HOW TONE MAPPING INFLUENCES THE BIT-RATE AND THE BIT DEPTH OF CODED SEQUENCES

Lenzen L., Hedtke R., Christmann M.

RheinMain University of Applied Sciences, Germany

ABSTRACT

In the last few years, High Dynamic Range (HDR) has made major steps forward to become the next big broadcast technology. It is generally accepted that HDR will need a higher bit-rate because of its minimum quantisation and the fact that the images have much more details in the highlights and shadows. However, it is insufficiently taken into account that most of these details will be preserved when performing an HDR down-conversion using tone mapping. It remains unclear how the SDR (Standard Dynamic Range) bit-rate is influenced by HDR production. Therefore, PSNR measurements are performed and a detailed explanation of the reasons is given. Moreover, it is not known for certain if the strong manipulation of the luminance component at down-conversion will produce artefacts like banding. Information on the incoming bit-depth of the HDR with subsequent tone mapping is derived from a viewing test.

INTRODUCTION

One of the main difficulties of introducing HDR is compatibility with the SDR-TV-sets already installed. To benefit from the higher dynamic range of modern TV cameras and to achieve better picture quality on legacy TV sets, tone mapping techniques such as the Sectional Tone Mapping introduced at IBC2016 [1], can be used. An intense manipulation of the luminance component in response to neighbouring pixels, enables a reproduction of a high scene contrast range on SDR displays. Compared to that of today's SDR images, the down-converted pictures show much more detail in the shadows and highlights. This leads to two important issues which will be considered in this paper:

First: An SDR picture after tone mapping, which was down-converted from HDR, could require a higher bit-rate to cope with the increased detail in the highlights and shadows. The bit-rate could even be higher than with PQ (Perceptual Quantizer) or HLG (Hybrid Log Gamma) signals [2] because of a non-perfect bit-allocation at gamma. This could not only be important for SDR-TV, but also for HDR-TV when using the base- and the enhancement-layer concept. In this case, the base-layer is designed to replace the existing SDR picture – but it is not the same for the aforementioned reasons.

Due to the complexity of coding algorithms, it is difficult, if not impossible, to predict quantitative results. When thinking about an interframe coding structure, the bit-rate of the I-frames could increase significantly. But at P- or B- frames, an opposite effect could be observed due to the higher homogenization over time e.g. no changes in aperture. Therefore, motion-compensation can be more effective. In this paper, the results of objective and subjective tests of different test scenes and compression rates are given, that show which parameters in the picture influence the compressed bit-rate.

Second: The Perceptual Quantizer (PQ) is based on the Barten ramp [3] modelling the contrast sensitivity function of the human visual system (HVS) to get the best possible bit-allocation and to avoid visible artefacts, like banding. However, tone mapping adds another nonlinear function to the PQ curve so that the size of the steps and the OETF (Opto-electronic Transfer Function) changes can lead to undesirable effects, especially when encoded with 8 bits. The test results reveal the critical parts in the picture and the corresponding compression rate at which these artefacts are recognized by the viewers.

RELATED WORK

There are only a few relevant studies investigating the influence of HDR on the bit-rate. The most relevant one is by Litwic *et al* [4] where four different test sequences were used. They were graded by subjective test using SDR and HDR displays. Afterwards, they were coded using 1, 3 and 5 Mbit/s in 10-bit depth. In a subjective evaluation, the reference was compared with the coded version and the difference in Mean Opinion Score (MOS) was measured. In this way, the problem concerning different luminance levels when directly comparing SDR to HDR, was avoided.

At the low bit-rate (1 Mbit/s) the MOS score was rather high for SDR whereas HDR performed better for the high bit-rate (5 Mbit/s), but there was a high dependency on the content. Summing up all three different bit-rates, HDR achieved a higher score in two sequences, an equal score in one, and a lower score in another sequence. The authors conclude that in the case of contents with low luminance levels, increasing the nonlinearity of the EOTF has a positive impact on picture quality and significantly removes banding because of its smaller steps in the dark area. For high luminance levels, a opposite effect could be observed.

However, to conclude that HDR is more efficient for relevant broadcast bit-rates, would be a misapprehension because in this case SDR was down-converted from HDR, even if it is done in a manual way. Therefore, SDR has a higher picture quality than in (live) broadcast today.

Mantiuk *et al* [8] measured the overhead by using an enhancement-layer which is called 'residual layer' in the paper. Compared with the base-layer, the bit-rate increases by about 20-25%. But there is no comparison betwee the base-layer and today’s broadcast signal.

Fröhlich *et al* [5] investigated how 12-bit images can be mapped to 10-bit most skilfully to fit them into current image storage and transmission pipelines. In this context, it becomes evident that the necessary bit-depth has a nearly perfect correlation with the noise. Adding a slight dither could help to reduce the requirements in bit-rate without negatively affecting the image quality. However, most natural sequences will have some noise already so that 10 bits are sufficient. In this paper, we want to verify if this statement also applies when the signal is used for tone mapping.

EXPERIMENT: BIT-RATE

test sequence	fps	duration	contrast range	motion	av. luminance
football_longshot	50p	10.00 s	9.10 f-stops	0	80.6 cd/m^2
allianz_arena	50p	10.00 s	7.00 f-stops	4	739.9 cd/m^2
counterattack	50p	27.76 s	7.49 f-stops	8	173.5 cd/m^2
football_fans	50p	30.00 s	9.61 f-stops	7	198.7 cd/m^2
bistro_2 [6]	25p	25.92 s	13.96 f-stops	3	5.4 cd/m^2
bistro_3 [6]	25p	6.80 s	11.26 f-stops	4	19.7 cd/m^2
fishing_longshot [6]	25p	33.96 s	10.55 f-stops	6	7.7 cd/m^2
cars_fullshot [6]	25p	17.68 s	11.08 f-stops	5	17.1 cd/m^2

Table 1 – An overview of the test sequences used

In our experiment, several test sequences are used which differ: in their time duration; their estimated relevant scene contrast range; their motion inside the picture; and in their camera motion (pan). A closer look at these scene-describing parameters will be given later on. An overview can be found in Table 1.

All the test sequences were available in the openEXR file format dealing with the highest dynamic range possible. From these linear luminance values, four different versions were generated, namely:

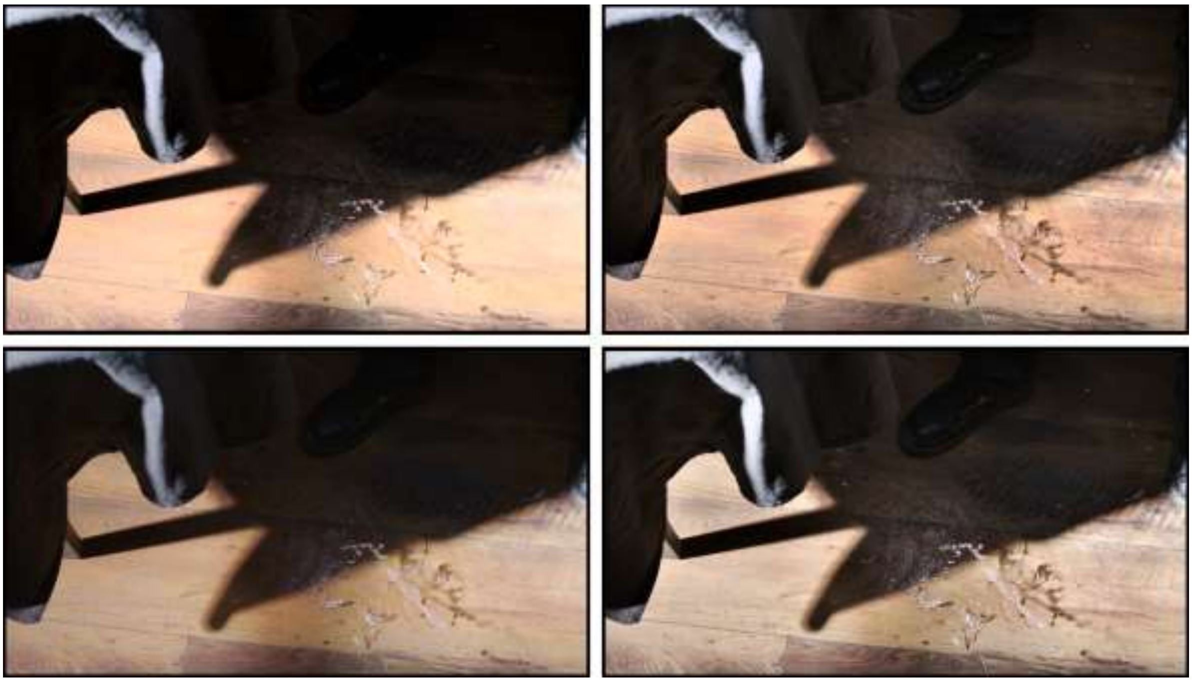

Figure 1 – From top left to bottom right: TV simulation, EVI+Gamma, HLG simulation and EVI+HLG showing bistro_03

TV simulation: As the name suggests, this version should simulate today's SDR-TV look dealing with a dynamic range of about 8.5 f-stops which is more of an upper limit. In scenes with big differences in exposure due to a camera pan, the aperture is controlled in a slightly visible way.

EVI+Gamma: Like the TV simulation the EVI+Gamma version is based on the gamma curve and therefore designed to be played on SDR screens. However, this time, tone mapping is used as it was defined in [1].

HLG simulation: The third version represents the HDR broadcast using HLG. Compared to the TV simulation, a dynamic range of over 12 f-stops is preserved. In scenes with big differences in exposure due to a camera pan, the aperture is controlled in a slightly visible way.

EVI+HLG: Although HLG is created to be backward-compatible with SDT-TV-sets, the look is not ideal. Compared with today's broadcast images, these suffer from low local contrast. So, we combined the EVI algorithm with HLG leading to a better local contrast and a slight homogenization. In consequence, the image has a higher compatibility with SDR, but also produces improvements at HDR. Compared with HLG, there is more spatial information which could result in a higher bit-rate.

A visible comparison of all the images can be found in Figure 1. For HEVC coding the HM 16.3 software [7] was used. The versions using gamma were set to 8-bit, the versions using HLG were set to 10-bit. In the first round, the bit-rate was fixed to 10 Mbit/s by looking at the PSNR. In the second round, the Quantisation Parameter (QP) was fixed by looking at the changing bit-rates. As with Litwic [4] the coded signal is only compared to its reference. In Figure 2 the differences between the four versions as well as the different test sequences can be seen.

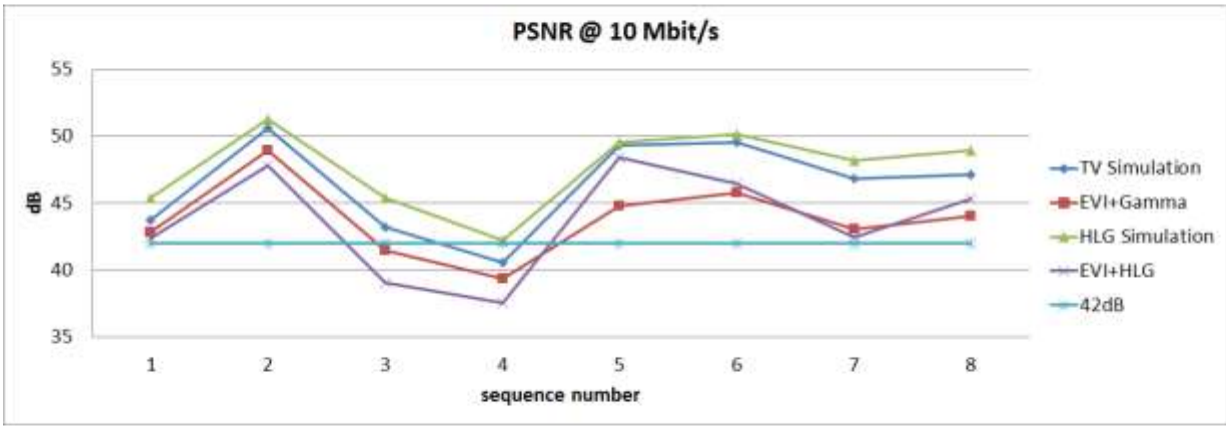

Figure 2 – the PSNR of the four different versions plotted against all sequences

It becomes obvious that EVI has a lower PSNR compared with the TV and the HLG simulation. Checking the two gamma versions the mean difference is 2.5 dB. The largest difference is 4.6 dB and the smallest 0.9 dB. But only some of the EVI versions fall below the important 42 dB threshold. On average, the differences at HLG are 1.5 dB larger. Surprisingly the PSNR at HDR could be higher than at SDR because the better bit allocation seems to compensate the 10-bit.

However, it would be a misinterpretation to conclude that the TV simulation has a better overall picture quality, because the coded versions were always compared with their own reference. Consequently, in the first instance, we can only say that the EVI versions are more difficult to compress due to a higher local contrast and due to their better scene contrast reproduction. Although of course, the second question is whether the mentioned positive effects of tone mapping will be equalised if you don't want to spend more bit-rate. We measured 40% on average for our case to compensate the mentioned 2.5 dB. Such a similar question was raised at High Frame Rate (HFR) where a smooth motion and a sharper image were created. It could be shown by performing viewing tests that the reduced PSNR has nearly no impact on the MOS scale [9]. Unfortunately, a viewing test to verify this assumption has not yet been possible due to time constraints. However, due to the fact that most sequences stayed above the 42 dB threshold, where visible differences can hardly been detected by subjective tests, it is very plausible.

In a second step, we tried to analyse the influence of tone mapping on video coding more deeply. Therefore, we split the average bit-rate (still limited to 10 Mbit/s) into the I- and B-frames. The results are given in Figure 3. For all sequences it can be observed that the bit-rate of the I-frames at EVI increases significantly, whereas the bit-rate of the B-frames decreases slightly. The bit-rate ratio between the I- and B-frames changes respectively. Keeping with the hypothesis, EVI images are more difficult to compress but have a higher correlation over time. Unfortunately, the first effect has more impact on bit-rate.

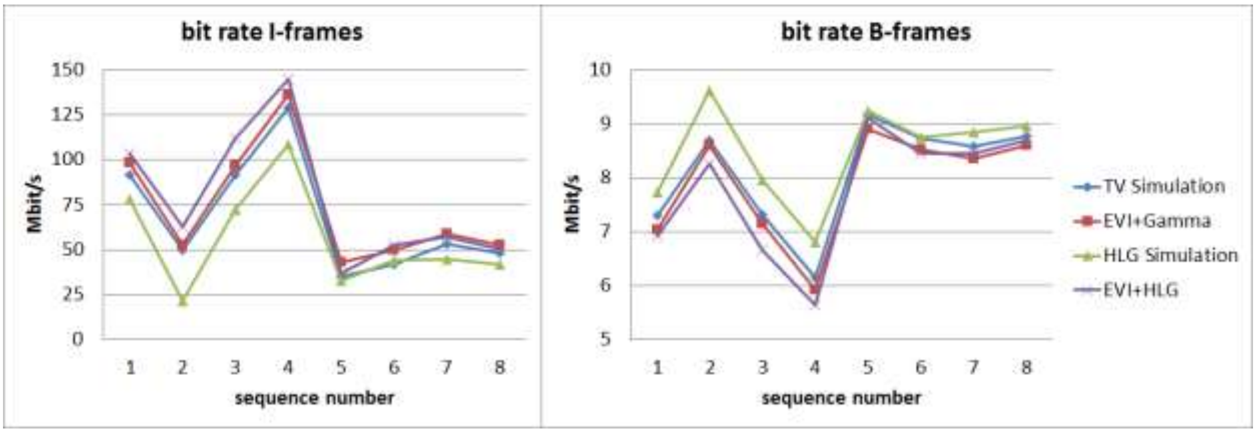

Figure 3 – the average bit-rate of the I- and B-frames plotted against all sequences

In a third step, we tried to find correlations between scene-describing parameters and the previously mentioned effect in the bit-rate ratio. That is why we characterised the sequences regarding to their contrast range, their motion and their average luminance as shown in Table 1.

For the contrast range calculation, the analysis results of EVI were used. One of the main aims of EVI is to detect the highest and lowest luminance values of an image that are really relevant for the viewers. So, we overcame the problem by taking the highest and lowest luminance values which are distorted by outliers and noise as shown in [10]. The contrast range had been calculated on every frame and was averaged afterwards. For the motion, four factors - namely pan speed, pan duration, object speed and object size were calculated and weighted resulting in an expert value between 0 (no movement) and 10 (very high movement). For the brightness, the average luminance of the scene was measured using the logarithmic mean value, to see if a correlation could be found as in [4].

In Figure 4 every point represents a sequence. On the y-axis the changed ratio between the bit-rates of the I- and B-frames is given in percentiles, whereas on the x-axis the scene-describing parameters are scaled. Dealing with an ideal correlation, the points should be spread on a diagonal line. The correlation coefficient measured was 0.91 for the contrast range, -0.59 for the average luminance and -0.33 for the motion factor. Accordingly, it can be deduced that the contrast range of the scene is the most important factor for the increasing complexity of the EVI I-frames and therefore also for the reduced PSNR. At the same time, the result shows how much contrast range can be reproduced in SDR using tone mapping. A possible explanation for the negative correlation with the average luminance could be that very bright scenes tend to have a lower contrast range.

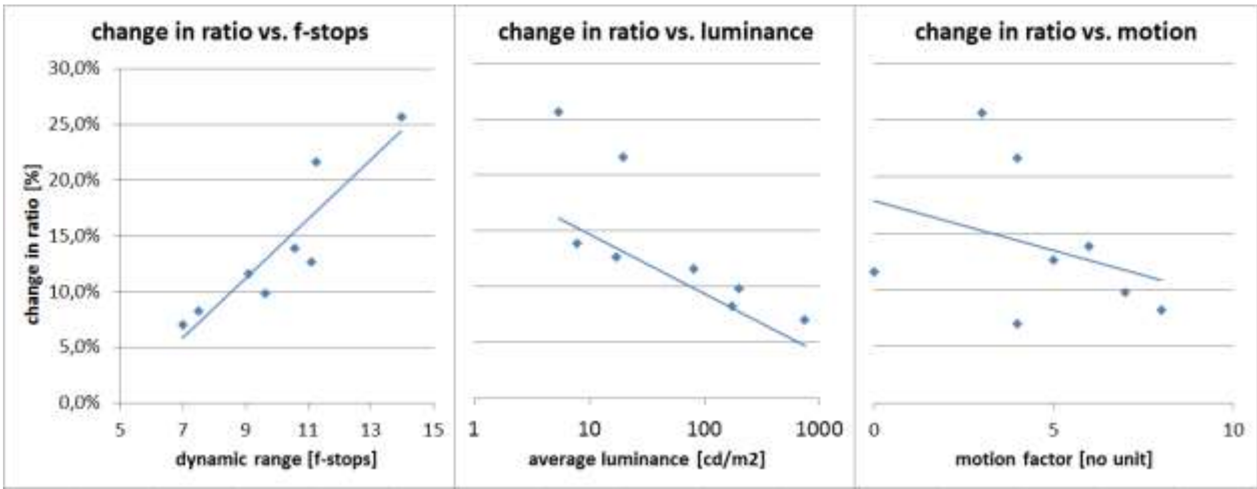

Figure 4 – correlations with scene-describing parameters

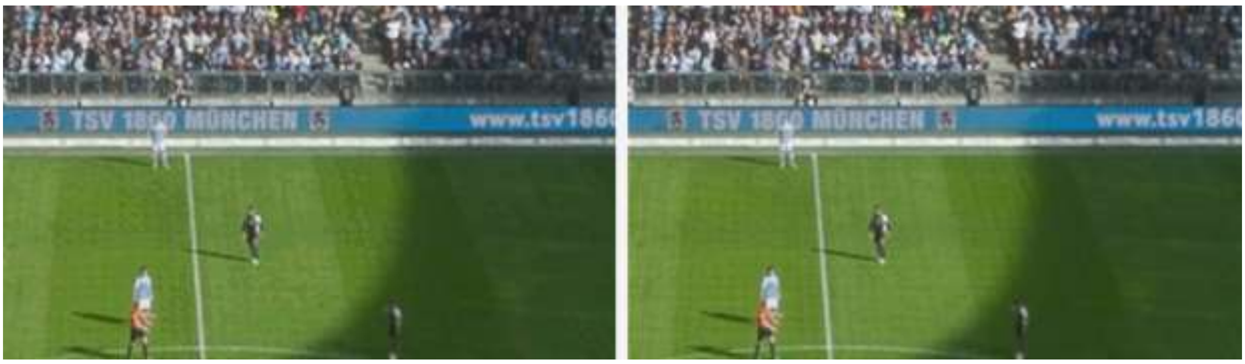

Figure 5 – A greatly magnified image from the sequence football_longshot. The input signal on the left side has 6 and on the right side 12 bits before tone mapping.

EXPERIMENT: BIT-DEPTH

In this experiment, the effect of different bit-depths at the recording side was evaluated. Therefore, three test sequences (football_fullshot, cars_fullshot, fishing_longshot), which were scaled in linear openEXR data, were transformed into PQ first. Afterwards, the bit-depth was limited to 6, 8, 10 and 12 bits. As a last step, the tone mapping was performed by inverting the PQ and a 10-bit signal with EVI+HLG, was generated. Two examples can be found in Figure 5.

In the viewing test, subjects were asked if they could distinguish between the different versions. The different bit-depths were shown in ascending order with a short grey blank. The question was how many bits it would take to fall below a Just Noticeable Difference (JND). For a more meaningful result, the experiment was repeated with three different types of display namely Panasonic TX-65dxw904 (SDR mode), Panasonic TX-65dxw904 (HDR mode) and Sony BVM-E250A. 30 people participated in this test.

For all sequences and all displays, only one of the participants said that he could distinguish between 10 and 12 bits. The averaged values can be found in Figure 6. They span a range from 8.5 to 9.4 but without a clear tendency based on the sequence or the display. This can also be seen in the relatively large confidence intervals. All things considered, the values are relatively low and it can be concluded that a 10-bit production is sufficient for a local or sectional tone mapping.

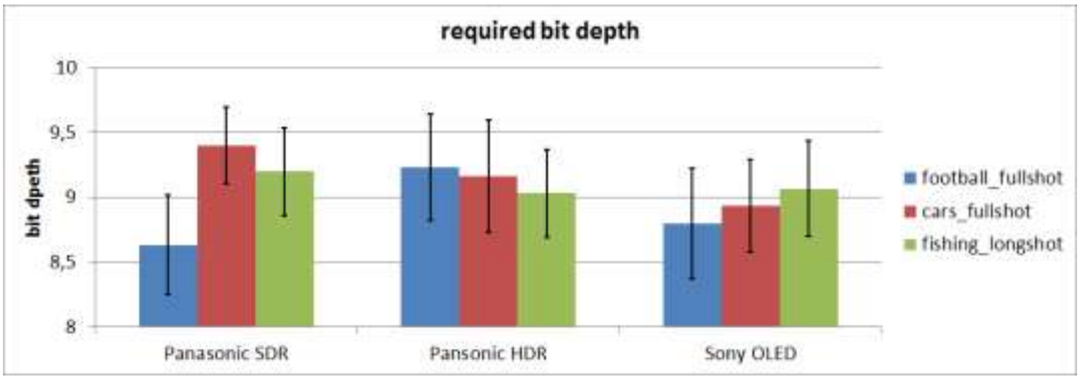

Figure 6 – The averaged results for the required bit depth at tone mapping

CONCLUSION AND OUTLOOK

In this paper, the influence of tone mapping on the bit-rate and the bit-depth has been analysed. The evaluation has shown that using a powerful HDR down-conversion leads to a fundamental different SDR signal compared to today. Although the convergence in time has increased and the B-frames needed a lower bit-rate, the I-frame encoding became much more complex negating the effect. Dealing with more details and a higher local modulation, led to an increasing bit-rate and a decreasing PSNR. This effect correlates almost perfectly with the scene contrast range. However, the positive impact of tone mapping on the overall subjective picture quality will be larger than a change of about 2.5 dB at gamma, since the 42 dB threshold (where no differences can be subjectively detected), is mostly achieved. Surprisingly, the PSNR at HDR could be higher than at SDR because the better bit allocation seems to compensate for the 10 bits.

In the case of the bit-depth, a clear recommendation for production can be given. When using 10 bits in conjunction with a HDR transition curve (e.g. PQ), no banding can be seen, even when performing tone mapping and manipulating the luminance values.

To put it all in a nutshell, it could be concluded that the bit-rate and the bit-depth are not hurdles in bringing HDR to the consumers' homes. We hope that these findings will motivate broadcasters to make HDR productions even if they cannot distribute them directly. HDR and tone mapping is a powerful combination.

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.

REFERENCES

1. Lenzen, L., 2016. HDR for legacy displays using Sectional Tone Mapping. Proceedings of the IBC conference 2016, September 2016

2. ITU-R, BT.2100:0 - Image parameter values for high dynamic range television for use in production and international programme exchange

3. Barten, P., 1992. Physical model for the contrast sensitivity of the human eye. SPIE/IS&T, pp. 57-72.

4. Litwic, L., Baumann, O., White, P. and Goldman, M., 2016. Bit-rate Requirements for High Dynamic Range Video. SMPTE Motion Imaging Journal. pp. 52-60.

5. Froehlich, J., Su, G.-M., Daly, S., Schilling, A. and Eberhardt, B., 2016. Content aware quantization: Requantization of high dynamic range baseband signals based on visual masking by noise and texture. ICIP. pp. 884--888.

6. 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

7. Joint Collaborative Team on Video Coding (JCT-VC), 2015. HM Reference Software, https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.2+SCM-3.0rcl/

8. 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.

9. Lenzen, L., Hedtke, R., 2015. HFR für UHDTV – Wie stark muss die Bildwechselfrequenz steigen?. FKT Fachzeitschrift für Fernsehen, Film und Elektronische Medien. vol.69. pp. 393-398.

10. Hulusic, V., Valenzise, G., Debattista, K. and Dufaux, F., 2017. Robust Dynamic Range Computation for High Dynamic Range Content. Proceedings of the IS&T International Symposium on Electronic Imaging 2017. February 2017.