The preferred system gamma is primarily determined by the ratio of dynamic range of the original scene and the displayed image.

David Kane
Antoine Grimaldi
Emin Zerman
Marcelo Bertalmío
Vedad Hulusic
Giuseppe Valenzise

Abstract

The dynamic range of real world scenes may vary from around 10^2 to greater than 10^7, whilst the dynamic range of monitors may vary from 10^2 to 10^5. In this paper, we investigate the impact of the dynamic range ratio (DR_{ratio}) between the captured scene and the displayed image, upon the value of system gamma preferred by subjects (a simple global power law transformation applied to the image). To do so, we present an image dataset with a broad distribution of dynamic ranges upon various subranges of a SIM2 monitor. The full dynamic range of the monitor is 10^5 and we present images using either the full range, 75% or 50% of this, while maintaining a fixed mid-luminance level. We find that the preferred system gamma is inversely correlated with the DR_{ratio} and importantly, is one (linear) when the DR_{ratio} is one. This strongly suggests that the visual system is optimized for processing images only when the dynamic range is presented correctly. The DR_{ratio} is not the only factor. By using 50% of the monitor dynamic range and using either the lower, middle or upper portion of the monitor, we show that increasing the overall luminance level also increases the preferred system gamma, although to a lesser extent than the DR_{ratio}.

Introduction

The luminance range of a monitor and a real world scene rarely coincide. To illustrate this we plot the luminance range of images from the high dynamic survey by Mark Fairchild [11] in Figure 1, alongside 4 monitor types of increasing capability; the CRT (red), today only used by odd-ball scientists, a typical LCD display (green); a Samsung Quantum Dot display released for the consumer market in 2016 (yellow); and the professional SIM2 monitor used in this paper (cyan). The image set was captured using multiple exposure photography to avoid the risk of over or under exposure and allow the capture a theoretically infinite dynamic range. All the images were of static scenes, the majority were daytime outdoor locations and a small minority are of indoor or nighttime scenes. As can be seen, only in a few instances does the luminance range of an scene/image match those of our target display devices. A common way to describe the luminance range of a monitor is in terms of the dynamic range DR and some measure of the overall luminance level, such as the peak luminance level P common in commercial specifications.

$$DR = \frac{max}{min} \qquad P = max \qquad (1)$$

Arguably the greatest concern when displaying an ungraded real world image upon a given monitor is the mismatch between the dynamic range of a monitor and the original scene/image. For the professional SIM2 monitor it is common for images to be of a substantially lower DR than the monitor (the inverse tone-mapping problem), and for monitors that are currently in widespread usage, images frequently have a much greater dynamic range (the tone-mapping problem). The multiplication required to increase the dynamic range of the lowest dynamic range image in the Fairchild database for presentation on the SIM2 is 323, whilst the division needed to present the highest dynamic range image on a CRT is 152500. This may in part explain why a simple power law appears to be sufficient to solve the inverse problem [5], but not the tone-mapping problems where an array of more complex algorithms has been proposed. For the rest of this paper we shall define the dynamic range ratio DR_{ratio} as the dynamic range of the original scene over the dynamic range of the display device.

$$DR_{ratio} = \frac{DR_{scene}}{DR_{display}} \qquad (2)$$

Additionally, we compute the log mid luminance level M as

$$log(M) = log(min) + \frac{log(max) - log(min)}{2} \qquad (3)$$

Aim

In this paper, we ask what the impact of the dynamic range ratio is upon the preferred system gamma where the system gamma is simple a power law that describes the relative relationship between the original image and the displayed image. More precisely, if I_n is the image normalized between zero and one, the processed image is simply

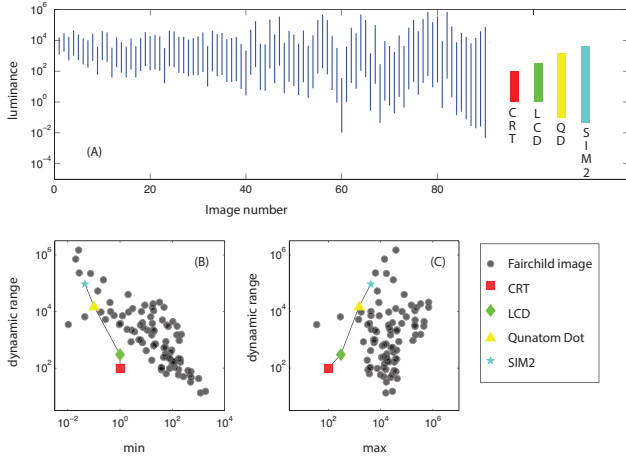$$I_p = I_n^{\gamma_{sys}} \qquad (4)$$

Figure 1. *(A) The luminance range of images for the Fairchild high dynamic range survey and various monitor types. The images are sorted from lowest dynamic range or highest. (B) Scatter plot of the minimum luminance against dynamic range (C) Scatter plot of the maximum luminance against dynamic range.*

and the displayed image

$$I_d = min + I_p(max - min) \tag{5}$$

where *max* and *min* describe the maximum and minimum luminance levels of a given monitor. Early display engineers using relatively low dynamic range displays and low dynamic range single exposure content (both between two to three orders of magnitude), found that the value of system gamma that gave rise to the most pleasing image was nonlinear ($\gamma_{sys} \neq 1$) and that the strength of the nonlinearity depends on the viewing condition; images viewed with a dark surround typically use a gamma of around 1.5 whilst images viewed with a light surround use a gamma of around 1.1 [8, 15].

This work is a direct extension of [3, 4] which investigated the preferred system gamma for images presented on three monitor types; a CRT, an OLED and an LCD monitor. The theory developed in this paper formed the basis for a patented tone-mapping operator which is close to market [9]. The original work was limited by the monitors used, which were of relatively low brightness and/or dynamic range. In this study we use the professional SIM2 monitor which is one of the highest dynamic range monitors available today. The aim of this study is to investigate the impact of the dynamic range ratio upon the preferred system gamma. To do so, subjects were asked to rate the perceived image quality of images presented with various levels of system gamma and upon either the full range, or four sub-ranges of the SIM2 monitor (see Table 1) using the the distribution of images originally used in [3, 4] which were selected to cover a broad range of dynamic ranges.

Computing dynamic range

In the introduction, we defined the (physical) dynamic range of an image or scene as the ratio between the maximum and minimum luminance of a monitor or scene. Dynamic range is a global image statistic, but computed in the simplest manner relies on just two pixels and thus may be subject to noise. In a previous study [3] we used the 0.01 and 99.99 percentiles to compute dynamic range, however, as images, and particularly high dynamic range images, have extremely long tails at the high luminance levels [3], the use of the upper percentile can lead to a large decrease in the estimated dynamic range. Given that noise levels are typically very low for the highest luminance levels in an image [18] we can be relatively certain of the higher values, but not of the lower values. As such we now use a percentile for the estimate of the lowest luminance level (0.01%) and use the maximum for the upper estimate. Unfortunately, without access to a ground truth database to evaluate the reliability of a given measure, we acknowledge that this procedure remains a best guess. We note that the study of perceived dynamic range is a current research topic [12] and may provide insight towards a more perceptually relevant measure.

The measure of monitor dynamic range is also tricky as it is not simply defined by the light emitted by the display but also the reflectivity of a monitor surface and the ambient light conditions in a room. One is advised to read [17] for an evaluation of the impact upon the effective dynamic range of cinema displays. For the SIM2 monitor that we use in this study, we compute the lowest luminance level as the minimum luminance level recorded by a photometer on the display under the ambient light levels. This level was found to be 0.05 nits. The highest luminance level is found to be 4249 nits. Thus for mathematical simplicity, we define the minimum as 0.04249 and the maximum as 4249 nits.

Methods

The psychophysical procedure, image processing and image set used here are identical to [3], although we shall review the major points below. The only major difference is the monitor used to present the images.

Subjects

8 subjects took part in the experiment, all had corrected to normal vision.

Apparatus

The experiment took place in a dark and quiet room. The stimuli were displayed at full HD (1920×1080 pixels) resolution on the SIM2 HDR47ES4MB 47" screen. The monitor was run in DVI Plus (DVI+) mode, which allows for direct and independent control of the backlight LEDs and LCD pixel values, based on the dual-modulation algorithm [21]. The viewing distance was fixed to three heights of the display, with the eyes in the middle of the display, both horizontally and vertically.

SIM2 HDR screen is able to generate high luminance values up to 4249 nits; however, finding proper LED and LCD values for each image becomes an optimization problem due to the light diffusing layer of the screen. The aforementioned dual-modulation algorithm finds LED and LCD values for each image using an iterative scaling approach. This method was found to be able to render the HDR image as close as possible to the intended luminance values [22]. Using this dual-modulation algorithm, each HDR image is ensured to have the intended luminance values.

Presentation protocol

Images were presented on various subranges of the SIM2 monitor. There were five ranges in total as shown in Table 1. For

conditions 1 to 3 the dynamic range decreases from 100% to 75% to 50% of the total monitor range, but the mid log luminance level is fixed at 13.43. For conditions 4, 3, and 5 the dynamic range is fixed at 50%, but the mid log luminance level varies from 0.76 to 13.43 to 238.94 cd/m^2 respectively.

#	Name	min	max	DR	M
1	Full range	0.04249	4249.0	100000	13.43
2	75% range	0.1792	1007.6	5623	13.43
3	50% center	0.7556	238.9	316	13.43
4	50% dark	0.04249	13.4	316	0.76
5	50% bight	13.4365	4249.0	316	238.94

Table 1 - Images were displayed on five sub-range of the SIM2 monitor.

All images were presented with values of system gamma from 2^{-2} to 2^2 at half log_2 intervals. All images and all conditions were randomly interleaved. To investigate the potential impact of sub-optimal adaptation condition 4 (50% dark) was repeated in a separate experiment and compare to the fully interleaved condition. Images were presented using Matlab and the psychtoolbox.

The image database

All the stimuli were high dynamic range images from the HDR Photographic Survey [10]. The images were captured using multiple exposure photography using between 8 to 18 exposures levels. Images were combined using Adobe's built-in software to produce a *.exr* file which is a linear estimate of the relative luminance levels in a scene. In total, 105 images exist in the database. Of these, 98 have a corrective factor used to convert the values to absolute luminance levels. These were computed by checking the luminance level using a photometer used at the original scene. The images selected for this experiment are those used in [3, 4]. These were originally chosen to span a broad range of dynamic ranges and are detailed in Table 2.

#	Name	min	max	DR
1	AmikeusBeaverDamPM2	0.37	3220	8680
2	CanadianFalls	41.09	17390	420
3	DelicateArch	269.80	24160	90
4	DevilsBathtub	74.15	173960	2350
5	DevilsTower	53.66	464750	8660
6	FourCornersStorm	16.67	3300	200
7	HancockKitchenOutside	0.70	81130	115790
8	LetchworthTeaTable2	1.08	4440	4090
9	MackinacBridge	135.75	27330	200
10	MirrorLake	55.88	18710	330
11	PeckLake	21.19	28430	1390
12	SmokyTunnel	1.98	38220	19320
13	WallDrug	5.11	26520	51910

Table 2 - Table of images used in this experiment.

The task

The subject was instructed to rate each image according to the perceived image quality. The subject did so by manipulating a sliding scale with a mouse. Values on the left hand side indicated poor image quality scores while images on the right hand side indicated good quality. Subjects were initially shown a random selection of the images to allow them to judge the range of quality images and to scale their answer appropriately. The val-

ues of system gamma evaluated were the same for all images and conditions tested as shown in Figure 2. Due to the heterogeneous nature of the image database, the preferred system gamma varies substantially between images and thus many of the presented values of system gamma are very sub-optimal. At the extremes, this could lead to images that were almost totally black or white. Subjects were instructed to only use the far left of the scale if no visible information about the scene was available. Otherwise, the subject was given no further instructions about how to score the images. In this manner, a floor effect could be avoided. Subjects naturally avoided a ceiling effect at the upper end of the scale (no maximum values were recorded). Subjects had unlimited time and proceeded to the next trial with a key press.

Data processing

The preferred system gamma is taken to be the value that maximizes the image quality score. The image quality scores were averaged for all subjects and the resulting function was found to be unimodal, although the shape varies substantially between images. To obtain a sub-interval estimate, a fourth order polynomial was fit to each function. The image quality score was then the peak of this function and the preferred system gamma was taken to be the system gamma value at which the peak was located. This procedure is identical to [3] and example fits are shown in figure 2. This procedure is not ideal, but is sufficient so long as the effect sizes noted are greater than the potential bias caused by an inadequate fit to the data.

Results
An example image

We begin by illustrating the image quality functions obtained for an individual image: *DevilsBathtub* in Figure 2. The data for the three dynamic range conditions is shown on the first graph (bottom left) and the data for the luminance level conditions on the second graph (bottom right). Note that the data indicated by the red is identical in both. The solid dots denote the average image quality score at each system gamma level and the solid lines, the best fitting polynomial. The color denotes the condition as shown by the legends. These color scheme shall be used throughout the paper.

A control experiment

In the main experiment we interleaved the five conditions. This meant that subjects may be sub-optimally adapted for the viewing condition in question. To evaluate this we run a control experiment by running condition 4 (50% range, dark condition) in isolation from the other conditions. The results are shown in Figure 3. On the left, we plot the system gamma for the interleaved condition against the system gamma for the control conditions. As can be seen, the values lie either side of the identity line and a t-test revealed no significant difference ($n = 13, p > 0.5$). In contrast, on the right, we see that the image quality scores are always lower in the interleaved condition than the control condition ($n = 13, p < 0.001$). Overall, this is encouraging as the primary research goal is to investigate the system gamma. The lowering of the image quality scores in the interleaved condition has two potential explanations (a) subjects may be rescaling their answers with regard to those images presented within a single session or (b) image quality scores may be higher in the control condition
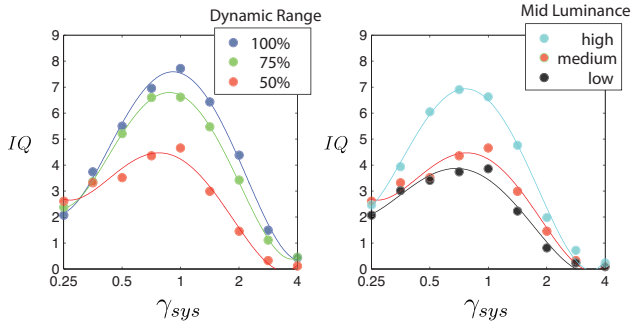
Figure 2. *Top. The image Devil's Bath tub from the Fairchild HDR photographic survey. Bottom left, image quality scores for conditions, 1,2, and 3. Bottom right, image quality scores for the conditions, 3, 4, and 5.*

Figure 4. *Results averaged across all images: System gamma as a function of dynamic range for conditions 1, 2, & 3 (top left); System gamma as a function of the mid-luminance level for conditions 3, 4, & 5 (top right); Images quality scores obtained at the preferred system gamma (bottom).*

due to better visual adaptation. It should be noted that in the previous study [3, 4] which used an identical paradigm, obtained clear differences in image quality on different monitors (each monitor was evaluated in a separate session), indicating that that any rescaling was not complete. The lack of an effect for the system gamma data is theoretically intriguing as it indicates that this particular adaptation is rapid.
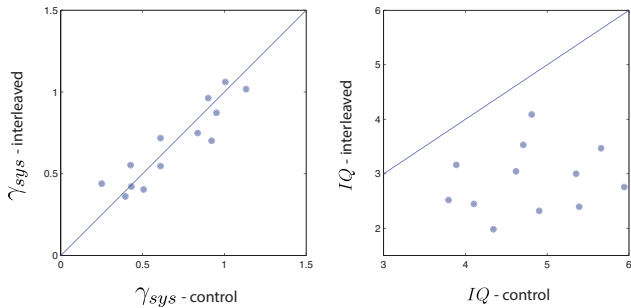

Figure 3. *Results for condition 4 (dark, 50%) with trails either interleaved with the other conditions or investigated in isolation: a scatter plot of system gamma (left), and a scatter plot of image quality (right).*

Evaluating the different conditions

In Figure 4 we present the average results for each condition. On the left, the data is plotted as a function of the dynamic range, while on the right as a function of the mid luminance level. The results demonstrate that both image quality and system gamma increase with both dynamic range and the mid luminance level.
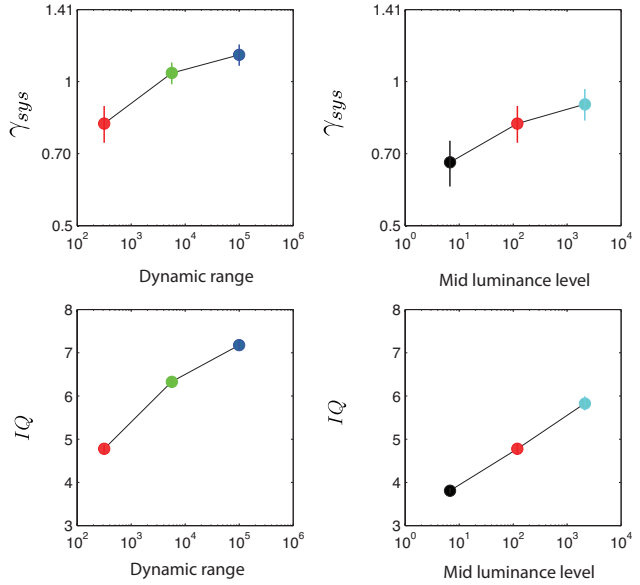
As a function of the dynamic range ratio

In Figure 5, we plot the preferred system gamma as a function of the dynamic range ratio (Equation 2). On the left, the three dynamic range conditions are presented, whereas on the right the three mid-luminance levels are used. If one considers all three dynamic range conditions as one, then the major conclusion is that when the ratio of the dynamic range of the monitor and the image are approximately matched, the preferred system gamma is approximately one (linear). We speculate in the discussion that this is due to the visual system being well tuned to the statistics of natural images.

When considering all five conditions it is clear that the dynamic range ratio is not the only factor impacting system gamma values. Firstly, the lines of best fit for the three dynamic range conditions do not have the same gradient or constant. Second, there is a clear impact of the luminance level, even when the dynamic range is fixed.
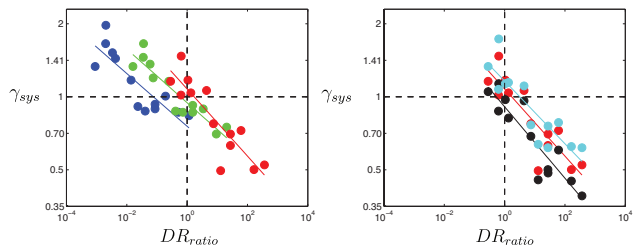

Figure 5. *The preferred system gamma plotted as a function of the dynamic range ratio R (equation 2). Each dot represents an individual image. For the dynamic range conditions 1, 2, & 3 (left); for the luminance level conditions 4, 3, & 5 (right).*

In Figure 6, we plot the image quality achieved at the optimal system gamma. Note that the data is taken from the interleaved

conditions, thus there may be mal-adaptation and subjects will have evaluated quality relative to the preceding images' quality. Thus these results are of no use to someone who wants to understand how good an image looks on a given monitor presented in isolation from others. Nonetheless, it is interesting to find that the low dynamic range images presented on the 50% dynamic range, but high brightness condition (monitor condition 5) achieve and image quality comparable to that of the full dynamic range condition. A similar finding has also been reported in [1] and together suggests that there is no benefit to the high dynamic range monitor for images with a low dynamic range. Thus the benefit of high dynamic range monitor appears is that they allow high dynamic range images to be presented with relatively little processing applied. Possible explanations for why there is little benefit of presenting low dynamic range scenes on a high dynamic range monitor are the role of monitor reflectance [17] and/or optical scatter [14] in reducing the effective dynamic range of the perceive image. Both factors plays a greater role when bright (low dynamic range) images are presented.
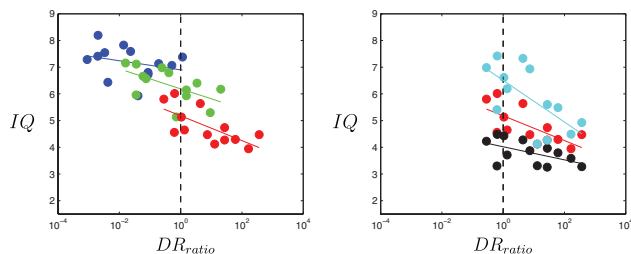

Figure 6. *The image quality scores, plotted in the same manner as in Figure 5.*

Discussion

Cameras taking a single exposure image can store images either as an unprocessed RAW images or processed JPEG. The processed images should automatically achieve a good visual appearance. Typically the algorithms used to do this are proprietary and considered to be of critical important to the camera makers 'brand', with some brands known for having a particular style, popular with some consumers. This problem has had to be revisited with the advent of high dynamic range technologies and monitors and this paper is an attempt to develop a framework from which the problem of tone-mapping high dynamic range content can be understood. In this regard, we should be clear that HDR content refers to images captured via a method that does not suffer from the potential for over and under exposure, however, the actual dynamic range of the captured images may vary greatly depending on the scene captured. Thus we distinguish between *dynamic range content* and *dynamic range*. This critical distinction is necessary for the development of effective algorithms as the role of the dynamic range ratio in this study would suggest.

System gamma is a simple, but powerful transformation that is easily studied due to its single parameter. Nonetheless, studies of the preferred system gamma using single exposure photography and low dynamic range monitors report that it is difficult to predict the image dependent variability in the preferred system gamma [16]. However, when one uses a broad distribution of dynamic ranges it is clear that the preferred system gamma is largely predictable [19, 3]. This is due to the fact that the median

luminance of an image is strongly and inversely correlated with dynamic range [3]. This effect is so strong that for very high dynamic range images over 95% of pixels values may occur with in the first 1% of the total luminance range, whilst for low dynamic range images the luminance distribution is relatively well distributed. As such the value of system gamma needed to achieve even a reasonable looking image varies hugely between images and becomes the dominant factor.

The fact that the preferred system gamma is approximately one when the dynamic range ratio is one, suggests that only limited preprocessing of the image is needed when the dynamic range is correctly reproduced. This in turn suggests that the visual system is well tuned to the statistics of real world images and in particular, the negative correlation between the normalized median luminance of dynamic range described in figure two of [3]. To counter this effect, the nonlinearity applied by the visual system must be increasingly compressive for high dynamic range images. There are two lines of evidence that the human visual system is increasing compressive as the dynamic range of an image increased. First, in a direct study of perceived lightness, Radonjic et al. [13] report increasingly compressive functions as dynamic range increases. Second, Fechner integration of detection thresholds also predicts increasingly compressive functions when a greater dynamic range is evaluated [2, 20].

References

[1] Akyz AO. Fleming R. Riecke BE. Reinhard E. and Blthoff HH. Do HDR displays support LDR content?: a psychophysical evaluation. ACM Transactions on Graphics (TOG), 26(3), 38. (2007).

[2] Blackwell HR. Contrast thresholds of the human eye Journal of the Optical Society of America, 1946 JOSA, 36(11), 624-643.

[3] David K. Bertalmío M. System Gamma as a Function of Monitor- and Image Dynamic Range. Journal of Vision, 2016

[4] Is there a preference for linearity when viewing natural images? Proc. SPIE 9396, Image Quality and System Performance XII, 939614 (8 February 2015);

[5] Bist C. Cozot R. Madec G. and Ducloux X. Style aware tone expansion for hdr displays In Proceedings of Graphics Interface 2016 Canadian Information Processing Society, Victoria, B.C., Canada 2016.

[6] Bartleson CJ. Breneman EJ. Brightness reproduction in the photographic process. I Photographic Science and Engineering, 11 (4), 254262. (1967b).

[7] Bartleson CJ. and Breneman EJ. Brightness perception in complex fields. Josa 57.7 (1967): 953-957.

[8] Bartleson CJ. Optimum image tone reproduction. Journal of the SMPTE 84.8 (1975): 613-618.

[9] P Cyriac, M Bertalmio, D Kane, J Vazquez-Corral A tone mapping operator based on neural and psychophysical models of visual perception Proc. SPIE 9394, Human Vision and Electronic Imaging.

[10] Fairchild MD. Considering the surround in device-independent color imaging. Color Research & Application 20.6 (1995): 352-363.

[11] Fairchild MD. The HDR photographic survey. In J. Larimer & M. Moroney (Eds.), 15th Color and Imaging Conference (pp. 233238). Albuquerque, NM: Society for Imaging Science and Technology. 2007

[12] Hulusic V. Debattista K. Valenzise G. Dufaux F. A model of perceived dynamic range for HDR images. Signal Processing: Image Communication, 51, 26-39, 2017

[13] Radonjic AA, Allred SR. Gilchrist AL. Brainard DH. The dynamic

range of human lightness perception. Current Biology, 2011 21(22) 1931-1936.

[14] Stiehl WA. and McCann JJ and Savoy RL. Influence of intraocular scattered light on lightness-scaling experiments. Journal of the Optical Society Of America, 1983 73(9), 1143-1148

[15] Poynton C. Digital video and HD: Algorithms and Interfaces. Elsevier, 2012

[16] Roufs J. A. J., Goossens I. M. J. (1988). The effect of gamma on perceived image quality. In Conference Record of the 1988 International Display Research Conference (pp. 2731). San Diego, CA: IEEE.

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

[18] Seybold T. Cakmak O. Keimel C. Stechele W. Noise characteristics of a single sensor camera in digital color image processing Color and Imaging Conference, pp. 53-58, 2014.

[19] Singnoo J., Finlayson G. D. (2010). Understanding the gamma adjustment of images. 18th color and imaging conference, Vol. 2010 (pp. 134139). San Antonio, TX: Society for Imaging Science and Technology.

[20] Nezamabadi M., Miller S., Daly S., Atkins R. (2014). Color signal encoding for high dynamic range and wide color gamut based on human perception. In Proceedings of SPIE 9015, color imaging XIX: Displaying, processing, hardcopy, and applications, 9015C, doi:10.1117/12.2042893.

[21] Zerman E. Valenzise G. De Simone F. Banterle F. and Dufaux, F. Effects of display rendering on HDR image quality assessment. In SPIE Optical Engineering+ Applications. International Society for Optics and Photonics, 95990R-95990

[22] Zerman E. Valenzise G. and Dufaux, F. A dual modulation algorithm for accurate reproduction of high dynamic range video. In Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), 2016 IEEE 12th.

Author Biography

David Kane received his PHD from the University Collage London studying motion perception and the aperture problem. From there he went on to apply his skills to applied problems in the lab of Marty Banks at UC Berkeley, investigating stereo-3D vision. David now works at Universitat Pompeu Fabra investigating the issues surrounding the capture and presentation of high dynamic range content in the lab of Marcelo Bertalmío.