo.bertalmio@upf.edu

Written for presentation at the

SMPTE 2016 Annual Technical Conference & Exhibition

Abstract. *Cameras automatically apply non-linear transformations to the sensor data, allowing for perceptually-uniform quantization suited to standard dynamic range displays in dim conditions. In the cinema industry, data is recorded in raw (linear) format and non-linearly corrected in post-production by a skilled technician who optimizes image appearance for cinema (dark) conditions. We propose a method that automatically performs this non-linear transformation taking into account the intended viewing conditions. It is based on visual perception models and produces results that look natural, without any spatio-temporal artifacts. User preference tests show that our method outperforms state of the art approaches. The technique is fast and could be implemented on camera hardware. It can be used for on-set monitoring on regular displays, as a substitute for gamma-correction, and as a way of providing the colorist with content that is both natural looking and has a crisp, clear image.*

Keywords. High dynamic range, tone mapping, human visual system, psychophysics, natural image statistics, contrast grading.

Introduction

Cameras normally apply automatic non-linear transformations to the sensor data. These transformations allow for a perceptually-uniform quantization that is suited to standard dynamic range displays in dim conditions. The non-linearity applied in most digital cameras is well approximated by a simple power law, and while this may perform well on average, in general when dealing with high dynamic range scenes it is suboptimal.[1]

In the cinema industry, camera sensor data is recorded in raw (linear with respect to light intensity) form and is later non-linearly corrected by a skilled technician in post-production to optimize image appearance for cinema (dark) conditions. The literature provides a number of image-dependent automatic and semi-automatic non-linearities, some based on models of the human visual system.[2,3] Additionally, there are more complex, local tone mapping algorithms that could, in theory, transform the sensor data, but these tend to be computationally more expensive, sensitive to image fluctuations, and harder to tune for specific applications. As a result, manual contrast grading is always preferred in professional cinema.

Previously, we proposed a method[4] that automatically performs this non-linear transformation taking into account the image content and which has the following properties:

- It is automatic (no need for user-selected parameters) and fast, such that it can work in-camera.
- It accurately reproduces the detail and contrast that are visible in the original scene.
- It produces image that looks natural, without any visible artifacts and color distortions.
- It works on video sequences and results do not show temporal artifacts.

The method[4] is based on visual perception models and is well suited to the statistics of natural scenes, having been optimized and validated through psychophysical tests that confirm that it outperforms other state of the art algorithms in terms of users' preference on an LCD display in an office environment.

It is well known that viewing conditions and the display's capabilities in terms of contrast and brightness can significantly affect the perceived image quality. Ideally, contrast grading should take into account all these factors, although a conventional simplified solution is to apply a system gamma depending on the intended viewing condition.

Our contribution here is threefold. First, we conduct psychophysical experiments and determine the adequate gamma adjustments needed for the results produced by the method[4] to look best in different environment and display types. Second, we record some luminance measurements and compute sequential and ANSI contrast of three display types in two different surround conditions. We show that the effective contrast produced by a display depends not only on its maximum contrast ratio but also on the surround and the reproduced image content. Last, based on psychophysical data, we develop a mathematical model and show that the subjets' choice of gamma adjustment can be predicted from the ANSI contrast and peak display brightness.

The proposed method can provide the colorist with a fast and automatic graded content that is both natural looking and has a crisp and clear appearance. In addition the method can be implemented in-camera as a substitute for gamma correction. The method can also be used as an off-line tone mapping method for converting high dynamic range (HDR) images into low dynamic range (LDR) ones, with applications to cinema shoots (on-set use of LDR monitors with an HDR camera), television broadcast (making HDR signals compatible with LDR equipment), and rendering in computer graphics (for video games, 3D animation, the integration of CGI onto real footage, etc.)

Model Explanation

The efficient coding hypothesis[5] and neuroscience experiments[6] indicate that the visual system, when presented with natural scenes, transforms input signals to ensure a uniform distribution of output levels[7]. In addition, the psychophysical study[8] showed that subjects tend to prefer images with a flat lightness histogram, however complete histogram equalization may produce results with unnatural appearance due to sharp changes in contrast.

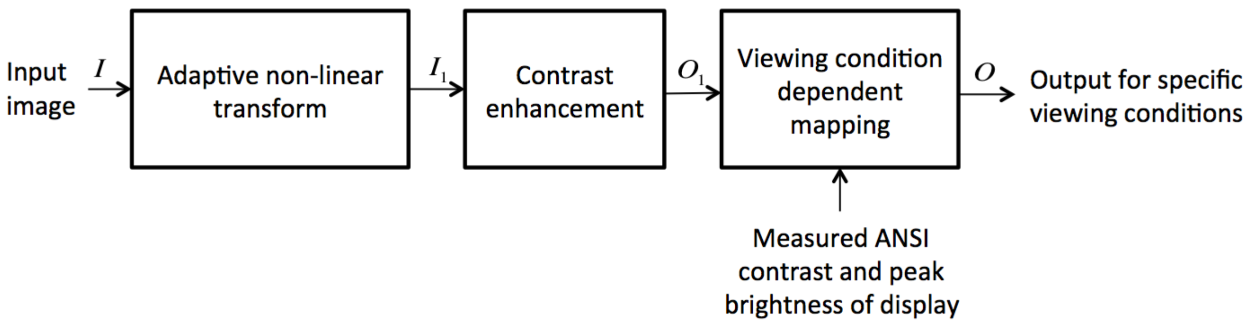


Figure 1. Block diagram of the proposed method

Figure 1 shows the block diagram of our proposed model based on the behavior of visual perception just mentioned. Blocks 1 and 2 were introduced by Cyriac et al.[4] The first stage is an adaptive non-linear transform that performs constrained histogram equalization by modeling the cumulative histogram as a smooth function based on natural image statistics. Studies in natural image statistics[9,10] reported that, on average, natural images have a triangular shaped histogram in log-log coordinates. Thus, the cumulative histogram (*H*) can be modeled as a piecewise linear function with two slopes (γ_L and γ_H) in log-log coordinates; therefore, in linear-linear coordinates *H* has the form

$$H(I) = I^{\gamma(I)}$$

where I is the normalized image and γ is a slope function with $\gamma \simeq \gamma_L$ for small intensities, $\gamma \simeq \gamma_H$ for large intensities and a smooth transition at M_{lin} with a slope n as follows:

$$\gamma(I) = \gamma_H + (\gamma_L - \gamma_H)\left(1 - \left(\frac{I^n}{I^n + M_{lin}^n}\right)\right)$$

The parameter values of γ_L, γ_H and M_{lin} are automatically computed from the cumulative histogram of the input intensity image to corroborate well with values obtained in a subjective study (detail explanation in article[4] by Cyriac et al.).

So the first stage, that performs constrained histogram equalization based on natural image statistics, is given by:

$$I_1 = \left(I^{\gamma(I)}\right)C(I)$$

where C is an adaptive clipping function incorporated to preserve the global contrast of the image. The parameter values of the function C is also computed from the cumulative histogram and from γ_L , γ_H and M_{lin}.

The second stage performs contrast normalization based on the neurophysiological evidence[11,12] which explains that the visual system performs normalization of the contrast by a factor depending on the standard deviation of the light intensity. The second stage is given by:

$$O_1(x) = \mu(x) + (I_1(x) + \mu(x))\frac{k}{\sigma}$$

where x is the pixel position, I_1 is the output of the previous stage, μ and σ are the local mean and global standard deviation of I_1 respectively, k is a constant and $O_1(x)$ is an intermediate output, optimized for a LCD display in an office environment.

The final output of the method is

$$O(x) = (O_1(x))^{\gamma_{adj}}$$

The following sections detail how to determine the optimal value for γ_{adj} automatically, depending on the display characteristics, ambient light intensity and surround luminance.

The method can also be applied to video if we incorporate temporal coherence in the parameter estimation of the first stage: the initial parameter values are modified by averaging with the values of the previous frames.

Viewing condition dependent mapping

In order to develop a general model that determines the necessary non-linear adjustment value γ_{adj} needed for our results to look optimal under some given conditions, we perform the following steps:

1. Conduct psychophysical experiments to determine the optimal non-linear adjustment the subjects prefer for different surround conditions and displays.

2. Record several physical measurements including minimum and maximum luminance of the display, surround luminance and ambient illuminance.

With the above data in hand we develop a general formula to predict the subject’s choice of non-linearity value of γ_{adj} for a specific surround environment and display type.

Psychophysical experiment

We explain the details of the psychophysical experiment in this section. The viewing conditions that we consider are

Two surround environments:

Office room: ambient illuminant of 47 nits and average near surround luminance of 65 nits

Dark room: ambient illuminant of 0.3 nits and average near surround luminance ≈ 0 nits

and three display types:

LCD: ASUS VS197D LCD monitor set to sRGB mode

OLED: Sony Trimaster PVM

HDR: SIM2 HDR47ES4MB monitor set to HDR mode

The stimuli for the experiment are 20 tone mapped versions of the HDR images from the high dynamic range survey by Mark Fairchild[13]. The images were chosen to cover a variety of scenarios: night images, indoor scenes, bright outdoor scenes, landscapes, etc. The HDR images were tone mapped by the method of Cyriac et al.[4]

Experiment procedure: A schematic of the experiment setup is shown in figure 2. Subjects were asked to adjust the gamma non-linearity via a scroll bar such that the image achieves an optimal appearance. Table 1 shows the average subject choice for γ_{adj}, for two observers, 20 images, 3 display types and two surround environments.

Figure 2. Psychophysical experiment setup

Display	Dark room	Office room
LCD	1.11	1
OLED	1.09	0.98
HDR	-	1.5

Table 1. Subject choice for γ_{adj}

Contrast measurement

For each combination of display type and surround condition we measured both the sequential and ANSI contrast. Both contrasts are given by the ratio of the average white luminance to the average black luminance. In the case of the sequential contrast, the luminance measurement is recorded when the whole display is black or white. On the other hand, in the case of the ANSI contrast the luminance measurement is recorded while displaying the pattern shown in figure 3.

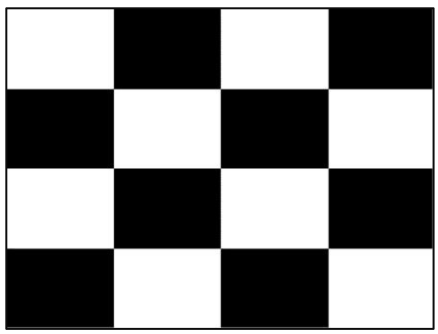


Figure 3. Checker pattern to measure ANSI contrast

A Konica Minolta LS 100 photometer was used to measure the luminance and the reading was taken at a distance at which an observer views the display (approximately 3 times the display height). Results are shown in tables 2 and 3.

Model to predict the experiment result

Analyzing the results of the psychophysical experiment and the contrast measurement, we found that the subject preferred non-linear adjustment value γ_{adj} can be predicted from the ANSI contrast and the maximum brightness of the display by the following formula:

$$\gamma_{adj} = (1 + 0.2\,|C|)^{sign(C)} \qquad (1)$$

where

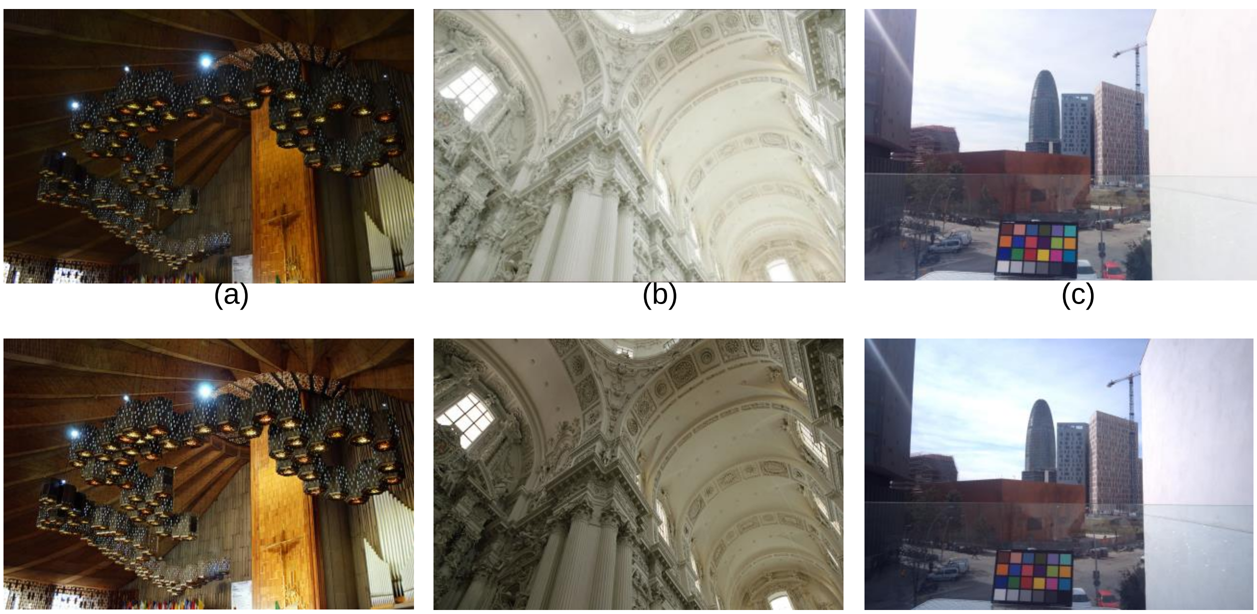
$$C = log_{10}\left(\frac{L_{vc}^{peak}}{L_{grad}^{peak}}\right) + log_{10}(ANSI_{vc}) - log_{10}(ANSI_{grad}),$$

L_{vc}^{peak} and $ANSI_{vc}$ are the peak brightness and the ANSI contrast of the intended display in which the image is to be viewed. L_{grad}^{peak} and $ANSI_{grad}$ are the peak brightness and the ANSI contrast of the grading display: an LCD display in an office environment in our experiments, since the parameter values for the first stage of our model were selected to optimize image appearance in that scenario. Table 4 shows that the model predicts well the result of the psychophysical experiment.

Results and discussion

In this section we first present some sample results of our method when applied to the RAW camera sensor data and also when applied to HDR images. Then we show the luminance and contrast measurements recorded by setup detailed in the contrast measurement section. Last we show the result of the psychophysical experiment and compare it with the values predicted by equation 1.

In figure 4 we show the potential of our method to be used for the in-camera non-linear processing and its advantage over the conventional transformations used in a camera-imaging pipeline. Three sample images each from a consumer camera, smart phone and cinema camera, are shown along with the results of applying the method of Cyriac et al.[4], which is optimized for LCD office condition, to the corresponding RAW sensor values. Our results look natural in appearance, with enhanced overall contrast and without any visual artifacts.



(a) (b) (c)

(d) (e) (f)

Figure 4. Top 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 a LogC image. Bottom row: results of applying our method[4] to the corresponding RAW images. Camera models: left column, Nikon D3100 consumer photography camera; middle column, ARRI Alexa digital cinema camera[14]; right column, Nexus 5 smartphone camera. Figure courtesy of Cyriac et al.[4]

In figure 5 we illustrate some results of our method that is optimized for LCD office condition when used as a tone mapping operator and applied to HDR images. The input HDR images are from Fairchild dataset[13](top row) and ARRI dataset[15] (bottom row). When applied to HDR video sequences our method produces results without flicker or any sort of spatiotemporal artifacts (see http://ip4ec.upf.edu/tonemapping for image and video example results).

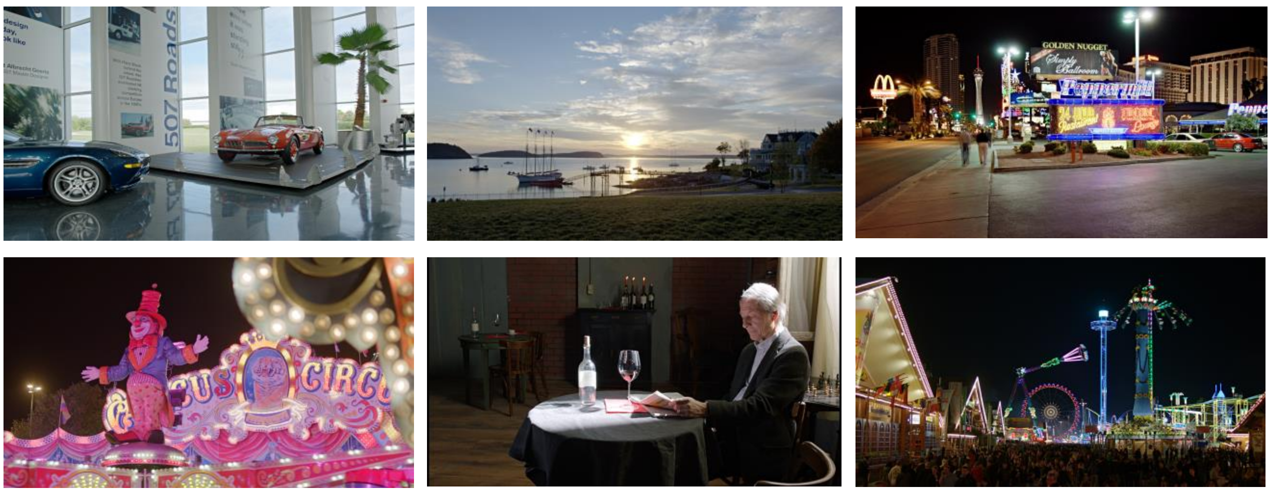


Figure 5. Results of our method applied to HDR images from the Fairchild dataset[13] (top row) and to video frames from the ARRI dataset[15] (bottom row).

In tables 2 and 3 we show the luminance and contrast measurements taken under the different display-surround setups designed above (unfortunately, for the HDR display we could only collect data from the office room condition). In general, the ANSI contrast is substantially smaller than the sequential contrast in the dark surround condition. This is mainly because some of the light emitted by the display is reflected back onto the screen by objects in the surround, resulting in raising the effective minimum value of the display. For example, the minimum luminance measured on the OLED is 0.001 nits in a dark surround when the full screen is black, whereas the measured luminance is 0.3 nits when the pattern shown in figure 3 is displayed. This result is consistent with the result of Schuck and Lude[16], and Tydtgat et al.[17], where they measured both sequential contrast and the effective contrast by varying the image white content on cinema projectors. We may interpret from these measurements that the effective contrast produced by a display depends not only on its maximum contrast ratio but also on the surround and the reproduced image content.

Display	Dark room			Office room		
	Min luminance	Max luminance	Sequential contrast	Min luminance	Max luminance	Sequential contrast
LCD	0.35 nits	170 nits	486	2.3 nits	170 nits	74
OLED	0.001 nits	97 nits	97000	1.2 nits	97 nits	80
HDR	-	-	-	1.7 nits	2700 nits	1588

Table 2. Sequential contrast measurement.

Display	Dark room			Office room		
	Min luminance	Max luminance	ANSI contrast	Min luminance	Max luminance	ANSI contrast
LCD	0.7 nits	170 nits	242	2.6 nits	170 nits	65
OLED	0.3 nits	97 nits	323	1.3 nits	97 nits	74
HDR	-	-	-	2 nits	2700 nits	1350

Table 3. ANSI contrast measurement.

Now we discuss the results of the psychophysical experiment, in which the subjects adjusted the gamma non-linearity γ_{adj} in each of the viewing conditions so as to optimize image appearance. Table 4 shows the average subjects' choice and the value predicted by the model (equation 1) in different viewing conditions. The results indicate that the model predicts well the viewer's preference.

Display	Dark room		Office room	
	Subject choice	Model prediction	Subject choice	Model prediction
LCD	1.11	1.12	1	1
OLED	1.09	1.1	0.98	0.98
HDR	-	-	1.5	1.51

Table 4. Comparison between user chosen and model predicted gamma adjustment.

In figure 6 we show the result produced by our method for three different viewing conditions. As the effective contrast of the display (ANSI contrast) increases, our method tries to compensate it by decreasing the contrast of the lower mid-intensities of the input image. Note that these images will look optimal only under the intended viewing conditions and display type.

LCD office (65) OLED dark room (323) SIM2 office (1350)

Figure 6. Results of our method applied to HDR video frame from the ARRI dataset[15] for three different viewing conditions. ANSI contrast given in parenthesis.

Conclusion

We have developed a fast and automatic contrast grading method based on visual perception that can accurately reproduce the details and contrast visible in the real scene. We recorded luminance measurements of several display types and surround conditions and showed that the effective contrast (ANSI contrast) produced by a display depends not only on its maximum contrast capability but also on the surround and the reproduced image content. We have conducted psychophysical experiments and developed a mathematical model to predict the users' chosen non-linear adjustment for our results to look best in each viewing condition. The applications of the proposed approach, apart form providing a fast and automatic graded content to the colorist, include: replacing the gamma correction stage in the camera image processing pipeline; use as a tone mapping operator; on-set monitoring of HDR cameras; and real time grading of HDR content for television broadcast.

Currently we are extending our psychophysical experiment by testing more displays, viewing conditions and subjects.

Acknowledgements

A patent application based on the research in this article has been filed at the European patent office, Application no 15154172.9-1906. We thank Kadi Bouatouch, Remi Cozot and Cambodge Bist for allowing us to run experiment on their SIM2 display. This work was supported by the European Research Council, Starting Grant ref. 306337, by the Spanish government and FEDER Fund, grant ref. TIN2015-71537-P(MINECO/FEDER,UE), and by the Icrea Academia Award.

References

[1] D. Kane and M. Bertalmío, "System gamma as a function of image- and monitor-dynamic range", Journal of Vision 16(6):4, 1-13, 2016.

[2] P. Cyriac, D. Kane, and M. Bertalmío, "Perceptual dynamic range for in-camera image processing", In BMVC, 2015.

[3] S. Ferradans, M. Bertalmío, E. Provenzi, and V. Caselles, “An analysis of visual adaptation and con- trast perception for tone mapping”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(10):2002– 2012, 2011.

[4] P. Cyriac, D. Kane, and M. Bertalmío, “Optimized tone curve for in-camera image processing”, In SPIE/IS&T Image Quality and System Performance, 2016.

[5] H. Barlow, "Possible principles underlying the transformation of sensory messages", in Sensory Communication, MIT Press, 1961.

[6] W.E. Vinje, J.L. Gallant, "Sparse coding and decorrelation in primary visual cortex during natural vision", Science. 287 (5456): 1273–1276, 2000.

[7] B. A Olshausen and D. J Field, “Vision and the coding of natural images”, American Scientist, 88(3):238–245, 2000.

[8] D. Kane and M. Bertalmío, “Is there a preference for linearity when viewing natural images?” In SPIE/IS&T Image Quality and System Performance XX, 2015.

[9] J. Huang and D. Mumford, “Statistics of natural images and models”, In Computer Vision and Pattern Recognition, 1999 volume 1

[10] D. L Ruderman, “The statistics of natural images”, Network: computation in neural systems, 5(4):517–548, 1994.

[11] N. Brenner, W. Bialek, and R. V. Steveninck, “Adaptive rescaling maximizes information transmission”, Neuron, 26(3):695–702, 2000.

[12] M. Carandini and D. J Heeger, “Normalization as a canonical neural computation”, Nature Reviews Neuroscience, 13(1):51–62, 2012.

[13] M. D Fairchild, “The hdr photographic survey”, In Color and Imaging Conference, volume 2007, pages 233–238.

[14] S. Andriani, H. Brendel, T. Seybold, and J. Goldstone, “Beyond the kodak image set: A new reference set of color image sequences”, ICIP, pages 2289– 2293, 2013.

[15] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel, “Creating cinematic wide gamut hdr video for the evaluation of tone mapping operators and hdr displays” 2014.

[16] M. Schuck and P. Lude, “An analysis of system contrast in digital cinema auditoriums”, SMPTE motion imaging journal, 2016.

[17] C.Tydtgat, D. Maes, G. Stojmenovik and A. Grillet, “Modelling of achievable contrast and its impact on HDR projection in commercial cinema environments”, SMPTE Annual Technical Conference, 2015.

