Temporally Coherent Local Tone Mapping of HDR Video

Tunç Ozan Aydın* Nikolce Stefanoski* Simone Croci† Markus Gross*† Aljoscha Smolic*
* Disney Research Zürich † ETH Zürich

Figure 1: *Our method allows local tone mapping of HDR video where the scene illumination varies greatly over time, in this case more than* $4\ log_{10}$ *units (c). Note that our method is temporally stable despite the complex motion of both the camera and the actor (b), and the resulting tone mapped video shows strong local contrast (a). The plots show the logmean luminance of the HDR video and the mean pixel values of the tone mapped video.*

Abstract

Recent subjective studies showed that current tone mapping operators either produce disturbing temporal artifacts, or are limited in their local contrast reproduction capability. We address both of these issues and present an HDR video tone mapping operator that can greatly reduce the input dynamic range, while at the same time preserving scene details without causing significant visual artifacts. To achieve this, we revisit the commonly used *spatial* base-detail layer decomposition and extend it to the *temporal domain*. We achieve high quality spatiotemporal edge-aware filtering efficiently by using a mathematically justified iterative approach that approximates a global solution. Comparison with the state-of-the-art, both qualitatively, and quantitatively through a controlled subjective experiment, clearly shows our method's advantages over previous work. We present local tone mapping results on challenging high resolution scenes with complex motion and varying illumination. We also demonstrate our method's capability of preserving scene details at user adjustable scales, and its advantages for low light video sequences with significant camera noise.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Display Algorithms I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Time-varying Imagery

Keywords: Video Tone Mapping, Edge-Aware Video Filtering

Links: DL PDF

1 Introduction

New video technologies keep improving the quality of the viewing experience: display resolutions are moving up from *HD* to *4K*, frame rates in cinemas are increasing to *48fps* and beyond, and stereoscopic *3D* is introducing depth as an additional dimension. While through these advances the fidelity of the content in terms of spatial and temporal resolution and depth has been gradually increasing, the dynamic range aspect of video has received little attention until recently. However, this can be expected to change quickly as modern high-end cameras (such as the Red Epic Dragon, Sony F55 and F65, and ARRI Alexa XT) now can natively capture High Dynamic Range (HDR) video up to 14 f-stops. Creatives are highly interested in HDR video because it allows them to show more visual detail and extend the limits of artistic expression. Consequently, the entertainment industry is working towards HDR video pipelines for delivering content to the end user, including related distribution standards (MPEG and JPEG). However, the final element of the pipeline is slightly lagging behind: despite the existence of a small number of products and impressive research prototypes, consumer level HDR display technology is not yet on the horizon.

Still, in the absence of displays that are capable of fully reproducing the captured dynamic range, *tone mapping* of HDR video provides a means of visualization and artistic expression. More specifically, it may be desirable to reduce the dynamic range of captured HDR content while both maintaining most of the visual details and not hampering the picture quality by introducing visible artifacts. Re-

cent subjective experiments on the state-of-the-art in HDR video tone mapping [Eilertsen et al. 2013; Petit and Mantiuk 2013] revealed that none of the current methods (including a camera response curve) could achieve both goals at the same time.

In this work we propose a new HDR video tone mapping operator (TMO) to help close the gap between the captured dynamic range and displayed dynamic range. We build upon prior work in *image* tone mapping that utilize base and detail layers. Such a decomposition allows compressing the base layer's dynamic range while the local contrast remains intact in the detail layer. Different from prior work, our decomposition utilizes *spatiotemporal* filtering through per-pixel motion paths. This way, our method enforces temporal coherence and significantly reduces temporal artifacts such as brightness flickering and camera noise without introducing ghosting. As a result, we enable local HDR video tone mapping that can be art-directed without introducing visually significant artifacts.

The two main contributions of our work are the following:

- A temporally coherent and local video tone mapping method that can maintain a high level of local contrast with fewer temporal artifacts compared to the state-of-the-art (validated by a subjective study).
- A practical and efficiently parallelizable filtering approach specifically designed for tone mapping, that reduces halo artifacts by approximating a global solution through iterative application (with a formal analysis of the filter's halo reduction property).

We show our method's advantages both qualitatively through examples (Sections 5.1 and 5.4), and quantitatively through a controlled subjective study (Section 5.2). We also demonstrate that our method allows creative control over spatial and temporal contrast (Section 5.5) through a simple user interface (Section 5.3). Finally, due to temporal filtering we show that our method works especially well for low light shots with significant camera noise (Section 5.6). Next section continues with a brief discussion of the related work.

2 Background

HDR Image Tone Mapping has been extensively studied in computer graphics. Here we give a brief introduction and refer the reader to Reinhard et al. [2010] for a comprehensive overview of the field. Since the early days of photography, reproducing the dynamic range of natural scenes on chemically limited negative has been a big challenge. This problem has been reduced by adopting an S-shaped tone curve that gradually compresses highlights and shadows [Mantiuk et al. 2008], as well as utilizing techniques such as *dodging* and *burning* during the photographic development process to locally adjust the image exposure [Reinhard et al. 2002]. The disconnect between the technical aspects of tone reproduction and the artistic concerns of photography has been addressed by the *Zone System* [Adams 1981], which has been widely used since its inception [Reinhard et al. 2002].

Reinhard et al. [2002] attempted to faithfully model this photographic tone reproduction process for digital HDR images. Their method is known for its tendency to preserve the natural look of the original scene. Another class of TMOs [Ferwerda et al. 1996; Pattanaik et al. 2000; Reinhard and Devlin 2005] takes the *naturalness* aspect one step further by modeling various aspects of the human visual system and aiming to reproduce the scene from the eyes of a hypothetical human observer.

A number of operators take an alternative approach where they focus on reproducing the original scene by preserving its local contrast. Images produced by such local (spatially-varying) TMOs can often easily be recognized due to their characteristic look with reduced brightness differences between highlights and shadows, but greatly enhanced fine scale details (our teaser shows an example).

It is worth noting that results of local and global TMOs can look quite different, and some people have strong aesthetic preferences for one type of tone mapping over the other. In this work, even though we do not advocate any particular visual style, we focus on local tone mapping because it is currently unresolved for video, and is the more general problem among the two.

One way of achieving local image tone mapping is by selectively compressing the input HDR image in the gradient domain, and reconstructing a reduced dynamic range image from the modified gradients [Fattal et al. 2002; Mantiuk et al. 2006]. Another approach is to decompose the HDR image into a *base* and a *detail* layer, which respectively contain coarse and fine scale contrast [Durand and Dorsey 2002]. The dynamic range reduction is achieved by recombining the detail layer with a tone compressed base layer. The decomposition is often done using edge-aware filters in order to reduce the effect of the well-known *halo* artifacts, which appear near strong image edges if the base layer compression is too strong. In principle, our method is also based on a base-detail layer decomposition using edge-aware filters (although we compute these layers in a novel way). Therefore we review edge-aware filtering in the following paragraph from a tone mapping perspective.

Edge-Aware Filtering is a fundamental building block for several computer graphics applications including tone mapping. Milanfar [2013] gives an excellent overview of the many filtering approaches. HDR image tone mapping has been a popular application among the edge-aware filtering methods published since the establishment of edge-aware filtering based tone mapping [Durand and Dorsey 2002]. However, the treatment of tone mapping among these works has (understandably) been often brief, since the main contributions of these works lie elsewhere. In the absence of extensive comparisons, it is difficult to reach any conclusions on how well various edge-aware filters perform on HDR tone mapping.

Nevertheless, if the base layer is compressed strongly during tone mapping, the bilateral filter [Tomasi and Manduchi 1998] has been known to produce visible halo and ringing artifacts in the resulting image. The Weighted Least Squares (WLS) filter [Farbman et al. 2008] has the advantage of preventing halo artifacts by minimizing a function whose data term penalizes the distance between the original and filtered image. On the downside, this approach requires solving large numerically-challenging linear systems [Fattal 2009], and involves the complex machinery of a multi-resolution preconditioned conjugate gradient solver [Paris et al. 2011]. Still, WLS has been used as the reference method in follow-up work due to the high quality of its results [Fattal 2009; Gastal and Oliveira 2011]. The computation of the Edge-avoiding Wavelets [Fattal 2009] is much simpler, however they have been shown to suffer from aliasing and generating irregular edges [Paris et al. 2011]. The Local Laplacian Filter is capable of producing high quality results, however its authors point out the running time as the method's main shortcoming [Paris et al. 2011]. While recently a faster implementation became available [Aubry et al. 2014], the temporal stability aspect is still an open question, and it is not clear how to extend the method to the temporal domain. A number of other recent filtering techniques [Gastal and Oliveira 2011; He et al. 2013] have been shown to produce plausible tone mapping results in a limited number of images and configurations, but have not been thoroughly tested for halos and other tone mapping artifacts. Likewise, recent work involving spatiotemporal filtering has neither been designed nor tested for video tone mapping [Lang et al. 2012; Ye et al. 2014].

In this work we propose a practical and parallelizable filter designed

specifically for tone video mapping. The key properties of our filter are its natural extension to the temporal domain and the reduction of halo artifacts through iterative application (Section 4). Importantly, we also present a formal discussion of our method's halo reduction property and its theoretical similarities to WLS (Section 4.4).

Video Tone Mapping has been a far less active field than image tone mapping. A major obstacle for video tone mapping research has been the absence of high quality HDR content available to the research community, which has been partially addressed by the recent developments in high-end cameras and other experimental systems [Tocci et al. 2011].

A number of TMOs comprise temporal components that allow them to process HDR video. Among them, the global operators [Ferwerda et al. 1996; Pattanaik et al. 2000; Irawan et al. 2005; Van Hateren 2006; Mantiuk et al. 2008; Boitard et al. 2012] including the S-shaped camera response curve generally produce results with good temporal coherency but low spatial contrast. On the other hand, the local operators [Ledda et al. 2004; Bennett and McMillan 2005; Benoit et al. 2009; Reinhard et al. 2012] often maintain high contrast at the cost of more temporal artifacts. The advantages and shortcoming of these operators have been discussed and subjectively evaluated by Eilertsen et al. [2013] (we present a complementary subjective study in Section 5.2).

Other work in video tone mapping [Kang et al. 2003; Ramsey et al. 2004; Kiser et al. 2012] focused on extending the photographic TMO's tone curve [Reinhard et al. 2002] with temporal filtering for computing a temporally coherent key value. For such global TMOs, the temporal coherence of the tone mapping results is shown to be enhanced by simply ensuring the temporal smoothness in the TMO parameter space (they could also benefit from automatic flicker detection [Guthier et al. 2011]). However, like other global TMOs, these methods are inherently limited in local contrast reproduction. Recent work by Boitard et al. [2014b] extends their previous work [Boitard et al. 2012] on temporally coherent global tone mapping. The extension consists of segmenting each video frame (typically to $2-4$ segments) and applying a global tone curve to each segment individually. This way they introduce local adaptation at a segment level at the cost of more complex processing involving video segmentation. While they claim local brightness coherency, the effectiveness of their method remains unclear since their user study measures *subjective preference* rather than assessing temporal artifacts and/or local contrast reproduction. It is important to note that the parameter smoothing approach employed in the aforementioned methods is not an option for local tone mapping, where temporal coherence has to be achieved in the image domain. Such a temporal extension has been proposed for the gradient domain image TMO [Lee and Kim 2007] that utilizes a block matching algorithm between two consecutive frames. However, the effectiveness of the extension is limited to the highest frequency temporal artifacts due to the two-frames temporal neighborhood. Also, the Poisson reconstruction step of gradient domain operator is costly at high resolutions.

Subjective studies and surveys that evaluated the state-of-the-art in HDR video tone mapping [Eilertsen et al. 2013; Petit and Mantiuk 2013] came to the conclusion that the problem is far from being solved. As for future work, temporal models for avoiding artifacts, and local processing for maintaining a high level of detail and contrast have been pointed out as two venues for improvement. In this paper we address both challenges and propose a novel method for temporally coherent local HDR video tone mapping.

3 Local HDR Video Tone Mapping

In this section we give an overview on the various aspects of the HDR video tone mapping problem and describe the main computational steps of our method.

3.1 Visual Trade-offs in Local Video Tone Mapping

Since HDR tone mapping often significantly reduces the input dynamic range, some of the scene contrast is inevitably lost during the process. As such, local image tone mapping involves a visual trade-off between fine and coarse scale contrast. If coarse scale contrast is emphasized, the luminance difference between large image regions, as well as highlights and shadows become more pronounced at the cost of the visibility of the fine scale details. If, on the other hand, fine scale contrast is emphasized, the relative reduction of coarse scale contrast often results in a "flattening" effect.

Additionally in local video tone mapping, another similar trade-off exists between spatial and temporal contrast. Consider an HDR video clip that transitions from bright sunlight to a darker hallway, and therefore has strong temporal contrast (Figure 2). One strategy for tone mapping such a clip is to utilize the full dynamic range independently at every video frame, such that both frame (a) and frame (b) have sufficient brightness to reproduce most of the scene details. This way one can maintain the tone mapped video's spatial contrast, with the side effect of reducing the sensation that the hallway is much darker than the outside. Another (complementary) strategy for tone mapping the same video is maintaining a certain amount of the temporal contrast at the cost of less visible spatial details during the hallway part.

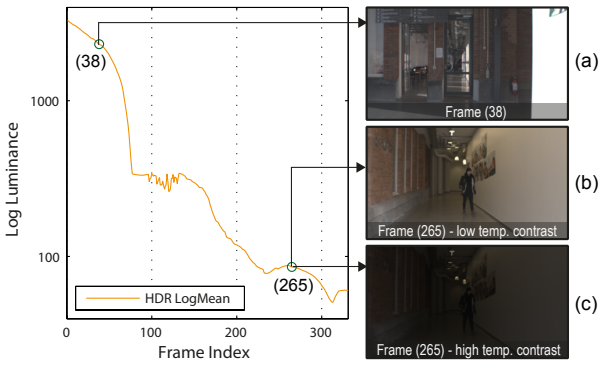

Figure 2: *The visual trade-off between emphasizing spatial contrast (a, b) and temporal contrast (a, c). While in both settings frame 38 remains the same (a), frame 265 can be adjusted to either maintain spatial (b) or temporal contrast (c).*

In our view, the aforementioned trade-offs are context dependent and ultimately artistic decisions, and no tone mapping strategy is inherently better than others in all possible scenarios. It is therefore desirable that the tone mapping algorithm offers the required artistic freedom by providing explicit control over spatial contrast at different scales, as well as temporal contrast. Importantly, the tone mapping method should maintain a consistent level of image quality, since the artistic freedom makes little practical sense if a certain set of edits create visually noticeable artifacts. In that sense, video tone mapping is especially challenging because the temporal dimension emerges as a new source of artifacts. In fact, high temporal frequency artifacts such as brightness flickering, camera noise, as well as any temporal inconsistencies of the TMOs are immediately noticeable because of the human visual system's proper-

ties. In particular, Eilertsen et al. [2013] noted that even very small amounts of ghosting and flickering in the tone mapped HDR videos are unacceptable in practice.

3.2 Tone Mapping Pipeline

Recent subjective studies revealed that local TMOs (which usually maintain local contrast) are temporally unstable [Eilertsen et al. 2013]. Similarly, naïvely applying image TMOs to individual video frames has been found to produce strong temporal artifacts [Eilertsen et al. 2013; Boitard et al. 2014a]. Global TMOs with relatively simple processing steps are found to be less susceptible to temporal artifacts, however they lack the contrast reproduction capabilities of local operators. The temporal instability of local TMOs underlines the necessity of temporal filtering for local video tone mapping.

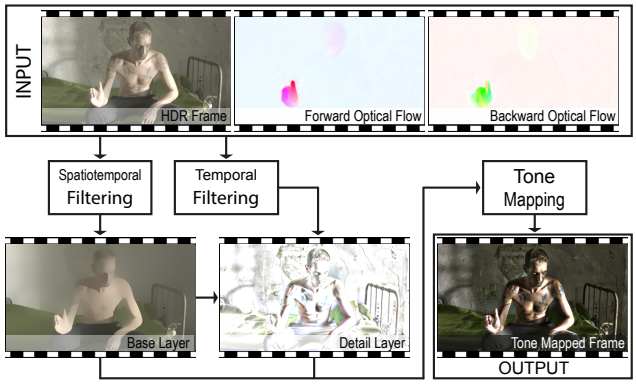

Figure 3: *Major processing steps of our method. Refer to the text for details.*

Since filtering in the temporal domain is notably more challenging than in the spatial domain, one can assume that the scene is static and filter through a straight line in the temporal dimension [Bennett and McMillan 2005]. However this approach generates strong ghosting artifacts as in practice the static scene assumption rarely holds. In contrast, Lang et al. [2012] propose a general purpose approach that filters through each pixel's path over time. The downside of their method is that the path computation is performed globally over the whole video sequence, which therefore needs to be kept in memory. As such, this approach becomes infeasible for longer video sequences at high resolutions. Similarly, for such sequences Ye et al.'s [2014] method is prohibitively expensive as it is reported to require over a minute to process a 800×600 frame (moreover this method is not designed for tone mapping).

Our approach, similar to Durand and Dorsey [2002] and many others, uses the idea of decomposing each frame into a base and detail layer by utilizing an edge-aware filter, but with the key difference of filtering in the temporal dimension by utilizing optical flow. The main processing steps of our method are depicted in Figure 3. In our approach, the base layer is obtained by edge-aware spatiotemporal filtering of an HDR video cube consisting of a center frame and a number of its temporal neighbors. This way, we reduce the effect of illumination changes over time and enforce temporal coherence in the base layer. The computation of the detail layer also involves filtering the input HDR sequence, but only in the temporal dimension. This way we effectively reduce temporal artifacts with low amplitude and high temporal frequency (such as camera noise and brightness flicker) without sacrificing spatial resolution. While still retaining a contrast reproduction capability comparable to Durand and Dorsey's image tone mapping framework, we also suppress halo artifacts commonly observed in such frameworks through our shift-variant filter that approximates a global solution.

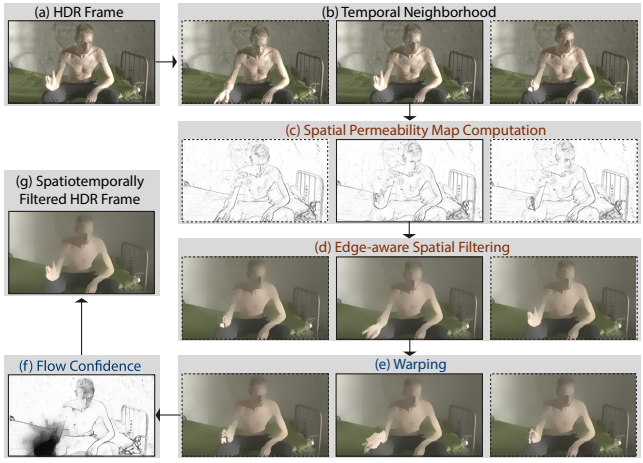

Figure 4: *Spatiotemporal filtering process of a single HDR frame within a temporal window. Brighter colors in (c) and (f) indicate higher permeability. Steps (c, d) comprise the spatial filtering, whereas (e, f) comprise the temporal filtering.*

The spatiotemporal filtering is performed on a temporal neighborhood and uses forward and backward optical flow estimates to warp each frame's temporal neighborhood, such that their pixels are aligned on a straight line in time dimension. The temporal filtering proceeds only if the temporal alignment is predicted to be reliable. This way temporal consistency is enhanced while minimizing ghosting. While our method is not bound to a specific optical flow algorithm, in our implementation we use Zimmer et al.'s approach [2011]. We mostly used the method's standard parameters and spent no special effort for fine-tuning the optical flow results, but rather focused on developing measures that detect optical flow errors (discussed in Section 4.3).

Our method's spatiotemporal filtering process is depicted in Figure 4. For a single input HDR frame I_t in the log domain (a), we consider a number of consecutive past and future frames as the temporal neighborhood (b). The optimal size of the temporal neighborhood depends on the content and flow quality, however we empirically found that 10 frames in each direction works well in our implementation. Next, for each frame in the temporal neighborhood we compute a *permeability map* (c) which controls the spatial diffusion between neighboring pixels. The permeability map is used as an input to the spatial filtering process (d) and prevents filtering across strong image edges. Specifically, the spatial part of our filtering involves iterative application of the shift-variant spatial filter h^{s} to the frame $J_t^{(k)}$, where k denotes the iteration, and $J_t^{(0)} = I_t$. The $(k+1)^{th}$ iteration result is computed as:

$$J_t^{(k+1)} = J_t^{(k)} * h^{\mathsf{s}} + \lambda \left(I_t - J_t^{(k)} \right), \qquad (1)$$

where λ is a spatially and temporally varying weighting factor that introduces a bias towards the original image (important for halo reduction), and $*$ denotes an ordinary shift-variant convolution with filter h^{s}. Our filtering approach is discussed in detail in Section 4.

Next, as the first step of the temporal part of the process in Figure 4, we warp (e) the frames in the temporal neighborhood such that all their pixels are aligned with the corresponding pixels of the input HDR frame I_t. The warping process uses forward and backward

optical flow[1]. For each pixel of a frame $J_{t'}$ with $t' > t$ we compute the cumulated forward vector

$$d_{t\to t'} := \sum_{\tau=t}^{t'-1} d_{\tau\to(\tau+1)} \tag{2}$$

based on optical flow vectors $d_{\tau\to(\tau+1)}$ between neighboring frames. Next, we warp all pixels of $J_{t'}$ by shifting them in the opposite direction $-d_{t\to t'}$ which temporally aligns the pixels of $J_{t'}$ with those in J_t. This process is repeated in 2D for all pixels of all future frames that are in the temporal neighborhood of J_t. An analogous warping process is also repeated in the past direction using the backward optical flow.

An inevitable problem with the warping process is the inaccuracies due to the errors and limits of the optical flow. In order to prevent visual artifacts we compute a *flow confidence* map (f), which is analogous to the permeability map in the sense that it prevents the filter support from growing across certain pixels. Differently, the flow confidence map is used in temporal filtering, and is computed by combining a photo constancy measure and a flow derivative measure (discussed later in Section 4.3). Finally, the spatiotemporally filtered HDR frame (g) is obtained by applying the same filter that performs the spatial filtering in the temporal domain, by simply using the flow confidence (f) as the permeability map.

We compute our base layer by applying all the steps depicted in Figure 4 to the *log luminance* of the input HDR video. Additionally, we also apply the temporal filtering steps (e-f) independently to all color channels of the input HDR video, which is then used for computing the detail layer (refer to Figure 3) similarly in the log domain. Formally, we express the temporal filtering combined with the aforementioned warping process as a simple shift-variant temporal convolution:

$$\left(I \star h^{\mathtt{t}}\right)_t := \left(\tilde{I}_t * h^{\mathtt{t}}\right)_t, \tag{3}$$

where $\tilde{I}_t(t', p) = I(t', p + d_{t\to t'})$, i.e. $\tilde{I}_t$ is an image sequence, in which each frame t' is warped to t using the flow field $d_{t\to t'}$. The $\star$ operator denotes the shift-variant convolution with image warping. The spatiotemporal filtering $I \star h^{\mathtt{st}}$ differs from the temporal filtering in Equation 3, in that the warped image $\tilde{I}$ is computed from the spatially filtered image J instead of I:

$$\left(I \star h^{\mathtt{st}}\right)_t := \left(\tilde{J}_t * h^{\mathtt{t}}\right)_t, \tag{4}$$

where $\tilde{J}_t\,(t', p) = J\,(t', p + d_{t\to t'})$. Having described and formally expressed the temporal and spatiotemporal filtering operations, we can compute the base (B) and detail (D) layers as follows:

$$B_t = \left(I \star h^{\mathtt{st}}\right)_t \quad \text{and} \quad D_t = \left(I \star h^{\mathtt{t}}\right)_t - B_t. \tag{5}$$

Note that for computing D_t in color images, B_t is subtracted separately from each color channel of the temporally filtered source HDR video. Once we compute the base and detail layers, we can reduce the dynamic range of the input video without affecting the scene details at selected scales. To achieve this, we first apply a compressive tone curve to the base layer, and then recombine both layers. A simple way of compressing the base layer proposed by Durand and Dorsey. [2002] involves multiplying the log luminance of the base layer by a compression factor. On the other hand, using tone curves such as Drago et al. [2003] or Reinhard et al. [2002] gives greater control over the final look. In this work we used either a compression factor c, or Drago et al.'s tone curve (controlled by the $bias$ parameter)[2] based on personal aesthetic judgement. That said, other tone curves can be easily integrated to our method.

[1] See supplemental material for an illustration of the warping process.

4 Edge-Aware Video Filtering

In the previous section, we presented our tone mapping pipeline and discussed the steps involved in the spatiotemporal filtering process, while treating the underlying filter formulation as a black box. In this section we start by introducing a class of iterative filtering methods and analyse them from a tone mapping perspective (Section 4.1). We show that this practical formulation can easily be adapted to perform the spatial (Section 4.2) and temporal (Section 4.3) filtering steps we described earlier in Section 3. We also discuss our filter's relation to the WLS filter (Section 4.4), which is often used as the reference against which new methods are compared in terms of image quality. The main advantage of our parallelizable filter with sub-linear runtime is that it generates visually similar results to WLS even with a small number of iterations[3], and that it can be easily extended to the temporal dimension.

4.1 Filter Formulation

A number of anisotropic smoothing filters that are used for image tone mapping [Farbman et al. 2008; Gastal and Oliveira 2011; Durand and Dorsey 2002] can be expressed as a class of iterative filtering methods, which can be represented by the iteration shown in Equation 1. This iteration can be expressed in a slightly different form as follows:

$$J_p^{(k+1)} := \sum_{q\in\Omega} h_{pq}\, J_q^{(k)} + \lambda\, h_{pp} \left(I_p - J_p^{(k)}\right), \tag{6}$$

where I_p is the input image intensity at pixel position p at frame t (we omit the frame indices in this and the following equations for brevity), $J_p^{(k)}$ is the corresponding diffusion result after k iterations with $J_p^{(0)} = I_p$, matrix $\mathtt{H} := \{h_{pq}\}$ is a row-stochastic matrix, i.e. all matrix coefficients h_{pq} are between 0 and 1, inclusive, with $\sum_q h_{pq} = 1\ \forall p$, and the set Ω contains all pixel positions of a frame. The iteration consists of two parts, a sum that computes a weighted average at pixel position p which we denote as *diffusion estimate*, and a *fidelity term* [Mrázek et al. 2004; Nordstrom 1989] whose impact is controlled by a parameter λ and introduces a bias towards the input image.

The fidelity term significantly contributes to the reduction of halo artifacts in tone mapping applications. To provide some intuition for this property, let us introduce the intermediate term *diffusion strength* at a pixel p, and define it as $\kappa_p := 1/h_{pp}$. If the diffusion strength κ_p is large ($h_{pp} \ll 1$), then $J_p^{(k+1)}$ is computed by strong contributions from pixels $J_q^{(k)}$ with $q \neq p$. However, if the diffusion strength κ_p is small ($h_{pp} \approx 1$), then $J_p^{(k+1)} \approx (1-\lambda)J_p^{(k)} + \lambda I_p$, which means that the result does not depend on pixel positions other than p and it represents a weighted average. In particular, if the diffusion strength is small and $\lambda = 1$, then

$$J_p^{(k+1)} \approx I_p. \tag{7}$$

In other words, a parameter $\lambda = 1$ leads to a strong shift towards the original image in areas with low diffusion strength.

As discussed earlier, local tone mapping methods often suffer from halo artifacts that appear near strong image edges, especially if the input dynamic range is strongly compressed. The property shown in

[2] Formula presented in supplemental material.

[3] Refer to the supplemental material for a visual comparison.

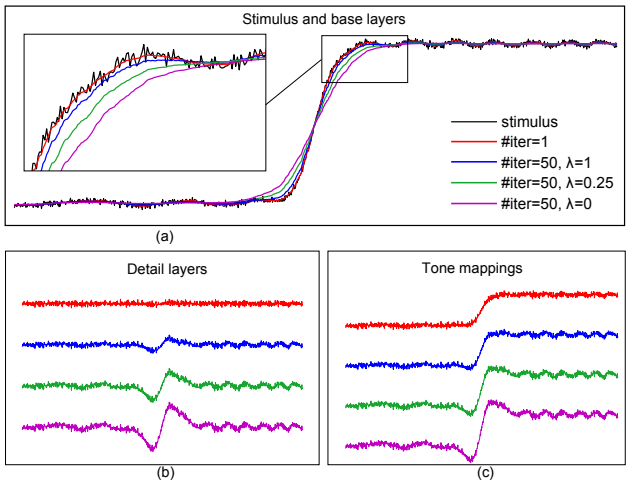

Figure 5: *Base layers computed from a stimulus signal (a). The stimulus signal is composed by a step function that is modulated with a sine wave of increasing frequency (a chirp signal) and Gaussian noise. In (b) and (c), corresponding detail layers and tone mapping results are shown, respectively. Tone mapping is performed by compressing the base layer from (a) and keeping the corresponding details from (b). After one iteration, only fine scale details (the Gaussian noise) are present in the detail layer (this result is identical for all λ), while after 50 iterations also medium scale details (from the sinusoidal wave) are added to the detail layer. Note that $\lambda = 1$ significantly reduces the halo artifact, i.e. the overshooting on both sides of the step edge in (c).*

Equation 7 is therefore highly desirable near strong edges, because it significantly reduces halo artifacts by shifting the filtered result towards the original input. Based on these observations we conclude that our instantiation of the filter shown in Equation 6 should use $\lambda = 1$ and it should have a low diffusion strength in areas close to strong image edges.

Figure 5 shows filtering results for different settings of λ on a representative 1D example (we use our instantiation of matrix $\mathtt{H}$ that will be introduced in the following section). Note that in the case $\lambda = 0$, strong edge-aware smoothing creates significant halo artifacts in the tone mapping result. These artifacts are substantially reduced in the case $\lambda = 1$, since the filtering result is biased towards the stimulus signal in the area close to the step edge.

The bilateral filter [Durand and Dorsey 2002] and the domain transform [Gastal and Oliveira 2011] correspond to an instantiation of Equation 6 with $\lambda = 0$. However, the WLS filter [Farbman et al. 2008] uses $\lambda = 1$ as will be shown in Section 4.4.

Note that the result of each iteration step in Equation 6 can be efficiently computed in parallel since every $J_p^{(k+1)}$ can be computed independently of other $J_q^{(k+1)}$. Such a parallelized implementation leads to an overall runtime of $\mathcal{O}(k|\Omega|)$ where $|\Omega|$ is the total number of pixels and k is the number of iterations. Even faster runtimes are possible depending on the structure of $\mathtt{H}$, which we discuss in the following section. Furthermore, Perron-Frobenius and Markov chain theory [Meyer 2001] show that the iteration converges to a non-trivial solution for $0 < \lambda \leq 1$ if all $h_{pp} \neq 0$. This allows us to approximate the convergence result by stopping the iteration after a reasonable number of iteration steps, which depends on the particular instantiation of matrix $\mathtt{H}$ in Equation 6.

4.2 Spatial filtering

In this section, we discuss how to perform the spatial filtering step described in Section 3 using the filter formulation from Equation 6. More specifically, while considering our conclusions from the previous section, we construct an iteration matrix $\mathtt{H}$ that leads to the smoothing results required for tone mapping after a low number of iterations. Thus, $\mathtt{H}$ has to perform a strong edge-aware diffusion per iteration, while having a low diffusion strength close to significant image edges, and λ should be 1. We observe that a strong edge-aware diffusion can be achieved even with a single iteration, if all pixels within a connected region of similar color contribute to the diffusion result at all pixels of that region [Cigla and Alatan 2013; Lang et al. 2012; Gastal and Oliveira 2011].

We derive $\mathtt{H}$ using *permeability weights* (mentioned earlier in Figure 4-c). We define the *permeability* between a pixel $p = (p_x, p_y)$ and its right neighbor $p' = (p_x + 1, p_y)$ as a variant of the Lorentzian edge-stopping function [Durand and Dorsey 2002]

$$\tilde{\pi}_p := \left(1 + \left|\frac{I_p - I_{p'}}{\sigma}\right|^{\alpha}\right)^{-1}. \tag{8}$$

The permeability between p and its right neighbor pixel is close to 0 if the absolute value of the corresponding color difference is high, and it is 1 if the difference is low. The parameter σ indicates the point of transition from large to low permeability, while α controls the slope of the transition around σ. In this work we used $\alpha = 2$ and a σ in the range $0.1 - 1$. Similar to [Cigla and Alatan 2013], we use these permeability weights to define the permeability between two arbitrary pixels p and q as

$$\pi_{pq} := \begin{cases} 1 & : \quad p = q \\ \prod_{n=p_x}^{q_x - 1} \tilde{\pi}_{(n,p_y)} & : \quad p_x < q_x,\ p_y = q_y \\ \prod_{n=q_x}^{p_x - 1} \tilde{\pi}_{(n,p_y)} & : \quad p_x > q_x,\ p_y = q_y \\ 0 & : \quad \text{otherwise} \end{cases}. \tag{9}$$

Thus, the permeability between two pixels p and q of the same row is large if the permeability between each pair of neighboring pixels on the path between p and q is large. In particular, the permeability between p and q is significantly reduced, if there is a pair of neighboring pixels along the path with low permeability (e.g. a strong vertical edge between pixels p and q). The permeability between pixels of different rows is defined to be zero. To obtain a stochastic matrix $\mathtt{H} := \{h_{pq}\}$, we normalize these weights according to

$$h_{pq} := \frac{\pi_{pq}}{\sum_{n=1}^{w} \pi_{(n,p_y),q}}, \tag{10}$$

where w is the image width. Note that the diffusion strength $1/h_{pp}$ tends to become smaller in the neighborhood of strong vertical edges since the permeability to all pixels on the other side of the edge is close to zero. The iteration matrix $\mathtt{H}$ is derived from weights describing inter-pixel permeability in the horizontal direction (Equation 8). Analogously, we also derive another matrix for permeability in the vertical direction. We denote the corresponding matrices $\mathtt{H}_\mathtt{h}$ and $\mathtt{H}_\mathtt{v}$.

Spatial filtering is conducted according to Equation 6 by alternating between $\mathtt{H}_\mathtt{h}$ and $\mathtt{H}_\mathtt{v}$ for each k and using $\lambda = 1$. After each iteration, the direction of diffusion is changed which allows a fast diffusion of intensities in connected regions of similar color [Lang et al. 2012; Gastal and Oliveira 2011]. Figure 6 shows a spatial filtering result and a visualization of the corresponding permeability map.

Due to the structure of the iteration matrices, diffusion is conducted either only in horizontal or only in vertical direction. Thus the overall runtime (on a parallel architecture) is $\mathcal{O}(kw + kh)$ where w is

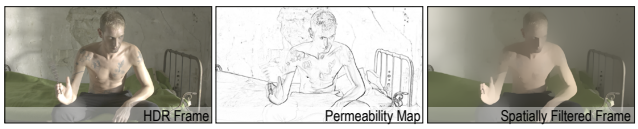

Figure 6: *Edge-aware spatial filtering using permeability map. Brighter colors in the permeability map indicates higher permeability.*

the width and h is the height of image I and k is the number of iterations. Thus, in cases where the total number of iterations is significantly smaller than $(wh)/(w+h)$, the runtime becomes sub-linear with respect to the total number of image pixels.

In practice, we observed that after a small number of iterations the filtering results became visually equivalent[4]. Our unoptimized Matlab implementation (using the Parallel Computing Toolbox) requires 19 ms for computing the permeability map, and 12 ms for a filtering iteration (comprising both vertical and horizontal steps) for a megapixel color image on a current consumer PC.

4.3 Temporal filtering

An important advantage of our filter formulation from Equation 6 is that it can easily be adapted to also perform the temporal filtering step in our tone mapping pipeline discussed in Section 3. Differently in temporal filtering, diffusion is conducted along the temporal dimension over the warped temporal neighborhood (Figure 4-e) with only one iteration. The permeability weights are computed in the same way as in the spatial filtering using the Lorentzian edge-stopping function variant (Equation 8). While their computation is

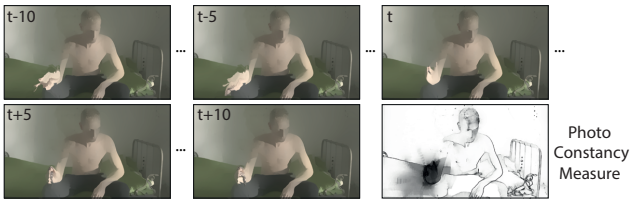

Figure 7: *Photo-constancy measure limits permeability near warping errors, but still allows filtering over temporal paths with small color differences.*

the same as in the spatial domain, the interpretation of the permeability weights is slightly different in the temporal domain. In the temporal domain the permeability is additionally affected by the warping errors within the temporal neighborhood. The direct extension of the permeability weights to the temporal domain, where the weights are obtained by applying Equation 8 to the color differences between consecutive frames, can therefore interpreted as a *photo-constancy* measure (Figure 7). The photo-constancy measure limits the permeability at strong temporal edges, as well as regions that are erroneously warped due to incorrect optical flow.

While we observed that the photo-constancy measure can be tuned to stop temporal filtering at most warping errors, we found that such a conservative configuration also stops at the temporal artifacts, which defeats the purpose of the temporal filtering. Our solution is to introduce another measure that penalizes flow vectors with high gradients (Figure 8), as it is a good indication of complex motion and the flow estimation tends to be erroneous in such regions [Mac Aodha et al. 2013].

[4] A visual comparison is presented in supplemental material.

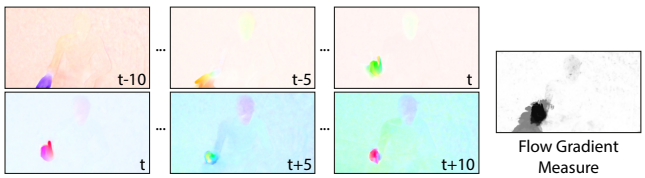

Figure 8: *Flow gradient measure penalizes flow vectors with high temporal gradient, providing an additional means to prevent warping artifacts.*

Both the photo-constancy and the *flow gradient* measures are normalized by Equation 8, and the final permeability is obtained by multiplying the two. This way, the photo-constancy constraint can be relaxed which allows regions with temporal artifacts to be permeable. The parameter values of both constraints depend on the tolerance to warping errors, desired level of temporal smoothness and the quality of flow. In our implementation, we empirically found that setting σ to 0.1 for photo-constancy, and to 2 for flow gradient measures, while keeping $\alpha = 2$ results in sufficient temporal filtering while still suppressing warping artifacts.

4.4 Relation to the WLS filter

Finally, we investigate the theoretical relation between the WLS filter, which is known to create anisotropic smoothing results of high quality [Fattal 2009; Gastal and Oliveira 2011], and our permeability weights-based smoothing method. WLS smoothing is uniquely defined as the solution of the following linear system

$$I_p = J_p + \sum_{q \in \mathcal{N}_4(p)} a_{pq} \left(J_p - J_q \right), \tag{11}$$

where I_p and J_p are pixel values of the input image and the smoothed output image, respectively, $\mathcal{N}_4(p)$ is the 4-neighborhood around pixel p, and a_{pq} are the smoothness weights as defined in [Lischinski et al. 2006]

$$a_{pq} := \frac{\lambda}{|I_p - I_q|^\alpha + \varepsilon}. \tag{12}$$

Apart from differences in notation, this is the same equation system as shown in the original article [Farbman et al. 2008]. The authors solve the equation system using a multi-resolution preconditioned conjugate gradient solver, but they also indicate that the same solution can be obtained iteratively with a Jacobi iteration. The corresponding Jacobi iteration is [Meyer 2001]

$$J_p^{(k+1)} = \frac{I_p + \sum_{q \in \mathcal{N}_4(p)} a_{pq} \, J_q^{(k)}}{1 + \sum_{q \in \mathcal{N}_4(p)} a_{pq}}. \tag{13}$$

This iteration is an instance of Equation 6 where λ is 1 and the iteration matrix $\mathtt{H}$ has the coefficients

$$h_{pq} := \begin{cases} a_{pq} / \left(1 + \sum_{r \in \mathcal{N}_4(p)} a_{pr} \right) & : \quad q \in \mathcal{N}_4(p) \\ 1 / \left(1 + \sum_{r \in \mathcal{N}_4(p)} a_{pr} \right) & : \quad q = p \\ 0 & : \quad \text{otherwise} \end{cases}. \tag{14}$$

Hence, WLS uses a bias term with $\lambda = 1$ which reduces halo artifacts, and it smoothens the input image by iteratively diffusing intensities in a small 4-neighborhood. Our method uses also a bias term with $\lambda = 1$, but in contrast to WLS it uses a significantly larger diffusion range per iteration which is controlled by permeability

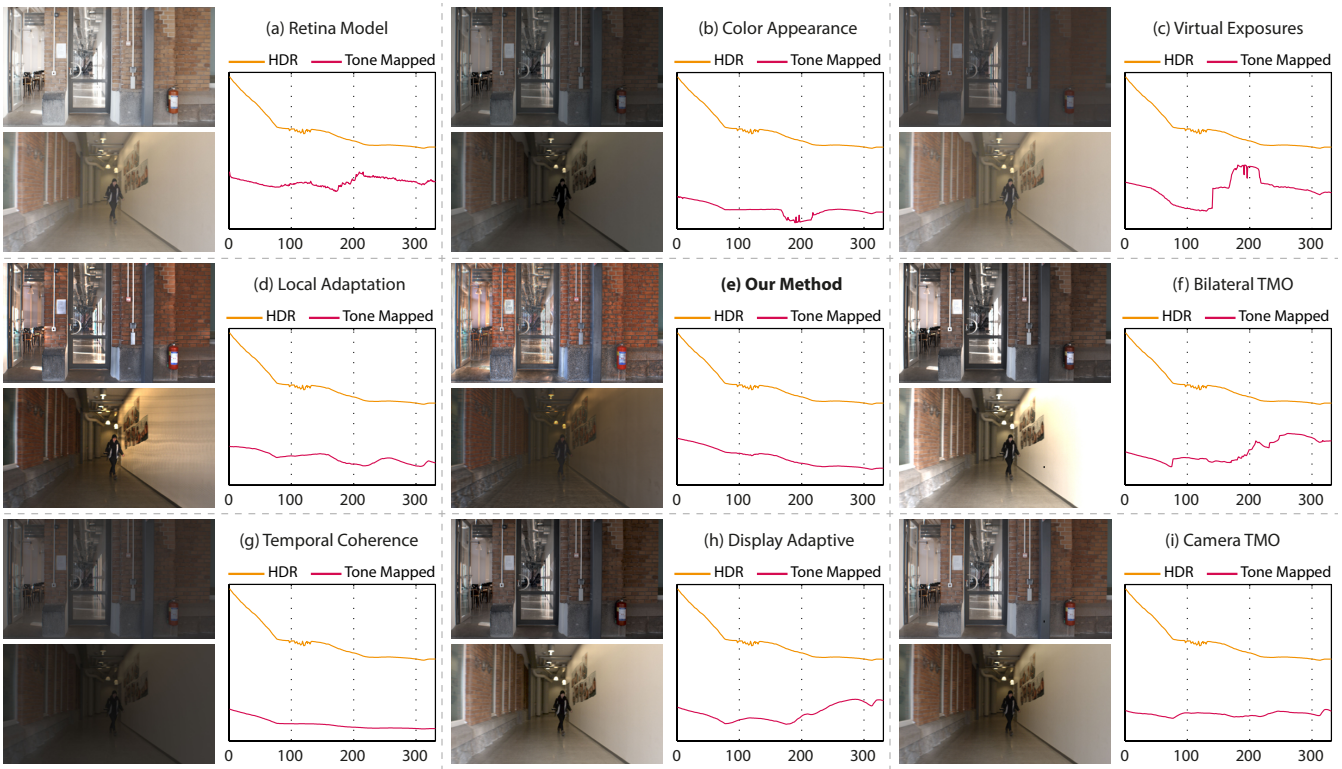

Figure 9: *Comparison of our method with current TMOs capable of processing videos. For each method, we show two representative frames and plot the mean pixel values (tone mapped) and logmean luminance (HDR) over time. Note that the HDR plots are shifted and compressed by an exponent for presentation.*

weights. In our method, by alternating the direction of diffusion between horizontal and vertical directions with each iteration, pixels are diffused in potentially large image regions after only a few iterations. However, although both smoothing methods converge against mathematically different solutions, filtering results are visually similar in practice[4].

5 Results

Our video tone mapping method was developed by necessity for a film production. Similar to Eilertsen et al. [2013], our experience with the state-of-the-art in video tone mapping yielded unsatisfactory results. Our newly developed method has been successfully used in a professional post-production workflow at FullHD and 2880×1620 resolutions. Excluding I/O, our unoptimized Matlab implementation of the full tone mapping pipeline requires on average 3.445 seconds to process a color frame at FullHD resolution, and 6.465 seconds at 2880×1620 resolution on a current desktop computer. The computation time of our method scales approximately linearly with the resolution of the input HDR video.

The results in this section have been generated by using the default parameter values presented in Sections 3 and 4, with the exceptions of σ^s (that controls spatial permeability) and the tone curve parameters (either the compression factor c or the $bias$ parameter of the adaptive logarithmic tone curve, depending on which method was used). Since the spatial permeability and tone curve parameters mainly govern the aesthetics of the tone mapped content, their values can be set up to personal taste. For generating the results in this section and the supplemental video we used $\sigma^s = [0.1, 1]$ and $bias = [0.2, 0.5]$ whenever we used the adaptive logarithmic tone curve. When we desired a stronger dynamic range compression, we utilized the log curve with the compression factor $c = [0.2, 0.4]$. We also conservatively set the number of spatial filtering iterations to 20. (Performance benchmarks have been obtained with these settings and temporal neighborhood size 21.)

In the remainder of this section, we first qualitatively compare our method's results on the publicly available HDR video data provided by Eilertsen et al. [2013] (Section 5.1). On the same public data set we also performed a controlled subjective study, which validates our initial claim that our method maintains a high level of local contrast with fewer temporal artifacts compared to the state-of-the-art (Section 5.2). After briefly discussing user interaction (Section 5.3), we present our main results obtained from processing HDR footage shot with an Arri Alexa XT camera at FullHD or 2880×1620 resolutions[5] (Section 5.4). Finally we present examples on how our method allows control over spatial and temporal contrast (Section 5.5), and demonstrate the advantage of using our method in low light HDR sequences (Section 5.6).

5.1 Comparison with Other Methods

In Figure 9 we show a qualitative comparison between a number of current tone mapping operators on the publicly available *Hallway* [Eilertsen et al. 2013; Kronander et al. 2013a; Kronander et al. 2013b]. In addition to the provided tone mapping results, we generated another sequence by running the Bilateral TMO using standard parameters. Presenting a concise and meaningful visual comparison between multiple video TMOs is challenging for various reasons. In Figure 9, we show for each video two representative

[5]Refer to the supplemental video for more results.

frames to demonstrate its ability to show local contrast, as well as a plot of the mean brightness of both the HDR and the tone mapped sequences over time, which is helpful for investigating temporal coherence.

The brightness flickering problems of local methods Retina Model [Benoit et al. 2009] (a), Color Appearance [Reinhard et al. 2012] (b) and Virtual Exposures [Bennett and McMillan 2005] (c) can be observed from their mean brightness plots. The Local Adaptation [Ledda et al. 2004] (d) operator performs better in terms of temporal coherence due to its temporal filtering. However, since the temporal filtering does not utilize motion paths, its results show strong ghosting artifacts (See the wall at the bottom image in Figure 9-d). Consistent with the observation that local image tone mapping operators generate strong temporal artifacts [Eilertsen et al. 2013], the mean brightness plot of the Bilateral TMO [Durand and Dorsey 2002] shows notable fluctuations over time. The mean brightness plots of the remaining operators Temporal Coherence [Boitard et al. 2012], Display Adaptive [Mantiuk et al. 2008] and Camera TMO [Eilertsen et al. 2013] suggest less temporal artifacts. The downside of these operators is their limited local contrast reproduction capability due to their spatially-invariant processing. On the other hand, we configured our TMO to emphasize local spatial contrast while also maintaining the temporal contrast between the beginning and the end of the sequence. Our tone mapping result remains temporally coherent similar to the global operators even though we strongly amplify the local contrast of the sequence (e).

We also compared our technique with Boitard et al.'s [2014b] recent segmentation based temporal coherency method applied to Ramsey et al. [2004] and the recursive domain transform filter [Gastal and Oliveira 2011] (Figure 10)[6]. The results show that Boitard et al.'s [2014b] method tends to underutilize the available display dynamic range and generates dark frames with relatively low contrast similar to their earlier work [Boitard et al. 2012] (see Figure 9-g). The mean pixel value plots show fluctuations (notably between frames 100 and 130) even for the Ramsey et al. [2004] version which is a global TMO. Note that our method is relatively stable at the same interval despite being local and therefore more susceptible to temporal artifacts. Finally, our method reproduces temporal contrast (Section 3.1) of the source HDR video better. Our result preserves the transition from the bright exterior to the darker hallway by modulating the mean luminance accordingly, whereas this transition is diminished or reversed in Boitard et al.'s [2014b] results.

5.2 Subjective Study

While the qualitative comparison in Section 5.1 is useful as an overview and helps to put the many TMOs into perspective, it certainly does not capture all the aspects relevant to the evaluation of video tone mapping. For example, while not apparent from Figure 9, most operators tend to show camera noise in their results. To that end, we did a pair of controlled subjective rating studies to compare our method with others in terms of local contrast reproduction and the absence of temporal artifacts. The studies respectively had 20 and 19 paid participants mainly consisting of college students. The stimuli consisted of 11 tone mapped sequences for each of the 4 HDR video sequences in our test set. The HDR video sequences and, whenever available, the tone mapped results were obtained from the public data set by Eilertsen et al. [2013] (*Hallway*, *Hallway2*, *ExhibitionArea* and *Students* sequences were used). The tone mapped videos for the Bilateral TMO [Durand and Dorsey 2002], the Gradient Domain TMO [Fattal et al. 2002] and the WLS TMO [Farbman et al. 2008] were generated from the source HDR videos using default parameters. Also, the results of our method have been generated using identical parameters for each video.

[6]Boitard et al.'s [2014b] results for the Hallway sequence were made public after the conditional acceptance of this publication and therefore have been evaluated separately in Figure 10.

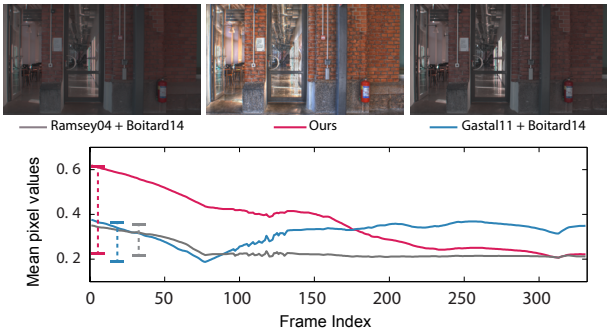

Figure 10: *Comparison of our method with Ramsey et al. [2004] and Gastal et al. [2011] (recursive filtering version) combined with Boitard et al.'s [2014] segmentation based temporal coherency method (both provided by Ronan Boitard). The dashed lines denote the min and max of each frame's averaged pixel values over the test sequence. Our method uses the display dynamic range more efficiently and causes less temporal artifacts. Note also that Gastal et al. [2011] with Boitard et al. [2014] reverses the temporal contrast of the sequence.*

The study was performed through an interface similar to Petit et al. [2013], and the subjects were asked to rate each video (presented on a NEC MultiSync 2490WUxi[2] display) on a scale from 0 to 10, in a first experiment for their local contrast reproduction, and in a second and separate experiment for the absence of temporal artifacts. Note that these questions directly assess our primary claim from Section 1 that our method is capable of maintaining a high level of local contrast with fewer artifacts compared to the state-of-the-art. Thus, we collected 88 data points for each subject. In a training session before each experiment, the participants were explained the concepts of local contrast and temporal artifacts. While there were no time limitations to our study, the average subject finished in approximately 30 minutes.

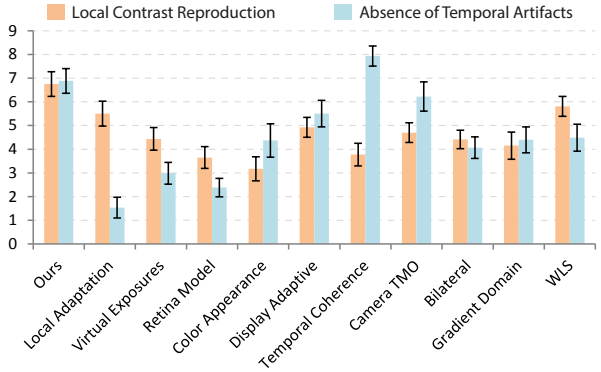

Figure 11: *Average ratings and* 95% *confidence intervals from our user study.*

The average ratings and the bars showing 95% confidence intervals are shown in Figure 11. We performed One-factor Analysis of Variance (ANOVA) with repeated measures to analyze both experiments. The main effects were statistically significant for both

the local contrast ($p < .001$) and temporal artifacts ($p < .001$) using Greenhouse-Geisser adjusted degrees of freedom. Pairwise comparison of video processing methods was conducted using Bonferroni's post-host test. For the contrast reproduction experiment, our method was assessed to have a significantly higher ($p < .05$) contrast than all methods except WLS. For the absence of temporal artifacts experiment, our method similarly was assessed to have significantly fewer ($p < .05$) artifacts than all methods except the Temporal Coherence TMO. The similarity in local reproduction with WLS is expected because of our filtering approach's connection to WLS (Section 4.4). Since WLS is an image TMO however, it scores significantly lower with respect to the absence of temporal artifacts. On the other hand, we note that the Temporal Coherence TMO, that scores highest in the second experiment tends to produce dark frames (See Figure 9-g), which might have effected the visibility of the temporal artifacts. For the same reason, it gets a low average score in the local contrast reproduction experiment.

To conclude, the findings of the subjective study as well as the qualitative comparison in Section 5.1 suggest that our method is capable of producing results with high local contrast with a similar level of temporal coherence as the global operators.

5.3 User Interaction

HDR tone mapping is an artistic process that often involves experimentation through trial and error. As such, any tone mapping software is expected to give the user visual feedback at interactive rates while tone mapping parameters are being changed. The dilemma is that interactive video editing on current standard hardware at high resolutions is challenging to achieve beyond simple operations. Our solution for interactive editing consists of the offline pre-computation of the base and detail layers required for tone mapping, followed by and interactive editing session where the user adjusts tone mapping parameters through a simple user interface (Figure 12). While most of the artistic decisions can be deferred until after the pre-computation step, the user needs to select the σ^s parameter that controls the base layer smoothness a priori. That said, we found that our two-step tone mapping workflow to work well in practice, as σ^s and temporal filtering parameters can efficiently set by trial and error on a small number of representative frames. Overall, despite its simplicity, we observed that user interface notably facilitated the creative side of the tone mapping process in a post-production environment.

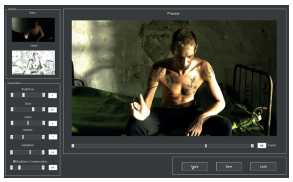

Figure 12: *A screenshot from our user interface during editing process.*

5.4 Local HDR Video Tone Mapping

The main application of our method is local video tone mapping. Figure 13, our teaser and the supplemental video show our method's results produced from HDR sequences obtained with an Arri Alexa XT camera at FullHD or 2880×1620 resolution[7]. The input video sequences contain complex motion and strong changes in scene illumination. Despite that, our method maintains a temporally stable mean brightness over time, as well as high local contrast.

[7]Due to their high contrast, our results are best viewed on a computer display. Printed versions may not be representative of the intended results.

5.5 Spatial and Temporal Contrast Adjustment

Local tone mapping allows producing various visual styles by adjusting the spatial scale of the detail layer. Figure 14 shows results of two σ^s values where one emphasizes fine scale spatial details (right), and the other produces a more balanced result (left).

Figure 14: *Similar to local image TMOs, our method allows the user to adjust the emphasis given to fine scale details. Left image is generated with $\sigma^s = 1$, whereas the right image with $\sigma^s = 0.1$ (extreme parameter settings have been used for illustration). In the right image, note how coarse scale details (e.g. shadows) become less pronounced at the cost of emphasized fine scale details.*

Our framework also allows explicit control over the temporal contrast (discussed in Section 3.1) of the tone mapped video. One possibility when tone mapping HDR sequences with high temporal contrast to apply strongly compressive tone curve, which reduces the dynamic range of the whole sequence such that it fits into the display dynamic range without generating any over or under exposed regions. However, such a tone mapping that significantly reduces temporal contrast may sometimes be undesirable. The other option of applying a less compressive tone curve, on the other hand, may generate strongly over or under exposed frames (Figure 15, top row). This can easily be remedied in our framework using a brightness compensation factor that controls the brightness of each frame. Since our method's base layer is temporally stable, such a brightness compensation factor can be modulated by the inverse of the logmean luminance of the spatiotemporal base layer without causing brightness flickering (Figure 15, bottom row).

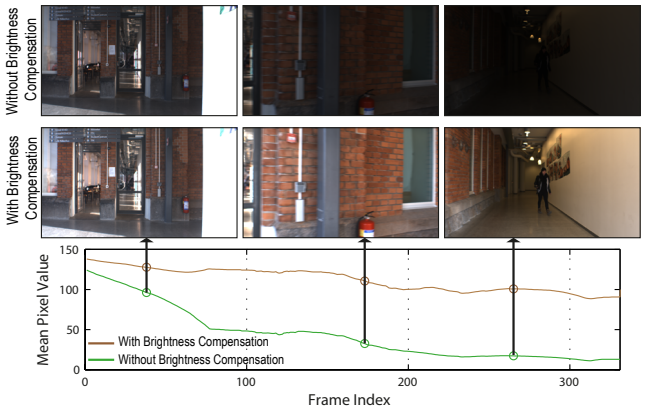

Figure 15: *The brightness compensation applied as a function of logmean value of our method's temporally coherent base layer can be used to limit the temporal contrast in tone mapping configurations where a strongly compressive tone curve is not desired.*

5.6 Tone Mapping of Low-light Sequences

Tone mapping video sequences shot in low-light scenes is often difficult due to the camera noise. TMOs can not distinguish camera

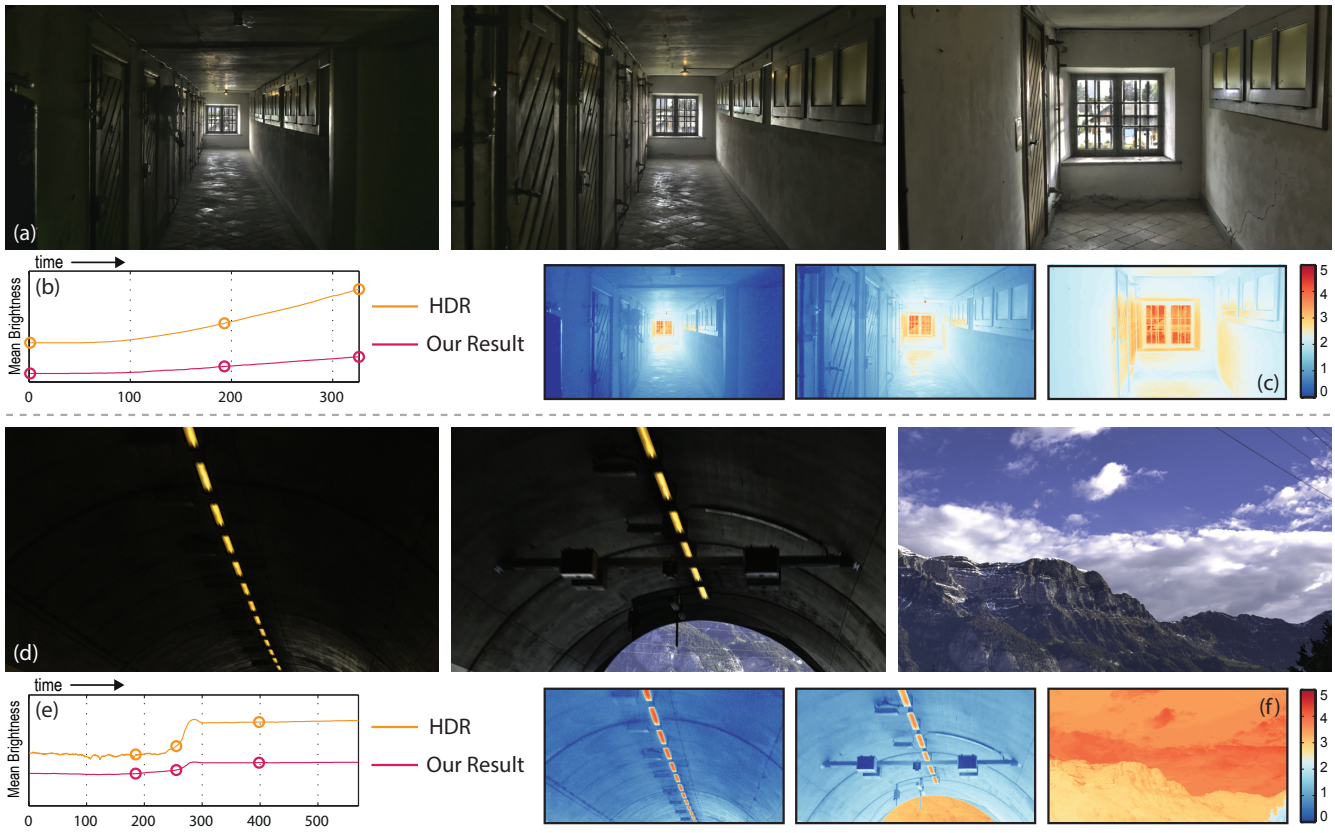

Figure 13: *Representative local tone mapped frames from two sequences (a, d). Despite the high local contrast and strong dynamic range compression, our method maintains a temporally stable logmean luminance over time (b, e). The input HDR videos that span more than* 4 *log_{10} units are visualized in false color (c, f).*

noise from scene details, which means that as scene details become more visible during the tone mapping process, so does the noise. Due to the denoising effect temporal filtering involved in our tone mapping pipeline, our method is especially suitable for such low-light sequences. Figure 16 shows an example where the TMO significantly amplifies camera noise (a), which can be somewhat suppressed by denoising the HDR frame before applying tone mapping (b). The advantage of our method is that our results are comparable to the latter without the need of an extra denoising step (c).

6 Discussion

While our method does not offer the final solution to the open problem of local HDR video tone mapping, we believe that it takes a big step forward by significantly advancing the state-of-the-art and paves the way for the (much needed) future work in this field. As the entertainment industry is moving towards HDR video pipelines, we think that the relevant research can be stimulated by further advances in filtering techniques geared towards tone mapping applications, and the emergence of high quality public HDR video data.

Our method is not without limitations. As an example, for artistic reasons it is desirable to have multiple detail layers at different contrast scales. Due to the many technical challenges involved in obtaining even a single detail layer, we deferred a multi-scale approach to future work. Also, as with every optical flow based method, our technique is always susceptible to visual artifacts due to the inevitable errors in the flow computation. Even though the outcome of our subjective study (Section 5.2) suggests that these artifacts can be suppressed to a reasonable level, we still feel that our method can be improved by using additional and more sophisticated flow confidence measures (e.g. by utilizing machine learning [Mac Aodha et al. 2013]). Finally, our method currently provides only a basic means for controlling chrominance in the form of a saturation slider, which we hope to address in future work.

7 Conclusion

We presented a local HDR video tone mapping method, that can significantly reduce the input dynamic range while preserving local contrast. Our key difference is the use of a temporal filtering through per-pixel motion paths, which allowed us to achieve temporal stability without ghosting. We formulated an edge-aware filter that is applied by iterative shift-variant convolutions, and shares the same halo suppression properties with the WLS filter, which allowed us to efficiently realize our tone mapping framework. We presented qualitative and subjective comparisons to the state-of-the-art which resulted favorably for our method. We showed results produced with various tone mapping configurations from challenging HDR sequences, and presented a simple yet effective user interface. Finally we noted the advantage of our method in low-light HDR sequences.

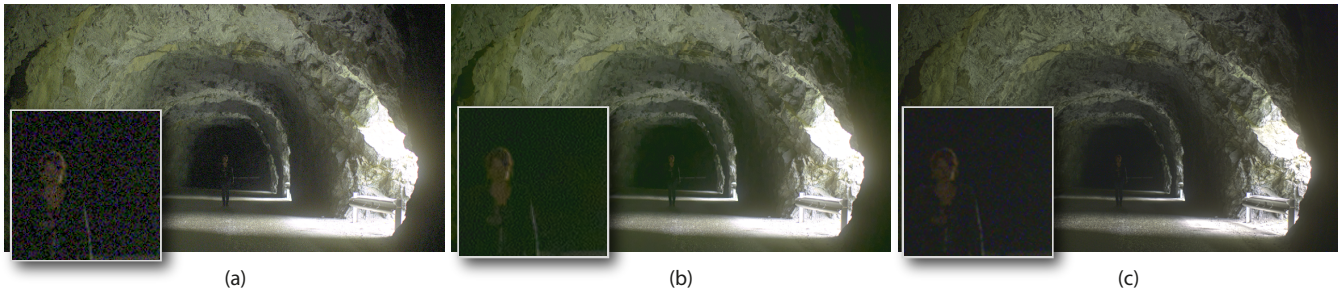

Figure 16: *A frame-by-frame tone mapping using the bilateral TMO results in significantly visible camera noise (a), which can be reduced by denoising each frame before tone mapping using external video editing software, as an example we use The Foundry NUKE (b). On the other hand, our method can achieve comparably low noise levels without an extra denoising step.*

Acknowledgements

The authors would like to thank the anonymous reviewers, the participants of our user study, the *Lucid Dreams of Gabriel* crew, Michael Schaffner, Steven Poulakos, Maurizio Nitti and Ronan Boitard.

References

ADAMS, A. 1981. *The Print, The Ansel Adams Photography Series 3*. New York Graphic Society.

AUBRY, M., PARIS, S., HASINOFF, S. KAUTZ, J., AND DURAND, F. 2014. Fast local laplacian filters: Theory and applications. *ACM Trans. Graph. (To Appear) 33*, 4.

BENNETT, E. P., AND MCMILLAN, L. 2005. Video enhancement using per-pixel virtual exposures. *ACM Trans. Graph. 24*, 3.

BENOIT, A., ALLEYSSON, D., HERAULT, J., AND CALLET, P. L. 2009. Spatiotemporal tone mapping operator based on a retina model. In *CCIW*, Springer, vol. 5646 of *Lecture Notes in Computer Science*, 12–22.

BOITARD, R., BOUATOUCH, K., COZOT, R., THOREAU, D., AND GRUSON, A. 2012. Temporal coherency for video tone mapping. *Proc. SPIE 8499*.

BOITARD, R., COZOT, R., THOREAU, D., AND BOUATOUCH, K. 2014. Survey of Temporal Brightness Artifacts in Video Tone Mapping. In *HDRi2014*.

BOITARD, R., COZOT, R., THOREAU, D., AND BOUATOUCH, K. 2014. Zonal brightness coherency for video tone mapping. *Signal Processing: Image Communication 29*, 2, 229 – 246.

CIGLA, C., AND ALATAN, A. A. 2013. Information permeability for stereo matching. *Signal Processing: Image Communication 28*, 9, 1072–1088.

DRAGO, F., MYSZKOWSKI, K., ANNEN, T., AND CHIBA, N. 2003. Adaptive logarithmic mapping for displaying high contrast scenes. *Computer Graphics Forum 22*, 3, 419–426.

DURAND, F., AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. Graph. 21*, 3, 257–266.

EILERTSEN, G., WANAT, R., MANTIUK, R. K., AND UNGER, J. 2013. Evaluation of tone mapping operators for hdr-video. *Computer Graphics Forum 32*, 7, 275–284.

FARBMAN, Z., FATTAL, R., LISCHINSKI, D., AND SZELISKI, R. 2008. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graph. 27*, 3, 67:1–67:10.

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. *ACM Trans. Graph. 21*, 3, 249–256.

FATTAL, R. 2009. Edge-avoiding wavelets and their applications. *ACM Trans. Graph. 28*, 3, 22:1–22:10.

FERWERDA, J. A., PATTANAIK, S. N., SHIRLEY, P., AND GREENBERG, D. P. 1996. A model of visual adaptation for realistic image synthesis. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, 249–258.

GASTAL, E. S. L., AND OLIVEIRA, M. M. 2011. Domain transform for edge-aware image and video processing. *ACM Trans. Graph. 30*, 4, 69:1–69:12.

GUTHIER, B., S., K., M., E., AND W., E. 2011. Flicker reduction in tone mapped high dynamic range video. *Proc. SPIE 7866*.

HE, K., SUN, J., AND TANG, X. 2013. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence 35*, 6, 1397–1409.

IRAWAN, P., FERWERDA, J. A., AND MARSCHNER, S. R. 2005. Perceptually based tone mapping of high dynamic range image streams. In *Proceedings of Eurographics Conf. on Rendering Techniques*, EGSR'05, 231–242.

KANG, S. B., UYTTENDAELE, M., WINDER, S., AND SZELISKI, R. 2003. High dynamic range video. *ACM Trans. Graph. 22*, 3, 319–325.

KISER, C., REINHARD, E., TOCCI, M., AND TOCCI, N. 2012. Real time automated tone mapping system for HDR video. In *Proceedings of the IEEE International Conference on Image Processing*, 2749–2752.

KRONANDER, J., GUSTAVSON, S., BONNET, G., AND UNGER, J. 2013. Unified HDR reconstruction from raw cfa data. In *Computational Photography (ICCP), 2013 IEEE International Conference on*, 1–9.

KRONANDER, J., GUSTAVSON, S., BONNET, G., YNMERMAN, J., AND UNGER, J. 2013. A unified framework for multi-sensor HDR video reconstruction. In *Signal Processing: Image Communications*.

LANG, M., WANG, O., AYDIN, T., SMOLIC, A., AND GROSS, M. 2012. Practical temporal consistency for image-based graphics applications. *ACM Trans. Graph. 31*, 4 (July), 34:1–34:8.

LEDDA, P., SANTOS, L. P., AND CHALMERS, A. 2004. A local model of eye adaptation for high dynamic range images. AFRIGRAPH, 151–160.

LEE, C., AND KIM, C.-S. 2007. Gradient domain tone mapping of high dynamic range videos. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 3, III – 461–III – 464.

LISCHINSKI, D., FARBMAN, Z., UYTTENDAELE, M., AND SZELISKI, R. 2006. Interactive local adjustment of tonal values. *ACM Trans. Graph.*, 646–653.

MAC AODHA, O., HUMAYUN, A., POLLEFEYS, M., AND BROSTOW, G. J. 2013. Learning a confidence measure for optical flow. *IEEE Trans. Pattern Anal. Mach. Intell. 35*, 5, 1107–1120.

MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2006. A perceptual framework for contrast processing of high dynamic range images. *ACM Trans. Appl. Percept. 3*, 3, 286–308.

MANTIUK, R., DALY, S., AND KEROFSKY, L. 2008. Display adaptive tone mapping. *ACM Trans. Graph. 27*, 3, 68:1–68:10.

MEYER, C. D. 2001. *Matrix Analysis and Applied Linear Algebra*. SIAM: Society for Industrial and Applied Mathematics.

MILANFAR, P. 2013. A tour of modern image filtering. *IEEE Signal Processing Magazine 30*, 1, 106–128.

MRÁZEK, P., WEICKERT, J., AND BRUHN, A. 2004. On robust estimation and smoothing with spatial and tonal kernels. In *Proc. Dagstuhl Seminar: Geometric Properties from Incomplete Data*, Springer, 3–88722.

NORDSTROM, K. N. 1989. Biased anisotropic diffusion–a unified regularization and diffusion approach to edge detection. Tech. rep., EECS Dept., UC Berkeley.

PARIS, S., HASINOFF, S. W., AND KAUTZ, J. 2011. Local laplacian filters: Edge-aware image processing with a laplacian pyramid. *ACM Trans. Graph. 30*, 4, 68:1–68:12.

PATTANAIK, S. N., TUMBLIN, J., YEE, H., AND GREENBERG, D. P. 2000. Time-dependent visual adaptation for fast realistic image display. In *Proc. of Conf. on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, 47–54.

PETIT, J., AND MANTIUK, R. K. 2013. Assessment of video tone-mapping: Are cameras s-shaped tone-curves good enough? *Journal of Visual Communication and Image Representation 24*, 7, 1020–1030.

RAMSEY, S. D., JOHNSON III, J. T., AND HANSEN, C. 2004. Adaptive temporal tone mapping. In *Proceedings of the 7th IASTED International Conference on Computer Graphics and Imaging*, 124–128.

REINHARD, E., AND DEVLIN, K. 2005. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics 11*, 13–24.

REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. *ACM Trans. Graph. 21*, 3, 267–276.

REINHARD, E., WARD, G., PATTANAIK, S., DEBEVEC, P., HEIDRICH, W., AND MYSZKOWSKI, K. 2010. *HDR Imaging - Acquisition, Display, and Image-Based Lighting, Second Edition*. Morgan Kaufmann.

REINHARD, E., POULI, T., KUNKEL, T., LONG, B., BALLESTAD, A., AND DAMBERG, G. 2012. Calibrated image appearance reproduction. *ACM Trans. Graph. 31*, 6 (Nov.), 201:1–201:11.

TOCCI, M. D., KISER, C., TOCCI, N., AND SEN, P. 2011. A versatile hdr video production system. *ACM Trans. Graph. 30*, 4, 41:1–41:10.

TOMASI, C., AND MANDUCHI, R. 1998. Bilateral filtering for gray and color images. In *ICCV*, 839–846.

VAN HATEREN, J. H. 2006. Encoding of high dynamic range video with a model of human cones. *ACM Trans. Graph. 25*, 4, 1380–1399.

YE, G., GARCES, E., LIU, Y., DAI, Q., AND GUTIERREZ, D. 2014. Intrinsic Video and Applications. *ACM Trans. Graph. (To Appear) 33*, 4.

ZIMMER, H., BRUHN, A., AND WEICKERT, J. 2011. Optic Flow in Harmony. *International Journal of Computer Vision 93*, 3, 368–388.