FULLY-AUTOMATIC INVERSE TONE MAPPING PRESERVING THE CONTENT CREATOR'S ARTISTIC INTENTIONS

Gonzalo Luzardo*†, Jan Aelterman*, Hiep Luong*, Wilfried Philips*, Daniel Ochoa† and Sven Rousseaux‡

*imec-IPI-UGent, Ghent University
Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium
†Facultad de Ingeniería en Electricidad y Computación, ESPOL Polytechnic University
Campus Gustavo Galindo Km. 30.5 Vía Perimetral, Guayaquil, Ecuador
‡ Vlaamse Radio -en Televisieomroeporganisatie
Auguste Reyerslaan 52, Brussels, Belgium

Abstract—**High Dynamic Range (HDR) displays can show images with higher color contrast levels and peak luminosities than the common Low Dynamic Range (LDR) displays. However, most existing video content is recorded and/or graded in LDR format. To show this LDR content on HDR displays, a dynamic range expansion by using an Inverse Tone Mapped Operator (iTMO) is required. In addition to requiring human intervention for tuning, most of the iTMOs don't consider artistic intentions inherent to the HDR domain. Furthermore, the quality of their results decays with peak brightness above 1000 nits. In this paper, we propose a fully-automatic inverse tone mapping operator based on mid-level mapping. This allows expanding LDR images into HDR with peak brightness over 1000 nits, preserving the artistic intentions inherent to the HDR domain. We assessed our results using full-reference objective quality metrics as HDR-VDP-2.2 and DRIM. Experimental results demonstrate that our proposed method outperforms the current state of the art.**

I. INTRODUCTION

Dynamic range refers to the ratio between the whitest whites and darkest blacks present at the same time in an image. This is often measured in "stops", which is the logarithm base-2 of this ratio. While conventional display technology is capable of displaying brightness ranges from 1 cd/m^2 to 300 cd/m^2 (nits) and 8 stops of dynamic range, objects captured in sunlight can easily have brightness values up to 10000 (nits). Considering that the human eye can see up to about 14 stops of dynamic range in a single view, clearly conventional display technology is unable to display luminance-realistic images. To improve the dynamic range of an image, the number of bits for its representation needs to be increased. Hence, a wider range of luminance levels can be encoded to produce a higher dynamic range through heightened contrast.

Images associated to display technology with narrow dynamic range and a low brightness have been retroactively called Low Dynamic Range (LDR) images. Furthermore, a High Dynamic Range (HDR) image refers to an image that encodes a higher dynamic range and a larger amount of brightness than a reference LDR image. Recent developments in so-called HDR displays have yielded prototypes capable to reach a peak brightness of more than 1000 nits and 14

stops of dynamic range [1]. Hence an HDR image encodes dark areas darker and bright areas brighter, which gives an increased sense of detail, sharpness, clarity, color, and saturation. Unfortunately, a large amount of video content has been recorded and/or graded in LDR. To display LDR content on an HDR display, an inverse/reverse tone mapping operator (iTMO) is required.

A large number of iTMOs have been proposed in the past few years. Most approaches use a parametrized non-linear curve to expand the luminance intensity values and bit-depth of an input LDR image, increasing in this way its dynamic range. The main drawbacks of these iTMOs are: 1) they need human intervention to decide the most suitable parameters to use for the expansion procedure, 2) the quality of their results decays when expanding LDR images to more than 1000 nits, and 3) they are designed to preserve the appearance of the original LDR image without considering the artistic intentions inherent to the HDR domain.

In this paper, we propose a fully-automatic inverse tone mapping operator, based on mid-level mapping, that does not need any input parameters from the user. Furthermore, it can expand LDR images into HDR with peak brightness over 1000 nits, while preserving the artistic intentions inherent to the HDR domain. This paper is organized as follows: An overview of previous works is given in Section II. Then, the proposed inverse tone mapping algorithm is described in Section III. Experimental results are presented in Section V. Finally, concluding remarks are discussed in Section VI.

II. RELATED WORK

Several methods for automatic inverse tone mapping have been proposed in recent years. Kuo et al. [2] propose a content-adaptive inverse tone mapping algorithm, that consists of two steps: scene-adaptive inverse tone mapping and over-exposure enhancement. The first step includes an inverse tone mapping function as proposed by Schlick [3] which is adjusted depending on the type of scene previously identified. The second step allows reconstructing the truncated information

lost in over-exposed regions. Despite this approach works well in many cases, its main disadvantage is related to the use of an inverse tone mapping function that does not consider the enhancement of aesthetic appeal of the output HDR image, but only the accuracy of the resulting HDR image, which usually appears aesthetically dull.

In more recent studies, Kovaleski and Oliveira [4] propose a reverse-tone-mapping operator for images and videos, which can deal with images with a wide range of exposure conditions. Their technique is based on the automatic computation of a brightness enhancement function (BEF), also called expand-map, that determines areas where image information may have been lost and fills these regions using a smooth function. Like-wise, Masia et al. [5] propose a dynamic inverse tone mapping algorithm based on image statistics. The main premise of this method is that input LDR images are not always correctly exposed. In this way, the authors propose to use a simple gamma curve as inverse tone mapping operator in which the gamma value is estimated from a multi-linear model that incorporates the key value [6], the number of overexposed pixels, and the geometric mean luminance, as input parameters. Likewise, Bist et al. [7] propose a method, based on the conservation of the lighting style aesthetics. As with previous approaches, these authors propose to use a simple gamma correction curve; however, this time the gamma value is computed based on the median of the luminance in the input LDR image. Despite these methods produce good results across a wide range of exposures because they are based on preserving the aspect of the original LDR image, they might not always reflect the artistic intentions intrinsic to the HDR domain. Furthermore, their quality decays with peak brightness over 1000 nits.

In this paper, we propose a fully-automatic inverse tone mapping operator based on mid-level mapping, which allows expanding LDR images to HDR domain with peak brightness over 1000 nits. For this, a simple non-linear expansion function that takes only one parameter, namely the middle-gray value in the output HDR inverse tone mapped image (m_o) controlling its overall brightness, is proposed. This parameter is automatically estimated using simple first-order image statistics. Unlike [5] and [7], our proposed expansion function offers better capabilities to increase the perceived image quality of the resulting HDR image than a simple gamma expansion. Furthermore, it is capable to preserve the content creator's artistic intentions inherent to the HDR domain.

III. PROPOSED EXPANSION FUNCTION

To expand the dynamic range of an LDR image (I_{LDR}) and obtain its inverse tone mapped HDR image counterpart (I_{HDR}), a simple non-linear function is proposed as follows:

$$f(L) = L_w = \frac{L^a}{L^{(ad)}b + c} \quad \text{s.t.} \quad L \in [0,1] \qquad (1)$$

Where L is the display luminance of I_{LDR}, L_w is the expanded luminance of I_{HDR} in the real-world domain, and $L_w \in [0, L_{w,max}]$. The parameter $L_{w,max}$ is the maximum luminance in which I_{LDR} wants to be expanded, and its value

usually depends on the peak brightness of the HDR display where I_{HDR} will be displayed.
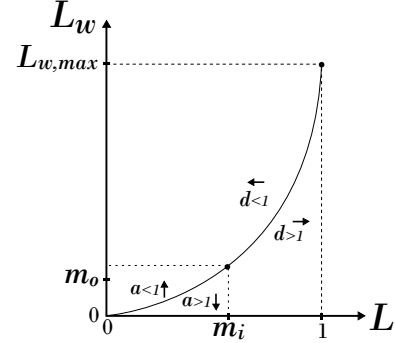

Fig. 1. The shape of the proposed inverse tone mapping function. LDR input values are normalized between 0 and 1. The maximum output HDR value depends on the peak luminance in nits of the HDR display e.g. 6000, or 1 for expressing that we intend to achieve the maximum luminance supported by the display.

Figure 1 shows the shape of the proposed function. The parameters a and d allow controlling the shape of the so-called toe (contrast) and shoulder (expansion speed) of the curve, respectively. This offers the possibility to fine-tune the resulting inverse tone mapped HDR image to improve its perceived image quality, for example by increasing the contrast [8]. The parameters m_i and m_o act as an anchor point for the curve, and they represent the middle-gray value defined for the input I_{LDR} and the middle-gray value desired in the output I_{HDR}, respectively. The middle-gray value is a tone that is perceptually about halfway between black and white on a lightness scale. These parameters enable us to control the overall luminance of I_{HDR}. For this, the parameter m_i is fixed to the middle-gray predefined for I_{LDR} (e.g., 0.214 for sRGB linear LDR images), and m_o is adjusted to the middle-gray value desired in I_{HDR}. This operation can be seen as mid-level mapping operation. A low value for the m_o causes I_{HDR} to become darker, and a higher value brighter.

Considering the following constraints: $f(m_i) = m_o$ and $f(1) = L_{w,max}$. The parameters b and c can be calculated as follows:

$$b = \frac{m_i{}^a L_{w,max} - m_o}{m_o(m_i{}^{ad} - 1)L_{w,max}} \qquad (2)$$

$$c = \frac{m_i{}^{ad}m_o - m_i{}^a L_{w,max}}{m_o(m_i{}^{ad} - 1)L_{w,max}} \qquad (3)$$

As it may be seen, the proposed function in (1) is applied only to the luminance channel in order to leave the chromatic channels of I_{LDR} unaffected. Thus, the color image I_{HDR} is reconstructed by preserving the ratio between the red, green, and blue channels as follows [9]:

$$C_{HDR} = \left(\left(\frac{C_{LDR}}{L} - 1 \right)s + 1 \right)L_w \quad \text{s.t.} \quad s \geqslant 1 \qquad (4)$$

Where C_{LDR} and C_{HDR} denote one of the color channel values (red, green, blue) of the I_{LDR} and I_{HDR}, respectively.

The parameter s is a color saturation parameter and allows to compensate the loss of saturation during the luminance expansion operation. As it can be seen, fixing the saturation (s), contrast (a) and expansion speed (d), the expand operation of I_{LDR} can be performed effortlessly using (1) to (4). Likewise, the proposed function allows changing the peak brightness of the resulting HDR image without affecting its middle tones, which is controlled by the parameter m_o.

IV. FULLY-AUTOMATIC INVERSE TONE MAPPING

Our iTMO is based on a mid-level mapping approach. For this, it uses the proposed expansion function to expand the dynamic range of an input LDR image. Parameters s, a and d are fixed manually; then, the input m_o, value required for the expansion operation, is automatically estimated through a multi-linear function. This function uses simple image-statistics from the LDR image as input parameters to estimate m_o. This section describes how this multi-linear function was built by using multi-linear regression from a training set.

A. Image dataset

A set of 180 LDR-HDR pairs of corresponding frames from the following HDR video sequences were used to create our dataset: Tears of Steel [10], Stuttgart HDR image dataset [11] and a short film created by the Flemish Radio and Television Broadcasting Organization (VRT). These sequences were selected because they contain scenes with a wide range of contrast and lighting conditions. Furthermore, they were professionally graded to HDR with a peak brightness of 6000 nits using the SIM2-HDR47ES6MB display [1] as a reference. For this reason, they are considered as ground truth sequences with the most of the artistic intentions that content creators' consider when they work in the HDR domain. The frames selected to create the dataset were chosen in a systematic way; such that, it contains images with equally distributed contrast and lighting conditions. The LDR counterparts for the HDR frames selected from Tears of Steel were obtained from the project's website. Whereas, the other sequence LDR counterparts were obtained by professionally grading. For each LDR image in the dataset, simple first-order image statistics, most of them computed on the luminance (L), were extracted. These image statistics include the average (L_{avg}), variance (L_{var}) and median (L_{med}), and those included in Table I. Images in the dataset were normalized into the $[0, 1]$ range and linearized by using a gamma correction curve with a power coefficient value $\alpha = 2.2$, hence L was computed as follows:

$$L = 0.213R + 0.715G + 0.072B \quad \text{s.t.} \quad R, G, B \in [0, 1] \quad (5)$$

Where R, G, and B are respectively red, green and blue channel of the linearized input image. In Table I, L_{min} and L_{max} refer to the minimum and maximum luminance value, respectively. These values were computed after excluding 5% of the pixels, on the dark and bright sides, which were considered as outliers. N is the total number of pixels in the image without outliers, and N_{ov} is the total number of overexposed pixels. Overexposed pixels refers to those pixels

where at least one color channel is greater than or equal to $(254/255)$.

TABLE I
FIRST-ORDER IMAGE STATISTICS COMPUTED.

Geometric mean	$L_h = exp((1/N) \sum_{i=1}^{N} \log(L_i + \varepsilon))$ s.t. $\varepsilon = 0.0001$
Key value	$K = \dfrac{\log(L_h) - \log(L_{min})}{\log(L_{max}) - \log(L_{min})}$
Kurtosis	$K_u = \dfrac{(1/N) \sum_{i=1}^{N} (L_i - L_{avg})^4}{L_{var}^2}$
Skewness	$S_k = \dfrac{(1/N) \sum_{i=1}^{N} (L_i - L_{avg})^3}{L_{var}^{3/2}}$
Contrast	$C = \sqrt{(1/N) \sum_{i=1}^{N} (\log(L_i) - \log(L_{avg}))^2}$
% of saturated pixels	$P_{ov} = \dfrac{N_{ov}}{N}$

B. Subjective study

Six different middle-gray values (m_o) to inverse tone map each LDR were obtained in a subjective study. Six non-experts participants, between 25 and 40 years old, took part in this study. They were asked to tweak m_o until they found the subjective best match of the inverse tone mapped HDR with its corresponding professionally graded HDR image (ground truth) in the pair. Both, the luminance channel of the inverse tone mapped image and the ground truth, were displayed on a SIM2 HDR display side by side to facilitate the matching task and prevent that the differences on saturation between the two images affect our results. The contrast (a) and expansion speed (d) were set manually in 1.25 and 4.0 to increase the contrast and minimize that diffuse whites in the input produce artifacts in the output, respectively. Likewise, the saturation (s) was manually fixed in 1.25 as suggested in [7], and $L_{w,max}$ in 0.67 to specify that we want to reach only 67% of the maximum luminance of the SIM2 display (around 4000 nits). Using the value of m_o selected by the participant as the input parameter, HDR inverse tone mapped images were computed in real-time using the proposed function described in Section IV. A total of 1080 m_o values (180 values per participant) were collected. To avoid any light interference, this study was conducted in a light-controlled room with dim lights.

C. Multi-linear regression

Multi-linear regression was used to predict the middle-gray value in the output (m_o, dependent variable) from image first-order image statistics (independent variables). First, an outlier analysis was performed. In this, all observations that were deemed as extreme values were removed. Given the IQR (Interquartile Range) of the samples in a group (values of m_o obtained for the same pair of images), an extreme value is a

data point that is below $Q_1 - 1.5IQR$ or above $Q_3 + 1.5IQR$. Then, a multivariate outliers detection using Mahalanobis distance was performed [12]. At the end of the outlier analysis, 47 input values were removed from the dataset. Consequently, a multi-linear regression was carried out using a stepwise regression method for variable selection. In this method, a variable is entered into the model whether the significance level of its F-value is less than a so-called entry value, and it is removed if the significance level is greater than the removal value. For this, values of 0.05 and 0.10 were used as entry and removal value. After we applied this method, a multi-linear model was obtained. This model indicated that linear combination of the geometric mean (L_h), contrast (c_o), and the percentage of overexposed pixels (P_{ov}) were significantly related to the middle-gray output value (m_o), $R^2 = 0.702$, $F(3, 126.673)$, $p < 0.01$, trough the following equation:

$$m_o = 0.017 + 0.097L_h + 0.008C - 0.028P_{ov} \qquad (6)$$

As it can be seen on (6) and as was expected, the geometric mean (L_h), a parameter that represents perceptual overall brightness of the input image, has the highest positive relation with the middle-gray value in the output (m_o), followed by the contrast (C). This means that LDR images with a high perceptual brightness value tend to expand into HDR images with high brightness. This is also controlled by the contrast, in the sense that LDR images with high perceptual brightness but with lower contrast are not expanded to very bright HDR images. Another interesting finding was that the negative relation between percentage of overexposed pixels (P_{ov}) and m_o acts as a control parameter that avoids that an overexposed image is expanded to a very bright HDR image.

V. RESULTS

For the quality assessment of our inverse tone mapping method, we decided to use the following full-reference quality metrics: Calibrated Method for Objective Quality Prediction (HDR-VDP-2.2) [13], using the HDR image (ground truth) as a reference; and the Dynamic Range Independent Image Quality Assessment (DRIM) [5], using this time the LDR image as a reference. We compare our method with the following most recent full-automatic inverse tone approaches proposed by Kovaleski and Oliveira [4], Masia et al. [5], and by Bist et al. [7]. The sequences Stuttgart, Tears of Steel, VRT, and additionally the DML-HDR dataset [14], were used for the assessment.

HDR-VDP-2.2 allowed us to compare our resulted HDR inverse tone mapped HDR image with the professionally graded HDR image. For this, we compute the Quality Scores (Q-score) and Probability of Detection Maps (PMap). Figure 2 shows the average Q-score obtained by HDR-VDP-2.2 for each video sequence. As can be seen, our method produces on average better Q-score results in comparison with the other evaluated methods; therefore, it can better preserve artistic intentions than the other studied methods. PMaps and Q-scores of a part of the results are included in the supplementary material. Likewise, DRIM allowed us to compare our results
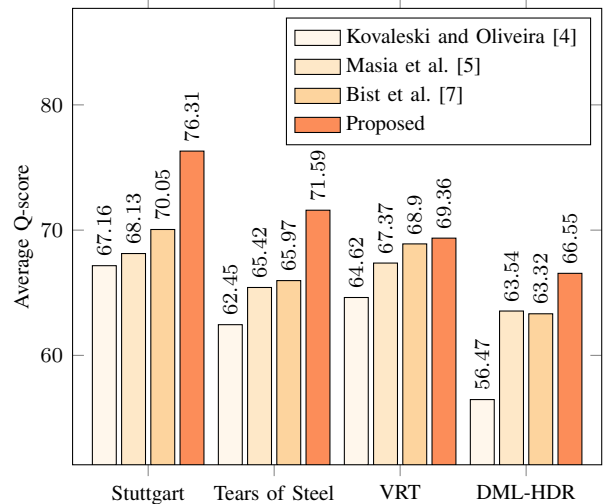

Fig. 2. Average Q-scores obtained by HDR-VDP-2.2 for each video sequence.

with the LDR in terms of structural changes (lost of visible contrast, amplification of invisible contrast and reversal of visible contrast) between the LDR image and the HDR image obtained by our proposed method. For this, DRIM generates so-called distortion maps, which show areas where the visible contrast is lost (green color), distorted (red color) or improved (blue color). Distortions maps obtained reveal that the proposed method causes far less reverse and loss of the contrast, and much more contrast amplification than the other tested methods. All these help to increase the perceived quality of our method. Figure 3 shows some distortion maps obtained in the assessment. More results obtained by DRIM are included in the supplementary material. Supplementary material is available at http://telin.ugent.be/~gluzardo/midtones-itmo

VI. CONCLUSIONS

We proposed a fully-automatic inverse tone mapping operator. We have provided experimental results demonstrating the effectiveness of the proposed method using two full-reference objective metrics to compare our method with most recent approaches. Compared to other methods, ours provides better image quality scores. Unlike other approaches, our method is focused on mid-level mapping that promises to effortlessly adjust to different "versions" of HDR, i.e. different levels of peak brightness. Initial experimental results are very promising, but an exhaustive research into effects of this parameter across a large range of peak brightness levels is an option for future work. Regarding this, it is important to know that using greater peak brightness leads to the generation of false contours. However, this can be mitigated using a method to remove false contours in HDR images [16]. Furthermore, because professionally-graded HDR images were used as ground truth during the subjective study, our proposed method better preserves the artistic intentions inherent to the HDR domain in a better way, which is reflected in better Q-scores obtained by HDR-VDP-2.2. Subjective evaluation of

Q-score = 67.56 | Q-score = 69.36 | Q-score = 69.34 | Q-score = 77.38

Q-score = 59.94 | Q-score = 65.92 | Q-score = 66.20 | Q-score = 73.16

Kovaleski and Oliveira [4]   Masia et al. [5]   Bist et al. [7]   Proposed method
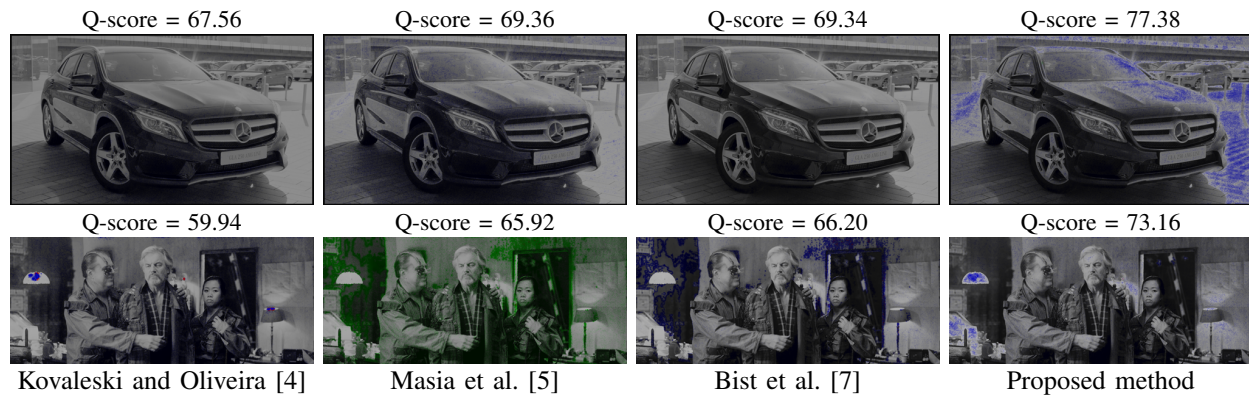
Fig. 3. Comparing our results with most recent automatic iTMO using DRIM [15]. Green, blue and red pixels in the images identify loss of visible contrast, amplification of invisible contrast and contrast reversal, respectively. Note that green and red pixels are considered as a negative result and blue pixels as a positive. Q-score value obtained by HDR-VDP-2.2 is shown above each image. *These images are best viewed in the electronic PDF version.*

the performance of proposed inverse tone mapping algorithm that confirms our objective results will be carried out as the next step in our research.

ACKNOWLEDGEMENT

REFERENCES

[1] (2017) SIM2 HDR display - HDR47ES6MB. [Online]. Available: http://hdr.sim2.it/hdrproducts/hdr47es6mb

[2] P. H. Kuo, C. S. Tang, and S. Y. Chien, "Content-adaptive inverse tone mapping," in *2012 Visual Communications and Image Processing*, Nov 2012, pp. 1–6.

[3] C. Schlick, "Quantization techniques for visualization of high dynamic range pictures," in *Photorealistic Rendering Techniques*. Springer, 1995, pp. 7–20.

[4] R. P. Kovaleski and M. M. Oliveira, "High-quality reverse tone mapping for a wide range of exposures," in *2014 27th SIBGRAPI Conference on Graphics, Patterns and Images*, Aug 2014, pp. 49–56.

[5] B. Masia, A. Serrano, and D. Gutierrez, "Dynamic range expansion based on image statistics," *Multimedia Tools and Applications*, vol. 76, no. 1, pp. 631–648, Jan 2017.

[6] A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bülthoff, "Do hdr displays support ldr content?: A psychophysical evaluation," in *ACM SIGGRAPH 2007 Papers*, ser. SIGGRAPH '07. ACM, 2007.

[7] C. Bist, R. Cozot, G. Madec, and X. Ducloux, "Tone expansion using lighting style aesthetics," *Computers & Graphics*, vol. 62, pp. 77–86, 2017.

[8] A. G. Rempel, M. Trentacoste, H. Seetzen, H. D. Young, W. Heidrich, L. Whitehead, and G. Ward, "Ldr2hdr: On-the-fly reverse tone mapping of legacy video and photographs," in *ACM SIGGRAPH 2007 Papers*, ser. SIGGRAPH '07. New York, NY, USA: ACM, 2007.

[9] R. Mantiuk, A. Tomaszewska, and W. Heidrich, "Color correction for tone mapping," in *Computer Graphics Forum*, vol. 28, no. 2. Wiley Online Library, 2009, pp. 193–202.

[10] (2017) Tears of Steal | Open Mango Movie Project. [Online]. Available: https://mango.blender.org

[11] 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," 2014.

[12] P. Filzmoser, R. G. Garrett, and C. Reimann, "Multivariate outlier detection in exploration geochemistry," *Computers & geosciences*, vol. 31, no. 5, pp. 579–587, 2005.

[13] M. Narwaria, R. Mantiuk, M. P. Da Silva, and P. Le Callet, "Hdr-vdp-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images," *Journal of Electronic Imaging*, vol. 24, no. 1, p. 010501, 2015.

[14] A. Banitalebi-Dehkordi, M. Azimi, M. T. Pourazad, and P. Nasiopoulos, "Compression of high dynamic range video using the hevc and h. 264/avc standards," in *Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine), 2014 10th International Conference on*. IEEE, 2014, pp. 8–12.

[15] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic range independent image quality assessment," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, p. 69, 2008.

[16] G. Luzardo, J. Aelterman, H. Luong, W. Philips, and D. Ochoa, "Real-time false-contours removal for inverse tone mapped hdr content," in *Proceedings of the 2017 ACM on Multimedia Conference*, ser. MM '17. New York, NY, USA: ACM, 2017, pp. 1472–1479.