Spatio-Temporal Constrained Tone Mapping Operator for HDR Video Compression

Cagri Ozcinar, Paul Lauga, Giuseppe Valenzise, Frédéric Dufaux

HAL Id: hal-01806470

https://hal.science/hal-01806470

Submitted on 10 Jan 2020

Spatio-Temporal Constrained Tone Mapping Operator for HDR Video Compression

Cagri Ozcinar[a,b], Paul Lauga[b], Giuseppe Valenzise[c], Frédéric Dufaux[c]

[a]*V-SENSE, School of Computer Science and Statistics (SCSS), Trinity College Dublin (TCD)*
[b]*LTCI, Télécom ParisTech, Université Paris-Saclay*
[c]*L2S, CNRS, CentraleSupelec, Université Paris-Sud*

Abstract

With the growing popularity of high dynamic range (HDR) imaging, efficient compression techniques are demanded, as HDR video entails typically higher raw data rate than traditional video. For this purpose, we introduce a hybrid spatially and temporally constrained content-adaptive *tone mapping operator* (TMO) to convert the input HDR video into a tone mapped video sequence, which is then encoded using the high efficiency video coding (HEVC) standard. The proposed TMO simultaneously exploits intra-frame spatial redundancies and preserves inter-frame temporal coherence of the tone mapped video sequence. Extensive experimental results show that the developed spatio-temporal TMO (ST-TMO) solution yields higher coding performance than existing frame-by-frame TMO's, and compares favorably with state-of-the-art methods based on a fixed transfer function.

Keywords: High dynamic range, video compression, tone mapping operator, convex optimization

1. Introduction

With the increased demand for realism in entertainment applications in recent years, high dynamic range (HDR) video has become ubiquitous [12, 19]. HDR video takes the users' viewing experience beyond the inherent limitations of standard dynamic range (SDR) video [7], such as limited luminance range and color gamut, resulting in substantially improved visual experience. To this end, HDR pixels encode the physical luminance captured from the real-world scene, with a peak brightness higher than 1000 candelas per square meter (cd/m^2).

The highly realistic visual experience of HDR video comes at the cost of significant bitrate requirements. Efficient video compression is therefore one important aspect to be addressed. With this objective, a substantial amount of work has been done in the last few

years to standardize HDR video compression algorithms. To this end, nine contributions were submitted as a response to the MPEG call for evidence (CfE) on HDR and wide color gamut (WCG) video compression [34]. Those solutions [20, 27] were based on the state-of-the-art video coding standard, high efficient video coding (HEVC) [28], and focused on HDR video compression efficiency using an electro-optical transfer function (EOTF) with *10 bit-depth* coding profile [21]. EOTF defines the mapping of digital code values to display luminance.

EOTFs can be used with video compression standards to provide a perceptually uniform representation that allows reducing the number of bits required for encoding. Especially, two perceptually optimized transfer functions, hybrid log-gamma (HLG) [10] and perceptual quantizer (PQ) [40], were recently introduced. Both transfer functions map absolute luminance values to perceptual codewords and share similarities with the perceptually uniform (PU) encoding introduced by Aydin *et al.* in [1]. However, they are mainly addressing two different applications. On one hand, HLG aims at providing a backward-compatible representation with 10 bit-depth video devices, especially suited for TV broadcasting services. On the other hand, PQ focuses on high bit precision representation coding, *e.g.,* 10 or 12 bit-depth video representation, which is *not* backward-compatible with the currently available SDR devices.

Unlike EOTF-based HDR video compression solutions, where a fixed curve is used for converting each HDR frame to a reduced dynamic range representation, in this work we consider a content-adaptive tone mapping operator (TMO). Therefore, our solution takes statistical characteristics of the input HDR frame into account. More specifically, we focus on exploiting spatial and temporal pixel redundancies and introduce spatio-temporal regularization terms to find the optimal rate distortion (RD) tone-mapping curves for each frame. Our work exploits results from Mai *et al.* TMO [35]. It extends our previous conference publications [32, 46] by providing a unified framework combining both spatial and temporal regularization, and presenting a more comprehensive performance assessment including comparisons with state-of-the-art methods and multiple objective quality metrics. It is shown that the proposed scheme leads to significant coding gains compared to the state-of-the-art methods based on a fixed transfer function.

The rest of this paper is organized as follows. Section 2 provides a brief summary of the most relevant studies on HDR video coding, tone mapping operators, and the main contributions of our work. Following that, the overview of the proposed method is described in Section 3. Next, experimental results are provided in Section 4, and the conclusion in Section 5.

2. Related Work

2.1. Studies related to HDR video compression

Applying an effective quantization on the raw HDR video format is one of the most used solutions for HDR video compression [9, 57]. Boitard *et al.* proposed a motion-guided quantization method in [9] to increase the correlation between successive frames. Their method relies on the motion compensation between frames to adapt the quantization during

tone mapping curve estimation. With a similar aim and different research perspective, Zhang *et al.* developed a quantization model in [57] that removes perceptually redundant pixels with the help of luminance masking. Their proposed approach, using the TMO in [31], was implemented in HEVC and compared with the PQ transfer function in [58]. Both objective and subjective evaluation results showed that it achieves significant bitrate savings compared to the PQ transfer function. It is noted that the TMO in [31] is one of the first TMO solutions for HDR images. Using a video compression optimized tone mapping algorithm, *e.g.,* our proposed spatio-temporal constrained TMO in this paper, can generally achieve higher coding efficiency.

Perceptually driven transfer functions are widely applied to the captured HDR video, in conjunction with high bit-depth profiles of standard video codecs. The PQ transfer function, which is proposed by Miller *et al.* in [40], achieves efficient HDR video coding gain compared with reference methods. This transfer function is based on the contrast sensitivity model developed by Barten [3]. It is more adapted to HDR video than the ITU-R Rec. BT.1886 [49], which is commonly referred to as the "gamma curve".

To standardize the compression algorithms of HDR and WCG video, a substantial amount of work has been done in the last few years by MPEG. Two parallel standardization efforts, normative codec changes and non-normative encoder optimization, were organized after the CfE responses. For the normative codec changes, technologies from five proponents were incorporated to form the exploratory test model (ETM) [41]. The ETM uses the forward and the inverse reshaping techniques with the delivered metadata to produce the HDR video output at the receiver end. It can also be used to generate SDR output for traditional displays. In the context of non-normative encoder optimization effort, its main objective was to reduce the color artifacts. For this aim, the chrome QP offset, for instance, was introduced in order to reduce color artifacts [52]. As a result of the HDR and WCG video compression standardization progress, HEVC Main 10 profile was determined to be sufficient and a short term standardization effort was stopped. The HDR and WCG video compression works will continue within the future video compression standards developments. A detailed overview of MPEG CfEs for HDR/WCG video compression can be found in the overview papers [20, 27].

Some previous academic works focus on high bit-depth encoding for HDR content [23, 40, 42, 47, 59]. For instance, Motra *et al.* in [42] extended the LogLuv transform [30] in order to use adaptive transformation parameters. In their method, the minimum and maximum luminance values are adaptively used to map the HDR floating point values to 14-bit quantized integer values. Additionally, Zhang *et al.* in [59], compressed HDR video by considering perceptual quality. For this, an optimized bit-depth transformation and perceptual model based wavelet transform denoising are introduced for HDR video compression. In [23], a weighted prediction method and quantization adaptation were proposed to compensate the dynamic contrast variation across frames. In that work, a frame-wise adaptive luminance-to-luma mapping and the usage of the weighted prediction tools of the video codec are described to maintain the temporal coherence of HDR video. Similarly, Le Pendu *et al.* proposed an adaptive re-quantization method [47], where the minimum and maximum luminance values are used to adapt the mapping of the data. Their work analyzes the complete
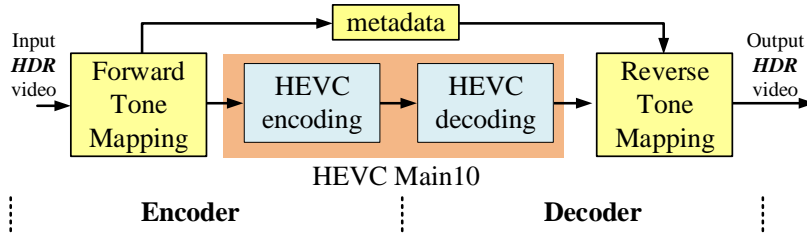
3

Figure 1: General diagram of the proposed HDR video coding method.

group of picture (GOP) in order to keep the temporal consistency of inter prediction and re-encodes frames accordingly. Conversely, here we consider a compression solution which takes the current and previous frames into account and uses a TMO to improve the HDR video coding performance.

2.2. Studies related to tone mapping operators

Even though most existing TMOs achieve good visual quality performance for still HDR images [17], they may not be adequate for video contents [44]. Notably, applying a frame-by-frame TMO may introduce spurious temporal discontinuities, *i.e.,* flickering, in the tone mapped video sequences. Straightforwardly, such discontinuities may negatively affect the video compression efficiency.

Several solutions were introduced to deal with temporal discontinuities when applying TMO on HDR video sequences [2, 8]. For instance, the quality impact of the TMOs is explored for video compression in [8]. In their work, the photographic tone reproduction [50] and temporal coherency method [5] were evaluated with respect to video compression performance. Their work demonstrates smooth temporal coherency results at a target bitrate. However, as shown in their experimental results, their solution cannot preserve the reconstructed quality of HDR video well. For this reason, Aydin *et al.* proposed a local TMO for HDR video in [2]. More precisely, each frame was decomposed into two layers, *i.e.,* base and detail, using an edge-aware filter. Such decomposition allows processing each layer's dynamic range separately. This way, spatio-temporal filtering was applied along per-pixel motion paths to prevent flickering artifacts between frames. Their proposed method performs well with different content types, but with the limitation of not considering the impact of video compression.

Besides flickering artifacts, applying TMO for each frame may cause temporal brightness discontinuities. To this end, a zonal brightness coherency method was introduced in [6] to tackle such artifacts. In order to reduce temporal discontinuities while maintaining local contrast, each frame was divided into segments, which were independently tone mapped. That solution focuses on the tone mapped HDR video quality on SDR displays. Conversely, our TMO solution is targeting high coding performance for HDR video, where the quality is measured on the reconstructed HDR signal.

4

2.3. Limitations of existing works and contributions of the present one

One of the earliest works considering the optimization of a TMO in the context of scalable HDR video coding is that of Mai *et al.* [35]. There, a global tone mapping curve was found by minimizing the mean squared error (MSE) between the original and the reconstructed HDR pixels, *independently* for each frame. A piecewise-linear parametrization of the tone mapping curve enables to compute the solution to this optimization problem in closed form. The resulting tone curve is optimal at high bitrates and close to optimal in the mid-to-low bitrates. The evaluation reported in [35] for H.264/AVC video demonstrates significantly better rate-distortion performance than popular TMOs such as [18, 38]. Since in this work we employ a similar tone curve parametrization as [35], we discuss it in greater detail in Section 3.1.

Later, Koz and Dufaux [29] improved the results of Mai *et al.*, by adding two additional constraints to the TMO optimization problem. First, they take into account the quality of the tone-mapped frames, which was not considered in the original formulation [35]. To this end, they assume that a visually pleasing tone-mapped image is the result of some perceptually motivated, global tone mapping curve, *e.g.,* [50]. Then, in the optimization process they penalize curves that are far away from the visually favorable one. The second constraint aims at reducing temporal flickering, by avoiding that consecutive frames differ in average brightness beyond a certain visible threshold. Both constraints are applied when HDR values are expressed in PU units [1] in place of logarithm, in order to reflect the lower contrast sensitivity of the human visual system at darker luminance levels.

Interestingly, the authors of [35] have independently come up with similar constraints in their follow-up work [36], though under a different optimization framework. Specifically, in order to reduce flickering an additional constraint is added by considering the difference between the average brightness of consecutive tone-mapped frames. In our work, we also target temporal dependencies but we consider local (per pixel) motion trajectories. We show that this finer granularity approach constitutes a more effective temporal regularization term.

In our previous work [32, 46], we have analyzed two drawbacks of [35]: the possible loss of intra-frame spatial smoothness and of the inter-frame temporal coherence after tone mapping. A spatial regularization term was introduced by Lauga *et al.* [32] to consider spatial pixel variations in HDR image compression. The resulting problem was shown to be convex and was efficiently solved using a proximal optimization algorithm. In [46], we formulated a temporal constraint for HDR video compression. In this work, we provide a unified framework by proposing a spatio-temporal constrained TMO, referred to as ST-TMO, which brings larger compression gains. We complete the work with an extended experimental evaluation, a comparison with state-of-the-art methods, and objective quality assessment using several HDR video quality metrics.

3. Content Adaptive Tone Mapping Operator

The general scheme of the proposed ST-TMO is illustrated in Figure 1. For each input HDR video frame, we compute two cost terms: an estimation of the MSE between the original and the reconstructed HDR frame; and a regularization term, which enforces spatial

or temporal coherence, depending on whether the frame is Intra or Inter predicted, respectively. Specifically, the temporal regularization relies on the knowledge of the motion field between the current and the previous frame (without loss of generality, we assume here that the temporal prediction at time t is obtained based on the frame at time $t-1$).

We express the unknown tone mapping curve to be found using the simple, piecewise-linear parametrization proposed in [35]. That is, the TMO is expressed as a vector s of slopes as described in Section 3.1. This enables to define the cost terms mentioned above as convex functions of s, and to solve the resulting convex optimization problem through a proximal optimization method. As a result, we obtain a vector s^* of optimal slopes, which can be used to tone map the HDR picture into an SDR frame f_t^{SDR}. The vector s^* is sent as metadata information to the decoder in order to invert the tone mapping and reconstruct the HDR information. A detailed description of the parametrization employed, the cost terms and the convex optimization follows.
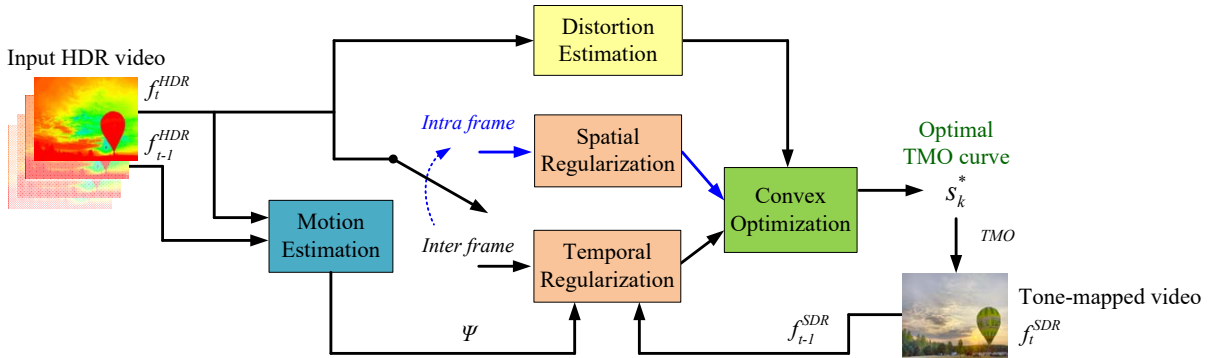

Figure 2: Block diagram of the proposed TMO. A content-adaptive spatially and temporally constrained tone mapping curve is obtained for each frame.

3.1. TMO parametrization and distortion estimation

Tone mapping algorithms can be broadly classified into global [18, 29, 32, 35, 36, 38, 46, 50] or local [2], according to whether their result changes depending on the content and characteristics of a local pixel neighborhood. Although local tone mapping is more flexible and can generally preserve better fine-level details of a picture, global operators are the most convenient option for HDR coding, as they require little side information to be sent to the decoder. Mai *et al.* [35] proposed a simple, yet general enough, parametrization of a global tone mapping curve as a piecewise-linear function in the logarithmic domain.

This parametrization is illustrated in Figure 3. The histogram of the log-luminance l is divided into N segments of equal length δ. Let $k \in [1, \ldots N]$ be the segment index, and p_k the associated probability mass. The TMO curve is then described by the set of nodes $\{l_k, v_k\}$, where v_k denotes the tone-mapped value; or, equivalently, by a vector $s \in \mathbb{R}_+^N$ composed by the non-negative slopes s_k for each segment, that is:

$$s_k = \frac{v_{k+1} - v_k}{\delta}. \tag{1}$$

6

Based on the s_k, the tone-mapped pixel values v are obtained as

$$v(l) = (l - l_k) \cdot s_k + v_k, \quad \forall l \in [l_k, l_k + 1), \tag{2}$$

where $l_1 = l_{\min}$ is the minimum luminance of the image. In order to inverse (2) and compute the reconstructed log-luminance, $\hat{l}$, the receiver needs to know as side information the slopes s_k, the value of δ as well as $l_{\min}$. Therefore, the size of the metadata depends on the number of segments, or equivalently on δ. Here we select $\delta = 0.1$ in $\log_{10}$ units to approximately match the Weber ratio (at least at high luminance), as proposed in [35]. For typical HDR content, this results in a number N of segments around $70 \sim 80$, which represents a modest overhead to overall transmission cost.

This parametrization can be used to find the optimal TMO curve, *i.e.,* the slopes s_k that minimize the squared reconstruction error $\varepsilon(s_k) = \|l - \hat{l}\|_2^2$. It is shown in [35] that, at least at high bitrates, the distortion is well approximated as a function of s_k as:

$$\varepsilon(s_k) \approx \sum_{k=1}^{N} \frac{p_k}{s_k^2}. \tag{3}$$

Based on this approximation, the authors of [35] find the slopes that minimize the squared-error distortion by formulating the following optimization problem:

$$\underset{s_k}{\text{minimize}} \ \varepsilon(s_k) \quad \text{subject to:} \ \sum_{k=1}^{N} s_k = \frac{v_{\max}}{\delta}; \ s_k > 0, \tag{4}$$

where the constraint guarantees that the TMO curve spans all the available standard dynamic range (*e.g.,* $v_{\max} = 255$ for 8-bit images). This problem can be solved in closed form, yielding:

$$s_k = \frac{v_{\max} p_k^{1/3}}{\delta \sum_{n=1}^{N} p_n^{1/3}}. \tag{5}$$

Notice that this result is optimal in the MSE sense, but does not take into account the excess of bitrate produced by the possible loss of spatial and temporal coherency in the v signal. In order to take these effects into account, in the following we propose a spatial and a temporal regularization terms to be added to the problem in (4).

3.2. Spatial regularization

In [32] we have shown that the s_k in (5) are indeed suboptimal from a video compression perspective, *i.e.,* they do not guarantee spatial smoothness in the resulting tone-mapped pictures. This, in turn, leads to higher coding bitrate.

In order to alleviate this effect, we modify problem (4) by adding a spatial regularization term, $C_{spa}(f_{intra}^{SDR})$, for tone-mapped Intra-coded images. C_{spa} is a real-valued convex function that models the spatial complexity, and f_{intra}^{SDR} is the intra frame of a given GOP of the SDR (*i.e.,* tone mapped) video. Since natural images usually exhibit a smooth spatial
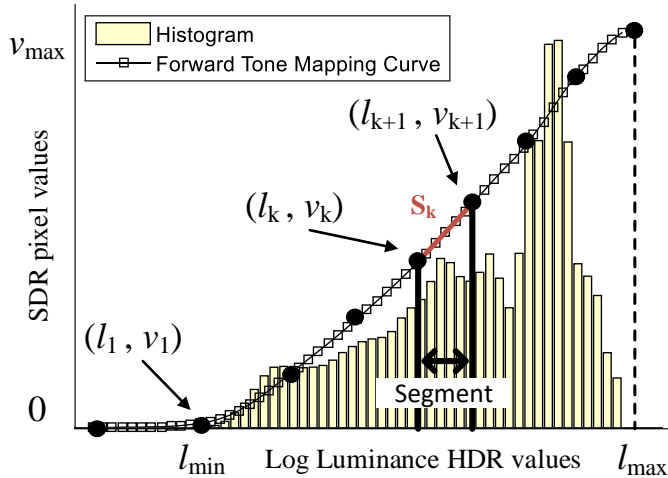
Figure 3: Piecewise parametrization of the optimal tone mapping curve as in [35].

behavior, except around some locations (*e.g.,* object edges) where discontinuities arise, popular regularization models tend to penalize the image gradient. In this context, a number of regularization terms have been proposed, *e.g.,* frame-based representations [37], Total Variation (TV) [51], etc. Here we adopt TV due to its simplicity and effectiveness [11]. Thus, we express the spatial regularization term as:

$$C_{spa}\left(f_{intra}^{SDR}\right) = \left\|\nabla f_{intra}^{SDR}\right\|_{1,2} = \sum_{i \in \Omega} \left\|(\nabla f_{intra}^{SDR})_i\right\|_2, \tag{6}$$

where Ω is the rectangular lattice over which the image f is defined, and $(\nabla f_{intra}^{SDR})_i$ is the 2-element vector denoting the gradient of f_{intra}^{SDR} at site i.

Notice that, since the log-luminance values l are constant for a given image, the tone mapping in (2) is linear in s, and can thus be conveniently rewritten as a matrix-vector multiplication $f_{intra}^{SDR} = Zs$. Specifically, for a given HDR image with M pixels, $Z = [z_1, \ldots, z_M]^T$ is an $M \times N$ matrix, where each row has the form:

$$z_m = [\delta, \ldots, \delta, l - l_k, 0, \ldots, 0], \tag{7}$$

with the term $l - l_k$ in the k-th position if $l \in [l_k, l_{k+1})$. This formulation expresses the tone mapping equation (2), in that an HDR pixel l_i, falling in the j-th bin of the histogram, is mapped as $v_i = \sum_{k=1}^{j-1} \delta s_k + (l_i - l_j)s_j$. We will use this formulation to express the spatial constraint in the convex optimization problem in Section 3.4.

3.3. Temporal regularization

For Inter-predicted tone-mapped video frames, applying the frame-by-frame tone mapping curve in (1) might lead to a loss of temporal coherence and a consequent increase of

coding bitrate. In this section, we describe how to enforce temporal smoothness in the tone-mapped video by proposing a temporal regularization term, C_{temp}, to add to problem (4).

In this work, C_{temp} is defined as:

$$C_{temp}(f_{inter}^{SDR}) = \sum_{i,j} \left(f_t^{SDR}(i,j) - \Phi \left[f_{t-1}^{SDR}(i,j); \Psi(i,j) \right] \right)^2 \tag{8}$$

where $\Phi[f_{t-1}(i,j); \Psi(i,j)] = f_{t-1}(i + \Psi(i,j)_x, j + \Psi(i,j)_y)$ is a function that gives the value of pixel at position (i,j) after motion compensation by the 2-element motion vector $\Psi(i,j)$ (x and y represent the horizontal and vertical components, respectively). The notation f_{inter}^{SDR} is used here to stress the fact that f_t is inter predicted. In order to get a precise per pixel motion field, we estimate Ψ by employing the optical flow algorithm in [13] directly on the original HDR frames f_t^{HDR} and f_{t-1}^{HDR}. This optical flow is then applied to obtain the motion compensated frame $\Phi \left[f_{t-1}^{SDR}(i,j); \Psi(i,j) \right]$.

Note that f_t^{SDR} is a function of s_k. By explicitly minimizing a temporal prediction residual, such a constraint leads to improved rate-distortion performance when encoding the tone mapped SDR video sequence. Instead of explicitly computing the sum of pixel-wise differences, as defined in Eq. (8), we notice that the temporal term C_{temp} is proportional to the expected temporal difference between two (motion-compensated) frames. By the definition of expected value, we can then compute (8) as the sum of all possible frame difference values, weighted by the probability of occurrence of each difference, that is:

$$C_{temp}(f_{inter}^{SDR}) = \sum_{k=1}^{N} \sum_{w=0}^{v_{max}} \left\{ \left(\frac{d_k}{2} + \sum_{i=1}^{k-1} d_i - w \right)^2 p_{k,w} \right\}, \tag{9}$$

where, for notational convenience, we define $d_k = \delta s_k$; $\frac{d_k}{2} + \sum_{i=1}^{k-1} d_i$ is then the SDR reconstruction value for pixels falling in the bin k of the log-luminance histogram as in Eq. (2), assuming a mid-tread quantizer on the real tone-mapped pixel values; $w \in [0, v_{max}]$ is the value of a pixel in the motion-compensated SDR frame $\Phi \left[f_{t-1}^{SDR}(i,j); \Psi(i,j) \right]$; and, finally, $p_{k,w} = \Pr\{f_t^{HDR} = l_k \wedge f_{t-1}^{SDR} = w\}$ is the joint probability that a pixel with log-luminance l_k in f_t^{HDR} has a motion-compensated predictor which has been tone mapped to the value w in f_{t-1}^{SDR}. Notice that, while tone mapping f_t^{HDR}, the previous SDR frame f_{t-1}^{SDR} has been already computed, i.e., w depends only on the (constant) motion vector field Ψ. In practice, we pre-compute $p_{k,w}$ after motion estimation, before computing Eq. (9). Finally, as shown in the Appendix, the temporal constraint in (9) can be rewritten as a quadratic function:

$$C_{temp}(f_{inter}^{SDR}) = s^T W_2 s + s^T W_1, \tag{10}$$

where W_1 and W_2 are constant matrices defined in Appendix.

3.4. Convex Optimization

Based on the spatial and temporal constraints defined above, we can redefine the optimization problem for each frame f_t^{HDR} in Eq. (4) as:

$$\underset{s_k}{\text{minimize}} \ \hat{\varepsilon}(s_k) + \mathcal{C} \quad \text{subject to:} \ \sum_{k=1}^{N} s_k = \frac{v_{\max}}{\delta}, \tag{11}$$

9

where the term:

$$\mathcal{C} = \begin{cases} \lambda_{spa} C_{spa}(f_{intra}^{SDR}) & \text{if } f_t^{HDR} \text{ is Intra-predicted} \\ \lambda_{temp} C_{temp}(f_{inter}^{SDR}) & \text{if } f_t^{HDR} \text{ is Inter-predicted.} \end{cases}$$

The weighting terms λ_{spa} and λ_{temp} define the relative importance of spatial/temporal smoothness with respect to MSE minimization, and are determined experimentally in Section 4. Notice that Problem (11) consists in minimizing the sum of two convex functions – $\mathcal{C}$ is either linear or quadratic for spatial and temporal regularization, respectively, as discussed in Sections 3.2 and 3.3 – over the convex set:

$$\Theta = \left\{ s \in \mathbb{R}^N \;\middle|\; \sum_{k=1}^{N} s_k = \frac{v_{max}}{\delta} \right\}. \tag{12}$$

Moreover, the term $\hat{\varepsilon}(s_k)$, defined as:

$$\hat{\varepsilon}(s_k) = \begin{cases} +\infty & \text{if } s_k \leq 0 \\ \varepsilon(s_k) & \text{otherwise,} \end{cases} \tag{13}$$

is also a convex function.

The solution of Problem (11) requires an efficient algorithm for dealing with non-smooth functions. Among the many approaches proposed to solve convex optimization problems [15], we resort here to proximal algorithms [37], since they provide a unifying framework that allows one to address both non-smooth functions and hard constraints. Within the large panel of existing proximal algorithms [14, 15, 48], we consider the primal-dual M+LFBF algorithm proposed in [16], which can address general convex optimization problems involving non-smooth functions and linear operators without requiring any matrix inversion.

4. Experiments

In this section, we evaluate the compression performance of the proposed ST-TMO through an extensive set of experiments. We start describing the experimental setup, followed by results and discussion.

4.1. Setup

The input to the proposed ST-TMO is an HDR video frame in linear (photometric) RGB domain. The required color conversion and chroma sub-sampling algorithms suggested in the MPEG CfE on HDR and WCG video coding [34] are used. We employ the optical flow algorithm (with the publicly available implementation) in [13, 24], with the configuration parameters reported in Table 1.

Table 1: Optical flow parameters.

Parameter	Description	Value
Lambda	Regularization weight	50
PyramidFactor	Gaussian pyramid scale	0.7
MaxIts	number of iteration per Gaussian level	50
SizeFilter	The size of the filter to remove outlier	7

(a) *Market3Clip4000r2*   (b) *FireEater2Clip4000r1*   (c) *SunRise*   (d) *EBU_04_Hurdles*

(e) *EBU_06_Starting*   (f) *Carousel_fireworks_03*   (g) *Carousel_fireworks_04*   (h) *Smith_Welding*

Figure 4: Sample frames from the eight selected test sequences used for the experiment. Images are tone-mapped using a tone mapping technique in [39] for visualization purposes.

4.1.1. Test Sequences

Experiments were carried out for eight high-definition (HD) test sequences. Figure 4 shows a sample thumbnail from the test videos, namely *Market3Clip4000r2*, *FireEater2Clip4000r1*, *SunRise*, *EBU_04_Hurdles*, *EBU_06_Starting*, *Carousel_fireworks_03*, *Carousel_fireworks_04*, and *Smith_Welding*.

In order to gauge the performance of the proposed ST-TMO for different input conditions, the selected test sequences have varied characteristics of spatial/temporal complexity and dynamic range, reported in Table 2. The video dynamic range, $DR = L_{max} / L_{min}$ where L_{max} and L_{min} are the average maximum and minimum relative luminance values of the video sequence. Spatial information (SI) and temporal information (TI) indexes were computed as described in the ITU-R P.910 recommendation [26] on the logarithmic luminance component of each content.

Table 2: Dynamic Range, Spatial Information and Temporal Information of the HDR video sequences used in the results.

Sequence	*Market3Clip4000r2*	*FireEater2Clip4000r1*	*SunRise*	*EBU_04_Hurdles*	*EBU_06_Starting*	*Carousel_fireworks_03*	*Carousel_fireworks_04*	*Smith_Welding*
DR	5421	55172	1613	873	9958	1718	2425	1790
SI	1.59	0.80	0.83	0.99	1.34	0.85	0.83	1.82
TI	0.31	0.17	0.17	0.11	0.04	0.48	0.50	0.41

4.1.2. Benchmarks and Coding Conditions

We compare the proposed ST-TMO with the frame-by-frame TMO of Mai *et al.* (2011) [35] and with the anchor solution based on the PQ EOTF proposed in the MPEG CfE [34], available in the MPEG HDRTools v.0.17 [54]. In addition, we implemented the temporally optimized TMO in [36], Mai *et al.* (2013), which is the most related approach to ours. The overhead from the proposed ST-TMO method (*i.e.,* metadata) is *included* in all reported results. Table 3 shows the average overhead of the required metadata for the proposed ST-TMO.

Table 3: Average metadata overhead for the proposed ST-TMO.

Sequence	Market3Clip4000r2	FireEater2Clip4000r1	SunRise	EBU_04_Hurdles	EBU_06_Starting	Carousel_fireworks_03	Carousel_fireworks_04	Smith_Welding
Bitrate (kbps)	132	138	126	123	126	138	135	138

In order to evaluate the compression performance, video test sequences were encoded using the HEVC reference model (HM) ver. 16.2 software. The low-delay HEVC encoder configuration we used is as follows: GOP length of 16, predictive coded (P) pictures, YCbCr 4:2:0 chroma sampling, and an internal bit-depth of 10. We set v_{max} to 1023 for TMOs. Variation in bitrates was achieved using different quantization parameter (QP) values.

Evaluation of HDR video distortion is more challenging than conventional SDR quality assessment, as only few ad hoc metrics have been proposed for the extended luminance range of HDR content. In order to provide a more informative comparison, we evaluate HDR video distortion using the following metrics: Peak signal to noise ratio (PSNR) and structural similarity index (SSIM) [53], both computed on either log-luminance or PU-encoded values [1]; HDR-Visible differences predictor (HDR VDP 2.2.1) [45]; the HDR-Video Quality Measure (HDR-VQM) [43]; HDR Metrics in the MPEG HDRTools [54], including tPSNR, L_{100} and L_{1000}. These metrics have been recently found to be sufficiently good predictors of image quality for the case of HDR image compression [25, 55]. The SSIM metric uses values between zero and one; the higher SSIM index indicates higher quality. The logarithmic encoding is based on the fact that the human visual system sensitivity approximately follows a logarithm function for light luminance values. The PU encoding, which is derived from the contrast sensitivity function, converts luminance values into perceptual code values [1]. Additionally, HDR VDP *ver.* 2.2.1 and HDR-VQM, which have been demonstrated to be well correlated with the visual perception in subjective studies [25, 55], were used to make a more robust comparison concerning perceptual quality. Both metrics require absolute luminance values as input. HDR-VDP metric is a calibrated metric which takes into account multi-scale and multi-orientation characteristics of the early visual system. Also, a neural noise block is defined to calculate per-pixel probabilities maps of visibility and the predicted quality metric. We used the HDR-VDP 2.2.1 metric with the following settings: 120 *cm* viewing distance, 1920×1080 display resolution, and 47-degree viewing angle. Differently, HDR-VQM was designed for quality assessment of HDR video content. HDR-VQM uses the PU space and relies on a multi-scale and multi-orientations analysis, similarly to HDR-VDP. We limit our choice to luma-only quality metrics based on a recent study conducted by some of the authors on the effect of color space on HDR video compression quality [56]. There,

it has been found that i) color artifacts due to using different color spaces have little or no effect on the overall quality of compressed video, which is instead dominated by structural distortion; and ii) that color metrics are not providing better prediction of mean opinion scores that color-blind metrics. Color metrics such as ΔE_{2000} have been found to be poorly correlated with subjective scores in [55]

The compression performance was measured for luma pictures using the Bjøntegaard delta (BD) metric [4], which describes the average quality difference between two RD curves.

4.2. Performance Evaluation

We present parameter selection of our proposed ST-TMO in this section, an analysis of the luminance temporal variation, the measured compression performances is evaluated using RD curves and the Bjøntegaard metric, and some qualitative visual examples.
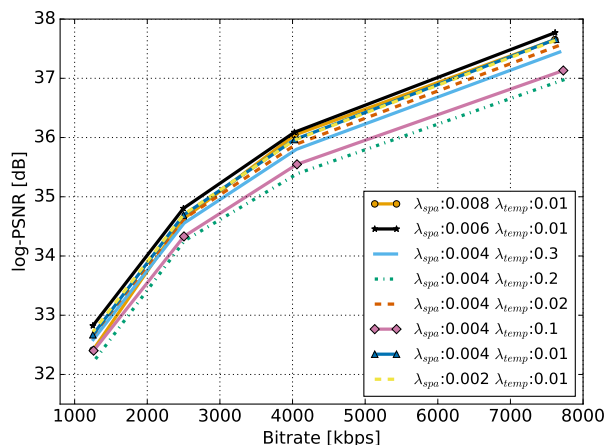
4.2.1. Determination of the parameter λ_{spa} and λ_{temp}

Figure 5: Optimum value selection of the λ_{spa} and λ_{temp} for the *Carousel_fireworks_04* sequence, which presents several temporal light variations. The video coding setup is the same as that described in Sec. 4.1.

Optimum λ_{spa} and λ_{temp} parameters are needed to be selected to achieve high overall compression gain. In our preliminary experiments, we found that varying λ_{temp} plays a major role in the overall coding performance than λ_{spa}. Therefore, in order to conservatively tune λ_{temp}, we optimize it using the video sequence with highest motion and flickering in our dataset, i.e., *Carousel_fireworks_04* which from Table 2 is found to have the highest temporal complexity. We report in Figure 5 rate-distortion curves (using the log PSNR metric) for several different values of λ_{spa} and λ_{temp}, for *Carousel_fireworks_04*. We observe that $\lambda_{temp} = 0.01$ and $\lambda_{spa} = 0.006$ provide the best results for this sequence. Accordingly, we use these parameters for all the video sequences in our experiments.

4.2.2. Temporal variation of luminance

Our proposed ST-TMO is supposed to reduce the temporal variation of luminance in the tone mapped frames, leading to rate-distortion gains. To analyze this luminance variation,

13

we calculate the mean luminance of each tone mapped frame, for each video sequence, as illustrated in Figure 6. Also, we compute the average value of the mean luminance differences between two consecutive tone mapped frames, for each video sequence, to observe the effect of the temporal regularization in our proposed formulation. Table 4 shows the temporal average luminance differences for the proposed ST-TMO and the references.

For four of the tested sequences, namely *Market3Clip4000r2*, *FireEater2Clip4000r1*, *Carousel_fireworks_03*, and *Carousel_fireworks_04*, the proposed ST-TMO method shows the lowest temporal variation. However, we notice that in other cases, the temporal variation in the ST-TMO is in-between the minimum and maximum values obtained by other methods. This suggests that our optimization, which mixes both per frame fidelity and temporal regularization, enables to trade the two terms in a non-trivial way; specifically, it avoids an excessive flattening of the luminance temporal profile, which might reduce further the bitrate but remove essential details. We can also notice from Figure 6 that our TMO tends to produce brighter tone-mapped pictures, while the PQ produces darker images (as observed in previous work). Finally, from Table 4 we notice that, differently from PQ and Mai *et al.* (2011), the temporal variation in our method does not change too much from content to content; this suggests that the ST-TMO has more consistent performance on varying content characteristics.

Table 4: Average variation of mean luminance across consecutive tone-mapped frames.

	Method			
Sequence	Mai *et al.* (2011)	Mai *et al.* (2013)	PQ	ST-TMO
Market3Clip4000r2	0.286	1.0	0.6	**0.357**
FireEater2Clip4000r1	3.501	2.857	2.786	**1.786**
SunRise	**0.428**	1.428	0.172	0.643
EBU_04_Hurdles	0.071	1.571	**0.052**	0.571
EBU_06_Starting	0.428	3.143	**0.214**	1.642
Carousel_fireworks_03	6.010	3.857	7.214	**3.571**
Carousel_fireworks_04	7.786	2.929	9.357	**2.786**
Smith_Welding	11.857	**3.429**	8.510	6.785

4.2.3. Assessment of coding performance

In order to assess the coding gains obtained with the proposed ST-TMO, we compute rate-distortion curves for the 8 video sequences reported in Table 2. Due to space limitation, we report the graphs only for HDR-VQM and HDR-VDP in Figures 7 and 8, respectively. Notice that these two metrics are HDR specific and are generally considered to be well correlated to human judgments [55]. For HDR-VQM, a lower value of the quality metric implies a higher visual quality, while the opposite is true for HDR-VDP. With respect to HDR-VDP, we observe that for all the tested contents the proposed ST-TMO provides important gains
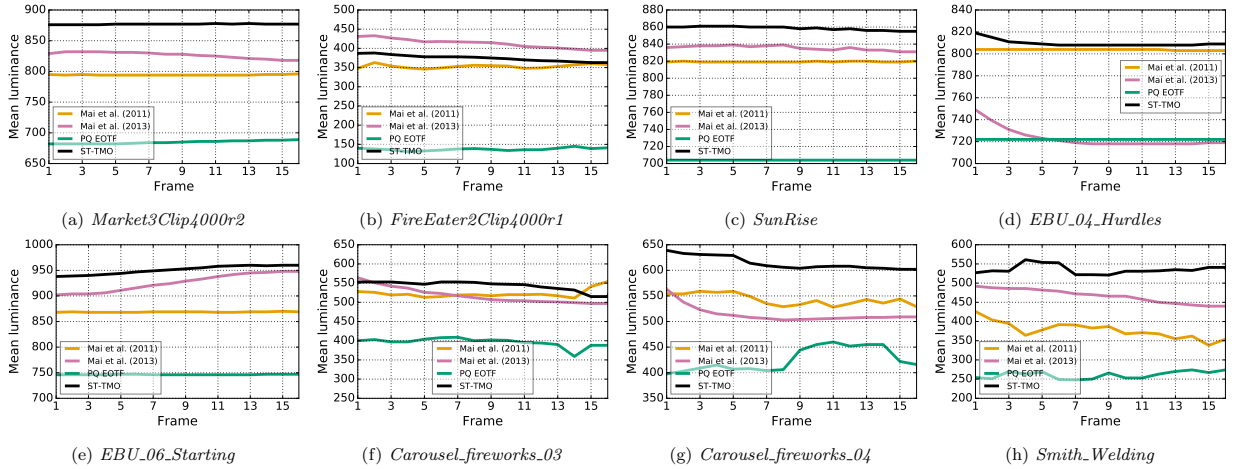
Figure 6: Comparison of luminance temporal variation between our proposed ST-TMO and the reference methods, for the first 15 frames of the tested video sequences. The x and y axes of the figure represent the tone-mapped video frame number and mean luminance of the tone-mapped video frames, respectively.

with respect to both tone mapping based solutions [35, 36] and PQ, across a wide range of bitrates. For *FireEater2*, sequences tone mapped with ST-TMO have substantially similar performance as the PQ solution and as [35]. This is somehow expected as *FireEater2* has low motion and spatial complexity, so the potential advantages of spatial and temporal smoothing are limited.

The results with the HDR-VQM metric (Figure 8) somewhat confirm those with HDR-VDP, although with some exceptions, in particular with respect to *Smith_Welding*, where PQ performs considerably better than all tone mapping based approaches. Notice that, with respect to HDR-VDP, the HDR-VQM pools errors both spatially and temporally, i.e., it is a video-specific metric; however, it employs a simpler model of the human visual system, which might justify the extra penalty assigned to errors in dark areas which are likely to be less visible in practice.

In order to complete the figures, we report average BD gains in Table 5 for 7 additional quality metrics. The reported values are the average gain in the corresponding quality metric. We opted for reporting BD quality instead of BD rate gains since the computation of the latter is unstable when the RD curves are not properly aligned on the quality axis, as it can be seen in Figures 7 and 8. From the table, it is evident that the proposed ST-TMO leads to superior coding performance in most cases, and on average yields consistent gains with all the considered quality metrics with respect to both a fixed transfer function and a state-of-the-art TMO-based HDR video compression scheme. In particular, our proposed TMO is beneficial for contents that display local motion and high spatial and temporal complexity.

4.2.4. Visual comparison

In order to provide qualitative visual comparison between methods, Figures 9 and 10 report some examples of reconstructed HDR frames for each method (tone mapped using
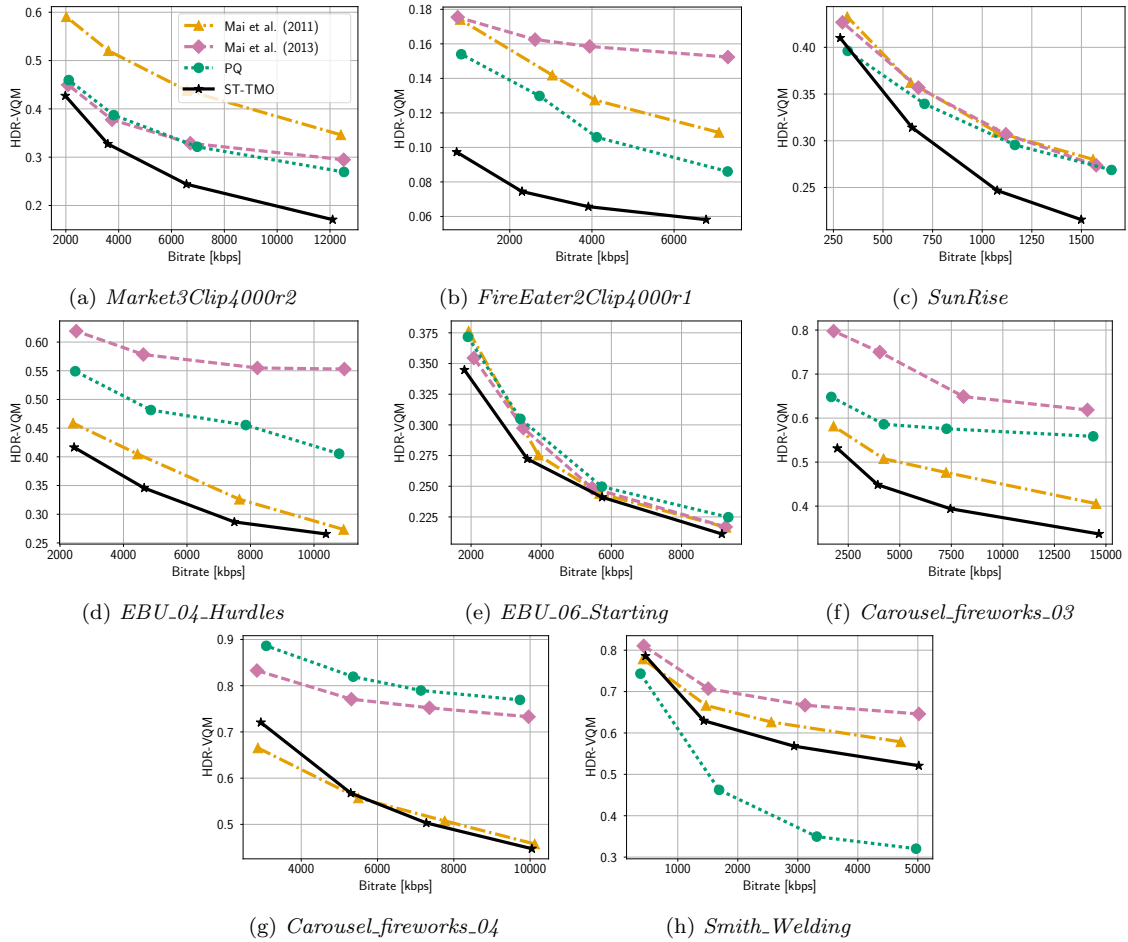
Figure 7: Performance comparison using the rate distortion curves computed with the HDR-VQM metric. The lowest HDR-VQM values indicate the best results.

(a) *Market3Clip4000r2*

(b) *FireEater2Clip4000r1*

(c) *SunRise*

(d) *EBU_04_Hurdles*

(e) *EBU_06_Starting*

(f) *Carousel_fireworks_03*
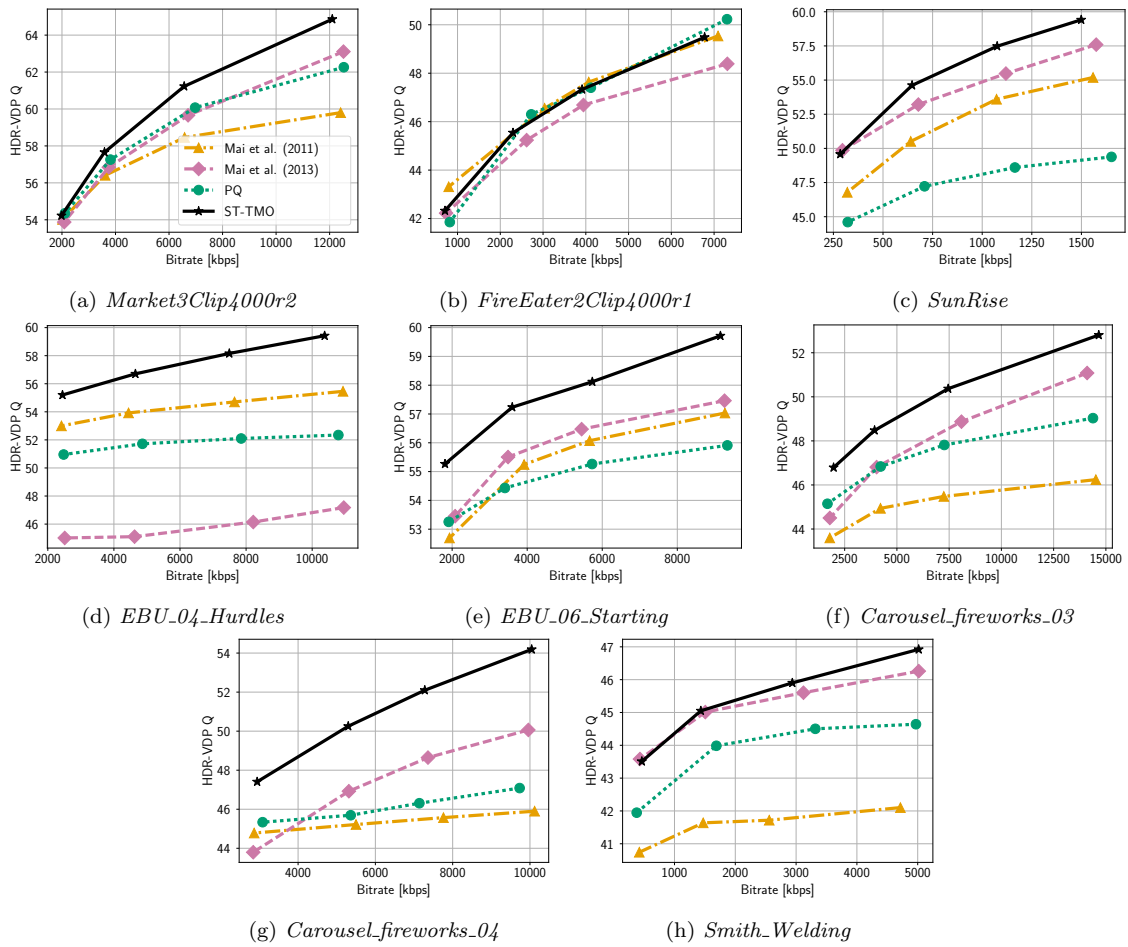
(g) *Carousel_fireworks_04*

(h) *Smith_Welding*

Figure 8: Performance comparison using the rate distortion curves computed with the HDR-VDP metric. The highest HDR-VDP Q values indicate the best results.

17

Table 5: Quality gain of the proposed ST-TMO in terms of BD quality (dB) gains. The highest BD-quality gains in **blue** and BD-quality losses in *red*.

Method	Sequence	log-PSNR	log-SSIM	PU-PSNR	PU-SSIM	HDR-VDP Q	HDR-VQM	tPSNR	L_{100}	L_{1000}
PQ		0.247	0.001	*-0.070*	*-0.001*	1.029	0.071	0.252	*-0.241*	*-0.272*
Mai *et al.* (2011)	*Market3Clip4000r2*	**1.507**	**0.011**	**0.85**	0.002	**2.119**	**0.185**	**1.468**	**0.237**	**0.216**
Mai *et al.* (2013)		0.813	0.003	0.543	**0.002**	1.198	0.070	0.687	0.066	0.061
PQ		0.794	0.008	0.002	*-0.001*	*-0.367*	0.056	4.469	1.911	1.903
Mai *et al.* (2011)	*FireEater2Clip4000r1*	**4.556**	**0.028**	**3.370**	**0.014**	*-0.309*	0.074	5.863	**2.299**	**2.104**
Mai *et al.* (2013)		0.221	*-0.001*	0.789	0.001	**0.579**	**0.088**	**10.180**	*-1.219*	*-0.985*
PQ		**5.337**	0.004	**6.838**	**0.003**	**7.265**	0.020	**4.537**	3.492	0.060
Mai *et al.* (2011)	*SunRise*	3.516	**0.006**	4.794	0.002	2.902	**0.030**	0.071	**3.867**	**3.932**
Mai *et al.* (2013)		*-0.896*	*-0.003*	*-0.892*	*-0.002*	0.638	*-0.030*	*-0.477*	*-0.154*	*-0.155*
PQ		0.128	*-0.006*	*-0.217*	*-0.008*	5.324	0.142	6.402	0.915	0.932
Mai *et al.* (2011)	*EBU_04_Hurdles*	1.477	*-0.001*	1.239	*-0.001*	2.901	0.042	2.138	**1.351**	**1.357**
Mai *et al.* (2013)		**2.514**	**0.017**	**2.437**	**0.021**	**11.581**	**0.240**	**6.685**	1.202	1.196
PQ		0.959	0.001	0.217	*-0.001*	**2.744**	0.017	**2.276**	1.358	1.369
Mai *et al.* (2011)	*EBU_06_Starting*	**2.025**	**0.006**	**1.595**	0.002	2.226	**0.095**	2.702	**1.389**	**1.401**
Mai *et al.* (2013)		0.952	0.003	0.744	**0.003**	1.716	0.011	1.485	0.912	0.919
PQ		3.688	0.042	4.770	0.008	2.172	0.158	3.598	1.288	1.363
Mai *et al.* (2011)	*Carousel_fireworks_03*	**7.529**	**0.147**	**10.294**	**0.040**	**4.323**	0.065	**7.438**	**4.209**	**4.798**
Mai *et al.* (2013)		1.249	0.020	1.441	0.01	1.670	**0.280**	1.837	0.865	0.915
PQ		4.819	0.021	5.679	0.013	4.523	**0.247**	5.543	1.558	1.465
Mai *et al.* (2011)	*Carousel_fireworks_04*	**7.862**	0.043	**9.074**	0.023	**5.254**	*-0.014*	**7.987**	**3.626**	**3.635**
Mai *et al.* (2013)		2.938	**0.052**	3.649	**0.040**	3.349	0.203	2.798	1.698	1.640
PQ		0.071	0.014	*-1.091*	*-0.001*	1.406	*-0.138*	*-0.335*	0.558	0.752
Mai *et al.* (2011)	*Smith_Welding*	**14.954**	**0.909**	**14.077**	**0.093**	**3.466**	0.024	**2.057**	**7.232**	**6.778**
Mai *et al.* (2013)		0.366	0.002	1.121	0.001	0.031	**0.072**	1.032	0.432	0.698
Average										
PQ		2.005	0.011	2.018	0.002	**3.012**	0.072	3.343	1.355	0.946
Mai *et al.* (2011)		**5.428**	**0.144**	**5.662**	**0.022**	2.860	0.063	**3.716**	**3.026**	**3.027**
Mai *et al.* (2013)		1.02	0.012	1.229	0.008	2.595	**0.117**	3.028	0.475	0.536

the method in [39]). We magnify some details to display the most significant differences. As can be seen in the figures, the proposed ST-TMO method is able to reproduce more precisely the high-frequency details in those regions, compared to the benchmark methods.

4.2.5. Computational complexity

To compare fixed transfer functions such as PQ and frame-based TMOs such as Mai *et al.* (2011) and PQ, we investigate average per frame computation times for each method. For this, MATLAB implementation of Mai *et al.* (2011), Mai *et al.* (2013), and ST-TMO were utilized, where each implementation was not well optimized for runtime. In contrast, we used an official C++ implementation of the PQ algorithm in MPEG HDRTools v.0.17 [54], which is well optimized. Each estimated run-time value was obtained in the preprocessing step, i.e., before the HEVC encoding, using an Intel(R) Core(TM) i7-3630QM CPU @ 2.40GHz with 16 GB of RAM. Table 6 reports average per frame computation times for each method.

Compared to fixed transfer functions such as PQ, or to the closed-form TMO of Mai *et al.* (2011), the proposed ST-TMO has clearly a higher complexity. Indeed, while the

Table 6: Average computation time of the proposed ST-TMO and reference TMOs.
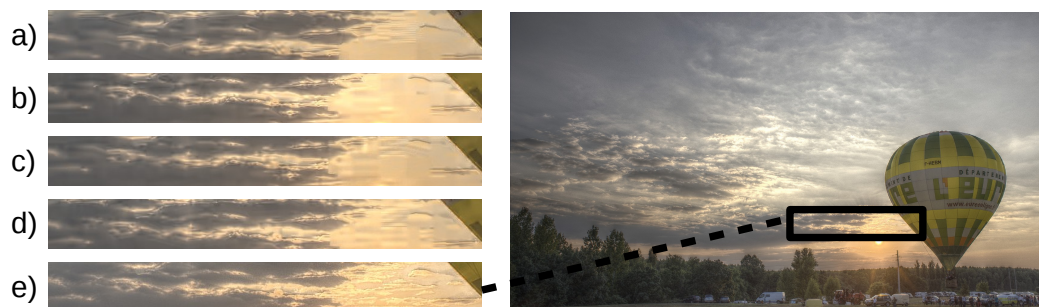
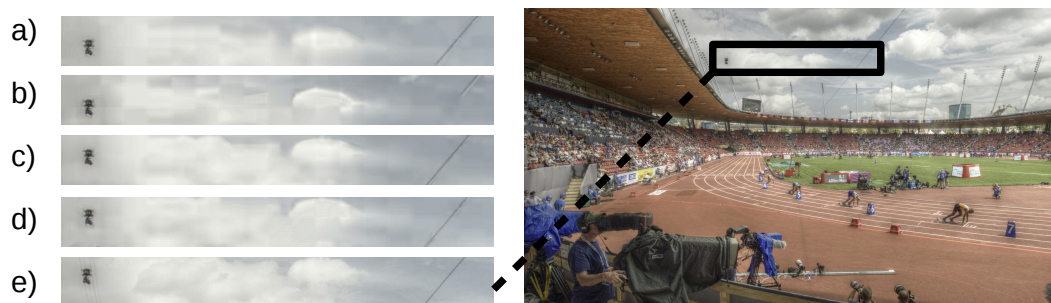| Method	Mai *et al.* (2011)	Mai *et al.* (2013)	PQ	ST-TMO
Run-time (average sec. per frame)	1.21	269.15	3.212	370.35

18

(a) *Market3Clip4000r2*, 3600 *kbps*
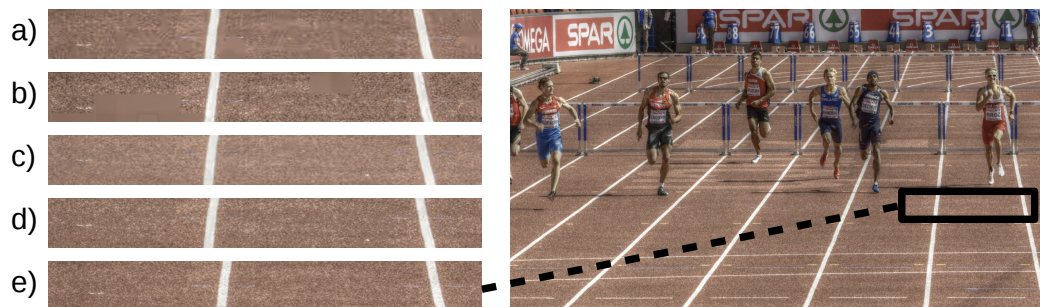

(b) *FireEater2Clip4000r1*, 3500 *kbps*
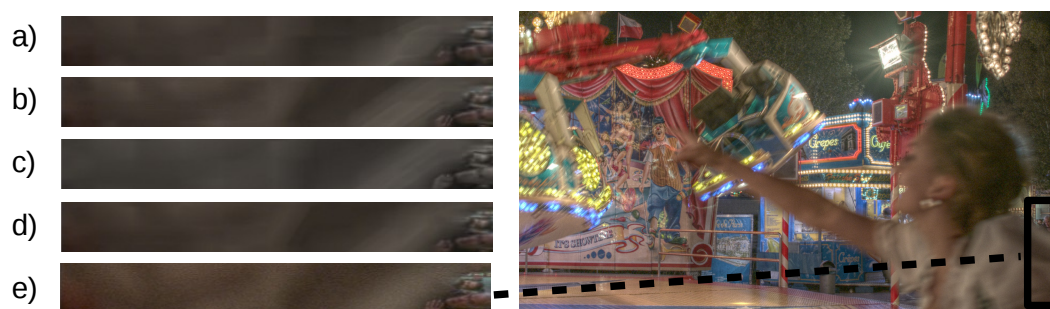

(c) *SunRise*, 640 *kbps*

(d) *EBU_04_Hurdles*, 4400 *kbps*

Figure 9: Examples of reconstructed HDR frames (details) using: **a)** Mai *et al.* (2011), **b)** Mai *et al.* (2013), **c)** PQ, and **d)** the proposed ST-TMO. **e)** represents the original HDR frame. Each HDR frame is displayed by the Luminance HDR software [33] using Mantiuk'06 tone mapping for visualization.

19

(a) *EBU_06_Starting*, 4000 *kbps*

(b) *Carousel_fireworks_03*, 4200 *kbps*

(c) *Carousel_fireworks_04*, 5500 *kbps*

(d) *Smith_Welding*, 1500 *kbps*

Figure 10: *(Continued)* Examples of reconstructed HDR frames (details) using: **a)** Mai *et al.* (2011), **b)** Mai *et al.* (2013), **c)** PQ, and **d)** the proposed ST-TMO. **e)** represents the original HDR frame. Each HDR frame is displayed by the Luminance HDR software [33] using Mantiuk'06 tone mapping for visualization.

tone mapping of Intra frame takes relatively little time (0.491 seconds on average in our experiments), the tone mapping time for Inter frames is much longer (383.10 seconds on average). About 30% of the time for Inter-frame optimization is actually due to optical flow motion estimation. However, this might probably be speeded up by more efficient motion estimation techniques, and indeed it is an interesting future research to investigate how the performance would be influenced by the choice of motion estimation approach. The remaining complexity is due to convex optimization, and is similar in magnitude to Mai *et al.* (2013), which also uses similar tools. Notice that our Matlab implementation is not optimized for performance. Much more efficient parallel implementations of proximal optimization methods exist, which can make use of GPUs with considerable time savings [22].

5. Conclusion

In this paper, to achieve efficient video coding performance for HDR video, a content-adaptive tone mapping algorithm is introduced that efficiently minimizes the spatial *and* temporal complexity of the tone mapped video. The proposed coding solution considers high bit-depth coding profile of the high efficient video coding (HEVC) standard, which uses 10-bit depth. To evaluate the performance of the proposed TMO, existing content-adaptive TMOs and the perceptual quantizer (PQ) transfer function were considered and extensive experimental results showed that the proposed TMO solution outperforms the reference methods. Despite the remarkable rate distortion (RD) gains achieved, the proposed TMO still has room for enhancement. Future work will focus to further increase the coding performance with the help of perceptual and color encoding techniques.

6. Appendix

By expanding Eq. (9) we obtain:

$$C_{temp}(f_{inter}^{SDR}) = \sum_{w=0}^{v_{max}} \sum_{k=1}^{N} \left\{ \left[\frac{d_k^2}{4} + d_k \sum_{i=1}^{k-1} d_i + \left(\sum_{i=1}^{k-1} d_i \right)^2 - d_k w - 2w \sum_{i=1}^{k-1} d_i + w^2 \right] p_{k,w} \right\}. \tag{14}$$

By recalling the definition of $d_k = \delta s_k$, we can rewrite Eq. (6) in terms of s_k:

$$\begin{aligned}
C_{temp}(f_{inter}^{SDR}) = \\
\delta^2 \sum_{w=0}^{v_{max}} \sum_{k=1}^{N} \left\{ \left(\frac{s_k^2}{4} + \left(\sum_{i=1}^{k-1} s_i \right)^2 + s_k \sum_{i=1}^{k-1} s_i \right) p_{k,w} \right\} \\
+ \delta \sum_{w=0}^{v_{max}} \sum_{k=1}^{N} \left\{ \left(-s_k w - 2w \sum_{i=1}^{k-1} s_i \right) p_{k,w} \right\} \\
+ \sum_{w=0}^{v_{max}} \sum_{k=1}^{N} w^2 p_{k,w}.
\end{aligned} \tag{15}$$

21

Factoring the s_k terms, we can write $C_{temp}(f_{inter}^{SDR})$ as a quadratic function of the vector s:

$$C_{temp}(f_{inter}^{SDR}) = s^T \overbrace{\left[\sum_{w=0}^{v_{max}} \sum_{k=1}^{N} A_k p_{k,w} \right]}^{W_2} s$$
$$+ s^T \overbrace{\left[\sum_{w=0}^{v_{max}} \sum_{k=1}^{N} b_{k,w} p_{k,w} \right]}^{W_1} \qquad (16)$$
$$+ \overbrace{\sum_{w=0}^{v_{max}} \sum_{k=1}^{N} w^2 p_{k,w}}^{W_0},$$

with:

$$A_k = \frac{1}{4} A_k^{(1)} + A_k^{(2)} + A_k^{(3)},$$
$$b_{k,w} = -(b_{k,w}^{(1)} + b_{k,w}^{(2)}), \qquad (17)$$

where $A_k^{(1)}$, $A_k^{(2)}$, and $A_k^{(3)}$ are $N \times N$ matrices and $b_{k,w}^{(1)}$ and $b_{k,w}^{(2)}$ are N-element column vectors, defined in terms of k and w as follows. $A_k^{(1)}$ is such that $A_k^{(1)}(i,j) = 1$ if $i = j = k$ and $A_k^{(1)}(i,j) = 0$ elsewhere. $A_k^{(2)}$ is such that $A_k^{(2)}(i,j) = 1$ if $i = k \wedge 1 \leq j \leq k-1$ and zero otherwise. $A_k^{(3)}$ is such that $A_k^{(3)}(i,j) = 1$ if $1 \leq i \leq k-1 \wedge 1 \leq j \leq k-1$ and zero otherwise. The vector $b_{k,w}^{(1)}$ is such that $b_{k,w}^{(1)}(i) = w$ if $i = k$ and zero otherwise. Finally $b_{k,w}^{(2)}(i) = 2w$ if $1 \leq i \leq k-1$ and is zero elsewhere.

By noting that the zero-order term W_0 in Eq. (16) does not depend on s, we obtain the cost function of Eq. (9).

Acknowledgment

References

[1] T. O. Aydin, R. Mantiuk, and H.-P. Seidel, *Extending quality metrics to full dynamic range images*, Human vision and electronic imaging xiii, 2008Jan.

[2] T. O. Aydin, N. Stefanoski, S. Croci, M. Gross, and A. Smolic, *Temporally Coherent Local Tone Mapping of HDR Video*, ACM Transactions on Graphics **33** (November 2014), no. 6, 196:1–196:13.

[3] G. J. Barten, *Formula for the contrast sensitivity of the human eye*, Electronic imaging 2004, 2003, pp. 231–238.

[4] G. Bjøtegaard, *Calculation of average PSNR differences between RD-curves (vceg-m33)*, Technical Report M16090, VCEG Meeting (ITU-T SG16 Q.6), Austin, Texas, USA, 2001.

[5] R. Boitard, K. Bouatouch, R. Cozot, D. Thoreau, and A. Gruson, *Temporal coherency for video tone mapping*, Society of photo-optical instrumentation engineers (SPIE) conference series, 2012oct.

[6] R. Boitard, R. Cozot, D. Thoreau, and K. Bouatouch, *Zonal brightness coherency for video tone mapping*, Signal Processing: Image Communication **29** (2014feb), no. 2, 229–246.

[7] R. Boitard, M.T. Pourazad, P. Nasiopoulos, and J. Slevinsky, *Demystifying high-dynamic-range technology: A new evolution in digital media.*, IEEE Consumer Electronics Magazine **4** (2015Oct), no. 4, 72–86.

[8] R. Boitard, D. Thoreau, R. Cozot, and K. Bouatouch, *Impact of temporal coherence-based tone mapping on video compression*, 21st European Signal Processing Conference (EUSIPCO), 2013Sept, pp. 1–5.

[9] ———, *Motion-guided quantization for video tone mapping*, IEEE International Conference on Multimedia and Expo (ICME), 2014July, pp. 1–6.

[10] T. Borer, *Non-linear opto-electrical transfer functions for high dynamic range television*, Technical Report ITU-R WP6C Contribution 369, British Broadcasting Corporation (BBC), 2014.

[11] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, *Image restoration: total variation, wavelet frames, and beyond*, Journal of the American Mathematical Society **25** (2012), no. 4, 1033–1089.

[12] A. Chalmers and K. Debattista, *HDR video past, present and future: A perspective*, Signal Processing: Image Communication **54** (2017), 49–55.

[13] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision **40** (2011may), no. 1, 120–145.

[14] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, *Epigraphical splitting for solving constrained convex formulations of inverse problems with proximal tools*, arXiv preprint arXiv:1210.5844 (2012).

[15] P. L. Combettes and J.-C. Pesquet, *Fixed-point algorithms for inverse problems in science and engineering* (H. Heinz Bauschke, S. Regina Burachik, L. Patrick Combettes, Veit Elser, Russell D. Luke, and Henry Wolkowicz, eds.), Springer New York, New York, NY, 2011.

[16] ———, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators*, Set-Valued and Variational Analysis **20** (2011), no. 2, 307–330.

[17] F. De Simone, G. Valenzise, P. Lauga, F. Dufaux, and F. Banterle, *Dynamic range expansion of video sequences: A subjective quality assessment study*, IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2014Dec, pp. 1063–1067.

[18] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, *Adaptive logarithmic mapping for displaying high contrast scenes*, Computer graphics forum, 2003, pp. 419–426.

[19] F. Dufaux, P. Le Callet, R. Mantiuk, and M. Mrak, *High dynamic range video: From acquisition, to display and applications*, Academic Press, 2016.

[20] E François, C Fogg, Y He, X Li, A Luthra, and A Segall, *High dynamic range and wide color gamut video coding in HEVC: Status and potential future enhancements*, IEEE Transactions on Circuits and Systems for Video Technology **26** (January 2016), no. 1, 63–75.

[21] E. Francois, C. Fogg, Y. He, X. Li, A. Luthra, and A. Segall, *High Dynamic Range and Wide Color Gamut Video Coding in HEVC: Status and Potential Future Enhancements*, IEEE Transactions on Circuits and Systems for Video Technology **26** (2016Jan), no. 1, 63–75.

[22] Raffaele Gaetano, Giovanni Chierchia, and Béatrice Pesquet-Popescu, *Parallel implementations of a disparity estimation algorithm based on a proximal splitting method*, Ieee int. conf. on visual communications and image processing (vcip), 2012, pp. 1–6.

[23] J U Garbas and H Thoma, *Temporally coherent luminance-to-luma mapping for high dynamic range video coding with H.264/AVC*, 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP), May 2011, pp. 829–832.

[24] GPU4vision, *http://gpu4vision.icg.tugraz.at/*, 2015.

[25] P. Hanhart, M. V. Bernardo, M. Pereira, A. MG Pinheiro, and T. Ebrahimi, *Benchmarking of objective quality metrics for HDR image quality assessment*, EURASIP Journal on Image and Video Processing **2015** (2015), no. 1, 1–18.

[26] P.910 ITU-T Recommendation, *Subjective video quality assessment methods for multimedia applications*, ITU-T, 1999.

[27] L Kerofsky, Y Ye, and Y He, *Recent developments from MPEG in HDR video compression*, 2016 IEEE international conference on image processing (ICIP), 2016, pp. 879–883.

[28] I.-K. Kim, K. McCann, K. Sugimoto, B. Bross, and W.-J. Han, *High efficiency video coding (HEVC) test model 10 (HM10) encoder description*, Technical Report N12242, ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, 2013.

[29] A. Koz and F. Dufaux, *Methods for improving the tone mapping for backward compatible high dynamic range image and video coding*, Signal Processing: Image Communication **29** (2014), no. 2, 274–292.

[30] G. W. Larson, *Overcoming gamut and dynamic range limitations in digital images*, Color and imaging conference, 1998, pp. 214–219.

[31] G. W. Larson, H. Rushmeier, and C. Piatko, *A visibility matching tone reproduction operator for high dynamic range scenes*, IEEE Transactions on Visualization and Computer Graphics **3** (1997Oct), no. 4, 291–306.

[32] P. Lauga, G. Valenzise, G. Chierchia, and F. Dufaux, *Improved tone mapping operator for HDR coding optimizing the distortion/spatial complexity trade-off*, 22nd European Signal Processing Conference (EUSIPCO), 2014Sept, pp. 1607–1611.

[33] Luminance HDR, *http://qtpfsgui.sourceforge.net/*, 2018.

[34] A. Luthra, E. Francois, and W. Husak, *Call for Evidence (CfE) for HDR and WCG Video Coding*, Technical Report MPEG2011/N12036, ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, 2015.

[35] 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** (2011June), no. 6, 1558–1571.

[36] Z. Mai, H. Mansour, P. Nasiopoulos, and R. K. Ward, *Visually Favorable Tone-Mapping With High Compression Performance in Bit-Depth Scalable Video Coding*, IEEE Transactions on Multimedia **15** (2013Nov), no. 7, 1503–1518.

[37] Stéphane Mallat, *A wavelet tour of signal processing*, Academic press, 1999.

[38] R. Mantiuk, S. Daly, and L. Kerofsky, *Display adaptive tone mapping*, ACM Transactions on Graphics (TOG) **27** (2008Aug), no. 3, 68.

[39] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, *A perceptual framework for contrast processing of high dynamic range images*, ACM Transactions on Applied Perception (TAP) **3** (2006), no. 3, 286–308.

[40] S. Miller, M. Nezamabadi, and S. Daly, *Perceptual signal coding for more efficient usage of bit codes*, Annual technical conference exhibition, SMPTE 2012, 2012Oct, pp. 1–9.

[41] K. Minoo, T. Lu, P. Yin, L. Kerofsky, D. Rusanovskyy, and E. Francois, *Description of the reshaper parameters derivation process in etm reference software*, Technical Report JCTVC-W0031 ITU-T Q6/16, ITU-T Q6/16, San Diego, CA, USA, 2016.

[42] A. Motra and H. Thoma, *An adaptive logluv transform for high dynamic range video compression*, 17th IEEE International Conference on Image Processing (ICIP), 2010Sept, pp. 2061–2064.

[43] M. Narwaria, M. P. Da Silva, and P. Le Callet, *HDR-VQM: An objective quality measure for high dynamic range video*, Signal Processing: Image Communication **35** (2015), 46 –60.

[44] M. Narwaria, M. P. Da Silva, P. Le Callet, and R. Pépion, *Impact of tone mapping in high dynamic range image compression*, International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), 2014Jan, pp. 1–6.

[45] M. Narwaria, R. K. 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 **24** (2015), no. 1, 010501–010501.

[46] C. Ozcinar, P. Lauga, G. Valenzise, and F. Dufaux, *HDR Video Coding based on a temporally constrained Tone Mapping Operator*, 2016 Digital Media Industry Academic Forum (DMIAF), 2016July, pp. 43–47.

[47] M. L. Pendu, C. Guillemot, and D. Thoreau, *Adaptive re-quantization for high dynamic range video compression*, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014May, pp. 7367–7371.

[48] J.-C. Pesquet and N. Pustelnik, *A parallel inertial proximal optimization method*, Pacific Journal of Optimization **8** (2012), no. 2, 273–305.

[49] Recommendation ITU-R BT.1886, *Reference electro-optical transfer function for flat panel displays used in HDTV studio production*, Recommendation ITU-R BT.1886, 2011.

[50] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, *Photographic tone reproduction for digital images*, ACM Transactions on Graphics **21** (2002July), no. 3, 267–276.

[51] Leonid I. Rudin, Stanley Osher, and Emad Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D **60** (November 1992), no. 1-4, 259–268.

[52] J. Samuelsson, M. Pettersson, J. Storm, and K. Andersson, *Using chroma qp offset on hdr sequences*, Technical Report m36581, MPEG, Warsaw, Poland, 2015.

[53] Z. Wang, A.C. Bovik, R. Sheikh Hamid, and E.P. Simoncelli, *Image quality assessment: from error visibility to structural similarity*, IEEE Transactions on Image Processing **13** (2004Apr), no. 4, 600–612.

[54] P. Yin, A. M. Tourapis, and W. Husak, *Comments on the HDR Demos at the 110th MPEG Meeting*, Technical Report MPEG2014/m35470, SO/IEC JTC1/SC29/WG11, Lausanne, Switzerland, 2014.

[55] E. Zerman, G. Valenzise, and F. Dufaux, *An extensive performance evaluation of full-reference HDR image quality metrics*, Quality and User Experience **2** (2017), no. 1, 5.

[56] Emin Zerman, Vedad Hulusic, Giuseppe Valenzise, Rafal Mantiuk, and Frédéric Dufaux, *Effect of color space on high dynamic range video compression performance*, Proc. 9th int. conf. on quality of multimedia experience, 2017, pp. 1–6.

[57] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak, and D. R. Bull, *High dynamic range video compression by intensity dependent spatial quantization in HEVC*, Picture coding symposium (PCSx), 2013, 2013Dec, pp. 353–356.

[58] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak, and D.R. Bull, *High dynamic range video compression exploiting luminance masking*, IEEE Transactions on Circuits and Systems for Video Technology **PP** (2015), no. 99, 1–1.

[59] Y Zhang, E Reinhard, and D Bull, *Perception-based high dynamic range video compression with optimal bit-depth transformation*, 2011 18th IEEE international conference on image processing, 2011, pp. 1321–1324.