https://doi.org/10.2352/lim.2022.1.1.12

Probing perceptual phenomena for color management

Trevor Canham, Marcelo Bertalmío; Spanish National Research Council (CSIC); Madrid, Spain

Abstract

Advancement of color management techniques is required to accommodate for emerging formats and devices. To address this, experiments were conducted by the authors in order to characterize and account for the effect of metamerism error, chromatic adaptation to the surround, contrast adaptation to display dynamic range, and viewing size dependent effects of retinal signal pooling. These topics were assembled to address perceptual representation inconsistencies which are becoming more common with the popularity of mobile, High Dynamic Range (HDR), and Wide Color Gamut (WCG) displays. In this paper, we briefly summarize the findings of these efforts and compile a series of takeaways which are key to the problem of perceptual color management. Furthermore, we discuss the implications of these takeaways on the advancement of the field.

Introduction

A digital artist at a solitary workstation could be led to believe that what they create could be reproduced in a lossless format and viewed in its intended state in perpetuity throughout the universe. If they then move to a different device, they may notice that the details they worked so hard to perfect have shifted in appearance. The artist's trust in their tools is degraded at this moment. Their perfectionism may also degrade, given the resulting cognitive dissonance from meticulously refining details while simultaneously suspecting that no one will see them in the same way. They will carry on though, ensuring that anything that appears under their name has passed their rigorous seal of approval. So we all must resign to trust state of the art tools to properly convey our ideas to recipients and receptacles unseen and unknown. It is the role of the discipline of color management to back this trust and eliminate systemic inconsistencies.

Unfortunately, systemic inconsistencies have proliferated in recent years. New formats and devices have stretched and broken the boundaries of standard use cases and foundational perceptual assumptions which color management systems have relied upon [1]. Narrow band primaries employed in WCG displays introduce error by focusing light energy in precise spectral regions where sensitivity may differ from person to person. Mobile devices have broken the expectation that screens will be used in a consistent and controlled environment, widening the range of potential viewer brightness/chromatic adaptation states. The HDR boom has led to the proliferation of display technologies, creating the possibility for viewers to encounter content in a wide range of renderings with different contrast properties. This has prompted exploration from many researchers into the relevant visual adaptation processes, to better understand their behavior and leverage them to account for physical differences between users, their devices, and the environments where they view them [2, 3, 4, 5, 6]. Chief among them are the spectral integration of photoreceptors, brightness/chromatic adaptation, contrast adaptation, and retinal signal pooling. Through modelling these processes, the color appearance of imagery can be better managed not only between varying display destinations, but also within imaging pipelines, allowing content graded for different contexts to be effectively homogenized.

In this work, we will review a series of studies conducted by the authors in order to characterize these mechanisms in the context of displayed imagery. Then, we consolidate general takeaways to aid the advancement of color management via the perceptual approach. Finally, we will discuss the implications of these takeaways with respect to future problems and potential solutions.

Methods

Metamerism

The Human Visual System (HVS) discriminates color stimuli not by their precise spectral shape, but by comparison between the integrated signals of three spectral bands. By extension of this, different scene spectra can result in the same color sensation to an observer, a phenomenon known as metamerism. This allows imaging systems to produce realistic reproductions of scene content by the same three channel mechanism. To predict these matches, color matching functions (CMFs), which aim to approximate the properties of these integration bands, are used to exploit this abbreviation and convert spectral readings into a device-independent, three channel representation. However, the use of a single average observer CMF has been shown to result in impactful color rendering errors, as there exists significant variation in the spectral absorption characteristics of color-normal observers. When this is crossed with the growing disparity between the spectral characteristics of emerging display technology, it becomes evident that this inter-observer variability should be accounted for.

In [7] Asano and Fairchild present a physiologically based individual observer model, as well as a method for separating a population of observers into a limited number of categorical CMFs. Building upon this, we conducted a simulation in [8] to find the inter-display matching error between each categorical observer and a standard observer (CIE 1931 2-degree) for a set of natural images and common display spectra (Sony CRT, Apple CCFL, Samsung LED, OLED, IMAX laser projector). To do so, we first simulated the pixel-wise cone excitation values that would be triggered in each observer by each color channel from each display. Then, by converting each cone excitation set back to $CIEL^*a^*b^*$ space, we could directly compare the ΔE_{2000} error between these renderings and the display pair match derived from a standard calibration process. The results showed that for our display, observer, and image set the metamerism error was over threshold ($\Delta E_{2000} > 2$) for 22% of image area on average. Looking at the white point, which is the most common source of calibration complaints from colorists, the error ranged from 1.90-6.16 ΔE_{2000} across the set of compared displays. The values for these two measures pretaining to each display combination are shown in Table 1.

Chromatic adaptation to the surround

The rise in prominence of portable viewing devices has presented a new challenge for color management in the last decade. Since they allow for a single observer to easily view their device

Table 1: Metamerism simulation results from [8]. White point calibration error averaged over observer categories and percentage of image area over threshold error, averaged over image set and observer categories

Displays	$\Delta E2000White$	% image area > 2 $\Delta E2000$
CRT Vs. CCFL	1.90	11.71
CCFL Vs. LED	2.77	24.78
CCFL Vs. OLED	3.39	19.29
pLCD Vs. IMAX	4.00	21.42
CCFL Vs. IMAX	4.89	25.88
CRT Vs. IMAX	6.16	30.85

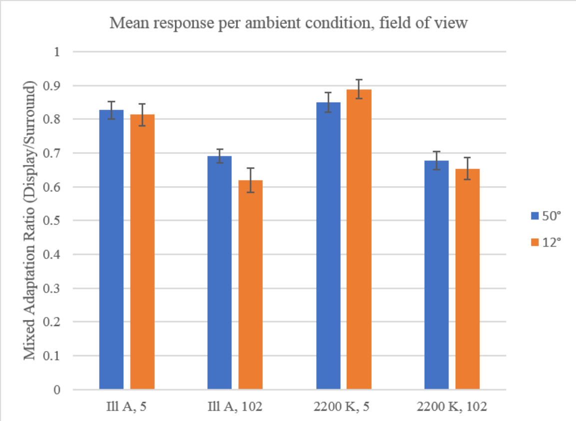


Figure 1: Experimental results, compared between results of [9] and [10] for ambient illumination brightness values and color temperatures tested, reported as degree of adaptation to the display. Interval bars represent 95% confidence about the mean, averaged over observers and repeat trials.

in a wide range of environmental conditions (and as a result, in a range of adapted states), many device manufacturers have explored solutions to adjust displayed imagery to account for these dynamics and render images with a consistent appearance [5]. Concurrently, a series of studies were conducted [9, 10] which aimed to characterize the impact of varying ambient brightness and chromaticity on displayed image appearance. In these studies, observers were shown images on displays of varying size first under SMPTE ST 2080 [11] standard viewing conditions (5 cd/m^2, D65), and later were asked to correct these images back to their original appearance after adapting to a series of lighting conditions. The task was simplified by the use of a control scheme where observers could adjust the white balance of images along a 1D vector in CIE 1976 u'v' space connecting the display white point to the illumination chromaticity. This control scheme was based on the assumption that the observers' adapted state would be somewhere between these points, given that they were the only white references currently present in the scene. So in all of the results moving forward, the observer adapted states are represented in terms of their chosen ratio between these points. Also, images with significant achromatic content were used such that observers could rely on not only their memory of the reference conditions, but also their achromatic preference.

The results showed that, between the white references of the display and the surround, observer adaptation ratios ranged from 0.26-0.9 adapted to the display. The relevant factors in this wide variance included ambient illumination brightness, chromaticity and stimulus type (simple stimuli vs. complex images). The two displays tested were viewed with visual angles of 12 and 50 degrees respectively, but we observed no significant difference between their influence on the observer adapted state. A subset of the results, representing the common ambient conditions tested between viewing sizes are shown in Figure 1.

Display contrast adaptation

High Dynamic Range displays and image formats allow for an increase in image contrast by extending the range of reproducible values in both shadows and highlights. The popularity of this format has triggered a technology rush in the past decade to produce displays which achieve this range to the highest specifications possible. As a result, different fundamental display technologies have been developed which are advantageous at one end of the range but disadvantageous at the other. For instance, OLED displays work better for dark images because their pixels can be driven at low values and completely turned off if necessary, while the back lights of LCD devices allow for higher peak values, but are disadvantageous for dim images as they cannot be completely doused by the filters at each pixel location. To accommodate for this, HDR industry standards like Rec. ITU-R BT.2100 [12] allow for this variation, but do not provide an adequate solution to account for appearance differences between these renderings. However, in all cases, contrast adaptation in the HVS allows for a relatively consistent visual experience between scenarios with vastly different ranges of light, despite the limited bandwidth of the system [13]. If modelled correctly, this mechanism could also be simulated in imaging pipelines to preserve appearance between renderings with different contrast properties.

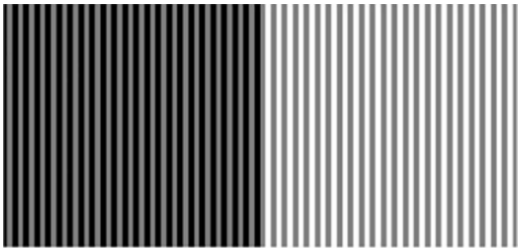
In [14] we proposed a 1D global tone mapping operator which featured a dual non-linearity with adaptive parameters based on image statistics. In addition to a version for rendering raw HDR images on SDR displays, we also proposed a method for adapting color graded content between displays of varying dynamic range. Considering that colorists learn an intuitive vision model through their work, we optimized and validated these methods with the help of these professionals. In order to avoid complications switching between adapted states in the limited experimental time frame, HDR and SDR versions of stimuli were manually pre-generated by the colorists, who aimed to make a visual match to the best of their ability. Then, the colorists compared these manually graded references to the results of our method, and adjusted its open parameters to make the best possible match. Finally, the results were validated against those of industrial and academic alternatives by a larger group of professionals, but instead with a collection of test images that were outside of the optimization set. The method outperformed the alternatives in both cases of converting SDR imagery for HDR viewing and vice versa.

Retinal signal pooling

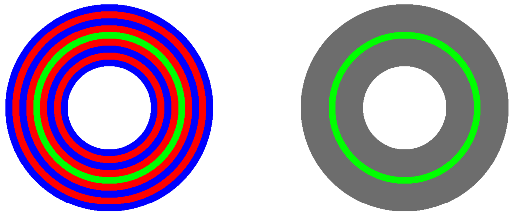
A color management inconsistency which is not new but is thus far unaccounted for is that of appearance differences with respect to viewing size. A potential explanation is related to the phenomenon of simultaneous contrast, or more generally, induction, which demonstrates the interdependent effect of adjacent image regions. These regions may influence each others' color attributes to be more similar, in the case of assimilation, or more different in the case of contrast. The effects are demonstrated in Figure 2, and have been theorized to be the result of signal pooling processes in the retina [15] which are dependent on spatial frequency and relative field size [16], and thus images can have subtle appearance differences when viewed at different sizes. In [17] we conducted a series of experiments in which observers reported their perception of brightness and chromatic induction illusions (square wave gratings and Shevell rings, respectively) when presented to them at mobile and cinema reference viewing sizes. We used this data to optimize a regularized neural field model of color appearance, in the form of a linear filter, which could predict and counteract the shift in appearance between viewing sizes. This constituted a novel use of neural field models for color management, in that other models are not invertible because they are based on the iterative minimization of energy functionals [18, 19]. An invertible model is required for any color management application, since it must not only predict the perceptual representation of the stimulus, but also be able to re-render it for a different viewing scenario.

The results of the brightness experiments showed a dependence on not only the full size and resulting spatial frequency of the pattern but also the relative size of its elements. The results of the chromatic experiments showed a dependence on hue, with fields high in short wavelength energy exhibiting a disproportionately large effect. Another interesting finding was that the perceived chromaticity of the central target ring in the Shevell patterns could exist outside of the gamut bounded by the physical chromaticities of its component fields. This implies that induction effects could be used to generate sensations which are outside of device gamut boundaries, which is also demonstrated in Figure 2.

By optimizing our model for this data, we were able to predict the observer responses from the experiment within the relatively large inter-observer confidence intervals. However, we found that when the model is applied to complex image stimuli, it resulted in an inconclusive correction benefit. This was observed through an evaluation experiment where observers viewed a reference image at full size, then down-scaled, and then were asked if the image shown had been altered. In theory, the down-scaled version which is uncorrected should exhibit an appearance difference from the reference of a greater magnitude than our corrected result, and thus observers should report in the majority of cases that the down-scaled original was changed and the successfully corrected image was not. Instead, observers reported the opposite, implying the effect was over corrected for by our model, was too subtle to be noticed by observers, or both.



(a)



(b)

Figure 2: Induction effect patterns. (a) Square wave grating, where the gray bars are the same value throughout, but change in appearance with white and black surroundings. (b) Shevell rings, where the central green ring on the left is the equivalent value to the central green ring on the right. All fields in this pattern are encoded at maximum red, green, and blue primary drive values, but the green ring on the left achieves a higher saturation at small viewing sizes than the green ring on the right, implying a sensation that lies outside of the gamut of the viewing device.

Discussion

Problems with colorimetry

The body of work on metamerism demonstrates that inter-observer differences in spectral integration properties lead to significant color rendering errors, not only between observers, but also between displays viewed by the same observer which are calibrated to match for a standard CMF. Therefore, this problem represents a fundamental limitation for practitioners who carry out standard calibration workflows in controlled environments. These CMFs are also problematic in the sense that they do not represent an effective average among the population, nor do they represent any single physiologically plausible observer [7]. This is especially unfortunate for perceptual color management in the sense that models begin with a conversion to a colorimetry based representation, so error is thus introduced at the very first stage of processing. However, if information on the observer's categorical CMF and display spectra were carried through metadata, their relative cone excitation to the stimuli could be reproduced for a different observer at a different display, provided the same information is available for the output conditions. In [8] we propose a pipeline to accomplish this in a way which is compatible with standard color management procedures. A solution of this variety would serve in lieu of a multi-spectral workflow, which would be able to better convey the physical properties of the scene without relying on the trichromatic simplification.

No direct comparisons

We found in our experiments that a fundamental limitation in modelling adaptive responses to asymmetric viewing conditions is the fact that these conditions cannot be simultaneously compared. We encountered this problem in both our efforts to account for chromatic adaptation differences between viewing conditions, as well as for contrast adaptation differences at different dynamic ranges. In the first case, observers were asked to memorize the appearance of a small set of images under reference viewing conditions. This process not only limited the number of images that could plausibly be tested in the experiment, but also introduced a degree of uncertainty resulting from the perceptual barrier of memory. Luckily, the observers could also lean on their achromatic preference in this case, but the same cannot be done in other asymmetric matching problems like those involving contrast adaptation. As a result, we had to rely on pre-generated asymmetric matches which could be viewed under the same contrast adaptation conditions for our tone mapping experiment. This strategy is not only problematic for being time consuming, but also in that it relies on the ability of a single observer and their tools to hone a perceptual match. An additional problem with the tone mapping experiment was that observers found it challenging to make comparative judgements between method outputs which were imperfect matches to the reference, requiring them to rely on aesthetic preferences to make decisions.

In the case of both research efforts, the experimental results can only be seen through a perceptual barrier, leaving room for one's aesthetic preference to have prominent influence. In some sense this simultaneously relieves pressure on the color management process, by preventing errors from being seen in plain sight, but also confounds the relation of perception between observers, given the wide ranging state of adaptation differences that could be present. These states can be impacted by the time of day, season, or life stage that observers are in. It can also be affected by their current location, the place they grew up, their current visual task or their occupation [20]. Since we cannot compare perception for a single observer in different viewing conditions, we certainly could not compare the perception of two different observers in different viewing conditions or the same ones for that matter. Additionally, to conceive of a color management system where all of these factors are parameterized and queried borders on the growing trend of invasive technology.

Diminishing returns

If we take the view that the visual pathway is physiologically arranged such that the sites of spectral integration mechanisms precede those of chromatic adaptation, followed by photoreceptor response non-linearity (a primary mechanism of contrast adaptation), followed by signal pooling processes, we can see that the results of our experiments follow the trend of increasing complexity and diminishing returns as we probe deeper into the visual pathway. At the stage of spectral integration, no perceptual signal processing is yet considered and thus one can physically measure the spectral properties of the eye's optics and photoreceptor pigments, or probe these qualities with direct color matching experiments. At the same time, the results of our experiment in this space showed that large color rendering errors could be avoided if this stage is handled correctly. Chromatic adaptation behavior can still be probed reliably in achromatic preference experiments, and correct accounting for the phenomenon can lead to obvious improvements with complex stimuli, given the perceptual prominence of the white point. In the case of tone mapping and contrast adaptation, the non-linear visual process starts to become more difficult to characterize and we lose the ability to easily evaluate complex images with a single reference point, unlike achromatic preference experiments. Finally, arriving at the task of accounting for differences related to retinal signal pooling processes, we see significantly higher inter-observer variability in the experimental process, as we are dealing with appearance shifts which are so subtle that the majority of observers do not notice them when modulating their viewing size.

Simple vs. complex stimuli

While of course not being the first to do so [21], our chromatic adaptation experiments showed that the type of stimuli (simple color patches vs. images) had a significant effect on the adapted state of observers. In this experiment and others described above, we showed that in some cases it is possible to get clean experimental results with complex images using simplified tasks. In other cases we relied on simple stimuli, like our experiments on retinal signal pooling, where the subtlety and complexity of the induction effect was prohibitive to characterize using images. Our validation experiments later showed that we could not easily apply our findings to complex images. Since we had to test on visual illusion patterns which exacerbate the effect, our optimized method over-corrected for the problem we aimed to solve. Additionally, recent research suggests that induction phenomena have a cognitive dependence, and could be employed by the HVS to assist in grouping or discrimination depending on the viewers understanding of the stimuli [22]. With that said, a large bulk of traditional perceptual models were derived from experimental data employing simple stimuli. This spans from the color matching experiments of Guild and Wright [23, 24], to the contrast sensitivity experiments of Barten which underlie the Perceptual Quantizer (PQ) function of Rec. ITU-R BT.2100 [12]. Researchers should thus tread lightly when applying these models to color management to ensure that key factors influencing their associated perceptual processes in the scenario of viewing images and videos are also considered in their foundational experiments.

Implications for emerging technology

When comparing the results of our tone mapping validation experiments to the predictions provided by state of the art image quality metrics, we found that none were able to effectively predict the choices of observers. It was also found in [2] that these metrics were unable to predict observer opinion on gamut mapping results. This demonstrates unstable footing, with respect to our understanding of relevant perceptual phenomena, for developing technology in the HDR/WCG space.

This is also problematic for the industrial development and adoption of these formats by users. For example, in this space, displays use more power to provide a larger canvas of potential brightness and saturation values to artists. However, our visual system still views this information through the same limited bandwidth as it does when encountering vastly different real-world dynamic range conditions. In the same way, the HVS may respond to a HDR or WCG image in the same way as it would for its SDR equivalent after adaptation. Following previous trends, as this technology is adopted by attention seeking parties like advertisers, they will likely try to use this full range without consideration of the adapted state of the viewer, potentially resulting in cancellation by the observer's visual system. Thus, accurate perceptual models are necessary not only for the development of technology but also to guide users to employ it in the most effective and energy efficient way possible, considering the intended viewing context. These models could be used to warn artists when the extra range they are using is likely to be cancelled out by adaptation. Alternatively, our findings on retinal signal pooling processes indicated that by exacerbating simultaneous contrast effects, sensations could be produced which correspond to chromaticities outside of the device color gamut. By the same token, it is possible that the local contrast properties of WCG and HDR images could be altered to stimulate equivalent sensations with a smaller dynamic range via a limiting function, allowing for images to be displayed in the most energy efficient way possible.

Conclusion

We conducted experiments exploring perceptual mechanisms in the context of emerging color management problems. Included among them were metamerism, chromatic adaptation to the surround, contrast adaptation for different display capabilities, and viewing size dependent simultaneous contrast effects. To address each problem, we formulated models with the intention of simulating the relevant visual processes, and conducted psychophysical experiments using complex image stimuli to optimize and validate them. Through this effort, we encountered a number of helpful general takeaways. Notably, we encountered problems with colorimetry, problems making direct simultaneous comparisons between viewing conditions, and the absence of reliable image quality metrics for HDR/WCG. Given these first hand encounters, we discuss additional problems which may arise due to our lack of understanding of the perceptual areas addressed. Moving forward, two promising avenues for color management are the development of multi-primary displays, which allow for the problematic simplifications of our current ecosystem to be avoided, and the spatio-temporal characterization of perceptual responses to HDR/WCG displays to guide artists to use them effectively and efficiently.

Acknowledgments

The authors would like to acknowledge the many observers who participated in the psychophysical experiments upon which this work is based, including the colorists of Deluxe, Spain. This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 952027 (project AdMiRe).

References

[1] M. Bertalmío. 2019. "Vision models for high dynamic range and wide colour gamut imaging: techniques and applications", Academic Press.

[2] S. Zamir, J. Vazquez-Corral, M. Bertalmío. 2019. "Vision Models for Wide Color Gamut Imaging in Cinema", IEEE Tras. PAMI, 43(5):1777-1790; doi:10.1109/TPAMI.2019.2938499

[3] A. Chapiro, T. Kunkel, R. Atkins, S. Daly. 2018. "Influence of Screen Size and Field of View on Percieved Brightness", ACM Trans. Appl. Perc., 15(3):1-13; doi:10.1145/3190346

[4] A. Sarkar, L. Blondé, P. Le Callet, F. Autrusseau, P. Morvan, Jürgen Stauder. 2010. "A Handy Calibrator for Color Vision of a Human Observer." 18th Color and Imaging Conference.

[5] Jiaying Wu, Lu Zhang, Serhan Isikman, Cheng Chen. 2019. "Enhanced Viewing Experience Considering Chromatic Adaptation", SID 2019 Digest, 61(1):857-860

[6] R. Mantiuk, S. Daly, L. Kerofsky. 2008. "Display adaptive tone mapping", ACM Trans. on Graph, 27(3):68.

[7] Y. Asano, M. Fairchild. 2016. "Categorical Observers for Metamerism", Color Res. Appl., 45(4): 576-585; doi:10.1002/col.22493

[8] T. Canham, D. Long, M. Fairchild, M. Bertalmío. 2022. "Physiologically Personalized Color Management for Motion Picture Workflows", SMPTE Motion Imaging Journal, 131(2): 8-16; doi:10.5594/JMI.2022.3143608

[9] T. Canham, M. Murdoch, D. Long. 2018. "Influence of Ambient Chromaticity on Portable Display Color Appearance", SMPTE Annual Technical Conference & Exhibition.

[10] T. Canham, M. Bertalmío. 2020. "The persistent influence of viewing environment illumination color on displayed image appearance", London Imaging Meeting.

[11] SMPTE ST 2080-1:2014, "Reference White Luminance Level and Chromaticity for HDTV."

[12] Recommendation ITU-R BT.2100-2:2018, "Image parameter values for high dynamic range television for use in production and international programme exchange."

[13] T. Gollisch, M. Meister. 2010. "Eye smarter than scientists believed: neural computations in the circuits of the retina", Neuron, 65(2): 150-164.

[14] P. Cyriac, T. Canham, D. Kane, M. Bertalmío. 2020. "Vision models fine-tuned by cinema professionals for High Dynamic Range imaging in movies", Multimedia Tools and Applications, 80: 2537-2563; doi:10.1007/s11042-020-09532-y

[15] C. Enroth-Cugell, J. Robson. 1966. "The contrast sensitivity of retinal ganglion cells of the cat", J Physiol, 187:517-552; pmid:16783910

[16] H. Helson. 1963. "Studies of Anomalous Contrast and Assimilation", Journal of the Optical Society of America, 53(1).

[17] T. Canham, J. Vazquez-Corral, E. Mathieu, M. Bertalmío. 2021. "Matching visual induction effects on screens of different size", Journal of Vision, 21(6); doi:10.1167/jov.21.6.10

[18] X. Otazu, C. Parraga, M. Vanrell. 2010. Toward a unified chromatic induction model. Journal of Vision, 10(5), 1-24

[19] A. Song, O. Faugeras, R. Veltz. 2019. A neural field model for color perception unifying assimilation and contrast. PLoS Computational Biology, 15(6)

[20] M. Webster. 2015. "Visual Adaptation", Annu. Rev. Vis. Sci. 1:547-67.

[21] Mark Fairchild. 1992. "Chromatic Adaptation to Image Displays", TAGA 2, 803–824.

[22] M. Murgia. 2016. "Perceptual belongingness determines the direction of lightness induction depending on grouping stability and intentionality", Vision Research, 126:69-79.

[23] W. Wright, 1928. "A Re-Determination of the Trichromatic Coefficients of the Spectral Colors", Trans. Opt. Soc., 30(4):141-164; doi:10.1088/1475-4878/30/4/301

[24] J. Guild, 1932. "The Colorimetric Properties of the Spectrum", Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 230:149-187, 1932.

Author Biography

Trevor Canham was born and raised in Rochester, NY, where he went on to receive his B.Sc in Motion Picture Science from the Rochester Institute of Technology in 2018. Since that time he has been researching perceptual problems in color management and color grading tools, and is currently affiliated with the Spanish National Research Council (CSIC), Instituto de Optica. His interests lie in the interaction between human visual perception and aesthetic imaging systems.

Marcelo Bertalmío (Montevideo, 1972) is a Researcher at the Spanish National Research Council (CSIC), in the Instituto de Optica. He received his Ph.D. in electrical and computer engineering from the University of Minnesota, USA, in 2001. His current research interests are in developing image processing algorithms for cinema that mimic neural and perceptual processes in the visual system, and to investigate new vision models based on the efficient representation principle.