HDR FOR LEGACY DISPLAYS USING SECTIONAL TONE MAPPING

Lenzen L.

RheinMain University of Applied Sciences, Germany

ABSTRACT

High dynamic range (HDR) allows us to capture an enormous range of luminance values within a still image or a sequence of video frames. But many consumers will not have the necessary displays to experience this in the near future. To allow these 'legacy' users to benefit, an adaptation using global tone mapping would be a possible solution. But the results tend to suffer from low subjective contrast and can product large-area flicker. To overcome these drawbacks, three enhancement steps are proposed. They are based on certain broadcast requirements as well as on viewer preferences, which were surveyed at the beginning of this study. The basic idea is to analyse each luminance value for its relevance in the image and discard unimportant ones. This 'virtual aperture' will be processed across the whole image and on image sections. Finally the tone mapping result will be composed with the chrominance values by using a modified IPT colour space.

INTRODUCTION

One of the main goals of future television is to create a more immersive experience. The viewer should get the feeling that he or she is inside the action. One key component to achieve this is HDR. By preserving details in highlights and shadows simultaneously, a visual sensation is created close to viewing the real scene. For efficient coding, Hybrid-Log-Gamma and Perceptual Quantizer (PQ) are discussed. This technology will produce a significant increase in subjective video quality as shown in several studies (e.g. 1, 2). But for the next few years we can expect that over 90% of viewers will still have legacy displays. So one of the main questions which comes with HDR is: 'Is there a way to let the viewers with legacy displays also benefit from HDR?' (Note that a faithful reproduction of the images in this paper is only available when viewed as a .pdf on a sRGB display.)

REQUIREMENTS AND GOAL

When we talk about showing HDR material on a standard dynamic range (SDR) display we have to think about how to handle the enormous dynamic range at capture compared to the small dynamic range of a particular display. The easiest way is to cut off all parts which are outside the range, but clipping would bring us back to the burned-out highlights and the undefined shadows. A more promising way is to do a contrast compression also known as tone mapping. Our goal is to create a tone mapping system which is capable of live broadcast and does not suffer from the typical problems of tone mapping which are discussed later on.

At first we define the following requirements for live broadcast:

- The most important point is to create a very pleasing image that matches the viewer's preference. Even though a very pleasing image is not equivalent to the most realistic and natural image. In addition, the look should not be too different from the familiar look of television today - in order to gain the viewer's acceptance.
- The system needs to be very stable so that it can be used for live broadcast without any post production. At no time should tone mapping artefacts like flicker, ghosting or halos be produced.
- Unlike an aperture which has discrete steps, a smooth adjustment to changing brightness during a scene should be performed. This transition should not take place if there is a scene change; in this case the exposure ought to adapt to the new situation immediately.
- To utilize the capabilities of legacy displays, the system should be sensitive to the display and environmental brightness and should adapt to these parameters depending on the scene.
- For live broadcast it is necessary that the system works in real-time.

RELATED WORK AND EVALUTATION

Before starting engineering, we ran a test letting 5 professional colourists grade 7 HDR sequences[1] for SDR displays. In a viewing session, 20 subjects (mostly non-expert viewers) would decide which grading he or she liked best. Therefore every grading was compared one-to-one with every other, so the subjects only had to decide which one of the two was the more pleasing. The experiment showed that the viewers liked high contrast and high saturation. This was confirmed in several previous studies (3,4,5). But moreover, sometimes excessive saturation was found and that high saturation does not necessarily lead to a pleasing image. It seemed rather important that all information within the image could be easily extracted. Or, in other words: the viewer likes to see all the details in all parts of the image without having to pay too much attention. Saturation and contrast often enhance this aspect. Clipping too, is acceptable in this case, and was sometimes even missed when it was not there. For example, to reproduce spectral highlights on SDR, clipping is an adequate process. Another outcome is that the white point is only of secondary importance. A warm grading was not preferred over a cold one, as one might expect. So the white point is only a creative decision and does not need to be adjusted by the tone mapping itself.

Discovering a viewer's preference in tone mapping is not an easy task. Tone mapping operators (TMOs) are often categorised into local and global. Local operators analyze the neighbourhood of each pixel. They create a spatial region of adaptation. That is why they can preserve small local contrast. But on the other hand this is the reason why they tent to produce visual artefacts such as: halo, ghosting or flicker. In addition, the real-time requirement does not make it any easier to use a local tone mapping. Furthermore, some of

---

[1] *bistro_01, bistro_02, carousel_fireworks_01, carousel_fireworks_08, carousel_fireworks_09, cars_fullshot and fireplace_01* filmed by Fröhlich et. al (7)

the results appear a little surreal and are far away from the familiar TV look. Also agreeing with the results of Petit and Mantiuk (3), Eilertsen *et al* (5) and Aydin *et al* (6), local operators do not seem to be the first choice for live broadcast.

The second category consists of global operators. They create one transfer curve for the entire image (like a camera transfer curve), mostly taking into account the brightest pixel and some kind of average luminance. In most cases this is less computationally intensive. Moreover, global operators are more stable. They can produce flicker too when only based on an intra-frame analysis, but smoothing by interpolation can be easily performed and could reduce, or even eliminate, flicker. However the key issue of global tone mapping is that the images suffer from low subjective contrast - they tend to look like log-signals. Not surprisingly the transfer curve is close to a log-curve.

Our idea is to improve subjective contrast, saturation and sharpness by combining classic global tone mapping with three enhancement steps, which we call “EVI” (Enhanced Video Information).

ENHANCEMENT STEPS

Step 1: Virtual Aperture

A log-curve has a flat gradation. Therefore a higher dynamic range can be coded with the same number of bits. Or in other words: you need a bigger change in luminance to produce the same change in coded values. If you display such a log-curve directly, the image looks like someone has overlain a grey veil. Contrast and 'crispness' are missing. The larger the difference between the brightest and darkest pixel, the flatter the gradation created by the global TMO. But is there really a need for such high dynamic range in every image? Which information is necessary and which can be seen on a SDR display?

For example let us take a look at the sequence filmed by Fröhlich *et al*. (7). The image is picked from the sequence called 'bistro' (Figure 10 first image - at the end of the paper). The highest luminance value is about 58253 cd/m^2 and the lowest about 0.16 cd/m^2 (with noise). So we are talking about a dynamic range of about 18.5 stops. If we remove only one per mill on each side, the luminance varies between 6653 cd/m^2 and 0.24 cd/m^2. It is reduced by nearly 4 stops. Of course this is a prime example but we can conclude that often extreme values which are unimportant for the image (sometimes only noise) influence the result of tone mapping. As a result a flat gradation is applied. Beyond that extreme values have high deviations. These cause the typical large-area flicker because brightness varies from one frame to another. So it is worth considering how to handle the extreme values.

Some TMOs perform clipping by cutting-off the highest and lowest percent. But such fixed values would not lead to the best results due to the fact that not every scene needs a steep gradation. For some images the impact of 1% might be subjectively less, for others subjectively more. That is why our system should be smarter when adapting automatically to the dynamic range distribution within a scene. A possible means of doing so would be to judge the importance of a luminance value based upon a histogram.

The luminance Y_{in} of the input is calculated e.g. by using the form mentioned below for Rec.709 in linear light. The calculation differs if another camera or input gamut is used.

$$\mathrm{Y_{in}} = 0.2126 \cdot \mathrm{R_{709}} + 0.7152 \cdot \mathrm{G_{709}} + 0.0722 \cdot \mathrm{B_{709}}$$

After that, the histogram $h(Y_{log})$ over all pixels of the incoming image, is computed (Figure 1). The histogram has about 13 orders of dynamic range (Figures 1 to 5 show a section of 8 orders). For the y-axis, the luminance Y_{in} is scaled in a logarithmic way:

$$Y_{log}(Y_{in}) = \begin{cases} 0, & \text{for} \quad Y_{in} \leq 10^{-k_1} \\ \log_{10}(Y_{in}) + k_1, & \text{for} \quad Y_{in} > 10^{-k_1} \end{cases}$$

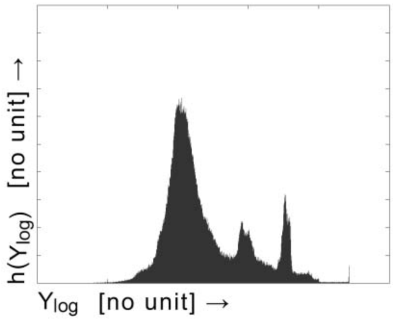

Figure 1 – histogram $h(Y_{log})$ of the incoming image

k_1 should be chosen so that no negative values are produced. As a next step we introduce a so-called 'contrast box'. The contrast box has a constant dynamic range independent from its position on the y-axis due to the log-scale. The width W is defined dependent upon the highest display luminance L_{dmax} (measured in cd/m²), because the brighter the display, the more dynamic range can be preserved without creating a flat gradation by using tone mapping. k_2 and k_3 are fixed scaling factors.

$$W = k_2 + k_3 \cdot L_{dmax}$$

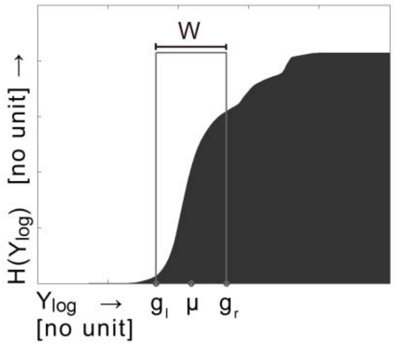

Figure 2 – cumulative histogram $H(Y_{log})$ with contrast box

This box is moved to the position where the most luminance values are inside the box (Figure 2). The region inside the box could be interpreted as the centre of interest. Here we would like to have no clipping but a steep gradation. Let us call the middle of the box μ (on the x-axis). It is useful to compute the cumulative histogram $H(Y_{log})$ and search for the position g_l with the highest slope at the distance W.

$$H(Y_{log}) = \sum_{k=0}^{Y_{log}} h_k$$

$$H(Y_{log} + W) - H(Y_{log}) = \max \rightarrow g_l$$

$$\mu = g_l + ½ \cdot W$$

The importance of the centre could be linked to the number of values inside the box. More values mean more importance. To normalize the number, a division by the width W is performed. This leads to the height H_a of the contrast box (Figure 3).

$$H_a = \frac{\sum_{g_l}^{g_l+W} h_k}{W} = \frac{H(g_l + W) - H(g_l)}{W}$$

When we define a standard deviation σ linked to H_a we can calculate a Gaussian function of the form

$$\sigma = \left(\frac{H_{av}}{H_a}\right)^{k_5} \cdot H_{av} \cdot k_6$$

$$g(Y_{log}) = \frac{H_a \cdot k_4 \cdot W}{\sqrt{(2\pi \cdot \sigma^2)}} \cdot e^{\left(\frac{-(Y_{log}-\mu)^2}{2 \cdot \sigma^2}\right)}$$

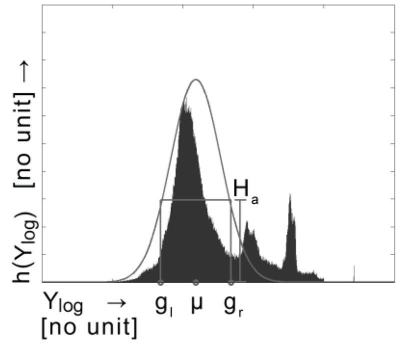

Figure 3 – histogram $h(Y_{log})$ with contrast box of height H_a and Gaussian function

dependent on the dynamic range distribution (Figure 3). k_4, k_5 and k_6 are fixed scaling factors for the height and width of the Gaussian function. H_{av} is an average height of the contrast box. In the case of a very centred dynamic range distribution, a high Gaussian function with very steep slopes is formed. In a case where the values are highly distributed over a large dynamic range, the slopes become very flat. The function $g(Y_{log})$ is used to weight the original histogram $h(Y_{log})$ by multiplication.

$$f(Y_{log}) = h(Y_{log}) \cdot g(Y_{log})$$

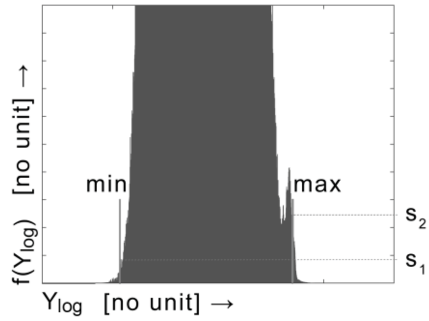

Figure 4 – $f(Y_{log})$ with overshooting around the center of interest

This leads to an overshooting around the centre of interest (Figure 4). By comparing $f(Y_{log})$ to the clipping thresholds for the shadows and highlights ($s_1 \cdot L_{room}$, $s_2 \cdot L_{room}$; where L_{room} is the environmental luminance) the centre would not lose information. The comparison leads to the darkest and brightest logarithmic luminance value, which is important for the image (Figure 5). Afterwards they can be transformed backwards into luminance values (Y_{max}, Y_{min}).

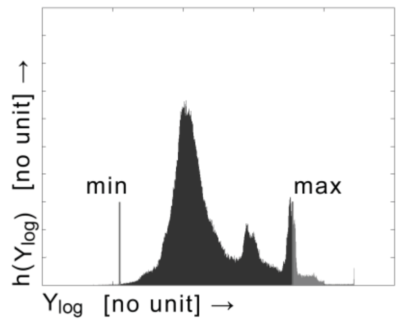

Figure 5 – histogram $h(Y_{log})$ with clipping thresholds

By avoiding extreme values, large-area flicker is reduced remarkably, but it has not been resolved completely. μ and σ can vary from one frame to another, too. Using a interpolation of the form

$$\mu_{gn} = \mu_{g(n-1)} \cdot z_\mu + \mu_n \cdot (1 - z_\mu)$$

$$\sigma_{gn} = \sigma_{g(n-1)} \cdot z_\sigma + \sigma_n \cdot (1 - z_\sigma)$$

helps to improve this. z_μ and z_σ have values between 0 and 1. The default value is 0.95. In this case the new μ and σ would have 5% impact on the overall μ (μ_{gn}) and overall σ (σ_{gn}), respectively. Such a transition should not take place if there is a scene change. In this case the exposure must adapt to the new situation immediately. A scene change detector is therefore needed. The scene detection should be coupled to the centre of interest and, therefore, to μ and σ. If one of the following conditions is not fulfilled, a new scene is assumed. s_μ and s_σ are the thresholds.

$$\left|\frac{\mu_n - \mu_{g(n-1)}}{\mu_{g(n-1)}}\right| \leq s_\mu \qquad\qquad \left|\frac{\sigma_n - \sigma_{g(n-1)}}{\sigma_{g(n-1)}}\right| \leq s_\sigma$$

Figure 6 shows a comparison with and without the virtual aperture. For global tone mapping the Drago operator (8) is used.

Figure 6 – Left: A typical live broadcast situation shown with a camera transfer curve. Middle: The same situation shown with the Drago global tone mapping operator. Right: Using the virtual aperture with the Drago operator.

Step 2: Sectional Tone Mapping

Global tone mapping has some big advantages compared with local tone mapping when considering (live) broadcast. But for scenes with a very high dynamic range, even the results obtained with the virtual aperture have problems at small local contrast. To overcome this, the virtual aperture has to be linked with some kind of local adaptation. For this purpose the image is separated into blocks. In experiments 16 x 9 led to the best results. The virtual aperture is now processed on every block separately. If we were to take more blocks there would not be enough luminance values to compute a good histogram for the virtual aperture. If we were to take less blocks the local impact is very small. We call this 'sectional tone mapping'. Compared with classic local tone mapping, we are able to avoid the major drawbacks.

To overcome visible steps at the borders of the boxes (Figure 7 middle), their results for Y_{max} and Y_{min} have to be smoothed with those of their neighbours. Therefore a non-linear function is used. At that time Y_{max} and Y_{min} are matrices of the size 16 x 9 and not scalars as with classic global tone mapping. Afterwards they are resized to the size of the whole image. The matrices will be passed over to the tone mapping operator (e.g. the modified Drago). Y_{in} and Y_{max} are scaled by a so-called 'world adapting luminance' (L_{wa}) as used e.g. at Drago *et al* (8) and therefore called Y_{win} and Y_{wmax} in the following equitation. L_{wa} can also be a matrix.

$$Y_{out}(x,y) = \frac{1}{\log_{10}(Y_{wmax}(x,y)+1)} \cdot \frac{\log(Y_{win}(x,y)+1)}{\log\left(2+\left(\left(\frac{Y_{win}(x,y)}{Y_{wmax}(x,y)}\right)^{\frac{\log(b)}{\log(0.5)}}\right)\cdot 8\right)}$$

b is set to 0.85 as in (8). For a smooth and stable Y_{max} it is recommended to blend the results of the global and the sectional virtual aperture. z is a value between 0 and 1.

$$Y_{max} = Y_{max\_global} \cdot z + Y_{max\_sectional} \cdot (1 - z)$$

Figure 7 – Left: Virtual aperture. Middle: Virtual aperture performed on 9 x 16 blocks.
Right: Sectional tone mapping with z = 0.25.

Step 3: Modified IPT Colour Space

When considering tone mapping, we perform a contrast compression of the luminance component only. So the question is :how should we apply these results to the chromaticity, too? The goal is not to influence the hue and saturation. Conventionally, the image would have been converted to the XYZ colour space and tone mapping would be applied on the Y component. After that Y_{in} would be compared with Y_{out} and the ratio would also be applied to X and Z. This approach does effect the colour impression.

For this reason our system is taking an easy but powerful approach by using the IPT colour space – an improvement of the CIELab colour space. The IPT is based on a decorrelation of luminance and chrominance and is designed to be perceptually uniform (9). 'I' stands for the intensity and 'P' and 'T' are colour-difference signals (called protan and tritan). So in our approach the incoming signal is always transformed to IPT space.

As usual the tone mapping gets applied to the luminance component ($Y_{in} \rightarrow Y_{out}$). By using step 2 Y_{out} is automatically normalized to the same range as 'I' (0 to 1). Y_{in} has to be normalized too, so the ratio of tone mapping can be easily applied on 'I'. If this is done the image looks overal,l too dark. This is caused by the fact that the LMS components (long-medium-short cone response data) were raised to the power of 0.43 before being transformed to IPT. So some kind of gamma correction needs to be applied to the ratio, too.

$$I_{out} = I_{in} \cdot \left(\frac{Y_{out}}{Y_{in}}\right)^{0.43}$$

At the end, the image is transformed backwards to RGB, for example according to Rec.709. Values outside the gamut get clipped. For display, a gamma has to be applied.

When using the IPT colour space as described above there is a big loss of saturation. The problem is that although chrominance is separated from luminance, the incoming luminance influences P and T. For higher luminance values P and T increase - namely at a factor of 2.6915 with every decade. In SDR images this would not have such an impact, but tone mapping manipulates the luminance very strong. The luminance after tone mapping does not correspond to P and T any longer. That is why we use a compensation, inspired by the traditional one in XYZ. However a full compensation leads to over-saturated results, which is why we added the factor 0.85 (based on a series of subjective tests. T is treated equivalently:

$$P_{out} = P_{in} \cdot \left(0.15 + 0.85 \cdot 2.6915^{\log_{10}\left(\frac{Y_{out}}{Y_{in}}\right)}\right) \quad \text{or} \quad P_{out} = P_{in} \cdot \left(0.15 + 0.85 \cdot \frac{I_{out}}{I_{in}}\right)$$

Compared with XYZ, IPT can provide a more realistic saturation, contrast and colour fidelity. In a direct comparison, as in Figure 8, XYZ looks washed-out. Using IPT also brings back structure in the highlights. In Figure 9 the whole workflow is shown.

Figure 8 – The left image shows the tone mapping in XYZ, the right one in IPT. The grass and the sky in XYZ have a little hue shift and look more burned-out.

TRANSMISSION

There are some possible applications for the proposed system in the production and transmission path. E.g. it could be implemented directly in the camera, so that a SDR-compatible signal (SDR with EVI) is at the output and everything else would remain the same. When using it for today's HD transmission, fixed L_{dmax} and L_{room} have to be chosen. 200 cd/m^2 and 10 cd/m^2 for display and environmental brightness respectively, are a good compromise which will match most viewing situations. For HDR transmissions in the future e.g. two fixed pairs of values could be used for creating a two-layer solution such as the Dolby Version method (10) or Mantiuk *et al*. (11). By generating metadata from the virtual mastering display, the home display can adapt the signal for its specification. For OTT, perfectly adjusted video could be generated at the server and sent out to the home display. Therefore the home display passes over its L_{dmax} and L_{room}. Last but not least, the algorithm could be implemented in the display - or for legacy displays, in the Set Top Box directly. So the automatic adaption can be done at home.

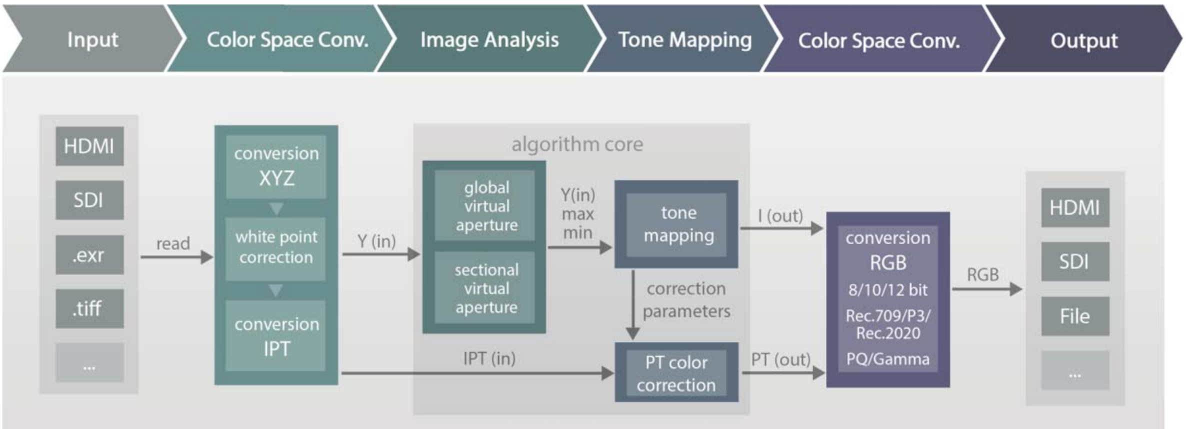

Figure 9 – The incoming HDR images are converted to XYZ and afterwards to IPT. The Y component is tone mapped using the three enhancement steps. P and T are corrected before composing these values with Y_{out} to a new IPT image. Finally, a transformation to the output colour space is performed.

Figure 10 – Example images: Left: TV as it is today. Right: EVI

CONCLUSION AND OUTLOOK

Global tone mapping has much more potential than is exploited in today's workflows. When putting it into an appropriate environment it is able to give a large improvement to broadcast pictures, even today. Our system processes HDR content to SDR without excessively influencing the look of the image. The impression of contrast, saturation and sharpness is close to today's television picture, so it will be accepted by the viewers. But at the same time, much more scene contrast is delivered. Two example images are shown in Figure 10. The viewer can easily follow the details in the shadows and is not dazzled by the highlights. The system is stable for all tested material so far and fulfils all requirements for live broadcast. A real-time implementation is currently under development. A further improvement is that the aperture does not need to be controlled. Also, the system is future-proof. Different display and environmental luminance conditions are considered as well as wide colour gamut. Beside the real-time implementation, further research will focus on a more complete viewer preference model and investigate different colour spaces as well as colour corrections for tone mapping to improve the enhancement system.

ACKNOWLEDGEMENTS

The author would like to thank his colleagues for their contributions to this work and DFL/Sky for the scenes captured in the Allianz Arena. The work was financed by the RheinMain University of Applied Sciences and the Federal Ministry for Economic Affairs and Energy Germany. TransMIT supports the patent process.

REFERENCES

1. Hoffmann, H., 2014. Investigation in an Image Dynamic Range Methodology NAB Show. April 2014.

2. Kunkel, T. and Reinhard, E., 2010. A Reassessment of the Simultaneous Dynamic Range of the Human Visual System. Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization. pp. 17-24.

3. Petit, J. and Mantiuk, R., 2013. Assessment of video tone-mapping: Are cameras' S-shaped tone-curves good enough? J. Visual Communication and Image Representation. vol. 23 pp. 1020-1030.

4. Akyüz, A. O., Fleming, R., Riecke, B. E., Reinhard, E. and Bülthoff, H. H., 2007 Do HDR Displays Support LDR Content? A Psychophysical Evaluation. ACM Trans. Graph. July 2007, articleno 38.

5. Eilertsen, G., Wanat, R., Mantiuk, R. K. and Unger, J., 2013. Evaluation of Tone Mapping Operators for HDR-Video. Comput. Graph. Forum. vol. 32 pp. 275-284.

6. Aydin, T. O., Stefanoski, N., Croci, S., Gross, M. and Smolic, A., 2014. Temporally Coherent Local Tone Mapping of HDR Video. ACM Trans. Graph. vol. 33 pp. 196:1-196:13.

7. Froehlich, J., Grandinetti, S., Eberhardt, B., Walter, S., Schilling, A. and Brendel, H., Creating Cinematic Wide Gamut HDR-Video for the Evaluation of Tone Mapping Operators and HDR-Displays. Proc. SPIE. vol. 9023

8. Drago, F., Myszkowski, K., Annen, T. and Chiba, N, 2003. Adaptive Logarithmic Mapping For Displaying High Contrast Scenes. Computer Graphics Forum. vol. 22, pp. 419-426.

9. Ebner, F., Fairchild, M. D., 1998. Development and Testing of a Colour Space with Improved Hue Uniformity. Proc. IS&T 6th Colour Imaging Conference. pp. 8–13.

10. Dolby, 2014. Dolby Vision White Paper. http://www.dolby.com/us/en/technologies/dolby-vision/dolby-vision-white-paper.pdf

11. Mantiuk, R., Efremov, A., Myszkowski, K. and Seidel, H.-P., 2006. Backward Compatible High Dynamic Range MPEG Video Compression. ACM SIGGRAPH 2006 pp. 713-723.