Art-Directable Continuous Dynamic Range Video

Alexandre Chapiro[1,2]   Tunç O. Aydın[2]   Nikolce Stefanoski[2]   Simone Croci[2]   Aljoscha Smolic[2]   Markus Gross[1,2]

[1]ETH Zurich   [2]Disney Research

Abstract

We present a novel, end-to-end workflow for content creation and distribution to a multitude of displays that have different dynamic ranges. The emergence of new, consumer level HDR displays with various peak luminances expected in 2015 gives rise to two new research questions: (i) how can the raw source content be graded for a diverse set of displays both efficiently and without restricting artistic freedom, and (ii) how can an arbitrary number of graded video streams be represented and encoded in an efficient way. In this work we propose a new editing paradigm which we call *dynamic range mapping* to obtain a novel *Continuous Dynamic Range (CDR)* video representation, where the luminance of the video content, instead of being a scalar value, is defined as a continuous function of the display dynamic range. We present an interactive interface where CDR videos can be efficiently created while providing full artistic control. In addition, we discuss the efficient approximation of CDR video using a polynomial series approximation, and its encoding and distribution to an arbitrary set of target displays. We validate our workflow in a subjective study, which suggests that a visually lossless CDR video representation can be achieved with little bandwidth overhead. Our solution can be implemented easily in the current distribution infrastructure and consists of transmitting two gradings and an additional meta-data stream, which occupies less than 13% current standard video distribution bandwidth.

Keywords: HDR, Continuous Dynamic Range, Dynamic Range Mapping, Lumipath

1. Introduction

After years of research and development, we are finally about to witness the emergence of High Dynamic Range (HDR) content distribution and display at the consumer level. While high-end cameras (such as the Red Epic Dragon, SonyF55 and F65, and ARRI Alexa XT) have been able to natively capture HDR video, up to now displaying HDR content has only been possible through research prototypes or custom built hardware. This landscape is rapidly changing as TV manufacturers including LG, Sony, Samsung, Panasonic and TCL have announced HDR displays with various peak luminances and black levels, which they plan to release in 2015. On the content creation side, Technicolor and the Sinclair Broadcast Group successfully demonstrated over-the-air broadcast of UltraHD HDR content and Technicolor now offers HDR grading services. Netflix and Amazon announced the upcoming start of HDR content streaming services. In the meantime, experimental HDR short films [1, 2] explored the creative use of HDR imaging in film making.

These efforts towards realizing an HDR content production and distribution pipeline from capture to display are fueled by the massive difference that HDR makes in viewing experience [3]. The significance of the increase in experience quality provided by HDR over *standard dynamic range* (SDR) is becoming widely accepted as common knowledge. As a result, the focus point of the next generation viewing experience is shifting from *more* pixels to obtaining *better* pixels by extending their dynamic range, among other factors.

The emergence of HDR displays from multiple vendors with different dynamic ranges creates some significant challenges for content production and distribution. Specifically, the *production challenge* is tailoring HDR content to a number of upcoming displays which are announced to have peak luminances ranging from 800-4000 nits, as well as future HDR displays with different dynamic ranges. The straightforward approach of grading content for each specific display dynamic range does not scale well due to the required additional manual labor. Methods proposed in literature, such as display adaptive tone mapping [4], can alleviate this issue, but do not allow for precise artistic freedom in the expression of brightness variations.

The *distribution challenge* is the task of efficiently coding and transmitting a large number of HDR streams graded for different display dynamic ranges. Previous work proposed distributing a single HDR stream efficiently as a residual signal over the SDR content [5]. This approach, however, is not well suited for application in the emerging landscape where numerous HDR streams are required to be transmitted simultaneously.

The content creation and distribution challenges force us to rethink the way raw source content is graded for a multitude of target displays, and how graded content can be efficiently represented. In this work we propose a new content creation paradigm which we call *Dynamic Range Mapping*, where raw source content is graded not only for a single display (as in
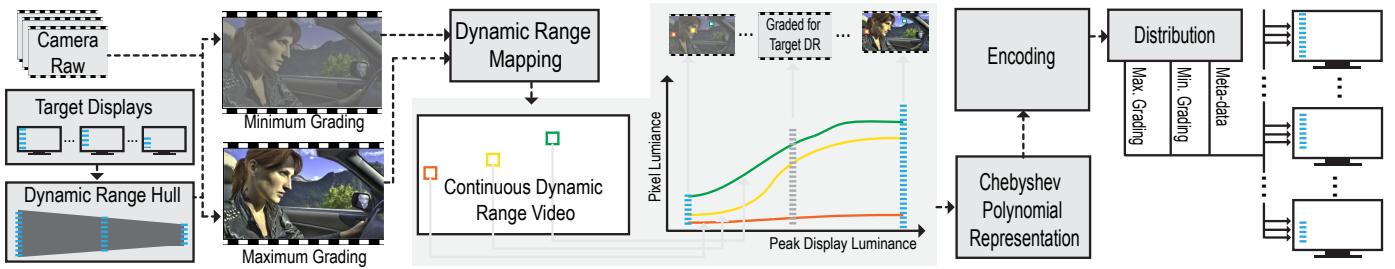
Figure 1: *We present an end-to-end pipeline for efficient content creation and distribution of HDR content for a multitude of target display dynamic ranges. Our workflow starts by color grading the source content for the largest and smallest of the target dynamic ranges among the set of targeted dynamic ranges. Next, through an interactive "dynamic range mapping" we obtain "continuous dynamic range" (CDR) video, where each pixel contains an art-directable function rather than a scalar value. We approximate this CDR video using Chebyshev Polynomials, and encode it for efficient distribution to the target displays.*

traditional tone mapping), but for a dynamic range continuum that entails the dynamic ranges of an arbitrary number of target displays. Unlike tone mapping where the resulting pixels have scalar luminance values, dynamic range mapped pixels are defined by art-directable functions of display dynamic range. We call this new data structure *Continuous Dynamic Range Video* and propose a method for its efficient representation and distribution. Specifically, our work makes the following contributions:

- A practical dynamic range mapping workflow, allowing the creation of continuous dynamic range video with full artistic control.

- An efficient representation of continuous dynamic range video using a polynomial series approximation.

- A demonstration of efficient encoding of continuous dynamic range video.

The individual components presented in this paper form an end-to-end solution for efficiently creating, representing and distributing content graded for an arbitrary number of target displays with different dynamic ranges. In our solution, continuous dynamic range content is efficiently represented by two video streams (graded for the highest and lowest target dynamic ranges) and an additional meta-data stream that occupies less than 13% of the current standard business-to-consumer video distribution bandwidth.

2. Related Work

In this section we discuss relevant work on HDR image and video tone mapping, HDR display and distribution.

Tone mapping of HDR images has been studied extensively in the literature. A comprehensive overview can be found in Reinhard et al. [6]. Early image tone mapping operators have been heavily influenced by the photographic film development process. The photographic tone mapping operator [7] utilizes an S-shaped curve to globally compress the input dynamic range, as well as dodging and burning operations to control local details. Another tone mapping approach aimed to produce natural looking results by modeling various mechanisms of the human visual system [8, 9, 10]. Durand and Dorsey [11] proposed decomposing HDR images into base and detail layers by utilizing edge-aware filtering. They showed that local image details can be preserved by restricting tonal compression to the base layer while keeping the detail layer intact. Similar effects were also achieved by processing the input HDR image in the gradient domain [12, 13].

While most tone mapping operators target a single hypothetical SDR display, the display adaptive tone mapping [4] approach tailors its outcome for a user-selected display dynamic range. Our CDR video representation can be thought of the union of content tone mapped for all possible displays. Additionally, our dynamic range mapping workflow does not restrict the user to a single tone mapping approach, as the source content graded for the smallest and largest dynamic range can be generated manually or using any tone mapping operator.

Tone mapping of HDR video has recently become an active field of research. The majority of the various video tone mapping operators have been discussed and subjectively evaluated by Eilertsen et al. [14]. More recently, Boitard et al. [15] proposed segmenting each video frame (typically to 2−4 segments) and applying a global tone curve to each segment individually. Local adaptation is introduced at a segment level at the cost of more complex processing involving video segmentation. Another recent operator achieved temporally coherent local tone mapping through efficient spatiotemporal filtering [16].

While tone mapping is a useful tool for displaying HDR content on SDR devices, the research community has long aspired to develop displays that can natively reproduce HDR images. While reproducing the entire range that the human eye can see may prove difficult, it has been subjectively shown that a luminance range from $0 − 10,000$ nits satisfies 90% of the viewers who were asked to select an ideal range [17]. The first prototype HDR display has been introduced by Seetzen et al. [18], which was then followed by multiple custom-built research prototypes [19, 20, 21, 22, 23]. For a detailed discussion on the various HDR display approaches we refer the reader to Reinhard et al. [6]. In parallel, experimental HDR displays have been introduced by private enterprises such as Brightside, SIM2 and Dolby. More recently, major TV manufacturers including LG, Sony, Samsung, Panasonic and TCL have announced the up-

coming release of their consumer-level HDR displays. While many of these displays are being marketed using the term HDR, their dynamic ranges are quite different from each other (peak luminances varying from $800 - 4000$ nits). As a consequence of these emerging displays, traditional content production and distribution methods have to be revisited.

The efficient distribution of HDR content has also been investigated by various researchers. Mantiuk et al. [5] proposed encoding HDR video as a residual stream over its SDR counterpart with an overhead of 30%. More recent work proposed an optimized bit-depth quantization and human visual system based wavelet transform denoising for HDR compression [24], and also investigated the distribution of HDR video using existing codecs such as H.264/AVC [25].

3. Continuous Dynamic Range Video

We propose Continuous Dynamic Range (CDR) as a novel way of representing video within a continuum of dynamic ranges. For practical purposes, it is important that the CDR representation is both efficient and allows full artistic freedom. In this section we will introduce key concepts and components of CDR video and discuss artistic control. The efficient approximation and encoding of CDR video will be discussed in Section 4.

3.1. Key Concepts

A high-level overview of our pipeline is illustrated in Fig. 1. The goal of our method is distributing the input source video to a number of *target displays*, where the grading for each of the target displays can be art directed. The first input to our method is the source video in camera raw format. While the dynamic range of the source video can be arbitrary, in this work we used HDR content with up to 14 f-stops. Formally, we denote a frame of the raw input video as $\mathtt{I}$, its color at pixel p as $\mathtt{I}^p$, and the corresponding luminance as $\mathcal{L}(\mathtt{I}^p)$.

Since the dynamic range continuum encompassed by a CDR video is a superset of the dynamic ranges of all target displays, we also require the user to specify a *dynamic range (DR) hull*. The DR hull defines a dynamic range continuum between the *minimum* and *maximum dynamic ranges*. The minimum dynamic range is bounded by the min peak luminance and max black level among the set of all target display dynamic ranges. Analogously, the maximum dynamic range is bounded by the max peak luminance and min black level (Fig. 2).

The first artistic interaction in our pipeline is the grading of the raw content for the minimum and maximum dynamic ranges to obtain *minimum* and *maximum gradings*, which we denote with $\mathtt{I}_\alpha$ and $\mathtt{I}_\beta$. Here, the user has full freedom in terms of tools to be used and edits to be performed, as long as the spatial correspondence between the pixels of the minimum and maximum gradings are preserved. We denote the minimum and peak luminance of $\mathtt{I}_\alpha$ by η_α and π_α, respectively, and the minimum and peak luminance of $\mathtt{I}_\beta$ by η_β and π_β, respectively.
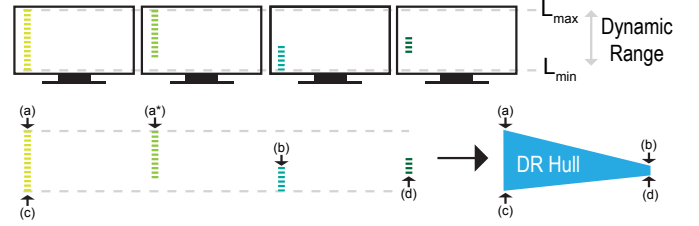

Figure 2: *The dynamic range hull is a superset of the dynamic ranges of all target displays. Colored bars represent the dynamic range of a display.*

After these traditional grading processes, the next artistic interaction step is *dynamic range mapping* where the user specifies how the pixel luminances change across the dynamic range hull. Functions of the following type have to be generated

$$h^p : [\eta_\alpha, \eta_\beta] \times [\pi_\alpha, \pi_\beta] \to [\mathcal{L}(\mathtt{I}_\alpha^p), \mathcal{L}(\mathtt{I}_\beta^p)], \tag{1}$$

which associate with each pixel p and dynamic range (η, π) a unique luminance value $h^p(\eta, \pi)$. To reduce the complexity of generating these functions and the amount of distributed data, we restrict the domain to $[\pi_\alpha, \pi_\beta]$ and define the associated minimum luminace for any $\pi \in [\pi_\alpha, \pi_\beta]$ by $\eta(\pi) := \eta_\alpha + (\eta_\beta - \eta_\alpha)\frac{\pi - \pi_\alpha}{\pi_\beta - \pi_\alpha}$. Thus, $(\eta(\pi), \pi) \ \forall \ \pi \in [\pi_\alpha, \pi_\beta]$ defines the considered DR hull.

Consequently, at each pixel a user-defined function, which we call a *lumipath*, represents the pixel's luminance value as a function of the peak luminance π of a target display. Formally, we define a lumipath as

$$g^p : [\pi_\alpha, \pi_\beta] \to [\mathcal{L}(\mathtt{I}_\alpha^p), \mathcal{L}(\mathtt{I}_\beta^p)], \tag{2}$$

where π_α and π_β are the peak luminances corresponding to the maximum and minimum dynamic ranges. The end result of the dynamic range mapping process, namely the *continuous dynamic range* video, stores a lumipath at each pixel rather than a scalar luminance value. In our current implementation we transform the graded image pair $\mathtt{I}_\alpha$ and $\mathtt{I}_\beta$ to the CIE YUV color space, where the variation of the Y channel across the dynamic range hull is controlled by the user defined lumipaths, and the chrominance channels are linearly interpolated.

3.2. Dynamic Range Mapping

We implemented a user interface (Fig. 4) for convenient dynamic range mapping. Users can select desired image regions by using masks and adjust the corresponding lumipaths by modifying the control points of a third degree polynomial spline interface. While we chose this particular representation on the based on the standard tone curve interfaces in commercial color grading software, other tools could be employed to a similar effect. Formally, given a series of image masks $\mathtt{M}_j$ with values $\mathtt{M}_j^p \in [0, 1]$, the user manually specifies functions $k_j : [\pi_\alpha, \pi_\beta] \to [\pi_\alpha, \pi_\beta]$ with the user interface. When applied to each pixel, the function is modulated at each pixel position by the mask, and k_j^p is obtained as follows:

$$k_j^p(\pi) = \mathtt{M}_j^p k_j(\pi) + (1 - \mathtt{M}_j^p)\pi. \tag{3}$$

This defines a blending between the artist defined curve and a linear curve based on the weights specified by the mask, allowing for smoothly varying edits. By employing n masks and specifying n such functions, the corresponding lumipaths g^p are obtained by applying all functions successively (layer based grading) and scaling the result:

$$g^p = \frac{k_1^p \circ ... \circ k_n^p - \pi_\alpha}{\pi_\beta - \pi_\alpha} \left(\mathcal{L}(\mathbf{I}_\beta^p) - \mathcal{L}(\mathbf{I}_\alpha^p) \right) + \mathcal{L}(\mathbf{I}_\alpha^p). \quad (4)$$

The lumipath $g^p : [\pi_\alpha, \pi_\beta] \rightarrow [\mathcal{L}(\mathbf{I}_\alpha^p), \mathcal{L}(\mathbf{I}_\beta^p)]$ is the desired curve defining the luminance of the pixel p for any display with maximum brightness between the two analyzed extremes. This process is illustrated in Fig. 3.

Figure 3: *A visual representation of the process of obtaining a numerical lumipath (Eqs. 3 and 4) is shown here. Lumipaths input by an artist are averaged with linear functions according to the weights specified in the interface and subsequently concatenated to obtain the final per-pixel lumipath g^p.*

In practice, the dynamic range mapping process begins by specifying maximum and minimum gradings to our tool. The user can additionally import multiple masks that can be generated using modern video editing software (e.g. Resolve, Nuke, etc.). Our user interface, which is rendered on a standard LCD display, provides interactive visual feedback on an external HDR display as the lumipaths for the selected region are modified. Visualization is provided by computing and rendering a user defined number of gradings (Fig. 4-right). Since (i) there are no restrictions in how the input gradings are obtained (except preserving pixel correspondences), (ii) any number of pixel-level masks for region selection can be used, and (iii) the lumipaths can be defined precisely using any number of control points, our method allows for significant artistic freedom during dynamic range mapping.

4. Efficient Approximation and Coding

CDR video in its raw form is represented by a considerable amount of data, where each frame f comprises (i) an LDR image $\mathbf{I}_\alpha^f$, (ii) an HDR image $\mathbf{I}_\beta^f$, and (iii) lumipaths $g^{p,f}$ for every pixel of a frame. In this section we describe how we solve the distribution challenge by efficiently approximating and coding CDR video.

LDR and HDR image sequences can be jointly compressed with dedicated coding methods like [5] or other methods which are currently the subject of intensive investigations in MPEG [26]. In this work, we assume that the image sequences of $\mathbf{I}_\alpha^f$ and $\mathbf{I}_\beta^f$ are already encoded. In this section a first exploration of the compressibility of the remaining data - the lumipaths - is performed.

Our compression approach can be subdivided into two parts: we begin by approximating the lumipath functions in a perceptually lossless way using a polynomial series, followed by a representation of the coefficients in an image-like format and encoding using a video compression method.

4.1. Approximation

The first step towards the efficient compression of this information is a suitable approximation of the individual lumipath functions. The goal is to find a representation of all lumipaths, which should be both compact and visually indistinguishable from the original. Our approach consists of approximating each lumipath by a series of functions. The series is truncated at a point where the resulting output is visually lossless based on a human visual system model. The result is a representation of each lumipath by a finite set of coefficients with respect to a polynomial basis. These coefficients are later further compressed with the help of a standard video codec.

Our human visual system model consists of a *threshold-versus-intensity (*tvi*)* function that computes an approximate threshold luminance, given the level of luminance adaptation ($\mathcal{L}_a$). The tvi function is computed by finding the peak contrast sensitivity at each luminance level as described in previous work [27, 28]:

$$\mathrm{tvi}(\mathcal{L}_a^p) = \frac{\mathcal{L}_a^p}{\max_x \left(\mathrm{CSF}\left(x, \mathcal{L}_a^p \right) \right)}, \quad (5)$$

where CSF denotes the contrast sensitivity function, and $\mathcal{L}_a^p$ denotes the adaptation luminance for pixel p. To avoid introducing visual artifacts, we make the conservative assumption that the human eye can adapt perfectly to a single pixel p. In practice, we found that even such a conservative threshold estimation can significantly reduce the number of required polynomial basis coefficients. In our experiments, the number of coefficients did not exceed 20.

In mathematical analysis, the *Weierstrass approximation theorem* shows that a continuous real-valued function $f : [a, b] \rightarrow [c, d]$ can always be uniformly approximated by a polynomial series. Approximation by simple functions is desirable because they can be easily computed and evaluated. Several bases of the space of polynomials can be used for such an approximation, but while all of them may converge, not all perform equally well for a given problem.

A common method for approximating functions with a polynomial basis consists of using Chebyshev series [29]. Chebyshev polynomials have some very useful properties that make them desirable for our problem, namely (i) they are guaranteed

Figure 4: *This figure shows our luminance grading interface. On the SDR display, masks with values in* $[0, 1]$ *can be loaded in the bottom left menu, and are displayed in the bottom-middle window. A cubic spline interface shown on the bottom right allows the user to manually input lumipaths. Different visualization options can be selected from the menu on the top. On the HDR display, users can visualize their edits in an interactive manner. To see a standard work session using our interface, we point the reader to the supplementary material of this paper.*

to minimize Runge's phenomenon when approximating in an interval (this is particularly important since in practice most displays are located near the minimum end of the examined dynamic range hulls), (ii) they can be quickly computed numerically, and (iii) the error of the approximated function as compared to the original can be estimated from the calculated coefficients, which is important as a stopping criterion.

Our goal is to approximate a lumipath $g^{p,f}$ at a given pixel in a perceptually lossless way by a truncated Chebyshev series $\bar{g}^{p,f}$ such that $\left\| g^{p,f} - \bar{g}^{p,f} \right\|_\infty < \mathrm{tvi}(\mathcal{L}_a^p)$, i.e. the deviation is smaller than the threshold computed by our model of the human visual system. The truncated Chebyshev series is represented by

$$\bar{g}^{p,f}(x) = \sum_{k=0}^{N_{p,f}} c_k^{p,f} \psi_k(x) \qquad (6)$$

where $\psi_k(x)$ is the k-th Chebyshev polynomial, $c_k^{p,f}$ the corresponding Chebyshev coefficient at pixel p of frame f, and $N_{p,f}$ is the smallest degree required to obtain an error $\left\| g^{p,f} - \bar{g}^{p,f} \right\|_\infty$ which is smaller than $\mathrm{tvi}(\mathcal{L}_a^p)$. This defines a perceptually lossless approximation of $g^{p,f}$ which is determined by $N_{p,f} + 1$ coefficients $c_0^{p,f}, \ldots, c_{N_{p,f}}^{p,f}$.

We implement our computation of the Chebyshev series as outlined by Broucke [30]. For simplicity we consider normalized lumipaths, i.e. the domain and the range of all lumipaths is scaled such that all of them lie in the standard Chebyshev domain $g^{p,f} : [-1, 1] \rightarrow [-1, 1]$. This normalization process can be easily inverted based on the provided peak luminances π_α and π_β and the images $\mathtt{I}_\alpha$ and $\mathtt{I}_\beta$.

Note that since each basis polynomial ψ_k has a domain $\mathcal{D} := [-1, 1]$ and its range $\psi_k(\mathcal{D})$ is also a subset of $[-1, 1]$, the total $\left\| g^p - \bar{g}^p \right\|_\infty$ error of the approximation is bounded by the sum of the absolute values of the infinite remaining coefficients of the series. When approximating a function with m continuous derivatives on $[-1, 1]$, the approximation error of a Chebyshev series truncated at n elements has a convergence rate of $O(n^{-m})$ when $n \rightarrow \infty$ [31]. As such, when operating on "well-behaved"

functions, a common stopping criterion is given by the sum of the absolute values of a small number of elements. In practice, our algorithm truncates the series when the absolute sum of the next three elements is below the allowed error threshold. An example of an approximation of a function by Chebyshev polynomials of different orders is illustrated in Fig. 5.

In our unoptimized Matlab implementation, computing lumipaths for every pixel of a FullHD image takes approximately 3-5 seconds. Decoding this information to reconstruct the represented functions for all pixels takes an additional 0.1-1 seconds. Importantly, this computation could be significantly sped up through parallelization as each pixel is independent of the rest of the image.

4.2. Coding

As discussed previously, an approximated but visually lossless representation of a lumipath $\bar{g}^{p,f}$ can be specified by a $N_{p,f}$ tuple of Chebyshev coefficients $\left(c_0^p, \ldots, c_{N_{p,f}}^{p,f} \right)$.

In practice, these coefficients are highly correlated over space and time which allows for further compression of the data. In this section we present a suitable coding approach, which quantizes Chebyshev coefficients and reorganizes them into monochrome video sequences (Fig. 6). H.264 is then used as a standard video coding method [32] for efficient compression. As our results show in the next section, we achieve very reasonable bitrates with this approach, which shows that CDR is a suitable solution for the distribution challenge. Our method for CDR video coding still leaves room for improvement, however. Better use of the nature of the provided data could provide improved data rates (see Sec. 6 for further discussion).

We represent all lumipaths in an image-like format, which then allows the application of a video codec. We compute the maximum degree $N := \max_{p,f} N_{p,f}$ and set $c_k^{p,f} := 0$ for $k > N_{p,f}$, which leads to a representation $\bar{g}^{p,f}(x) = \sum_{k=0}^{N} c_k^{p,f} \psi_k(x)$ of the function described in Equation 6, but with a fixed parameter N.
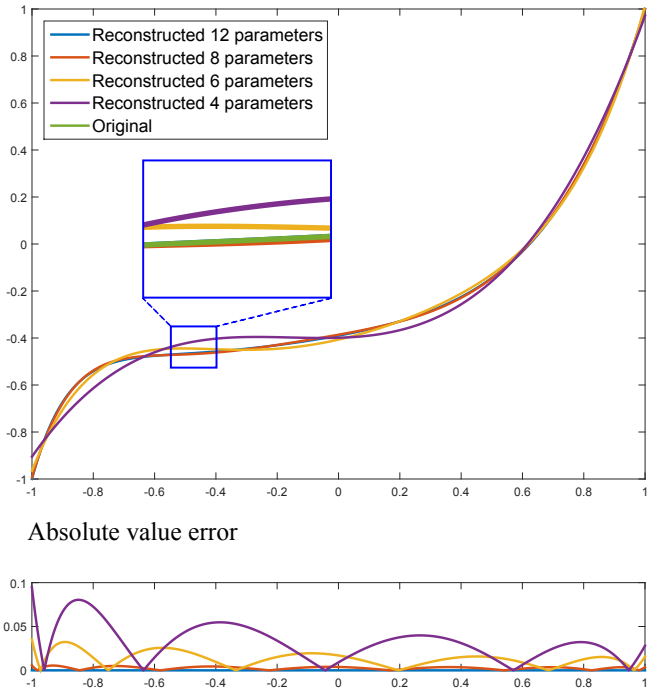
Figure 5: *A function is approximated with different numbers of parameters (top). The absolute value of the error between the original function and the reconstructed representation is shown in a larger scale (bottom).*

Figure 6: *Coefficient images $\bar{\mathsf{C}}_k^1$ of sequence Gunman.*

that the bitrate rapidly drops for coefficients with higher index values.

Figure 7: *Bit rates of individual coefficient sequences of sequence Gunman for a quantization parameter of 30.*

Each lumipath $\bar{g}^{p,f}$ is now specified by an *N*-tuple

$$\mathbf{c}^{p,f} := \left(c_1^{p,f}, \ldots, c_N^{p,f} \right). \tag{7}$$

To get an image-like representation, we represent the tuples $\mathbf{c}^{p,f}$ of all pixels of a frame by *coefficient* matrices $\mathsf{C}_k^{\mathsf{f}} \in \mathbb{R}^{\mathrm{h} \times \mathrm{w}}$ for k from 1 to N, which by construction have the same pixel resolution $\mathrm{h} \times \mathrm{w}$ as $\mathtt{I}_\alpha^f$ and $\mathtt{I}_\beta^f$. We uniformly quantize all entries of all matrices $\mathsf{C}_k^{\mathsf{f}}$ to 8-bit integers [33] obtaining N matrices $\bar{\mathsf{C}}_k^f$. A bitdepth of 8 is used because it corresponds to the maximum bit depth for images which are supported for compression by the main profile of H.264. Fig. 6 shows the first 8 coefficient images for a frame of sequence *Gunman*. It can be observed that the energy and variance in the coefficient images drops rapidly with increasing coefficient index. Most of the information is concentrated within the first coefficients. Coefficients often have uniform values over large image regions. We further observed very smooth behavior over time. Thus, the information content of such coefficient images and videos is relatively limited in practice as compared to the images and videos themselves, making them very well compressible.

A compressed representation of all lumipaths is obtained by storing (i) one integer value which represents the degree N, (ii) two floating point values representing the minimum and maximum value used for 8-bit quantization, and (iii) an encoded representation of the image sequences $\bar{\mathsf{C}}_k^1, \ldots, \bar{\mathsf{C}}_k^F$ for $k = 1, \ldots, N$ which is obtained with H.264. Fig. 7 shows the individual bitrates for each of the coefficient image sequences of a CDR video example. As suggested by Fig. 6, we can observe
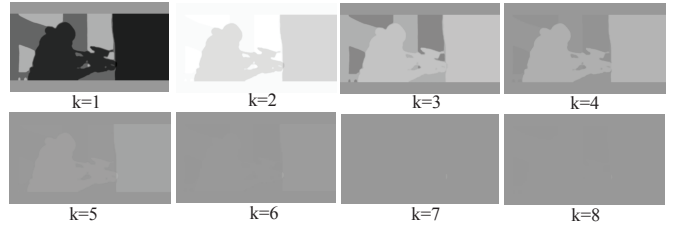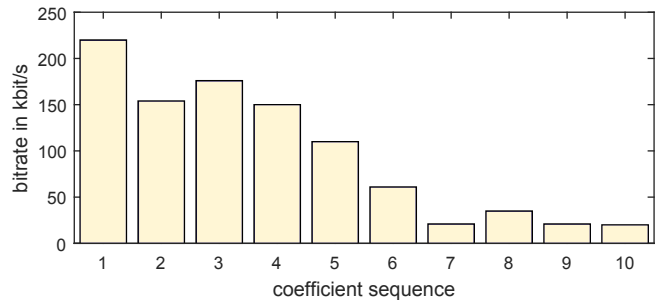
5. Results

We tested our system on a number of video sequences obtained from three short feature films, *Tears of Steel*[1], *Big Buck Bunny*[2] and *Lucid Dreams of Gabriel* [3]. As it is impossible to visualize HDR imagery with traditional SDR displays, in this work all results are presented with tonemapped images. It is important to note that these representations do not show the full extent of our method. We used Adaptive Logarithmic Mapping [34] to tonemap HDR frames for presentation as it is easily implemented and only requires a single input parameter.

5.1. Evaluation

Our system allows precise local control of luminance when grading for any display in the dynamic range hull. A comparison of a sample grading produced by the authors of this paper and automatic methods can be seen in Fig. 11, top. It is interesting to note that content creators may intend to maintain a particular luminance contrast in their scenes, which could be lost through global tone mapping operations. Notice the loss

---

[1] Tears of Steel - Old Man, Pannel, Gunman, Rockets scenes, copyright (CC) Blender Foundation (www.mango.blender.org).

[2] Big Buck Bunny - Bunny, Bird and Peach scenes, copyright (CC) Blender Foundation (www.bigbuckbunny.org).

[3] Lucid Dreams of Gabriel - Car scene, copyright Disney Research, ETH Zurich (www.disneyresearch.com/luciddreamsofgabriel).

of contrast between the background and foreground in the *Old Man* and *Pannel* scenes, and the excessively dark man in the *Gunman* scene when the views are interpolated linearly. In contrast, automatic methods such as Display Adaptive Tone Mapping [4] can preserve the appearance of the scene across multiple dynamic ranges (Fig. 9), but they do not allow the artistic freedom enabled by the localized editing of lumipaths as we do in our method. Such methods also do not account for the efficient approximation and distribution of the generated content to a multitude of different target displays. Another fundamental difference is that display adaptive tone mapping only utilizes single source grading for deriving any intermediate gradings. Our method in contrast uses the maximum and minimum grading.

To showcase some possibilities of artistic gradings that can be achieved using our method, we point the reader to Fig. 11, bottom. Notice that the *Rockets* scene can be graded to either emphasize the details near the rocket motor, or create a bloom effect to convey the brightness of the flames. The *Bird* and *Peach* scenes are graded to give greater emphasis on either the main object or the background of the scene. In the *Car* scene, grading can be used to create the sensation of a cloud above the scene, or that of a sunny day.

Gradings as presented above can be efficiently encoded using the method presented in Section 4. When using H.264 video coding, the FFmpeg library was employed and a group of pictures size of 24 was used for all sequences, while the quantization parameter (QP) was varied to control the loss of quality. Fig. 8 shows the data rates for the lumipaths of five sample scenes, with an average of 1.52 *Mbit/s* for the highest quality setting, which corresponds to approximately 13% of the current business-to-consumer distribution bandwidth used for 1080i50 television signals.

5.2. Perceptual Validation

We performed a perceptual experiment to test whether the distortions introduced due to lossy coding of the lumipath information would lead to visually noticeable artifacts. In our user study, we showed video content to our subjects on a SIM2 HDR display [35]. The CDR video, which was created with minimum and maximum grades at 100 and 4000 nits, respectively, was sampled over the continuous dynamic range at 700, 1500, and 3000 nits. We performed a 2 alternate forced-choice procedure (2AFC) on a set of 5 videos (*Bird, Car, Gunman, Peach and Rockets*). After a short training session where compression artifacts were explicitly pointed out, participants were tasked with selecting the video with better quality. The comparison was always performed between a reference uncompressed video sample with either itself (in which case the choice was at chance), or the same sample compressed using a quality parameter $QP \in \{30, 40, 50\}$. 16 volunteers participated (5F, 11M), aged 25 to 36 with normal or corrected-to-normal vision.

We performed *ANOVA* analysis on the results of the experiment and found a number of interesting interactions. Answers for

$QP = 50$ and $QP = 40$ were found to be significantly different from the reference ($\sigma \ll 0.001$ and $\sigma = 0.043$, respectively). In addition, both $QP = 30$ and $QP = 40$ were found significantly different from $QP = 50$ ($\sigma \ll 0.001$ and $\sigma = 0.012$). No difference was found between the reference and $QP = 30$. These results suggest that participants were unable to see the difference in quality between the uncompressed material and the $QP = 30$ version, but could clearly distinguish it from $QP = 40$ and $QP = 50$. The values presented above are shown in Fig. 10.

Participants were also more likely to see differences in the *Peach* sequence than the *Car*, *Gunman* and *Rockets* sequences ($\sigma < 0.05$). This could be explained by the fact that the video shown in *Peach* was computer generated and had a very clear image edge separating the slow-moving object of interest from a flat, motionless background - making compression artifacts stand out particularly strongly. No significant interactions were found for the *brightness* parameter.

These results indicate that the lumigraph data of CDR video can be compressed in a visually lossless way at QP30 to about 13% (1.52 *Mbit/s*) of the corresponding video bitrate on average.

QP	Bird	Car	Gunman	Peach	Rockets
30	0.768	2.903	1.110	1.318	1.477
40	0.400	1.411	0.787	0.397	0.871
50	0.300	0.659	0.534	0.221	0.552

Figure 8: *This figure shows the total bitrates of the lumipaths for five sequences. The bitrates are expressed in Mbit/s and obtained by encoding with different quantization parameters. These sequences are used for the perceptual experiment.*

Figure 9: *Display Adaptive Tone Mapping can be used to generate content for displays with different luminance levels, but does not allow for precise artistic control of content.*

6. Discussion

Our method is not without limitations. In this work we presented a formulation and implementation of a novel content creation and distribution paradigm. While we demonstrated that our method results in a low bandwidth overhead and allows full artistic freedom, many of the components that we
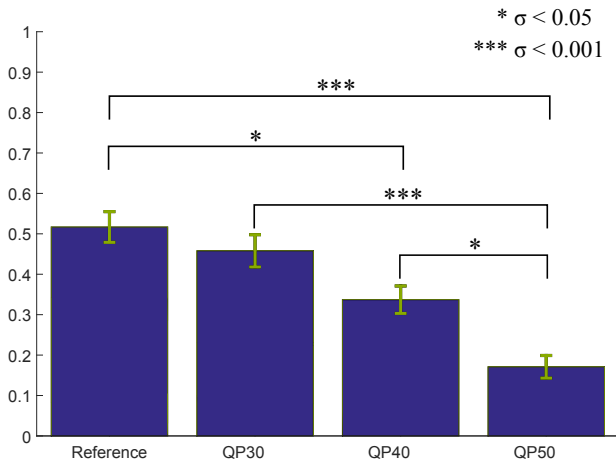
Figure 10: *This figure shows the results of our user study, averaged over all sequences and brightness levels. Our supplementary material contains the raw data and full statistical analysis for this experiment. The values on the Y-axis represent the ratio at which the reference was considered to have better quality, with a value of 0 meaning the reference was preferred in every trial.*

presented could be engineered for better performance. For example, a more sophisticated human visual system model could replace our current model for better predicting the threshold luminances, which could help reduce the number of polynomial basis coefficients while still maintaining a visually lossless representation (Section 4). While the editable lumipaths give the user full control over luminance during dynamic range mapping, our current implementation does not allow a similar control over chrominance, although the colors in the maximum and minimum gradings can be adjusted without any limitation. Formulating a representation for chrominance that is analogous to lumipaths, as well as extending our dynamic range mapping interface to support such a representation is left as future work. Further, more dedicated approaches for compression of lumipath image sequences (e.g. inspired by depth coding approaches [36]) could reduce bitrates even further.

7. Conclusion

We presented CDR video, a novel representation of pixel-level luminance as a function of display dynamic range. We introduced dynamic range mapping as a new approach for content creation targetting displays with different dynamic ranges. An efficient approximation of CDR video through a polynomial series approximation was presented, as well as coding that consumes only 13% of current standard business-to-consumer distribution bandwidth. Together, these components form an end-to-end solution for content production and distribution for the wide variety of emerging HDR displays.

Figure 11: *(Top) shows a comparison of results color graded with our method, as compared to a naive linear interpolation between the SDR and HDR graded versions. Notice that our system allows for local control of the grading at each point of the dynamic range hull. (Bottom) Two gradings obtained using our system are shown in contrast to showcase different artistic possibilities that can be achieved using our system.*

9

8. References

[1] Lukk H. Emma. http://www.emmathemovie.com; 2014.

[2] Schriber SA. Lucid dreams of Gabriel. http://www.disneyresearch.com/luciddreamsofgabriel; 2014.

[3] Hanhart P, Korshunov P, Ebrahimi T, Thomas Y, Hoffmann H. Subjective Quality Evaluation Of High Dynamic Range Video And Display For Future TV. In: International Broadcasting Convention (IBC). 2014,.

[4] Mantiuk R, Daly S, Kerofsky L. Display adaptive tone mapping. ACM Trans Graph 2008;27(3):68:1–68:10.

[5] Mantiuk R, Efremov A, Myszkowski K, Seidel HP. Backward compatible high dynamic range MPEG video compression. ACM Trans Graph 2006;25(3):713–23.

[6] Reinhard E, Ward G, Pattanaik S, Debevec P, Heidrich W, Myszkowski K. HDR Imaging - Acquisition, Display, and Image-Based Lighting, Second Edition. Morgan Kaufmann; 2010.

[7] Reinhard E, Stark M, Shirley P, Ferwerda J. Photographic tone reproduction for digital images. ACM Trans Graph 2002;21(3):267–76.

[8] Ferwerda JA, Pattanaik SN, Shirley P, Greenberg DP. A model of visual adaptation for realistic image synthesis. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH 1996; 1996, p. 249–58.

[9] Pattanaik SN, Tumblin J, Yee H, Greenberg DP. Time-dependent visual adaptation for fast realistic image display. In: Proc. of Conf. on Computer Graphics and Interactive Techniques. SIGGRAPH 2000; 2000, p. 47–54.

[10] Reinhard E, Devlin K. Dynamic range reduction inspired by photoreceptor physiology. IEEE Transactions on Visualization and Computer Graphics 2005;11:13–24.

[11] Durand F, Dorsey J. Fast bilateral filtering for the display of high-dynamic-range images. ACM Trans Graph 2002;21(3):257–66.

[12] Fattal R, Lischinski D, Werman M. Gradient domain high dynamic range compression. ACM Trans Graph 2002;21(3):249–56.

[13] Mantiuk R, Myszkowski K, Seidel HP. A perceptual framework for contrast processing of high dynamic range images. ACM Trans Appl Percept 2006;3(3):286–308.

[14] Eilertsen G, Wanat R, Mantiuk RK, Unger J. Evaluation of tone mapping operators for HDR-video. Computer Graphics Forum 2013;32(7):275–84.

[15] Boitard R, Cozot R, Thoreau D, Bouatouch K. Zonal brightness coherency for video tone mapping. Signal Processing: Image Communication 2014;29(2):229 –46.

[16] Aydın TO, Stefanoski N, Croci S, Seidel Gross M, Smolic A. Temporally coherent local tone mapping of hdr video. In: ACM Trans. Graph. (Proc. of SIGGRAPH Asia); vol. 33(6). 2014,Article 196.

[17] Dolby Laboratories I. Dolby vision white paper. http://www.dolby.com/us/en/technologies/dolby-vision/dolby-vision-white-paper.pdf; 2015.

[18] Seetzen H, Heidrich W, Stuerzlinger W, Ward G, Whitehead L, Trentacoste M, et al. High dynamic range display systems. ACM Trans Graph 2004;23(3):760–8.

[19] Wanat R, Petit J, Mantiuk R. Physical and perceptual limitations of a projector-based high dynamic range display. In: Theory and Practice of Computer Graphics, Rutherford, United Kingdom, 2012. Proceedings. 2012, p. 9–16.

[20] Ferwerda J, Luka S. A high resolution, high dynamic range display for vision research 2009;9(8):346.

[21] Zhang D, Ferwerda J. A low-cost, color-calibrated reflective high dynamic range display 2010;10(7):397.

[22] Guarnieri G, Albani L, Ramponi G. Image-splitting techniques for a dual-layer high dynamic range lcd display. J Electronic Imaging 2008;17(4).

[23] Kim MH, Weyrich T, Kautz J. Modeling human color perception under extended luminance levels. In: ACM Transactions on Graphics (TOG); vol. 28. ACM; 2009, p. 27.

[24] Zhang Y, Reinhard E, Bull D. Perception-based high dynamic range video compression with optimal bit-depth transformation. In: IEEE International Conference on Image Processing (ICIP). 2011, p. 1321–4.

[25] Touze D, Olivier Y, Thoreau D, Serre C. High dynamic range video distribution using existing video codecs. In: Picture Coding Symposium (PCS), 2013. 2013, p. 349–52.

[26] ISO/IEC MPEG . Call for Evidence (CfE) for HDR and WCG Video Coding. http://mpeg.chiariglione.org; 2015.

[27] Mantiuk R, Kim KJ, Rempel AG, Heidrich W. Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Trans Graph 2011;30(4):40:1–40:14.

[28] Aydın TO, Mantiuk R, Myszkowski K, Seidel HP. Dynamic range independent image quality assessment. In: ACM Trans. Graph. (Proc. of SIGGRAPH); vol. 27(3). 2008,.

[29] Davis PJ. Interpolation and approximation. Courier Corporation; 1975.

[30] Broucke R. Algorithm: ten subroutines for the manipulation of chebyshev series. Communications of the ACM 1973;16(4):254–6.

[31] Gil A, Segura J, Temme NM. Numerical methods for special functions. Siam; 2007.

[32] Wiegand T, Sullivan GJ, Bjontegaard G, Luthra A. Overview of the h.264/avc video coding standard. IEEE Trans Cir and Sys for Video Technol 2003;13(7):560–76.

[33] Sayood K. Introduction to Data Compression (2Nd Ed.). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2000.

[34] Drago F, Myszkowski K, Annen T, Chiba N. Adaptive logarithmic mapping for displaying high contrast scenes. Computer Graphics Forum 2003;22(3):419–26.

[35] Sim2 . Model HDR47ES4MB. http://www.sim2.com; 2015. Accessed: 01 June, 2015.

[36] Merkle P, Mueller K, Wiegand T. Coding of depth signals for 3d video using wedgelet block segmentation with residual adaptation. In: Multimedia and Expo (ICME), 2013 IEEE International Conference on. 2013,.