

Improved High Dynamic Range Video Coding with a Nonlinearity Based on Natural Image Statistics

Yasuko Sugito

Science and Technology Research Laboratories, NHK, Tokyo, Japan
Email: sugitou.y-gy@nhk.or.jp

Praveen Cyriac, David Kane, and Marcelo Bertalmío

Dept. of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

Abstract—High Dynamic Range (HDR) technologies support the capture and presentation of a wider range of luminance values than conventional systems. An important element of video processing is the transfer function which should emulate human perception and this needs to be revisited for HDR content and displays. In the paper, we adapt a nonlinearity designed for the tone-mapping problem to the problem of video coding. We test the nonlinearity using the Motion Picture Experts Group methodology and find it can outperform existing methods in terms of HDR video quality measure.

Index Terms—High Dynamic Range (HDR), HDR-television (HDR-TV), Transfer Function (TF), video coding, video compression

I. INTRODUCTION

High Dynamic Range (HDR) supports a wider range of luminance in images than the range supported by conventional, Standard Dynamic Range (SDR), systems. In television systems, luminance of SDR video is specified between 0.1 and 100 candela per square meter (cd/m^2) [1], while that of HDR is between up to 0.005 and out of 1,000 cd/m^2 [2].

In HDR video processing, the selection of a Transfer Function (TF) that transforms between scene linear luminance and perceptual non-linear signal value is one of the most important elements. An appropriate function successfully manages both darkness and brightness in luminance while maintaining color vibrancy in chroma. HDR-television (HDR-TV) systems employ two different standardized kinds of TFs: Perceptual Quantization (PQ) Electro-Optical Transfer Function (EOTF) and Hybrid Log-Gamma (HLG) Opto-Electronic Transfer Function (OETF) [2].

To deliver high-quality HDR videos, the Motion Picture Experts Group (MPEG) started an exploration of video compression methods using the latest video coding scheme, High-Efficiency Video Coding (HEVC)/H.265 [3], in 2013 [4].

This paper introduces a new non-linear function suited for HDR video coding. First, we introduce the HDR video coding method considered in MPEG. Then, we introduce existing HDR TFs. Subsequently, we explain our new function. Finally, we present and consider the results of the HDR video coding experiments.

II. HDR VIDEO CODING METHOD

In this section, we introduce the HDR video coding method investigated in MPEG meetings [5].

A. Overview

Fig. 1 depicts an overview of the HDR video coding method. The HDR video input uses linear light intensity captured by a camera sensor. The format is RGB 4:4:4, and the range of the pixel values is from 0 to the peak luminance L_p , measured in cd/m^2 . In terms of pre-processing, the input is converted to Y'CbCr 4:2:0 10 bit, which is the input format of the HEVC encoding. The coding TF expressed as a tone mapping is also included in this process.

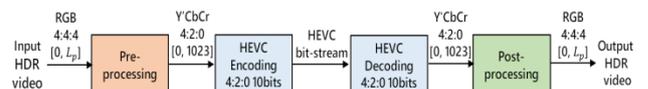


Figure 1. Overview diagram of the HDR video coding [4], [5].

Next, the HEVC reference software model [6] is applied for encoding and decoding. As a result of experimentation, MPEG decided to use HEVC Main 10 profile, a basic profile with 10-bit precision, for encoding. Particular configurations have been changed to effectively compress HDR videos within the existing standard: These configurations include the balancing of bits between luma and chroma components and that between darker and brighter areas [4]. The bit-stream may include metadata about the TF.

The output HDR video after decoding and post-processing has the same format as the input. The post-processing achieves the inverse operation of pre-processing, and the inverse coding TF is included. The objective of this is to reduce the bit-stream data quantity

while minimizing the distortion between the input and output videos.

B. Pre-and Post-Processing

Fig. 2 and Fig. 3 illustrate the pre- and post-processing steps, respectively.

The coding TF transforms each RGB component from linear luminance to non-linear signal value. The coding TF of SDR is conventionally expressed as a simple power function, called gamma, of luminance L and is defined as follows: $1.099L^\gamma - 0.099$ where $\gamma=0.45$ [7]. To manage a wider dynamic range, TFs for HDR are defined by more complicated functions. Examples of the HDR TFs (HLG and PQ) and a new non-linear function represented as tone curves at the 4,000 cd/m^2 system (the peak luminance of the system L_p is 4,000 cd/m^2) are shown in Fig. 4. In the graph, the horizontal axis represents the linear intensity of the input video while the vertical axis shows the perceptual non-linear signal value. The coding TF is represented by the mapping from the horizontal to the vertical axes, and the inverse coding TF is shown by the mapping from the vertical to horizontal axes. In this paper, we developed the new non-linear function in the pink line called the Natural Image Statistics Transfer Function (NISTF). We used these three curves (HLG, PQ, and NISTF) in the video coding experiments mentioned later. The curves are designed with the human visual system in mind as well as the output of the quantization process (Quant 10b in Fig. 2). We explain the details in section III and IV.

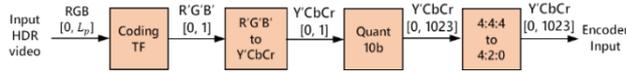


Figure 2. Diagram of pre-processing steps [4], [5].

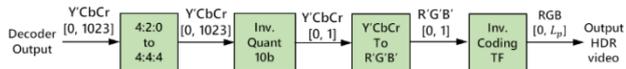


Figure 3. Diagram of post-processing steps [4], [5].

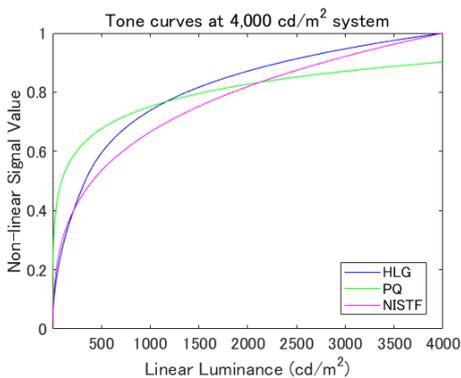


Figure 4. HDR TFs and the new non-linear function.

The latter processes are conducted to comply with the input format of the encoding. First, the RGB color space is transferred to Y'CbCr, a luma and two chroma components. Second, pixel values from 0 to 1 are increased by 1,023 times and rounded to an integer format to treat them as 10-bit precision values. Finally,

the chroma components Cb and Cr are subsampled to a 4:2:0 format. Their image sizes are decreased to half of the original image in both horizontally and vertically. However, this operation for PQ videos causes deterioration in the chroma data, and thus, its influence on the visual quality of the HDR videos and the necessity of luma adjustment have been pointed out [8].

The post-processing phase follows the inverse order of the pre-processing stage. It is desirable for the operations to be invertible except for the inverse quantization and chroma upsampling, whose order operations are essentially lossy.

III. EXISTING TFs

In this section, we introduce two existing TFs, PQ EOTF and HLG OETF which are standardized for HDR-TV [2].

A. PQ EOTF

The PQ EOTF written in the standard format is a type of inverse coding TF at the post-processing stage. The coding TF in the pre-processing stage, i.e., the inverse of the PQ EOTF, is illustrated by the following formula.

$$R' = \left(\frac{c_2 R_{01}^{m_1} + c_1}{1 + c_3 R_{01}^{m_1}} \right)^{m_2} \quad \begin{array}{l} R_{01} = R/10000 \\ m_1 = 0.1593017578125 \\ m_2 = 78.84375 \end{array} \quad \begin{array}{l} c_1 = 0.8359375 \\ c_2 = 18.8515625 \\ c_3 = 18.6875 \end{array}$$

In the above formula, R_{01} denotes a normalized R component. The same TF is also applied to the G and B components. Since the PQ system treats luminance as an absolute value and the peak luminance is 10,000 cd/m^2 , the normalization is conducted by dividing the input values by 10,000. The PQ tone curve is the green line in Fig. 4. In the figure, the maximum perceptual luminance of PQ is less than 1 because of the 4,000 cd/m^2 system.

B. HLG OETF

The HLG OETF is a kind of the coding TF in the pre-processing and described as the following formula.

$$R' = \begin{cases} \sqrt{3R_{01}} & 0 \leq R_{01} \leq 1/12 \\ a \cdot \ln(R_{01} - b) + c & 1/12 < R_{01} \end{cases} \quad \begin{array}{l} R_{01} = R/L_p \\ a = 0.17883277 \\ b = 0.02372241 \\ c = 1.00429347 \end{array}$$

Here, R_{01} describes a normalized R component, and the same TF is also applied to G and B components. The normalization value is calculated by division of the peak luminance of the system L_p . The reason for this is that the HLG system treats luminance as a relative value. The HLG tone curve is represented by a blue line in Fig. 4. The curve has compatibility with SDR displays and the low-intensity part of the above formula confirms this.

IV. DEVELOPED TF: NISTF

In this section, we explain the new non-linear function based on natural image statistics.

A. Definition of Non-linear Function NISTF

Our new function is defined as follows:

$$I' = I_{01}^{\gamma(I_{01})} \quad (1)$$

$$\gamma(I_{01}) = \gamma_{H'} + (\gamma_{L'} - \gamma_{H'}) \left(1 - \frac{I_{01}^n}{I_{01}^n + M_{lin}^n} \right) \quad (2)$$

where I_{01} is a normalized image with RGB components, $\gamma_{L'}=0.45$, $\gamma_{H'}=0.268$, $M_{lin}=4.47 \times 10^{-3}$ and $n=\gamma_{L'}=0.45$. The normalization is conducted the same manner as HLG. The tone curve is shown in Fig. 4 as NISTF (the pink line).

B. Development of NISTF for Video Coding

The developed function is based upon a recent tone mapping algorithm by Cyriac *et al.* [9]. The method has two stages; the first is a global nonlinearity which performs constrained histogram equalization; the second, performs local contrast enhancement. For the purpose of HDR video coding, we do not consider the second, local stage, however the first stage has several properties that make it applicable to the problem of video coding and the development of a global TF.

The first stage of [9] is developed according the theory developed in [10] and a model of the statistics of natural images [11]. Although the original approach produces tone curves that are image-dependent, instead, we have chosen a fixed and image-independent non-linear function that works well on various sequences for video coding. Also in this manner, our function doesn't require metadata for parameters.

As the adaptations of [9], we mainly considered the scaling of low luminance value in (2). In Fig. 5, we show examples of luma image Y' of Carousel Fireworks [12] after pre-processing. We noticed that HLG (Fig. 5 (b)) produces less noisy and darker images than PQ (Fig. 5 (a)) and a tone curve of [9] (Fig. 5 (c)) especially in low intensity areas, and this improves coding efficiency.

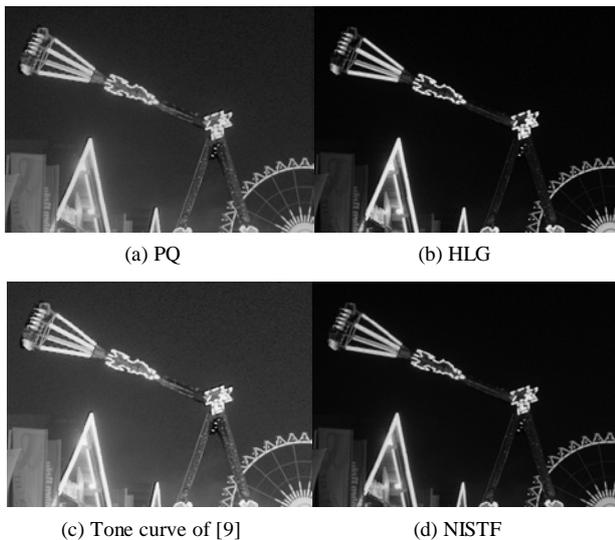


Figure 5. Examples of the luma image after pre-processing.

Fig. 6 shows the tone curves in log-log axes at the $4,000 \text{ cd/m}^2$ system. The tone curve of [9] (the red line) and PQ (the green line) allot a much larger value of non-linear luminance than HLG (the blue line) for darker linear luminance. Then, we modified to fix the parameter $\gamma_{L'}$ to 0.45, the conventional gamma value, as shown in (2). In Fig. 6, NISTF (shown by the pink line) is the result

of the adjustment of [9] (the red line). This is almost parallel to HLG in darker areas due to the slope of the HLG being 0.5 and close to $\gamma_{L'}$.

In Fig. 7, we show examples of the image Carousel Fireworks [12] after pre- and post-processing, not compressed by HEVC. The images were tone-mapped using the following formula:

$$R_{SDR} = \text{clip}(255 \times (R_{01} \times 2^c)^{\gamma}, 0, 255)$$

where R_{01} denotes a linear normalized R component of an HDR image, the exposure value $c=4$, and the slope $\gamma=0.45$. The function $\text{clip}(x, a, b)$ clamps the value x between a and b . The same equation was also applied to G and B components.

Due to the chroma subsampling in the pre-processing, some artifact can be seen in PQ (Fig. 7 (a)), which is not included in the original image (Fig. 7 (c)): In the upper left part, a pink dashed line and the surroundings were emphasized. To prevent this, a luma adjustment method was proposed [8]. On the other hand, our method NISTF (Fig. 7 (d)) produces a natural image like HLG (Fig. 7 (b)) even after the subsampling process, and this means that our method doesn't require the luma adjustment.

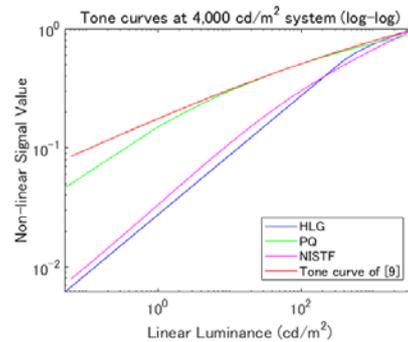
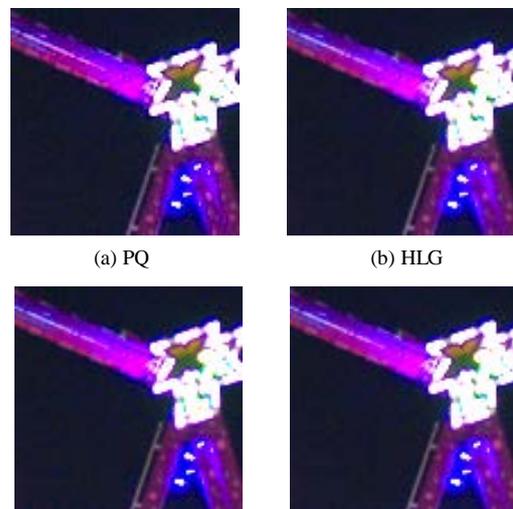


Figure 6. HDR TFs in log-log axes.



(c) original image (nonsubsampling) (d) NISTF

Figure 7. Examples of the image after pre- and post-processing.

V. HDR VIDEO CODING EXPERIMENTS

We conducted HDR video coding experiments with HEVC based on the MPEG common test condition [5].

We applied each NISTF and HLG OETF to PQ videos, and compared with a case with directly compressing PQ videos.

A. Test Sequences

As shown in Fig. 8, we selected four sequences from the color graded version of the HdM-HDR-2014 content [12]. Each sequence consists of 1920×1080 pixels per frame, 25Hz, and 10 seconds. The sequences were graded for ITU-R Rec. BT.2020 [13] primaries and from 0.005 to 4,000 cd/m². Each frame of the sequences is provided as a 16-bit tiff file, and inverse PQ EOTF was applied.

Fig. 9 shows the encoder input video generation process. To generate encoder input HLG and NISTF videos, we first applied PQ EOTF to the normalized input PQ video and got the linear input HDR video. We also confirmed that the input PQ video and the result after applying PQ EOTF and inverse PQ EOTF are nearly identical. Meanwhile, we used the normalized input PQ video to prepare the encoder input PQ video.

For the color space transform from R'G'B' to Y'CbCr, we used the matrix stated in ITU-R Rec. BT.2020 [13] bearing in mind the sequence's color space.

Fig. 10 shows the characteristics of the sequences, the dynamic range and the median of linear luminance in the logarithmic scale for each frame of the sequences. The dynamic range was computed as

$$Dynamic\ Range = \log_{10} \left(\frac{L_{max}}{L_{min}} \right)$$

where L_{max} and L_{min} respectively denote maximum and minimum linear luminance in cd/m² after discarding 0.5% brightest and darkest values. The linear luminance represents Y of CIE 1931 XYZ color space converted from the linear RGB which is shown in Fig. 9 as Input HDR video.

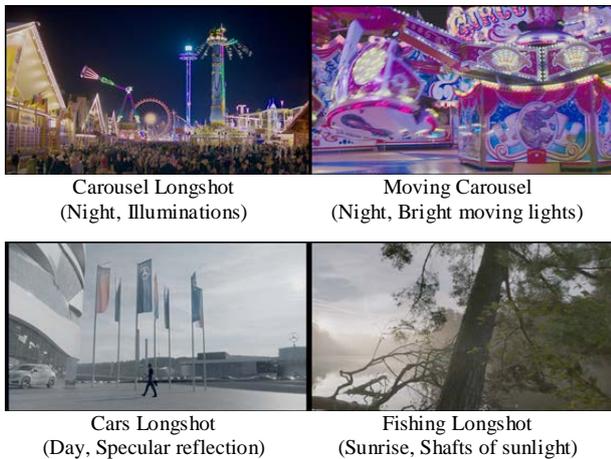


Figure 8. HDR test sequences used for video coding experiments.

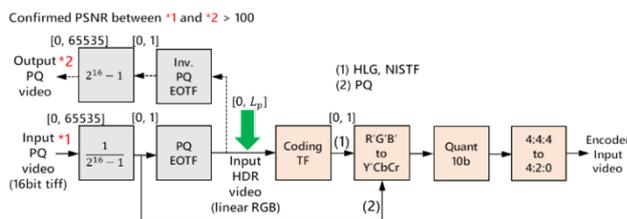


Figure 9. Diagram of encoder input video generation process.

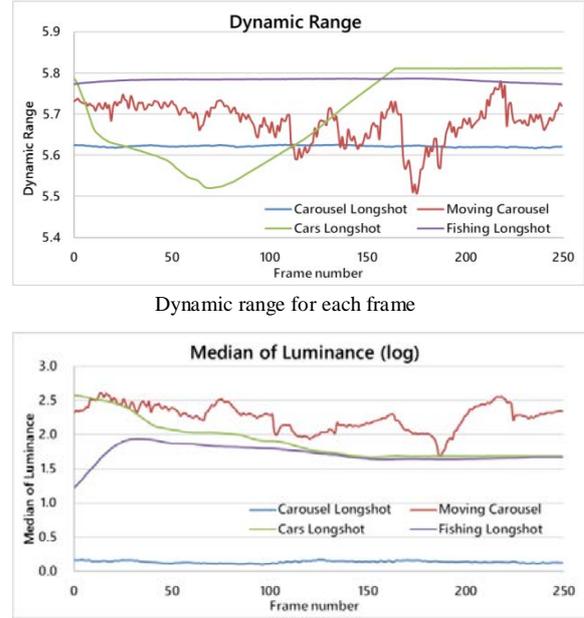


Figure 10. Characteristics of the sequences.

The characteristics of the four sequences are different. Carousel Longshot and Moving Carousel are night scenes, and self-illuminated objects and dark surroundings exist at the same time. In Carousel Longshot, the dynamic range and the median of luminance are relatively constant. On the other hand, both are fluctuant in Moving Carousel. Cars Longshot and Fishing Longshot are sunlight scenes, and their medians of luminance in the latter part of the sequences are close. The dynamic range of Cars Longshot is changing in the former part while that of Fishing Longshot is always flat.

B. Metrics

In terms of an image quality metric, we adopted HDR Video Quality Measure (VQM) [14]. Then, we calculated the Bjontegaard delta rate (BD-rate) [15] from the normalized HDR-VQM and the bit-rate of the encoding results. We compared the input and output HDR videos shown in Fig. 1.

1) HDR-VQM

HDR-VQM was developed to assess quality in HDR videos. The metric was designed based on the human visual systems, and it calculates the distortion between the original video and the encoded video while analyzing the difference of the videos in a spatio-temporal dimension.

We used the software provided by the authors of [14] in the preset SIM2 HDR display condition to evaluate the HDR videos whose peak luminance is 4,000 cd/m². We input linear luminance images, which consist of a component Y of CIE 1931 XYZ color space, and computed normalized HDR-VQM scores, which will be 1 for perfect quality and decrease as subjective quality decreases.

2) BD-rate

Fig. 11 shows an example of video coding results. In the graph, the horizontal and the vertical axes indicate

bit-rate in kilobit per second (kbps) and normalized HDR-VQM (VQM), respectively. Comparing the two results at VQM x , the bit-rate of the new method is two-thirds of the reference. In such a case, the bit-rate saving ratio of the new method based on the reference is -33.3% ($(2b/3-b)/b \cdot 100\%$).

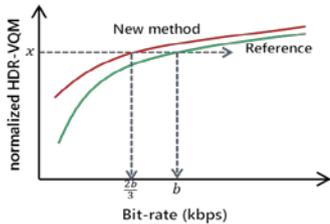


Figure 11. Examples of video coding results.

BD-rate [15] is a metric used to compute the average bit-rate saving ratio. To compute the BD-rate, we conducted video coding experiments with four different quantization parameters, which control the bit-rate of the encoded video. From the results of VQM and bit-rate, we calculated the BD-rate based on original PQ condition for HLG and NISTF applied conditions, which represents the bit-rate savings from PQ corresponding to each metric.

VI. RESULTS OF HDR VIDEO CODING EXPERIMENTS

Table I shows the BD-rate VQM of HDR video coding experiments. The values in bold refer to the best results among original PQ, HLG and NISTF applied conditions.

Fig. 12 shows the graphs of the normalized HDR-VQM with bit-rate for each sequence.

TABLE I. RESULTS OF EXPERIMENTS

	HLG	NISTF
Carousel Longshot	-8.2%	-15.6%
Moving Carousel	-11.6%	-24.3%
Cars Longshot	-29.2%	-32.0%
Fishing Longshot	-20.7%	-25.9%

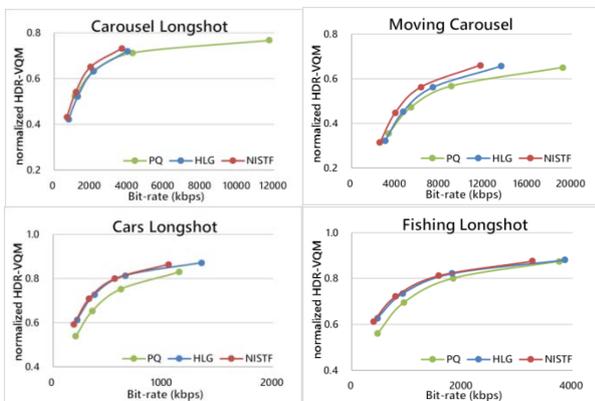


Figure 12. Results in the graph of HDR-VQM with bit-rate.

VII. CONSIDERATIONS

A. Video Coding Performance

In Table I, all the results represent minus values. This means that both HLG and NISTF are better than original

PQ for all conditions. Although HLG has a constraint of backward compatibility with SDR displays, it achieves excellent performance on video coding.

Comparing HLG and NISTF, all NISTF results show better results than HLG especially in night scenes. From this, applying NISTF to PQ video is very effective mapping for HDR video coding and it can preserve the quality of luminance even after the chroma subsampling process.

In the experiments, no metadata was added in terms of coding TFs. Since the NISTF tone curve is fixed, only mode information is required as data, therefore, the amount would be the same as HLG, and in terms of practical use, NISTF is also more efficient than HLG.

B. Complexity

To calculate the inverse of NISTF, we used a look-up-table generated by NISTF itself. This is because the inverse cannot be described as a simple formula as the definition expressed as (1) and (2) includes the input I_{01} .

Considering practical applications of a camera or other devices, both coding TF and its inverse would be calculated using a look-up-table. Thus, NISTF's complexity is comparable with that of existing methods.

VIII. CONCLUSIONS

We developed a new non-linear function for HDR video coding based on a tone mapping operator that takes into account natural image statistics. We conducted video coding experiments with various types of HDR sequences and confirmed that the function is applicable to the existing HDR systems and this makes them more efficient than their original performance. This paper compared performance to objective metrics calculated by only luminance component. In the future, we plan to verify performance levels using subjective, viewer evaluation.

The tone mapping operator on which NISTF is based was originally designed for contents adaptive processing; however, it performed well in video coding with a fixed tone curve. Our future work will focus on exploring more appropriate parameters and testing with more amount of HDR videos.

ACKNOWLEDGMENT

This research was conducted in the image processing for enhanced cinematography (IP4EC) group at the Universitat Pompeu Fabra, Barcelona, Spain between November 2016 and May 2017. This work was partially 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] Recommendation ITU-R BT.1886-0, "Reference electro-optical transfer function for flat panel displays used in HDTV studio production," Mar. 2011.

- [2] Recommendation ITU-R BT.2100-0, "Image parameter values for high dynamic range television for use in production and international programme exchange," Jul. 2016.
- [3] ISO/IEC 23008-2:2013, "High efficiency coding and media delivery in heterogeneous environments - Part 2: High efficiency video coding," Recommendation ITU-T H.265, High Efficiency Video Coding, 2013.
- [4] J. Ström and J. Samuelsson, "Progress report from MPEG," *SMPTE Motion Imaging Journal*, vol. 125, no. 7, pp. 80-84, 2016.
- [5] E. François, J. Sole, J. Ström, and P. Yin, "Common Test Conditions for HDR/WCG video coding experiments," *JCTVC-Z1020*, 2017.
- [6] High Efficiency Video Coding (HEVC). [Online]. Available: <http://hevc.hhi.fraunhofer.de/>
- [7] Recommendation ITU-R BT.709-6, "Parameter values for the HDTV standards for production and international programme exchange," Jun. 2015.
- [8] J. Ström, J. Samuelsson, and K. Dovstam, "Luma adjustment for high dynamic range video," in *Proc. Data Compression Conference*, 2016, pp. 319-328.
- [9] P. Cyriac, D. Kane, and M. Bertalmío, "Optimized tone curve for in-camera image processing," *Electronic Imaging*, 2016.
- [10] D. Kane and M. Bertalmío, "System gamma as a function of image- and monitor-dynamic range," *Journal of Vision*, vol. 16, pp. 4, 2016.
- [11] J. Huang and D. Mumford, "Statistics of natural images and models," in *Proceedings 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999, vol. 1, p. 547.
- [12] 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," in *Proc. SPIE*, 2014, vol. 9023, p. 90230X.
- [13] Recommendation ITU-R BT.2020-2, "Parameter values for ultra-high definition television systems for production and international programme exchange," Oct. 2015.
- [14] M. Narwaria, M. P. D. Silva, and P. L. Callet, "HDR-VQM: An objective quality measure for high dynamic range video," *Signal Processing Image Communication*, vol. 35, pp. 46-60, 2015.
- [15] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *VCEG-M33*, Apr. 2001.



Yasuko Sugito is currently with NHK (Japan Broadcasting Corporation) Science and Technology Research Laboratories, Tokyo, Japan, researching video compression algorithms and image processing on 8K Super Hi-Vision and developing 8K video codecs. Her current research interests focus on image quality improvements and speeding ups for full-featured 8K video, which includes WCG, HDR, and HFR technologies.



high dynamic range imaging and tone mapping.

Praveen Cyriac received the M. Tech degree in computer science from Kerala University, India, and the PhD degree in Information and Communication Technologies from Universitat Pompeu Fabra (UPF), Barcelona, Spain in 2017. His thesis entitled 'Tone mapping based on natural Image Statistics and Visual Perception Models'. He is currently a Post-doctoral fellow at the Image Processing for Enhanced Cinematography group at UPF. His research interests are in



David Kane is a visual psychophysicist working at the intersection of vision science and visual technologies. David is motivated by applied problems but strongly believes that a combination of basic and applied research is needed to develop effective visual technologies. David also believes that approaching vision science from applied perspective can help shed new light onto old problems.



Marcelo Bertalmío (Montevideo, 1972) received the B.Sc. and M.Sc. degrees in electrical engineering from Universidad de la República, Uruguay, and the Ph.D. degree in electrical and computer engineering from the University of Minnesota in 2001. He is an Associate Professor at Universitat Pompeu Fabra, Spain. Among other honors, he was awarded the Ramón y Cajal Fellowship, the ICREA Academia Award, and the 2012 SIAG/IS Prize of the Society for Industrial and Applied Mathematics (SIAM) for co-authoring the most relevant image processing work published in the period 2008-2012. Has an ERC Starting Grant for his project "Image processing for enhanced cinematography" and an ERC Proof of Concept Grant for a tone mapping method. Has written a book titled "Image Processing for Cinema". His current research interests are in developing image processing algorithms for cinema that mimic neural and perceptual processes in the human visual system.