HAL
open science

REGION-BASED TONE MAPPING FOR EFFICIENT HIGH DYNAMIC RANGE VIDEO CODING

Paul Lauga, Alper Koz, Giuseppe Valenzise, Frederic Dufaux

To cite this version:

Paul Lauga, Alper Koz, Giuseppe Valenzise, Frederic Dufaux. REGION-BASED TONE MAPPING FOR EFFICIENT HIGH DYNAMIC RANGE VIDEO CODING. 4th European Workshop on Visual Information Processing (EUVIP), Jun 2013, France. pp.208. hal-00973500

HAL Id: hal-00973500

https://hal.science/hal-00973500v1

Submitted on 4 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

REGION-BASED TONE MAPPING FOR EFFICIENT HIGH DYNAMIC RANGE VIDEO CODING

Paul Lauga, Alper Koz, Giuseppe Valenzise, Frédéric Dufaux

Télécom ParisTech, Dept. Traitement du Signal et des Images, 75014 Paris, France.
{paul.lauga, alper.koz, Giuseppe.Valenzise, frederic.dufaux}@telecom-paristech.fr

ABSTRACT

In this paper, we address the problem of efficient High Dynamic Range (HDR) video coding. For this purpose, we introduce a new representation which segments the HDR image into low-range and high-range regions which are independently tone mapped. This representation allows preserving more information, leading to a higher visual quality of the reconstructed HDR image.

***Index Terms*—** High Dynamic Range, video compression, tone mapping, Perceptually Uniform encoding, H.264/AVC

1. INTRODUCTION

High Dynamic Range (HDR) [1] is gaining a lot of interests nowadays, thanks to its ability to provide more appealing and immersive images and video sequences. An HDR image (or video) is scene referred, i.e. pixel values describe the real physical luminance values (in cd/m²) present in the scene. Therefore, HDR images can represent accurately information about the luminance in a very high contrast scene from the level of 10^{-6} cd/m² for extreme darkness to 10^{8} cd/m² for a bright sunshine. To represent this entire range, pixels have to be encoded with a higher numerical precision than the conventional imagery, e.g. up to 32 bits, in HDR formats such as OpenEXR or RGBE [F. Kainz, R. Bogart, "Technical Introduction to OpenEXR", Industrial Light & Magic, 18 February 2009.].

As a consequence, HDR images and video sequences require a very large amount of data to be stored and transmitted which is significantly larger than common Low Dynamic Range (LDR) content. Hence, efficient compression is required. However, standard coding technologies, such as H.264/AVC (Advanced Video Coding) codec [2], were not design to handle bit depths higher than 8 bits. Even if more recent extensions such as the High Profile of H.264/AVC can handle up to 14 bits per data sample, this is still not enough for HDR content.

Scalable (or layered) HDR video coding techniques have been proposed in 0[4][5]. In these techniques, a tone mapping operator is first applied on the HDR sequence in order to generate an LDR video. This LDR layer can be encoded using the usual H.264/AVC with 8 bits depth, providing backward compatibility with standard 8 bits displays. Next, the HDR layer can be predicted from the LDR layer in an inter-layer prediction stage. In turn, the prediction error (or residue) can also be encoded using H.264/AVC. Another approach consists in using the High Profile of H.264/AVC. This setup enables to use a higher number of bits (up to 14) to represent luminance values; however, this video content is not directly displayable on current LDR screens. The authors of [6] introduce the Adaptive LogLuv Transform, which is an extension of the LogLuv representation proposed in [7], to represent the luminance on 12 bits and then, to encode it using the High Profile of H.264/AVC. The Adaptive LogLuv Transform [6] compresses the range of HDR values using a logarithm function. This is motivated by the fact that the response for the luminance sensitivity of the Human Visual System (HVS) to the luminance can be approximated by a logarithmic function, as stated by Weber's law [x]. More recently, the Perceptually Uniform (PU) encoding [8] has also been proposed as a more precise model of the HVS.

A limitation of most of the tone mapping algorithms is that one has to choose which part of the luminance range would be reproduced with the best fidelity. Typically, tone mapping algorithms tend to reproduce the luminance range where most of the signal is concentrated as close as possible. The luminance values outside of this range would be mostly saturated. The resulting loss of information can produce annoying distortions when the reconstructed HDR image is viewed on an HDR display.

In this paper, we propose a new tone mapping operator for efficient HDR video compression using H.264/AVC. More specifically, the objective of the proposed tone mapping is to preserve as much information as possible from the HDR data, in order to enable the reconstruction of a high visual quality HDR image. However, this is in contrast with backward-compatible tone mapping operators, which aim at producing the most visually pleasing LDR image to display on LDR devices.

To this end, we propose a tone mapping algorithm which consists of two parts. In the first part, we apply a tone mapping on low luminance values and obtain an image where low values are well preserved and high values are saturated. Similarly, in the second part, we apply in parallel another tone mapping on the high luminance range. Both representations are then combined and

encoded using H.264/AVC High Profile. In this way, we have recorded the whole range of the HDR image into the available bit depth. At the decoder side, the HDR image is reconstructed with the help of a binary mask to identify both the low and the high luminance regions.

The rest of the paper is structured as follows. Section 2 explains in more details Adaptive LogLuv Transform [6], which is used as a base to our proposed method. In Section 3 we present our proposed approach for HDR representation and compression. We assess its performance in Section 4. Finally, we give the conclusions and future work in Section 5.

2. ADAPTIVE LOG LUV TRANFORM

The reference method used for the compression and decompression is the Adaptive LogLuv Transform by Motra and Thoma [6]. This method takes into account the minimum and the maximum luminance of the HDR image to be encoded to map the luminance input to the range $[0, 2^n - 1]$ where n is the bit-depth of the compressed luminance video (selected as 12 for the implementation).. Chrominance channels (u and v) are compressed in 8 bits.

More specifically, the Adaptive LogLuv Transformation is defined by the following steps. First, the RGB channels of the HDR input is transformed into the XYZ color space:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.497 & 0.339 & 0.164 \\ 0.256 & 0.678 & 0.066 \\ 0.023 & 0.113 & 0.864 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

After defining the transition variables, x and y, as :

$$x = \frac{X}{X+Y+Z} \quad (2)$$

$$y = \frac{Y}{X+Y+Z} \quad (3)$$

the Adaptive LogLuv representation is given by

$$u_e = \left\lfloor 410 \frac{4x}{-2x+12y+3} \right\rfloor \quad (4)$$

$$v_e = \left\lfloor 410 \frac{9y}{-2x+12y+3} \right\rfloor \quad (5)$$

$$L = \begin{cases} 0 & \text{if } Y = 0 \\ \lfloor a(\log_2(Y) + b) \rfloor & \text{otherwise} \end{cases} \quad (6)$$

with the parameters

$$a = \frac{2^n - 1}{\log_2 \frac{\max(Y)}{\min(Y)}} \quad (7)$$

$$b = \frac{1}{a} - \log_2(\min(Y)) \quad (8).$$

In the definition of a and b, min(Y) and max(Y) denote the minimum and maximum of the luminance component, Y, respectively. Note that both min(Y) and max(Y) are strictly positive values. The parameters a and b are sent as metadata to the decoder, as they are required to reconstruct the HDR content.

The backward transformation at the decoder is applied as follows:

$$L = \begin{cases} 0 & \text{if } L = 0 \\ 2^{\left(\frac{L+0.5}{a} - b\right)} & \text{otherwise} \end{cases} \quad (9)$$

$$u = \frac{u_e + 0.5}{410} \quad (10)$$

$$v = \frac{v_e + 0.5}{410} \quad (11).$$

The X and Z components are then obtained by

$$x = \frac{9u}{6u - 16v + 12} \quad (12)$$

$$y = \frac{4v}{6u - 16v + 12} \quad (13)$$

$$X = \frac{Yx}{y} \quad (14)$$

$$Z = \frac{Y(1 - x - y)}{y} \quad (15).$$

Finally, the reconstructed RGB values are given by

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 2.690 & -1.276 & -0.414 \\ -1.022 & 1.978 & 0.044 \\ 0.061 & -0.224 & 1.163 \end{bmatrix} * \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (16).$$

3. PROPOSED SCHEME

In this paper, we propose a new representation for HDR video content with the objective to better preserve the HDR information in the luminance signal. A block diagram of the proposed scheme is shown in Figure 1.

The Adaptive LogLuv Transform applies a logarithmic transformation to luminance values prior to the coding. Instead, in our algorithm, we use the PU luminance values [8] which are proposed as a closer approximation to the

luminance sensitivity of the HVS. The PU encoding is obtained from the Contrast Sensitivity Function (CSF), which models the contrast masking of the HVS. For low luminance values, PU encoding is closer to a logarithmic transformation. However, they diverge for higher luminance values.

With both PU encoding or logarithm transformation, a significant range of the luminance is well represented and preserved. However, the dim and the bright regions are typically saturated, causing a loss of information. Moreover, we observe that in most HDR image histograms, the energy of the signal is concentrated in a small range (most of the time in the low values). Nevertheless, a significant amount of information is contained in a much wider range of luminance values.

Taking into account these observations, the input HDR image is segmented into two regions, which will then be processed independently. To perform this binary segmentation, a threshold on the luminance is defined, in such a way to identify the small range which contains most of the signal energy. In this paper, we set this threshold so that 70 % of the luminance values are below the threshold. This threshold is decided empirically in our tests by observing the rate-distortion (RD) curves with respect to the different threholds values.

Applying a tone mapping algorithm on the low-range region, rather than the whole HDR image, will preserve more information. In parallel, another tone mapping algorithm is applied to the high-range regions. Hence, saturation is avoided in bright regions, which are conserving details. More specifically, both the low-range and the high-range regions are PU encoded. In order to stretch the histogram and fully exploit the whole range available, we next apply a logarithm on the PU values.

An example is illustrated in Figure 2. More precisely, the figure shows an original sample frame from the HDR video sequence Sun [9], which has been tone mapped by Adaptive LogLuv in (a), the binary segmentation mask defining the low-range and high-range regions for our proposal in (c), and the tone mapped image obtained by the proposed method in (b) and (d). Note that the objective of the proposed method is to preserve most information from the HDR data. Therefore, the tone mapped image resulting from the proposed method may appear unnatural. However, it carries more information and enables to reconstruct a higher quality HDR image.

Figure 3 (a) shows the histogram of the luminance, in the logarithmic scale for the first frame of Sun. Similarly, the histograms for the case of the Adaptive LogLuv Transform and the proposed method are shown in Figure 3 (b), (c), and (d) respectively. Notice that the proposed method produces a LDR image whose pixel values span all the available dynamic range. This is clearly visible in the histogram in Figure 3(c), which shows that

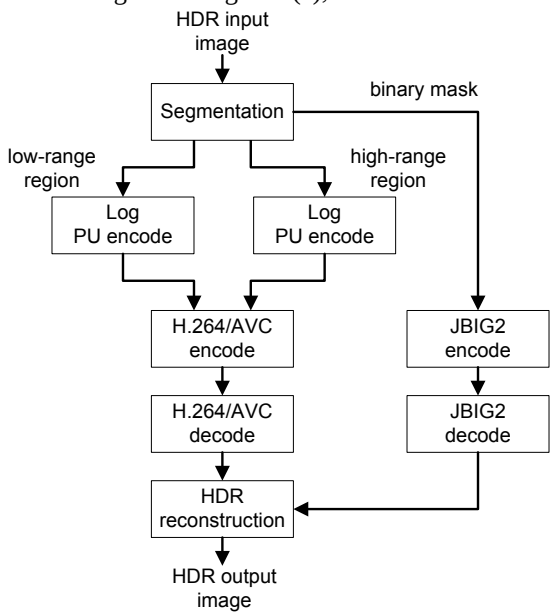


Figure 1: Block diagram of the proposed scheme.

both low and high range values in the LDR images have been mapped from the original HDR content.

After tone mapping, both regions have output values in the range $[0, 2^n - 1]$. To avoid the generation of a spurious edge at the border of the regions (figure 2.b), which would be costly to encode using H.264/AVC, the output values in the high-range region, initially in the range $[0, 2^n - 1]$, are remapped to the range $[2^n - 1, 0]$. In other words, the mapping $x \rightarrow (2^n - 1 - x)$, is applied.

In the following, we detail more formally the proposed segmentation-based tone mapping algorithm. First, the PU encoding is applied to the luminance values (Y) to obtain the luma values:

$$Luma = PUencode(Y) \tag{17}$$

$$\begin{cases} Luma_{Threshold} = PUencode(Y_{Threshold}) \\ L_{\min} = PUencode(\min(Y)) \\ L_{\max} = PUencode(\max(Y)) \end{cases} \tag{18}$$

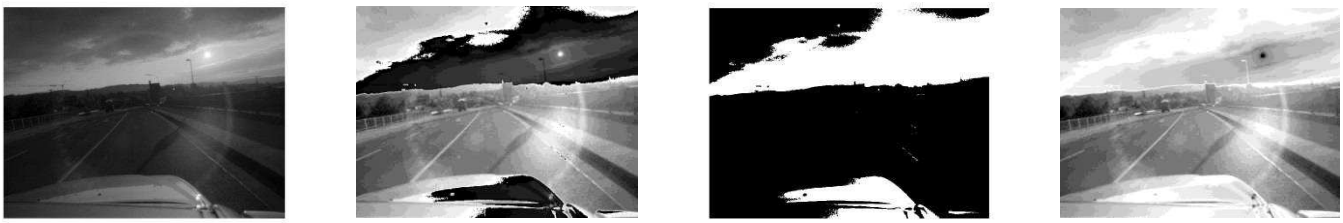
$$\begin{cases} L = \left\lfloor (2^n - 1)\left(\dfrac{\log_2 Luma - \log_2 L_{\min}}{\log_2 Luma_{Thres} - \log_2 L_{\min}}\right)\right\rfloor \text{ if Luma} \le \text{Luma}_{\text{Thres}} \\ L = \left\lfloor (2^n - 1)\left(1 - \dfrac{\log_2 Luma - \log_2 Luma_{Thres}}{\log_2 L_{\max} - \log_2 Luma_{Thres}}\right)\right\rfloor \text{ otherwise} \end{cases}$$

(19).

Compression is then applied on the Luv channels, using H.264/AVC.

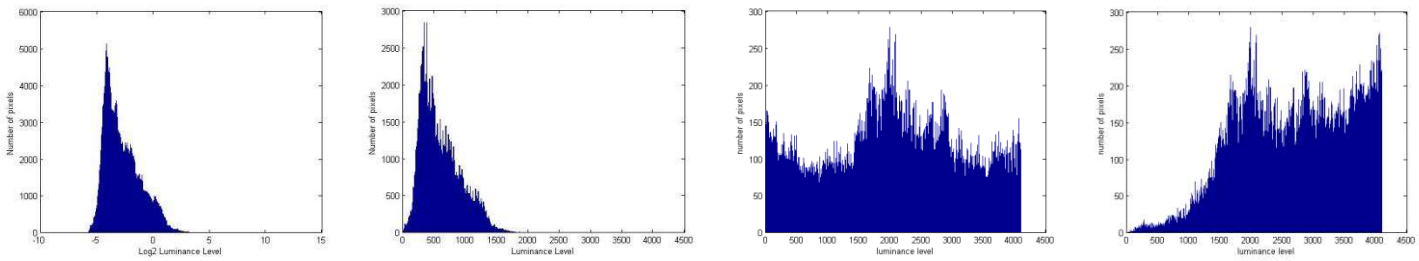
As they are required at the decoder to reconstruct the HDR content, the three parameters L_{min}, L_{max} and $Luma_{Thres}$ are

transmitted as metadata information to the decoder site.



(a) **(b)** **(c)** **(d)**

Figure 2 (a) The first frame of the HDR Sun sequence, which is tone mapped by using Adaptive LogLuv Transform,(b) the same image tone mapped by using our proposal without inversing the second part of the tone mapping curve, (c) the binary mask obtained with our proposal, (d) same image tone mapped using our proposal after inversing the second part of the tone mapping curve.



(a) **(b)** **(c)** **(d)**

Figure 3 (a) Histogram of the HDR image in logarithm domain. The image used is the first frame of Sun and we can see tone mapped versions on figure 2). (b) histogram of the tone mapped image using Adaptive LogLuv. (c) histogram of the tone mapped image using our propose without inversing the second part of the tone mapping curve. (d) histogram of the tone mapped image using our propose after inversing the second part of the tone mapping curve.

A binary segmentation map is also produced, defined by:

$$Mask = (Luma \geq Luma_{Thres}) \qquad (20)$$

This mask, which is required at the decoder side, is compressed using JIBG2 bi-level image compression standard [10] and transmitted to the decoder as illustrated in Figure 1.

At the decoder side, the backward transformation is applied as follows:

$$\begin{cases} Luma = 2^{\frac{\log_2\left(Luma_{Thresh}/L_{\min}\right)}{2^n-1}L+\log_2 L_{\min}} & \text{if Mask} = 0 \\ \text{Luma} = 2^{\frac{\log_2\left(L_{\max}/Luma_{Thres}\right)}{2^n-1}(1-L)+\log_2 Luma_{Thres}} & \text{if Mask} = 1 \end{cases} \qquad (21)$$

$$Y = PUencode^{-1}(Luma) \qquad (22).$$

Then, we use equation (16) to reconstruct RGB values.

4. EXPERIMENTAL RESULTS

To assess the performance of the proposed scheme, we have carried out experiments using the Sun and Tunnel sequences [9]. These are two HDR videos taken from inside a car which is driving on a highway during a sunny day for the first one and in a tunnel for the second one. The frame size of the video sequences is 640 × 480 pixels and the frame rate is 30 fps. In our experiments, we process the first 30 frames of the sequences. The overall dynamic range represented in these sequences is $10^7:1$ for Sun and $10^5:1$ for Tunnel.

We assess the compression performance of the proposed algorithm in comparison to, Adaptive LogLuv method. In both cases, *Y* is coded with 12 bits and *u* and *v* are coded with 8 bits. We use the H.264/AVC High Profile in Intra coding mode and with a fixed set of Quantization Parameters (QP). After decoding, the HDR images are reconstructed using the backward Adaptive LogLuv transform and our proposed scheme. In both cases, metadata are needed for reconstruction. Therefore, these two values have to be transmitted as metadata along with the compressed H.264/AVC stream.

Moreover, we compress the binary mask in our proposed scheme by using JBIG2 [10].

As our proposed scheme preserves more information in the tone mapped image compared to Adaptive LogLuv, it requires a higher bitrate for the same QP. However, it also produced reconstructed HDR images with better quality.

In order to compare the visual quality of the reconstructed HDR images, we use the perceptually-based HDR-VDP metric [11][12][13]. We choose this metric for its relevance to the HVS. The HDR-VDP metric takes as inputs both a processed and a reference HDR image. The output is an image of the same size, which estimates the probability for each pixel in the two HDR images to be seen different by a human observer. From this output image, we extract two numbers: the percentage of the pixels in the image which has a probability higher than 75% and 95 % to be detected. Finally, these percentages are averaged over all frames of the video sequence.

Figure 4 and 5 show the rate-distortion performance of the proposed method, compared to Adaptive LogLuv for Sun and Tunnel sequences. At both 75 % and 95 % probability thresholds, it can be observed that our proposed method consistently outperforms Adaptive LogLuv method.

5. CONCLUSIONS

In this paper we proposed a new scheme for HDR video content compression. It is based on a new representation which conserves more information in the HDR image. More precisely, the input HDR image is segmented into low-range and high-range regions. Each region is then tone mapped using PU encoding and a logarithmic scale. The resulting image is compressed using H.264/AVC.

Rate-distortion results, using the perceptually-based HDR-VDP metric, show the consistently superior performance of the proposed scheme when compared to the Adaptive LogLuv representation.

Future work will aim at an improved temporal coherence of the proposed representation in order to improve interframe coding performance in H.264/AVC. Moreover, we will focus on finding the optimal value for each frame using the HDR histogram.

ACKNOWLEDGEMENTS

This work was partly funded by *Région Ile de France* in the framework of the *NEVEx* project.

REFERENCES

[1] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, “Advanced High Dynamic Range Imaging,” AK Peters (now CRC Press), Natick, MA, 2011.

[2] T. Wiegand, G.J. Sullivan, G. Bjontegaard, A. Luthra, “Overview of the H.264/AVC video coding standard,” IEEE Trans. on Circuits and Systems for Video Technology, 13 (7), pp.560-576, 2003.

[3] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward, and W. Heidrich, “Optimizing a Tone Curve for Backward-Compatible High Dynamic Range Image and Video Compression,” IEEE Transactions on Image Processing, 20 (6), pp. 1558-1571, June 2011.

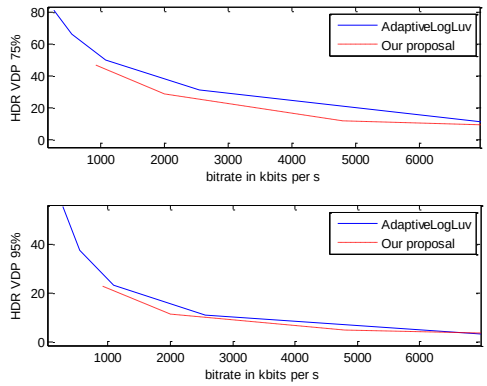


Figure 4: HDR-VDP rate-distortion results for Sun sequence. First row: 75 % probability to be detected; second row: 95 % probability to be detected.

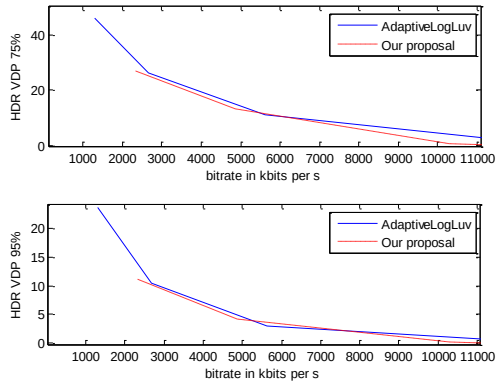


Figure 5: HDR-VDP rate-distortion results for Tunnel sequence. First row: 75% probability to be detected; second row: 95% probability to be detected.

[4] A. Segall, “Scalable Coding of High Dynamic Range Video,” in IEEE Proc. International Conference on Image Processing (ICIP), San Antonio, TX, Sept. 2007.

[5] S. Liu, W.-S. Kim, and A. Vetro, “Bit-depth scalable coding for high dynamic range video,” in SPIE Proc. Visual Communications Image Processing, San Jose, CA, January 2008.

[6] A. Motra, and H. Thoma, “An adaptive LogLuv Transform for High Dynamic Range Video Compression,” in IEEE Proc. International Conference on. Image Processing (ICIP), Hong Kong, September 2010.

[7] G.W. Larson, “Overcoming gamut and dynamic range limitations in digital images,” in Proc. Sixth Color Imaging Conference, 1998.

[8] T.O. Aydin, R. Mantiuk, and H.-P. Seidel, “Extending Quality Metrics to Full Luminance Range Images,” in SPIE Proc. Human Vision Electronic Imaging XIII, San Jose, CA, January 2008.

[9] HDR image Gallery [online]. Available: http://www.mpi-inf.mpg.de/resources/hdr/video

[10] F. Ono, W. Rucklidge, R. Arps, and C. Constantinescu, “JBIG2-the ultimate bi-level image coding standard,” in IEEE Proc. International Conference on Image Processing, Vancouver, BC, Canada, September 2000.

[11] T.O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, “Dynamic Range Independent Image Quality Assessment,” ACM Transactions on Graphics (TOG), 27 (3), August 2008.

[12] R. Mantiuk, S. Daly, K. Myszkowski, and H.-P. Seidel, “Predicting Visible Differences in High Dynamic Range Images – Model and its Calibration,” in SPIE Proc. Human Vision and Electronic Imaging X, San Jose, CA, January 2005.

[13] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, “Visible Difference Predicator for High Dynamic Range Images,” in IEEE Proc. International Conference on Systems, Man and Cybernetics, The Hague, Netherlands, October 2004.

[14] F. Kainz, R. Bogart, “Technical Introduction to OpenEXR”, Industrial Light & Magic, 18 February 2009.