DOI: 10.1111/cgf.12684
Eurographics Symposium on Rendering 2015
J. Lehtinen and D. Nowrouzezahrai
(Guest Editors)

Volume 34 (2015), Number 4

Motion Aware Exposure Bracketing for HDR Video

Yulia Gryaditskaya[1,2], Tania Pouli[2], Erik Reinhard[2], Karol Myszkowski[1] and Hans-Peter Seidel[1]

[1]Max-Planck Institute for Informatics, Saarbrücken, Germany
[2]Technicolor, Rennes, France

Figure 1: *The HDR video capture algorithm is shown on two consecutive frames (tonemapped for display), separated by a visualization of scene areas that have moved between these frames. These motion areas are analysed on the fly to guide a novel metering algorithm that determines the next exposure, thus minimizing ghosting artefacts. The false color images on either side show the output of the structural similarity index (SSIM) for each of the RGB channels of the reconstructed HDR frames relative to the ground truth, and indicate that the error in our method is small and well-controlled.*

Abstract

Mobile phones and tablets are rapidly gaining significance as omnipresent image and video capture devices. In this context we present an algorithm that allows such devices to capture high dynamic range (HDR) video. The design of the algorithm was informed by a perceptual study that assesses the relative importance of motion and dynamic range. We found that ghosting artefacts are more visually disturbing than a reduction in dynamic range, even if a comparable number of pixels is affected by each. We incorporated these findings into a real-time, adaptive metering algorithm that seamlessly adjusts its settings to take exposures that will lead to minimal visual artefacts after recombination into an HDR sequence. It is uniquely suitable for real-time selection of exposure settings. Finally, we present an off-line HDR reconstruction algorithm that is matched to the adaptive nature of our real-time metering approach.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—

1. Introduction

In 2014, sales of digital cameras amounted to 1.8 billion units. Of these, 95% were camera phones. In Q3 of 2014 alone, more than 300 million smartphones were sold (statistics from TomiAhonen Phonebook 2014). All smartphones and many other camera phones as well as tablets have at least one and often two cameras on board with which images and video can be taken. However, the visual quality obtained with such mobile devices is frequently impacted, as the range of luminance, commonly present in scenes, exceeds that which can be captured. The 'HDR mode' on many modern smartphones helps to ameliorate this problem for still images through techniques akin to exposure fusion [MKR07].

Capturing video in high dynamic range (HDR), however, remains a difficult problem with solutions few and far between. In this area, specialised professional (e.g., Arri Alexa XT and Red Epic Dragon) and custom-built HDR video cameras constitute the state-of-the-art [NB03, TKTS11, WLA*12, KGBU13, MRK*13, FGE*14], providing high-

end solutions that are mostly free of artefacts. Nonetheless, high cost and significant levels of required expertise currently limit their use and availability.

Mobile devices are particularly ill-suited to directly capture HDR video due to small sensor sizes that are optimized to capture ever higher pixel resolutions (as an example, mobile phones with 4K camera sensors such as the Sony Xperia Z2 and the Samsung Galaxy S5 are beginning to penetrate the market). This means that the capture of strongly dynamic scenes with complex occlusion relations and a high range of luminance values remains a challenge on mobile platforms.

Novel sensors such as the Aptina AR1331CP and Sony IMX135, increase the captured dynamic range by varying the electronic shutter time between alternate Bayer rows. For such specialized sensor architectures, the burden is on advanced off-line processing to avoid artefacts [GHMN10, HST*14, CKL14], as reconstruction errors such as rolling shutter distortion, motion artefacts, etc. are more likely. While such techniques usefully extend the dynamic range, we note that the quantum capacity of mobile sensors is low, necessitating multiple exposures to even approach, let alone surpass, the dynamic range of standard DSLR cameras.

HDR acquisition with several differently exposed shots is a common technique [RWD*10] that is starting to appear in some new smartphones, especially for still photography. In order to work on a mobile device these solutions should possess a certain level of computational efficiency, which often comes at the cost of low accuracy results. Only camera motion is accounted for and in some cases at most relatively small motion compared to the image size can be handled. As a result, the gain in dynamic range may be offset by strong motion artefacts such as ghosting. In addition, as several exposures are needed for each 'HDR' image, the 'HDR' frame-rate may be lower than the capture frame-rate in the case of video.

To compensate, one could use costly high frame-rate cameras [GKE13], or advanced post-processing could be applied [KUWS03, SKY*12, KSB*13]. Such algorithms synthesize the missing luminance information for each frame, often with the aid of optical flow techniques. Here, motion estimation is particularly complicated as pixel differences are due to a combination of motion and exposure time differences. This may complicate the detection of pixel correspondences and cause ghosting artefacts [SS12]. Additionally, such merging approaches are designed to require a regular and repeating sequence of exposures, and therefore cannot be directly combined with advanced metering techniques.

On the other hand, there are advantages to incorporate more advanced metering into the capture process. This would allow a more efficient use of the timing budget, i.e., no exposures would be wasted on going beyond the extremes of the scene's luminance range. Advanced metering is feasible for still images, and typically requires a sampling of the full dynamic range of the scene prior to capturing a small set of exposures [GTM*12]. Such an approach is too time consuming to extend to HDR video capture.

To capture viable HDR video, we therefore present a real-time metering algorithm, and combine it with an adaptive off-line HDR reconstruction algorithm (see Figure 1). To guide the design of our algorithm, we have conducted a psychophysical experiment to assess the relative importance of ghosting against dynamic range artefacts. We find that under comparable conditions, ghosting takes precedence over clipped luminance artefacts. Thus, metering is designed such that during capture we avoid ghosting rather than fix it afterwards, while capturing the full dynamic range when ghosting is unlikely to occur. Through an analysis of motion in the scene, we control the time difference between consecutive exposures, which allows us to minimize motion related artefacts in the resulting videos.

We demonstrate that HDR sequences generated with our system of on-line metering and off-line reconstruction generally feature fewer disturbing artefacts, while at the same time significantly improving the captured dynamic range of the output. The metering part of the algorithm is matched to the capabilities of current mobile devices, and as such, our system would be applicable to current mobile technology and would therefore serve a very large market. In summary, we present the following contributions:

- We show for the first time a system that allows efficient metering in the context of HDR video capture, taking into account both luminance and motion.
- The algorithm is tuned on the basis of a fundamental psychophysical experiment.
- We describe an HDR video reconstruction algorithm that takes an arbitrarily varying spacing of LDR exposures and produces high quality HDR video without losing frame-rate.

In the following sections, we describe related work, the perceptual experiment that informed our metering algorithm, the metering algorithm itself and the HDR video reconstruction algorithm. We end the paper by showing results and drawing conclusions.

2. Related Work

Our work combines real-time exposure metering with the inclusion of motion estimation, and off-line HDR reconstruction. While such a combination of techniques is novel (as are the constituent components we present), here we review existing work in either area.

Exposure Metering Multi-exposure HDR acquisition typically uses exponentially spaced exposure times [MPMP95, DM97,MN99,RBS03], usually with a fixed ISO setting. One could also tabulate exposure times [GN03]. However, such

approaches may use more exposures than necessary, especially if the scene histogram is sparse.

Spacing the exposures according to a Fibonacci sequence would be more efficient [GIN13], offering additional benefits for exposure registration. Other refinements are to incorporate noise models [Hir10] or vary ISO settings [HDF10]; the latter based on the observation that high ISO settings offer better signal-to-noise ratios.

Alternatively, metering algorithms can be applied so that a minimal number of exposures is selected to capture the scene with an optimal signal-to-noise-ratio [BHD08, GAW*10, GTM*12]. This can be extended to a user-assisted metering variant to capture a single LDR image [GHKE13]. These metering techniques are designed to optimize between a large number of possible exposures and require collecting multiple images/histograms of a scene to recover its irradiance distribution. As such, they are not suitable for HDR video capture.

One exception is a metering algorithm designed to find exposure parameters from a single frame [IM10], fitting a log-normal distribution to well-exposed pixels to estimate the scene's histogram. From this curve one can estimate the optimal exposure for the next frame.

Without resorting to high frame-rate capture, we offer a novel solution that features high reactivity to changes in the scene content and dynamic range. Moreover, our metering algorithm does not require the knowledge of the full scene irradiance distribution. On the basis that multi-exposure techniques have a propensity to suffer from ghosting artefacts, instead our metering algorithm focuses on minimizing the capture of exposures that could lead to ghosting errors during video reconstruction.

Exposure Registration HDR video can be reconstructed from LDR input by alternating frames with different exposure times [KUWS03, MG10]. Optical flow, often calculated to register such exposures, may lead to artefacts if objects exhibit non-rigid motion [KSB*13, Fig. 6]. In these cases, PatchMatch can ameliorate the ill-effects of scene-motion [SKY*12], allowing for HDR video reconstruction [KSB*13]. Moreover, this technique is amenable to reconstructing HDR video from sequences of more than 2 exposures. It is, however, relying on regular repeating patterns of exposure timing, precluding the use of metering strategies. Further, if over-exposed image areas do not follow the dominant motion between frames, this algorithm may still produce visible artefacts. We therefore extend this technique to allow for more freedom in exposure selection, and as such better match it to our metering algorithm.

Ghost Removal Artefacts due to exposure registration problems could in principle be reduced by employing ghost removal techniques (see [SS12, TAEE15] for a recent survey). However, such algorithms often require more exposures per reconstructed HDR image than would be available in an HDR video acquisition scenario. Exceptions are techniques for exposure registration [SKY*12, HGPS13] that employ PatchMatch [SCSI08, BSFG09]. Here, only well-exposed pixels from a reference exposure are used, while the remaining pixels are synthesized through analysis of similarity to corresponding regions in exposures captured before and after the reference exposure.

Perception of Ghosting Artefacts Objects moving through a scene are normally perceived as sharp due to the smooth pursuit of eye motion. This is possibly accompanied by retinal blur of other scene parts that undergo motion relative to the object that is attended to. In video, objectionable ghosting artefacts may be caused by rendering techniques and display technologies such as view blending in image-based rendering [VCL*11, BLL*10], insufficient frame rate [Wat13, DXCZ14], too short exposure times [SBE*15], or cross-talk between views in stereoscopic and light field displays [ZMD*06, Woo12, WS03].

In each case, ghosting is consistently judged as the key factor degrading image quality [VCL*11, Woo12, WS03]. The reduction of ghosting typically goes hand-in-hand with concessions along other dimensions, for instance a lower image contrast, an increase of motion blur, lower stereoscopic depth range, etc. Specific to our application, ghosting can be suppressed at the expense of admitting some over-exposed pixels. To understand the ramifications of this trade-off, here we investigate for the first time the interactions between those two image quality dimensions.

3. Perceptual Assessment of Ghosting and Over-Exposure Artefacts

A related previous study by Wilcox and Stewart [WS03] assessed the relative importance of brightness and ghosting in a 3D cinema scenario. Here, ghosting is caused by cross-talk between eyes, an issue specific to stereo 3D which depends on object depth and contrast, and is mostly visible on a dark background. Ghosting was simulated by adding various percentages of the left eye signal to the right eye signal. Different brightness levels were created by linearly scaling the original footage down to various target levels. A magnitude estimation task revealed that ghosting was a more critical artefact than reductions in brightness.

In our case, ghosting arises not as a consequence of 3D signal interaction, but as an artefact of HDR reconstruction. Moreover, we are interested in relative importance of ghosting to clipping of over-exposed regions and loss of details in under-exposed (while the above study was dealing with overall brightness).

Different exposure allocation strategies generally offer a trade-off between different temporal and luminance spacings. If the luminance range covered by the exposures is not

large enough, parts of the reconstructed HDR frame may remain over-exposed, while large temporal gaps between exposures are likely to create ghosting in the presence of motion. To evaluate the relative importance of these two artefacts, we perform a psychophysical experiment, based on the subjective assessment methodology for video quality (SAMVIQ) protocol, which is standardized in ITU-R recommendation BT.500 for assessing video quality [ITU12].

In this protocol, video sequences, including a reference, are shown to the observers. Observers are tasked with assigning a score of visual quality to each sequence with respect to the reference (the continuous range between 1 and 100 is used in our settings.) The reference is additionally shown as one of the ranking sequences. Unlike typical ranking or rating tasks, in SAMVIQ observers can revisit their scores for previously viewed sequences, allowing them to fine-tune their decisions based on later information. This tends to lead to more accurate assessments, reducing noise in the results.

3.1. Stimuli

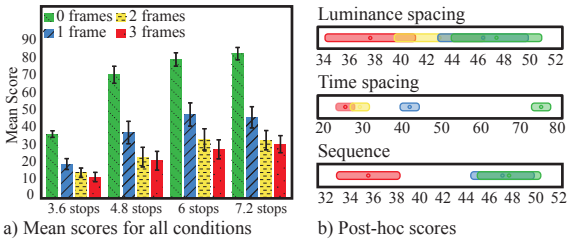
For our experiment, we processed three scenes taken from the Stuttgart HDR database [FGE*14] with the aim of creating new HDR versions of each, with varying amounts of ghosting or over-exposure. LDR frames were extracted from each sequence using different temporal and luminance spacings, which were then re-merged into HDR frames, mimicking standard multi-exposure techniques [RWD*10]. In all cases, three exposures were used per output frame. Either all three exposures were taken from the same frame (0 frames), or the shortest and the longest exposures were taken from the neighbouring frames with increasing spacing (1,2,3 frames). Additionally, four luminance spacings between the extreme exposures were used, set to 3.6, 4.8, 6.0 and 7.2 stops. The sequence created using 7.2 stops of luminance spacing and 0 frames of temporal spacing was selected as the reference. These parameters were selected so that ghosting and over-exposure artefacts affected approximately the same number of pixels at each level, as determined by a pilot study.

The LDR frame is obtained from the ground-truth HDR frame by scaling the latter accordingly with the exposure time. The scaled values are clipped to the range of values representable with 8 bits per channel image. The extracted exposures were then merged [DM97] using linear ramps between 0 and 25 and between 227 and 255 to reduce the weight of unreliable pixels. For each scene, 17 sequences were created.

3.2. Experimental Setup

Once all sequences for one scene were scored, following the SAMVIQ protocol, the observers could proceed to the next scene. The order of scenes and sequences was randomized.

The experiment was performed using a dual display setup.



a) Mean scores for all conditions b) Post-hoc scores

Figure 2: *Mean scores and errors for each luminance and time spacing condition, averaged over all scenes and participants are shown in (a). The results of post-hoc analysis are shown in (b).*

Comparison	F	p-value	Pearson ρ
Temporal	$F(3,764) = 284.5$	$p < 0.001$	-0.67
Luminance	$F(3,764) = 5.2$	$p < 0.005$	0.14
Scene	$F(2,765) = 17.2$	$p < 0.001$	0.18
Temp/Lum	$F(9,752) = 0.2$	$p = 0.989$	
Temp/Scene	$F(6,756) = 5.0$	$p < 0.001$	
Lum/Scene	$F(6,756) = 0.1$	$p = 0.997$	

Table 1: *ANOVA results for each of the main conditions and interactions between them. The last column shows the Pearson correlation of scores against each of the main variables.*

A laptop was placed on a low table in front of the observers, allowing them to select and score the different sequences, while a 42" HDR SIM2 monitor was used to show the stimuli. To achieve temporal consistency as well as to ensure comfortable display levels, each frame in each sequence was scaled by a constant of 500 to let the peak luminance on the SIM2 display not exceed 500 cd/m^2. Participants were placed approximately 1m away from the HDR monitor and were allowed to adjust their chair for optimal viewing. A total of 16 participants took part in this experiment (11 male, 5 female), all with normal or corrected to normal vision, with ages between 24 and 60 years ($\mu = 35.1, \sigma = 10.8$).

In total, each observer viewed 51 sequences (4 luminance spacings, 4 temporal spacings, 1 open reference, 3 scenes) and a score was recorded for each of them. Figure 2 (a) shows average scores for each of the luminance and temporal spacing combinations across all participants and scenes. Lower scores indicate a lower perceived quality.

3.3. Analysis

The obtained results were analysed for significance using ANOVA on each of the main variables (luminance spacing, temporal spacing, scene) as well as on pairwise interactions between them, outlined in Table 1. The Pearson correlation of scores against each of the main variables was also computed (last column of Table 1). Overall, all three conditions were shown to have a significant effect on the

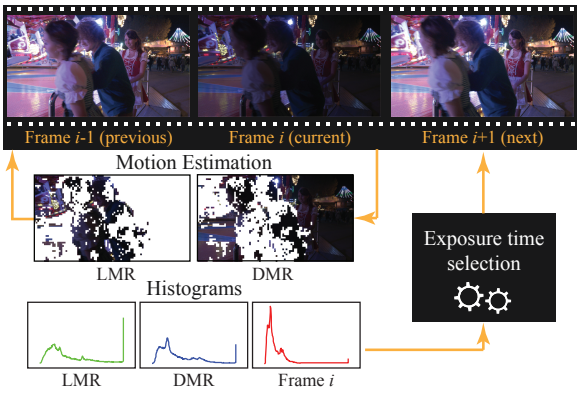


Figure 3: *A high-level flowchart of our metering algorithm.*

scores, albeit the ghosting artefacts were a more decisive factor in observers' scores, irrespective of the presence of over-exposure.

Post-hoc analysis was also performed using the Tukey-Kramer test with 95% confidence interval, shown in Figure 2 (b), further supporting the observation that ghosting is deemed a more severe artefact compared to over-exposure.

In HDR imaging, the goal is to extend the dynamic range encoded in the resulting images or videos. However, the above results suggest that this should not be done at the cost of ghosting artefacts, as their presence is likely to nullify the benefit of the extension in dynamic range. Our results complements the previous studies of Masia et al. [MAF*09], where the authors showed that perceived quality of HDR images depends more on the presence or absence of disturbing spatial artefacts than on the exact intensities. In their case the HDR images were obtained from LDR images using reverse tone-mapping operators, and the addressed artefacts are caused by these operators (noise, fringes, etc.). We consider the ghosting artefacts in videos and evaluate the presence of ghosting versus the increased number of details and reduced noise due to the high dynamic range.

Thus, to maximize the quality of our HDR content, we need to take into account the trade-off between ghosting and over-exposure. As such, to avoid ghosting, we detect and consider motion in the scene while capturing. Our metering algorithm is based on these observations and will be described in the following section.

4. Metering Algorithm

Our metering algorithm determines the optimal exposure time for every frame as a function of scene motion and its dynamic range. Based on the psychophysical experiment described in Section 3, we select exposures such that ghosting is minimized, while at the same time capturing the highest possible dynamic range.

To achieve that, our algorithm analyses previous frames to determine capture settings for the current frame. In particular, it gathers motion information from the two recent frames to divide the image into regions of local scene motion (local motion regions, or LMRs) and their counterparts, dominant motion regions (DMRs). The separation is performed to ensure that the number of well exposed pixels in DMRs in both frames is sufficient for dominant motion vector estimation. Its estimation precision influences the appearance of over-exposed regions after reconstruction. It is also important to ensure that the number of correctly exposed pixels in the LMRs is sufficient for its correct reconstruction.

The normalized histograms of each of these areas, as well as that of the full frame are then computed (normalized meaning that the sum of the values in the histogram bins is one). By examining the extreme bins of these three histograms (full, LMR, DMR) we can assess how well we have captured the dynamic range of the full scene and of the motion regions separately, and therefore decide how the next frame should be exposed. Additionally, as a by-product of the exposure time selection, our algorithm decides in advance whether the exposure for the following frame should be increased or decreased, initializing the next metering cycle. These steps are outlined in Figure 3 and will be described in detail in the following sections.

4.1. Initialization

To seed the metering algorithm, and to be able to derive the first motion vectors, we initialize the video capture by capturing two frames using the standard metering system provided by the camera or mobile device. In addition we will keep track of whether the next exposure will be longer or shorter, allowing us to successively step up and down through the luminance range. We set the initial direction to 'shorter' as over-exposed regions carry no useful information for correspondence computation in the HDR video reconstruction phase.

4.2. Motion Estimation

To estimate the motion between frames, we use the hierarchical diamond search (HDS) algorithm [UNR09], a real-time block matching algorithm, which performs a recursive search over a multi-level image decomposition with a diamond shape window around the current position. The HDS algorithm returns one motion vector for each block. Since subsequent frames are differently exposed, we adjust pixel intensities in the frame with the shorter exposure to the longer exposed frame.

As a distance measure between blocks we use an absolute difference between mean gamma corrected pixel intensities of the blocks, rather than the standard sum of absolute differences, as it is faster to compute and better in compensating for the differences between frame exposures.

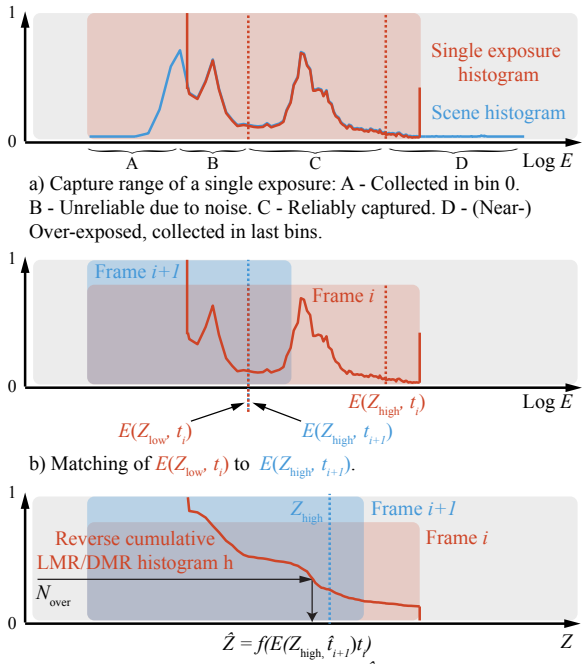


Figure 4: *Rationale for our initial exposure selection strategy (a - b), and selection adjustment step (c).*

The dominant motion vector for a frame is estimated with a block-based recursive weighted least-squares method after the motion field is computed [UNR09]. An affine model is used in which the dominant motion vector consists of 6 components (translation (2) + rotation (4)). Homogeneous blocks (particularly overexposed areas) are excluded from the computation of dominant motion vectors. A block is marked as homogeneous if the vertical or the horizontal gradient is below a set threshold (80 in our case). Blocks which do not follow the dominant motion are marked as LMRs.

4.3. Exposure Time Selection

The goal of this step is to find an exposure time t_{i+1} for the next frame $i+1$, for which we use the histograms computed for the current frame i, and so fulfil two conditions:

C1. If the pixels are over-exposed (under-exposed) in the current frame and we decrease (increase) exposure time, they should not be under-exposed (over-exposed) in the next frame. This means that these pixels will be properly captured at least in one of two subsequent frames.

C2. The next frame should contain enough meaningful information in DMRs and LMRs for dominant and local motion estimation. This means that a certain percentage of pixels should be properly exposed in both frames.

Effectively, this means that the dynamic range captured by two subsequent frames is extended as much as possible, while ensuring that sufficient correspondences will be found by the optical flow algorithm in the HDR reconstruction phase. To that end, we first calculate the largest exposure time increment (or decrement) possible, and then refine this initial estimate based on additional criteria.

Initial Exposure Time Estimate. Exposures with increasing exposure times capture increasing ranges of scene irradiance, albeit that all start at $0\,W/m^2$. Nonetheless, a good portion of this data is lost due to under-exposure, noise and over-exposure (Figure 4 (a)). We therefore define two thresholds $Z_{\text{low}} = 0.2 \cdot 255 = 51$ and $Z_{\text{high}} = 0.9 \cdot 255 \approx 230$, which approximate the lower and upper pixel value boundaries of the well-exposed region of the current frame's histogram (Figure 4 (a and b)). These boundaries are also matched to similar values used in the HDR video reconstruction algorithm. Pixels with values higher than Z_{high} will be reconstructed from frames with shorter exposure times, and pixels with values lower than Z_{low} will be obtained from frames with longer exposure times.

Here, we describe the search process to establish an exposure time t_{i+1} that will be longer than exposure time t_i. The search for a shorter exposure time is analogous.

The estimate of irradiance E, corresponding to pixel value Z in the frame with exposure time t, can be calculated as:

$$E(Z,t) = f^{-1}(Z)/t. \tag{1}$$

where $f^{-1}()$ is the inverse camera response curve (which can be computed using standard techniques [DM97]). Then, the following is required to fulfil condition C1:

$$E(Z_{\text{high}}, t_{i+1}) \geq E(Z_{\text{low}}, t_i) \tag{2}$$

From Equations (1) and (2) we calculate an upper bound $\bar{t}_{i+1}$ on the exposure time of the next frame:

$$t_i < t_{i+1} \leq \bar{t}_{i+1} = f^{-1}(Z_{\text{high}})t_i / f^{-1}(Z_{\text{low}}). \tag{3}$$

We initialize the estimate $\hat{t}_{i+1}$ of the optimal exposure time t_{i+1} with its upper bound: $\hat{t}_{i+1} = \bar{t}_{i+1}$.

Exposure Time Refinements. To prevent ghosting in subsequent reconstruction, we may have to adjust the current estimate $\hat{t}_{i+1}$ Here we describe how we ensure that the condition C2 is fulfilled.

From the histogram of the current frame we estimate $N_{\text{LMR,over}}(\hat{t}_{i+1})$ and $N_{\text{DMR,over}}(\hat{t}_{i+1})$ which are the expected percentages of over-exposed pixels in LMRs and DMRs of the next frame, as function of the new exposure time $\hat{t}_{i+1}$. We require these values to be below the predefined thresholds $\rho_{\text{LMR,over}}$ and $\rho_{\text{DMR,over}}$ (which are computed as described in Section 4.6):

$$\begin{aligned} N_{\text{LMR,over}}(\hat{t}_{i+1}) &\leq \rho_{\text{LMR,over}}, \\ N_{\text{DMR,over}}(\hat{t}_{i+1}) &\leq \rho_{\text{DMR,over}}. \end{aligned} \tag{4}$$

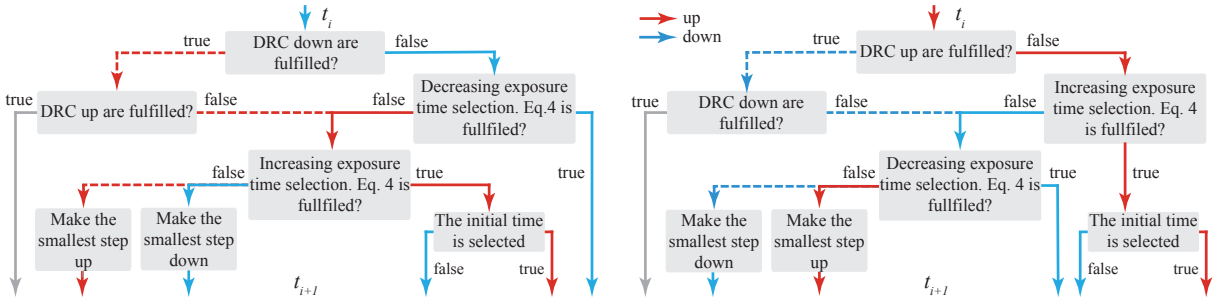


Figure 5: *The decision tree for selection of the next exposure time dependent on the initial direction: decreasing (left) or increasing (right). A by-product of exposure time selection for the next frame is the initial exposure direction for the frame with index $i+2$, and is given by the color of the output arrow. The grey arrows denote no changes in the exposure time and direction. Dashed lines denote the branch which corresponds to the case when the DRC (dynamic range conditions, Section 4.5) in the initial direction is fulfilled.*

The estimate $\hat{t}_{i+1}$ can take values between $\bar{t}_{i+1}$ and t_i. As such, starting at $\hat{t}_{i+1}$, the exposure time is estimated by testing all possible values within this range until Eq. 4 is fulfilled. If t_i is reached and Eq. 4 is not fulfilled, it means that the step in a given direction can not be performed.

The expected number of over-exposed pixels in the next frame equals the number of pixels with irradiance values higher than $E(Z_{\text{high}}, \hat{t}_{i+1})$. This corresponds in the current frame to the number of pixels with pixel values higher than $\hat{Z} = f(E(Z_{\text{high}}, \hat{t}_{i+1})t_i)$. Given the histograms of LMRs and DMRs, and the value of $\hat{Z}$, we can immediately get $N_{\text{LMR,over}}(\hat{t}_{i+1})$ and $N_{\text{DMR,over}}(\hat{t}_{i+1})$ by computing the backward cumulative histograms h_{LMR} and h_{DMR} (or standard cumulative histograms in the case of exposure decrease) with the number of bins equal to the frame bit depth. The number of over-exposed pixels is then equal to the histogram value at the bin with index $\hat{Z}$ (Figure 4 (c)): $N_{\text{LMR,over}}(\hat{t}_{i+1}) = h_{\text{LMR}}(\hat{Z})$ and $N_{\text{DMR,over}}(\hat{t}_{i+1}) = h_{\text{DMR}}(\hat{Z})$.

In many non-professional (particularly mobile) cameras, exposure settings can only be adjusted using a given exposure compensation step s. Based on that, the number N of the possible settings of $\hat{t}_{i+1}$ is predefined for a particular camera. The value of N can be found from $\overline{V}_{i+1} = V_i + Ns$, where $V_i = \log_2(t_i/t_0)$ is an exposure value, and t_0 is the time of the frame with zero exposure value (optimal exposure time for a scene). Then from Eq. 3 we derive:

$$\log_2\left(\frac{t_i}{t_0}\frac{f^{-1}(Z_{\text{high}})}{f^{-1}(Z_{\text{low}})}\right) = \log_2\left(\frac{t_i}{t_0}\right) + Ns. \qquad (5)$$

Note, that $\log_2(t_i/t_0)$ on the left and right hand side cancel out, and N depends only on the camera response curve, exposure compensation step and the values of Z_{low} and Z_{high}. The bin numbers $\hat{Z}_j$, $j \in [0,N]$ that correspond to each of these settings can be precomputed:

$$\hat{Z}_j = Z_{\text{low}} 2^{(N-j)s}. \qquad (6)$$

This simplifies the adjustment step to cumulative histogram computation, and in the worst case requires $2N$ comparisons.

4.4. Exposure Time Direction

To compute the exposure time for the next frame we first need to determine if we want to decrease or increase exposure time. This can be done by analysing the data from the current frame or it can be derived from the exposure selection step for frame $i-1$. Here we use the second approach, as it does not require any additional analysis.

Before executing the exposure time selection step for frame i as described previously, we check whether the full scene irradiance range is already captured in the specified direction by evaluating the last or first bin of the full frame histogram. Additionally, we check whether the camera settings allow us to make a step in that direction. We refer to these two conditions as *dynamic range conditions (DRC)* (Figure 5).

In the simplest case, if the scene's full dynamic range is already captured in both directions we do not change the exposure time and keep the initial direction. This would occur for scenes that can be captured with a single exposure.

Either the DRC or failure to fulfill the conditions in Eq. 4 may prevent an exposure time change in the current direction. In that case, the direction is reversed. If any condition in Eq. 4 prevents exposure adjustment in both directions, then the exposure time is changed in the current direction by the smallest step allowed by the camera. This helps to expand the captured dynamic range. Specifically, the smallest step in the initial direction is chosen if the DRC in that direction is not fulfilled, and in the opposite direction, otherwise.

Finally, if the current direction indicates an exposure time increase, we check if an adjustment from the initial time estimation is required, as explained in the preceding section. If that is the case, then for the next frame, the direction will

$\rho^{A_{high}}_{LMR,over}$	5%	$\rho^{A_{low}}_{LMR,over}$	90%	$\rho_{DMR,over}$	20%
$\rho^{A_{high}}_{LMR,under}$	85%	$\rho^{A_{low}}_{LMR,under}$	100%	$\rho_{DMR,under}$	90%

Table 2: *The threshold percentage ρ_{LMR} and ρ_{DMR} of under- and over-exposed pixels in LMRs and DMRs. Note that for LMRs different thresholds are used as a function of relative size A of these regions with respect to the whole frame.*

be reversed to exposure decrease to prevent too large over-exposed areas in LMRs with no data. For an exposure time decrease this step can be omitted. This asymmetry in operation is due to the fact that any exposure length will result in capture of information in the darks (albeit possibly with noise), whereas bright regions may get clipped, dependent on exposure duration. The full decision process for the direction of the next exposure change is illustrated in Figure 5.

4.5. Thresholds $\rho_{LMR,over}$ and $\rho_{DMR,over}$

We formulate condition C2 (Section 4.4) in terms of the threshold percentages ρ_{DMR} and ρ_{LMR} of under- and over-exposed pixels in LMRs and DMRs. Additionally, the thresholds for LMRs depend on the percentage A of pixels in LMRs with respect to all pixels in the frame. The rationale behind such threshold choice is that the exposure selection should not be optimized for tiny moving regions, where ghosting artefacts would be less disturbing than large clipped image areas. Note that the visibility of ghosting typically depends on the contrast of moving objects with respect to the background [DXCZ14, SBE*15]. This is implicitly covered by our algorithm. Whenever a dark (bright) object moves over a bright (dark) background, both the background and the object are included into the LMR. Since our algorithm gives higher priority for over-exposed regions, the ghosting artefacts in these regions will be minimized.

The threshold values used in our implementation are summarized in Table 2 and were chosen to minimize ghosting. This is consistent with the findings of our experiment in Section 3. Here, we compute the size A of the LMRs relative to the image size, and express this as a percentage. If this percentage is smaller than $A_{low} = 0.01\%$, then the LMRs are considered insignificant and therefore the algorithm will give priority to maximize the capture of the scene's dynamic range. Similarly, this percentage is greater than $A_{high} = 1\%$, then ghosting is likely to be noticeable and therefore priority is given to avoid it. Effectively, a higher priority of motion leads to a lower percentage of allowed under- and over-exposed pixels in LMRs. For LMRs with a relative area inbetween A_{low} and A_{high} the thresholds are determined through linear interpolation of the threshold values in Table 2:

$$\rho_{LMR} = (\rho^{A_{high}}_{LMR} - \rho^{A_{low}}_{LMR}) \frac{A - A_{low}}{A_{high} - A_{low}} + \rho^{A_{low}}_{LMR}. \quad (7)$$

4.6. Discussion

Our algorithm could either adjust the exposure time or manipulate the ISO value, which could be used, for instance, to reduce blur [NB12]. In that case the ISO and exposure time are adjusted based on the amount of motion. Thus, when strong motions happen in dark areas of a scene, instead of increasing exposure time the ISO value could be increased. Note that if motion occurs in mid to high luminance areas our algorithm will refrain from selecting long exposure times to avoid overexposure of moving objects.

4.7. Implementation

The algorithm was implemented in C++. To achieve real-time performance, the typically HD input resolution is downsampled to 40%. In the current implementation, motion estimation takes from 12–25 ms, exposure adjustment 2–3 ms, histogram computation 1 ms, and exposure time estimation less than 1 ms. All measurements have been performed on an Intel Xeon CPU E5-1607 processor.

Our algorithm incorporates a real-time block-based motion estimation solution as well as efficient histogram analysis. It is therefore well suited for mobile capture applications where resources are limited. Android 5 (Lollipop) allows capture settings to be adjusted per frame, whereas iOS 8 exposes manual camera controls through its API, which allows exposure times to be set for each frame, with probable delay of several frames from the moment the request about exposure time change is sent. Since the content of the scene typically changes smoothly, this delay should not have the strong affect on the performance. Our algorithm can readily be ported to these environments.

5. HDR Video Reconstruction

Our real-time metering algorithm allows the production of a sequence of exposures with exposure times that step up and down by some number of frames determined by an analysis of each frame. Moreover, the exposures do not have a regular spacing in exposure time, even if they are captured with a constant frame-rate. Current HDR video reconstruction algorithms on the other hand expect an evenly spaced sequence of exposures that only step in one direction every N frames. As such, they cannot cope with our configuration of exposures, and therefore there is a need to develop a new algorithm that is matched to this input.

We find, however, that we can repurpose an existing state-of-the-art algorithm, which has shown to perform well on evenly spaced exposure sequences. In particular, we adapt Kalantari's algorithm [KSB*13], which has the strong advantage that it is able to reconstruct an HDR video without loss of frame-rate relative to the input exposure sequence.

Algorithm. We refer to an input LDR frame as the *reference frame*. In its simplest form, one could view HDR reconstruction as an iterative two-step process. First, for a given time-step, we synthesize a set of *exposures* such that the reference frame becomes one exposure of an exposure stack. In the second step the exposure stack can then be merged into an HDR frame, using the previous estimate of the HDR frame to synthesize better exposures, which are merged into an improved HDR frame. This process is iteratively refined.

During exposures synthesis, firstly, all the *missing exposures* (the exposures for the frame, excluding the reference frame itself) are extracted from the previously reconstructed HDR frame. In the very first iteration, the missing exposures are created by adjusting the exposure of the reference frame. Then, the missing exposures are refined from the exposures in the exposures stacks of adjacent frames, using a search and vote procedure ([KSB*13, SKY*12, SCSI08]). The exposures used for the refinement are called *source frames*.

Exposure Arrangement. In Kalantari et al. [KSB*13] reference frames are captured using a repeating sequence of monotonically increasing exposure times. This means that every N frames, the exposure time resets, causing a large jump in exposure time. For those frames, optical flow may be computed less reliably. In contrast, we allow exposure times to oscillate, therefore ensuring that two consecutive reference frames will always have a sufficient number of overlapping pixel values to compute correct optical flow.

Exposure Times Stack Selection. Before the iterative process, the exposure times for the exposure stack of each frame should be determined. The exposure times of frames within the sequence to which the frame belongs to and two neighbouring sequences are immediately added to the exposure times stack. Additional exposure times are obtained by *recursively searching* separately for the shorter and longer exposure times of reference frames to the right and to the left of the current frame (direction to the right denotes the direction to the first reference frame and direction to the left corresponds to the direction to the last reference frame). The found exposure times are merged into a single stack. The source frames for each exposure are one or both adjacent frames, to the left or to the right. The directions to the source frames coincide with the directions to the reference frame at given exposure level.

Recursive Search. Here we describe how the recursive search for the shorter exposure times from frames to the right of the current frame i is performed. Around the current frame there exists a range of frames $K = [i, r]$ that have sufficient spatial overlap with the current frame. The search for the frame f^* with the shortest exposure time in this range, which is shorter than the exposure time of i, is performed. We store the found exposure time and set $r \leftarrow f^*$. The process is recursively repeated until either K is empty or found

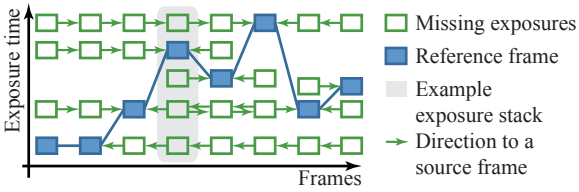


Figure 6: *Example reference frames arrangement with their exposure stacks.*

exposure time has sufficient overlap with exposure time of i (as defined in Section 4.4 C1). This process results in the exposure arrangement similar to Figure 6.

In the base version of our algorithm the range K is limited to the sequence to which the frame belongs to and two neighbouring sequences. This strategy is sufficient for most scenes, but might cause large bright areas to over-expose if there is motion in dark parts of the scene concurrently.

To obtain the temporal coherence with the base version all the frames could be clipped to the dynamic range of the frame with the smallest dynamic range .

Merging Process. In our exposure arrangement each exposure could have only one source frame, the right or the left one. At the same time, the same pixel could be captured properly in several exposures, and reconstructed better or worse from different source frames, as a result of the search and vote process. To account for this we modify the exposure merging scheme of Debevec and Malik [DM97].

After the search and vote process, for the current frame i, we have a set of complementary exposures $I_{n,i}$ with exposure times t_n, $n \in [1, N]$. At the first iteration, for each pixel p_n in each exposure $I_{n,i}$ we calculate a weight $\alpha_{p,n}$ that represents its level of unreliability. Setting these weights follows Kalantari's scheme [KSB*13], with $\alpha_{p,n} = 1$ for well-exposed pixels, a linearly decreasing ramp for pixel values above Z_{high}, and a non-linearly increasing ramp for pixels below Z_{low}. The non-linearity is adjusted based on the amount of motion of pixel p_n. Additionally, we have available the probability of a correct match R_n that stems from the patch-based search and varies at each iteration.

As discussed above, some exposures may have the same exposure time, with contributions coming from previous and next frames. We replace such exposures $I_{n_1,i}$ and $I_{n_2,i}$ on the exposure stack with a new exposure $I_{n',i}$ which is a weighted average of both input exposures:

$$I_{n',i} = \frac{\sum_{n \subset [n_1,n_2]} (1 - \alpha_{p,n}) R_n I_{n,i}}{\sum_{n \subset [n_1,n_2]} (1 - \alpha_{p,n}) R_n}. \quad (8)$$

The new weight for this exposure is $\alpha_{p,n'} = \min(\alpha_{p,n_1}, \alpha_{p,n_2})$. The resulting N' exposures are merged with a standard technique [DM97], albeit that we add our

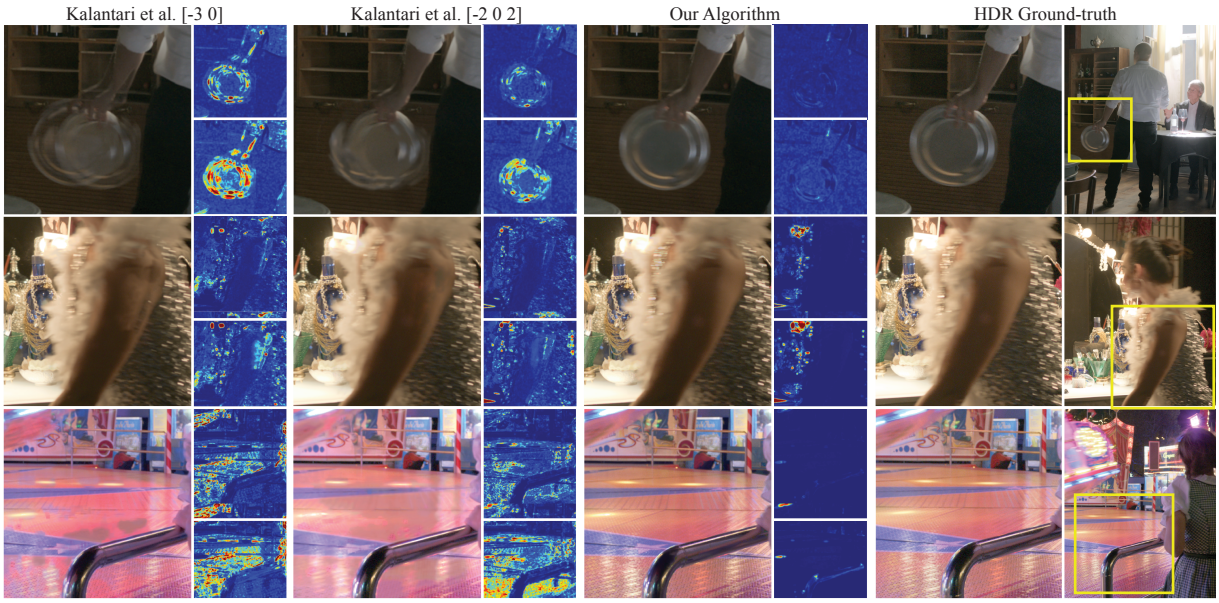


Figure 7: *Results comparing our method with Kalantari's (2 and 3 exposure variants), showing a visual comparison, as well as the SSIM metric applied to each result image (relative to the ground truth HDR). For each pair of false-color SSIM images, the top shows the structural component of SSIM only, whereas the bottom image shows the full metric.*

pixel reliability measure into the weighting scheme:

$$I_{\text{hdr},i} = \frac{\sum_{n\in N'}(1-\alpha_{p,n})\,\Lambda(I_{n,i})\,I_{n,i}/t_n}{\sum_{n\in N'}(1-\alpha_{p,n})\,\Lambda(I_{n,i})}. \qquad (9)$$

Here, $\Lambda()$ is a hat function that reduces the contribution for high and low pixel values. For frame i this constitutes the HDR output of each iteration, and after sufficient iterations have been performed, the final HDR output.

6. Results

For evaluation we apply our metering algorithm to HDR sequences from a publicly available database [FGE*14] and compare the reconstructed result with ground-truth HDR video. The exposures are extracted from the HDR frames as discussed in Section 3.1, where additionally gamma correction is applied before clipping. To compare the results, we used the structural similarity (SSIM) index [WBSS04], applied between any of the reconstructed HDR videos and the input HDR video. SSIM was chosen to assess structural differences between frames.

Before comparing against the current state-of-the-art, we first show the effect of some of our design decisions. Our merging approach has the option of adding extra exposures to increase temporal stability (as discussed in Section 5: Recursive Search). We computed the dynamic range for each frame in a reconstructed sequence as $\log(L_{\max}) - \log(L_{\min})$. This was done with the base and full versions of our algorithm for two sequences. The dynamic range in both cases is 11 stops, but the variance is more than 7.4 times smaller with the full version (0.126 against 0.017).

We compare our approach with Kalantari's [KSB*13] and following their evaluation, we experimented with two exposure values at -3 and 0 stops and with three exposures at -2, 0 and 2 stops. Here, zero indicates the optimal exposure value for each scene. Figure 7 (as well as the supplemental video and PDF) shows a collection of close-up frames for which the SSIM metric was computed. We also show the structural component of SSIM alone, where the differences in brightness and contrast are ignored. In all cases, our algorithm produces a closer match to the ground-truth. In most cases the differences with the ground-truth are structural in nature. To quantify these results, we calculated mean SSIM values for 334 frames drawn from 7 sequences for each method (Kalantari 2 and 3 exposures, and ours) and 2 SSIM scores (full metric, structural component only). These scores were subjected to a one-way ANOVA, showing highly significant differences between the three algorithms ($F(2,1001) = 25.54, p < 0.0001$ for the full SSIM metric and $F(2,1001) = 22.44, p < 0.0001$ for the structural part). The lower mean scores in Figure 8 indicate that the differences are due to a better performance for our method. Measured per scene, we found our approach to be statistically significantly better for 5 of the 7 scenes, while no significant difference was found for two scenes.

The supplemental materials confirm the main trend: our me-

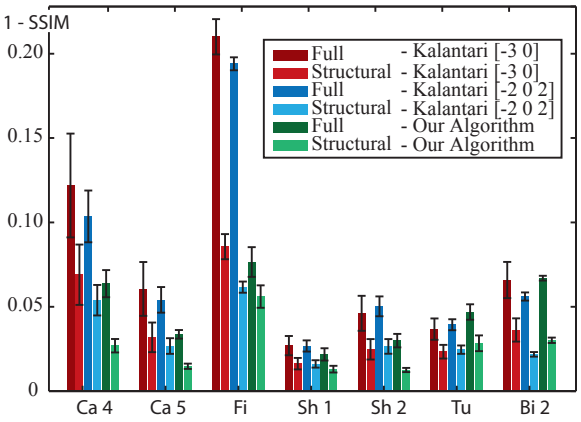


Figure 8: *Mean SSIM score (full metric and structural component alone) for each of 7 videos, comparing two and three-exposure reconstruction using Kalantari's method with our algorithm. Also shown are the 95% confidence intervals.*

tering algorithm adapts well to scenes with varying characteristics, showing a match to the ground-truth data that is consistently no worse than the current state-of-the-art, and in nearly all cases produces visually superior video.

7. Conclusions

A huge mobile phone market means that an enormous number of images and videos originate from mobile devices. Our research is oriented toward improving their visual quality by capturing a higher dynamic range than such cameras are natively capable off, while minimizing motion artefacts. To this end, we combine a novel real-time metering approach with off-line HDR video reconstruction.

Our hypothesis, backed by a psychophysical experiment, is that failing to capture the full dynamic range of a scene is a lesser evil than introducing motion artefacts that are seen as ghosting. The proposed metering algorithm is therefore designed to seamlessly adapt to the scene that is being captured: if sufficient motion warrants it, moving objects are captured with high fidelity; otherwise the scene's dynamic range is captured with high fidelity. This allows the algorithm to work within the potential hardware constraints imposed by mobile devices (sensor dynamic range, frame-rate and computing constraints). The resulting stream of exposures is then combined into a full HDR sequence in an off-line process that is well matched to our metering algorithm.

Analysis of our results shows that our capture strategy yields HDR video that is visually superior to the current state-of-the-art for all sequences that we have available. This is not to say that our results can never be improved upon. In certain rare cases, we are not able to avoid ghosting completely. For instance, in a high dynamic range scene with significant motion in both the bright and the dark parts, our algorithm will be able to lock onto one of the motion areas, while ignoring the other. This means that one of the two motion areas will be captured well, while the other may exhibit ghosting. We note, however, that competing techniques could produce ghosting in both motion areas (Figure 9).

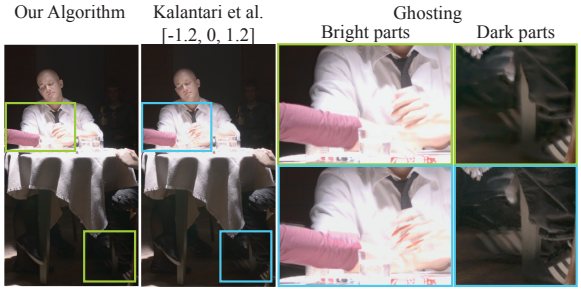


Figure 9: *Motion in both the bright and the dark parts.*

References

[BHD08] BARAKAT N., HONE A. N., DARCIE T. E.: Minimal-bracketing sets for high-dynamic-range image capture. *IEEE Trans. Image Process. 17*, 10 (2008), 1864–1875. 3

[BLL*10] BERGER K., LIPSKI C., LINZ C., SELLENT A., MAGNOR M.: A ghosting artifact detector for interpolated image quality assessment. In *Proc. 14th IEEE International Symposium on Consumer Electronics* (2010), pp. 1–6. 3

[BSFG09] BARNES C., SHECHTMAN E., FINKELSTEIN A., GOLDMAN D. B.: PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph. 28*, 3 (2009), 24. 3

[CKL14] CHO H., KIM S. J., LEE S.: Single-shot high dynamic range imaging using coded electronic shutter. *Comp. Graph. Forum 33*, 7 (2014), 329–338. 2

[DM97] DEBEVEC P. E., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH* (1997), pp. 369–378. 3, 4, 7, 10

[DXCZ14] DALY S., XU N., CRENSHAW J., ZUNJARRAO V. J.: A psychophysical study exploring judder using fundamental signals and complex imagery. *SMPTE Conference Proceedings 2014*, 10 (2014), 1–14. 3, 8

[FGE*14] FROEHLICH J., GRANDINETTI S., EBERHARDT B., WALTER S., SCHILLING A., BRENDEL H.: Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays. In *Proc. SPIE* (2014), vol. 9023. 1, 4, 10, 11

[GAW*10] GRANADOS M., AJDIN B., WAND M., THEOBALT C., SEIDEL H.-P., LENSCH H. P. A.: Optimal HDR reconstruction with linear digital cameras. In *Proc. IEEE CVPR* (2010), pp. 215–222. 3

[GHKE13] GUTHIER B., HO K., KOPF S., EFFELSBERG W.: Determining exposure values from HDR histograms for smartphone photography. In *Proc. ACM Int. Conf. Multimed.* (2013), pp. 425–426. 3

[GHMN10] GU J., HITOMI Y., MITSUNAGA T., NAYAR S.: Coded rolling shutter photography: Flexible space-time sampling. In *Proc. IEEE ICCP* (2010), pp. 1–8. 2

[GIN13] GUPTA M., ISO D., NAYAR S. K.: Fibonacci exposure bracketing for high dynamic range imaging. In *Proc. IEEE ICCV* (December 2013). 3

[GKE13] GUTHIER B., KOPF S., EFFELSBERG W.: Algorithms for a real-time HDR video system. *Pattern. Recogn. Lett. 34*, 1 (2013), 25–33. 2

[GN03] GROSSBERG M. D., NAYAR S. K.: High dynamic range from multiple images: Which exposures to combine? In *Proc. ICCV Workshop on CPMCV* (2003). 3

[GTM*12] GALLO O., TICO M., MANDUCHI R., GELFAND N., PULLI K.: Metering for exposure stacks. *Comp. Graph. Forum 31*, 2 (2012), 479–488. 2, 3

[HDF10] HASINOFF S. W., DURAND F., FREEMAN W. T.: Noise-optimal capture of high dynamic range photography. In *Proc. IEEE CVPR* (2010), pp. 553–560. 3

[HGPS13] HU J., GALLO O., PULLI K., SUN X.: HDR deghosting: How to deal with saturation? In *Proc. IEEE CVPR* (2013), pp. 1163–1170. 3

[Hir10] HIRAKAWA K.: Optimal exposure control for high dynamic range imaging. In *Proc. IEEE ICIP* (2010), pp. 3137–3140. 3

[HST*14] HEIDE F., STEINBERGER M., TSAI Y.-T., ROUF M., PAJAK D., REDDY D., GALLO O., LIU J., HEIDRICH W., EGIAZARIAN K., KAUTZ J., PULLI K.: FlexISP: A flexible camera image processing framework. *ACM Trans. Graph. 33*, 6 (2014). 2

[IM10] ILSTRUP D., MANDUCHI R.: One-shot optimal exposure control. In *Proc. ECCV* (2010), pp. 200–213. 3

[ITU12] ITU-R: *Recommendation ITU-R BT.500-13: Methodology for the subjective assessment of the quality of television pictures*. Tech. rep., ITU-R, 2012. 4

[KGBU13] KRONANDER J., GUSTAVSON S., BONNET G., UNGER J.: Unified HDR reconstruction from raw CFA data. In *Proc. IEEE ICCP* (2013), pp. 1–9. 1

[KSB*13] KALANTARI N. K., SHECHTMAN E., BARNES C., DARABI S., GOLDMAN D. B., SEN P.: Patch-based high dynamic range video. *ACM Trans. Graph. 32*, 6 (2013), 202:1–202:8. 2, 3, 9, 10, 11

[KUWS03] KANG S. B., UYTTENDAELE M., WINDER S., SZELISKI R.: High dynamic range video. *ACM Trans. Graph.* (2003), 319–325. 2, 3

[MAF*09] MASIA B., AGUSTIN S., FLEMING R. W., SORKINE O., GUTIERREZ D.: Evaluation of reverse tone mapping through varying exposure conditions. *ACM Trans. Graph. (Proc. SIGGRAPH Asia) 28*, 5 (Dec. 2009), 160:1–160:8. 5

[MG10] MANGIAT S., GIBSON J.: High dynamic range video with ghost removal. *Proc. SPIE 7798* (2010). 3

[MKR07] MERTENS T., KAUTZ J., REETH F. V.: Exposure fusion. *Comp. Graph. Forum (Proc. Pacific Graphics)* (2007), 382–390. 1

[MN99] MITSUNAGA T., NAYAR S. K.: Radiometric self calibration. In *Proc. IEEE CVPR* (1999), pp. 374–380. 3

[MPMP95] MANN, PICARD, MANN S., PICARD R. W.: On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proc. IS&T* (1995), pp. 442–448. 3

[MRK*13] MANAKOV A., RESTREPO J. F., KLEHM O., HEGEDÜS R., EISEMANN E., SEIDEL H.-P., IHRKE I.: A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging. *ACM Trans. Graph. 32*, 4 (2013), 47:1–47:14. 1

[NB03] NAYAR S. K., BRANZOI V.: Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *Proc. IEEE ICCV* (2003), pp. 1168–1175. 1

[NB12] NARASIMHA R., BATUR A. U.: Method and apparatus for producing sharp frames with less blur, June 5 2012. US Patent 8,194,141. 9

[RBS03] ROBERTSON M. A., BORMAN S., STEVENSON R. L.: Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *J Electronic Imaging 12*, 2 (2003), 219–228. 3

[RWD*10] REINHARD E., WARD G., DEBEVEC P., PATTANAIK S., HEIDRICH W., MYSZKOWSKI K.: *High Dynamic Range Imaging*. Morgan Kaufmann Publishers, 2nd edition, 2010. 2, 4

[SBE*15] STENGEL M., BAUSZAT P., EISEMANN M., EISEMANN E., MAGNOR M.: Temporal video filtering and exposure control for perceptual motion blur. *IEEE Trans. Vis. Comput. Graphics 21*, 3 (2015), 1–1. 3, 8

[SCSI08] SIMAKOV D., CASPI Y., SHECHTMAN E., IRANI M.: Summarizing visual data using bidirectional similarity. In *Proc. IEEE CVPR* (2008). 3, 9

[SKY*12] SEN P., KALANTARI N. K., YAESOUBI M., DARABI S., GOLDMAN D. B., SHECHTMAN E.: Robust patch-based HDR reconstruction of dynamic scenes. *ACM Trans. Graph. 31*, 6 (2012), 203:1–203:11. 2, 3, 9

[SS12] SRIKANTHA A., SIDIBÉ D.: Ghost detection and removal for high dynamic range images: Recent advances. *Signal Processing: Image Communication 27*, 6 (2012), 650–662. 2, 3

[TAEE15] TURSUN O. T., AKYÃIJZ A. O., ERDEM A., ERDEM E.: The state of the art in hdr deghosting: A survey and evaluation. *Comp. Graph. Forum* (2015). 3

[TKTS11] TOCCI M. D., KISER C., TOCCI N., SEN P.: A versatile HDR video production system. *ACM Trans. Graph. 30*, 4 (2011), 41:1–41:10. 1

[UNR09] URBAN F., NEZAN J.-F., RAULET M.: HDS, a real-time multi-DSP motion estimator for MPEG-4 H.264 AVC high definition video encoding. *Journal of Real-Time Image Processing 4*, 1 (2009), 23–31. 6

[VCL*11] VANGORP P., CHAURASIA G., LAFFONT P.-Y., FLEMING R., DRETTAKIS G.: Perception of visual artifacts in image-based rendering of façades. *Comp. Graph. Forum 30*, 4 (2011), 1241–1250. 3

[Wat13] WATSON A. B.: High frame rates and human vision: A view through the window of visibility. *SMPTE Mot. Imag. J 122*, 2 (2013), 18–32. 3

[WBSS04] WANG Z., BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process. 13*, 4 (2004), 600–612. 10, 11

[WLA*12] WAN G., LI X., AGRANOV G., LEVOY M., HOROWITZ M.: CMOS image sensors with multi-bucket pixels for computational photography. *IEEE J. Solid-State Circuits 47*, 4 (2012), 1031–1042. 1

[Woo12] WOODS A. J.: Crosstalk in stereoscopic displays: A review. *J Electronic Imaging 21*, 4 (2012), 040902. 3

[WS03] WILCOX L. M., STEWART J. A. D.: Determinants of perceived image quality: ghosting vs. brightness. *Proc. SPIE 5006* (2003), 263–268. 3

[ZMD*06] ZWICKER M., MATUSIK W., DURAND F., PFISTER H., FORLINES C.: Antialiasing for automultiscopic 3D displays. In *Proc. EGSR* (2006), pp. 73–82. 3