Evaluation of Tone Mapping Operators for HDR-Video

Gabriel Eilertsen[1], Robert Wanat[2], Rafał K. Mantiuk[2], and Jonas Unger[1]

[1]Linköping University, Sweden   [2] Bangor University, United Kingdom

Abstract
Eleven tone-mapping operators intended for video processing are analyzed and evaluated with camera-captured and computer-generated high-dynamic-range content. After optimizing the parameters of the operators in a formal experiment, we inspect and rate the artifacts (flickering, ghosting, temporal color consistency) and color rendition problems (brightness, contrast and color saturation) they produce. This allows us to identify major problems and challenges that video tone-mapping needs to address. Then, we compare the tone-mapping results in a pair-wise comparison experiment to identify the operators that, on average, can be expected to perform better than the others and to assess the magnitude of differences between the best performing operators.

Categories and Subject Descriptors (according to ACM CCS): I.3.0 [Computer Graphics]: General—

1. Introduction

One of the main challenges in high-dynamic-range (HDR) imaging and video is mapping the dynamic range of the HDR image to the much lower dynamic range of a display device. While an HDR image captured in a high contrast real-life scene often exhibits a dynamic range in the order of 5 to 10 log10 units, a conventional display system is limited to a dynamic range in the order of 2 to 4 log10 units. Most display systems are also limited to quantized 8-bit input. The mapping of pixel values from an HDR image or video sequence to the display system is called *tone mapping*, and is carried out using a *tone mapping operator* (TMO).

Over the last two decades an extensive body of research has been focused around the problem of tone mapping. A number of approaches have been proposed with goals ranging from producing the most faithful to the most artistic representation of real world intensity ranges and colors on display systems with limited dynamic range [RHP*10]. In spite of that, only a handful of the presented operators can process video sequences. This lack of HDR-video TMOs can be associated with the (very) limited availability of high quality HDR-video footage. However, recent developments in HDR video capture, e.g. [UG07, TKTS11, KGBU13], open up possibilities for advancing techniques in the area.

Extending tone mapping from static HDR images to video sequences poses new challenges as it is necessary to take into account the temporal domain as well. In this paper, we set out to identify the problems that need to be solved in order to enable the development of next generation TMOs capable of

robust processing of HDR-video.

The main contribution of this paper is the systematic evaluation of TMOs designed for HDR-video. The evaluation consists of three parts: a *survey* of the field to identify and classify TMOs for HDR-video, a *qualitative experiment* identifying strengths and weaknesses of individual TMOs, and a *pair-wise comparison experiment* ranking which TMOs are preferred for a set of HDR-video sequences. Based on the results from the experiments, we identify a set of key aspects, or areas, in the processing of the temporal and spatial domains, that holds research problems which still need to be solved in order to develop TMOs for robust and accurate rendition of HDR-video footage captured in general scenes.

2. Background and related work

Despite a large body of research devoted to the evaluation of TMOs, there is no standard methodology for performing such studies. In this section we review and discuss the most commonly used experimental methodologies.

Figure 1 illustrates a general tone mapping scenario and a number of possible evaluation methods. The physical light intensities (luminance and radiance) in a scene are captured with a camera or rendered using computer graphics and stored in an HDR format. In the general case, "RAW" camera formats can be considered as HDR formats, as they do not alter captured light information given a linear response of a CCD/CMOS sensor. In the case of professional content production, the creator (director, artist) seldom wants to show what has been captured in a physical scene. The camera-
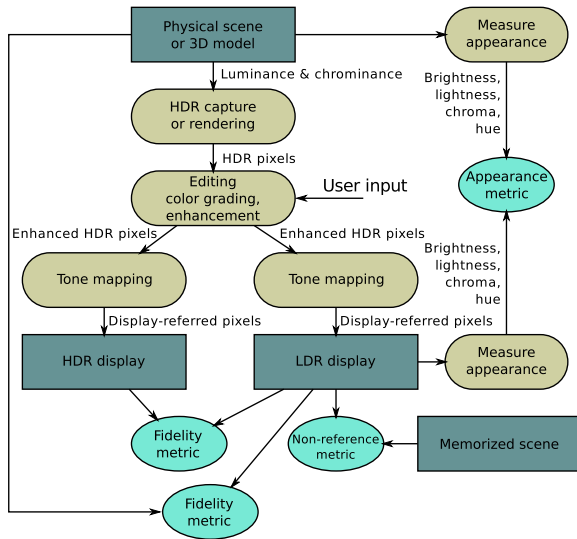
Figure 1: *Tone-mapping process and different methods of performing tone-mapping evaluation. Note that content editing has been distinguished from tone-mapping. The evaluation methods (subjective metrics) are shown as ovals.*

captured content is edited, color-graded and enhanced. This can be done manually by a color artist or automatically by color processing software. It is important to distinguish this step from actual tone-mapping, which, in our view, is meant to do "the least damage" to the appearance of enhanced content. In some applications, such as simulators or realistic visualization, where faithful reproduction is crucial, the enhancement step is omitted.

Tone-mapping can be targeted for a range of displays, which may differ substantially in their contrast and brightness levels. Even HDR displays require tone-mapping as they are incapable of reproducing the luminance levels found in the real world. An HDR display, however, can be considered as the best possible reproduction available, or a "reference" display. Given such a tone-mapping pipeline, we can distinguish the following evaluation methods:

Fidelity with reality method, where a tone-mapped image is compared with a physical scene. Such a study is challenging to execute, in particular for video because it involves displaying both a tone-mapped image and the corresponding physical scene in the same experimental setup. Furthermore, the task is very difficult for observers as displayed scenes differ from real scenes not only in the dynamic range, but they also lack stereo depth, focal cues, and have restricted field of view and color gamut. These factors usually cannot be controlled or eliminated. Moreover, this task does not capture the actual intent when the content needs enhancement. Despite the above issues, the method directly tests one of the main intents of tone-mapping (refer to VSS in the next section) and was used in a number of studies [YBMS05, AG06, YMMS06, ČWNA08, VL10].

Fidelity with HDR reproduction methods, where content is matched against a reference shown on an HDR display. Although HDR displays offer a potentially large dynamic range, some form of tone-mapping, such as absolute luminance adjustment and clipping, is still required to reproduce the original content. This introduces imperfections in the displayed reference content. For example, an HDR display will not evoke the same sensation of glare in the eye as the actual scene. However, the approach has the advantage that the experiments can be run in a well-controlled environment and, given the reference, the task is easier. Because of the limited availability of HDR displays, only a few studies employed this method: [LCTS05, KHF10].

Non-reference methods, where observers are asked to evaluate operators without being shown any reference. In many applications there is no need for fidelity with "perfect" or "reference" reproduction. For example, the consumer photography is focused on making images look possibly good on a device or print alone as most consumers will rarely judge the images while comparing with real scenes. Although the method is simple and targets many applications, it carries the risk of running a "beauty contest" [MR12], where the criteria of evaluation are very subjective. In the non-reference scenario, it is commonly assumed that tone-mapping is also responsible for performing color editing and enhancement. But, since people differ a lot in their preference for enhancement [YMMS06], such studies lead to very inconsistent results. The best results are achieved if the algorithm is tweaked independently for each scene, or essentially if a color artist is involved. In this way we are not testing an automatic algorithm though, but a color editing tool and the skills of the artist. However, if these issues are well controlled, the method provides a convenient way to test TMO performance against user expectations and, therefore, it was employed in most of the previous studies: [KYJF04, DZB05, AG06, YMMS06, AFR*07, ČWNA08, PM13].

Appearance match methods compare color appearance in both the original scene and its reproduction [MR12]. For example, the brightness of square patches can be measured in a physical scene and on a display using the magnitude estimation methods. Then, the best tone-mapping is the one that provides the best match between the measured perceptual attributes. Even though this seems to be a very precise method, it poses a number of problems. Firstly, measuring appearance for complex scenes is challenging. While measuring brightness for uniform patches is a tractable task, there is no easy method to measure the appearance of gloss, gradients, textures and complex materials. Secondly, the match of sparsely measured perceptual attributes does not need to guarantee the overall match of image appearance.

No existing method is free of problems. The choice of method depends on the application and what is relevant to the study. Here, we employ a non-reference method since

most applications will require achieving the best match to a memorized scene rather than a particular reference.

Almost all of the cited tone-mapping evaluation studies compared results of static image tone mapping rather than video tone mapping. The only exception is the study of Petit et al. [PM13], where 4 video operators, each at 5 different parameter settings, were compared on 7 video clips. The main observation of the study was that advanced TMOs perform better for selected scenes than a typical S-shape camera response curve. The study, however, used a low-sensitivity direct rating method and was limited to computer generated scenes and panning across static panorama images. In this paper we extend the scope of the study to 11 TMOs, use realistic HDR-camera captured video clips, and employ a much more extensive and sensitive evaluation methods.

3. Survey of TMO

For the evaluation of tone mapping algorithms designed for HDR-video, a number of different operators were considered. Here, we discuss some aspects in the selection of suitable candidates.

Intent of TMO. It is important to recognize that different TMOs try to achieve different goals [MR12], such as perceptually accurate reproduction, faithful reproduction of colors or the most preferred reproduction. After analyzing the intents of existing operators, we can distinguish three classes:

- *Visual system simulators (VSS)* – simulate the limitations and properties of the visual system. For example, a TMO can add glare, simulate the limitations of human night vision, or reduce colorfulness and contrast in dark scene regions. Another example is the adjustment of images for the difference between the adaptation conditions of real-world scenes and the viewing conditions (including chromatic adaptation).
- *Scene reproduction (SRP) operators* – attempt to preserve the original scene appearance, including contrast, sharpness and colors, when an image is shown on a device of reduced color gamut, contrast and peak luminance.
- *Best subjective quality (BSQ) operators* – are designed to produce the most preferred images or video in terms of subjective preference or artistic goals.

TMO selection. As the first step in the selection of operators, we identified which TMOs are explicitly designed to work with video. Since TMOs for static images do not ensure temporal coherence of pixel values, which *e.g.* can result in severe flickering artifacts, we restricted the evaluation to TMOs including a temporal model. To make a further selection, we classified the operators according to the method described above. As aiming for different goals will lead to different results, it is difficult to compare the performance of TMOs in a consistent way if their intent differs. Thus, our initial intention was to include only one class in the evaluation, namely the VSS operators. However, we observed

that some operators, which do not explicitly model the visual system, can potentially produce results that give better perceptual match than some VSS operators. Consequentially, the list of candidates was extended with a number of non-VSS operators. The final selection of operators is listed in Table 1.

4. Experimental Setup

Viewing conditions. All experiments, including pilot studies, parameter tuning, qualitative evaluation and pairwise comparisons, were carried out using the same viewing conditions. All clips were viewed in a dim room (25 lux) on a 24" 1920×1200 colorimetric LCD (Nec PA241W) set to the sRGB mode and a peak luminance of 200 cd/m^2. The observers sat at approximately 3 display heights (97 cm) distance, a typical viewing distance for HD-resolution content.

HDR-video sequences. Single frames from each of the video sequences used in the experiments are displayed in Figure 2. The sequences were selected to pose different challenges for the TMOs, and to represent a wide range of footage. These included both moderate and rapid intensity variations in the temporal and spatial domains, day and night scenes, skin tones, and varying noise properties. The sequences 2a - 2b and 2e - 2f were captured using a multi-sensor HDR camera setup similar to that described by [KGBU13], sequence 2c was captured using an RED EPIC camera set to HDR-X mode, and 2d is a computer graphics rendering. The captured sequences were calibrated by matching the luminance of test patches to the measurements made with a Photo Research PR-650 photo spectrometer.

5. Parameter selection experiment

It is well known that many TMOs are sensitive to the parameter settings and that extensive parameter tuning is often necessary to achieve a good result. However, it can also be argued that an automated algorithm should produce satisfactory results with a single set of parameters for a wide range of scenes (scene-independence). If the parameters need to be adjusted per scene, we are dealing with tone- and color-editing problems rather than automatic tone-mapping (refer to Figure 1). Therefore, in our experiments we use the same set of parameters for all tested scenes. Also, in addition to sensitivity to changes in the parameters, an important property of the TMOs are their sensitivity to calibration of the input data, as many TMOs require scene referred input. Some operators respond significantly to small changes in the scaling of the input, others are completely independent.

Ideally, we would like to use the default TMO parameters, which were optimized and suggested by the authors. However, it is not possible in a few cases. Both *Virtual exposures TMO* and *Camera TMO* do not offer default values for all parameters and require adjustment. We also found that *Maladaptation TMO* and *Cone model TMO* required fine tuning to produce acceptable results for our sequences. For these four operators we run a parameter adjustment experiment.

Operator	Processing	Intent	Description
Visual adaptation TMO [FPSG96]	Global	VSS	Use of data from psychophysical experiments to simulate adaptation over time, and effects such as color appearance and visual acuity. Visual response model is based on measurements of threshold visibility as in [War94].
Time-adaptation TMO [PTYG00]	Global	VSS	Based on published psychophysical measurements [Hun95]. Static responses are modeled separately for cones and rods, and complemented with exponential smoothing filters to simulate adaptation in the temporal domain. A simple appearance model is also included.
Local adaptation TMO [LSC04]	Local	VSS	Temporal adaptation model based on experimental data operating on a local level using bilateral filtering.
Mal-adaptation TMO [IFM05]	Global	VSS	Based on the work by Ward *et al.* [WLRP97] for tone mapping and Pattanaik *et al.* [PTYG00] for adaptation over time. Also extends the threshold visibility concepts to include maladaptation.
Virtual exposures TMO [BM05]	Local	BSQ	Bilateral filter applied both spatially for local processing, and separately in time domain for temporal coherence.
Cone model TMO [VH06]	Global	VSS	Dynamic system modeling the cones in the human visual system over time. A quantitative model of primate cones is utilised, based on actual retina measurements.
Display adaptive TMO [MDK08]	Global	SRP	Display adaptive tone mapping, where the goal is to preserve the contrasts within the input (HDR) as close as possible given the characteristic of an output display. Temporal variations are handled through a filtering procedure.
Retina model TMO [BAHC09]	Local	VSS	Biological retina model where the time domain is used in a spatio-temporal filtering for local adaptation levels. The spatio-temporal filtering, simulating the cellular interactions, yields an output with whitened spectra and temporally smoothed for improved temporal stability and for noise reduction.
Color appearance TMO [RPK*12]	Local	SRP	Display and environment adapted image appearance calibration, with localized calculations through the median cut algorithm.
Temporal coherence TMO [BBC*12]	Local	SRP	Post-processing algorithm to ensure temporal stability for static TMOs applied to video sequences. The authors use mainly Reinhard's photographic tone reproduction [RSSF02], for which the algorithm is most developed. Therefore, the version used in this survey is also utilising this static operator.
Camera TMO	Global	BSQ	Represents the S-shaped tone curve which is used by most consumer-grade cameras to map the sensor-captured values to the color gamut of a storage format. The curves applied were measured for a Canon 500D DSLR camera, with measurements conducted for each channel separately. To achieve temporal coherence, the exposure settings are anchored to the mean luminance filtered over time with an exponential filter.

Table 1: *List of tone mapping operators included in our survey.* Processing *refers to either global processing that is identical for all the pixels within a frame or local processing that may vary spatially.* Intent *is the main goal of the operator, see Section 3.*

Method. Four expert users tuned TMO parameters for three video clips using the method of adjustment; the clips used for that purpose were different from the ones in the other experiments. The clips were played in a loop, with the observer presented with a single slider to manipulate, allowing the adjustment of a single parameter at a time. Since it would be very time-consuming to generate a separate video for all parameter values represented by each possible position of the slider, only five video streams were pre-generated for different parameter values. They were then decoded at the same time, and the slider position indicated which clips were blended together to approximate results for intermediate parameter values. Because the parameter values were linearized prior to running the experiment, interpolation errors were found to be very small. To explore a multi-dimensional space of TMO parameters, we used Powell's conjugate direction method, [Pow64], for finding the minimum of a multi-dimensional non-differentiable function. At least two full

iterations were completed before the final values were accepted. Finally, the observer-averaged parameters for these four users were used for all of the following experiments.

6. Qualitative evaluation experiment

As a second step in our survey, we performed a qualitative analysis of the selected TMOs with the goal of identifying and tabulating their individual strengths and weaknesses. One of the main reasons for the evaluation was the results from a set of pilot studies, [EUWM13], showing the TMOs behaving very differently in the time domain, with some TMOs suffering from ghosting or flickering artifacts, making a comparison experiment difficult to interpret. To illustrate the TMOs temporal behavior, Figure 3 show their response over time at two pixel locations, denoted with a green and a red dot in 3a. Figures 3b - 3d show the responses for an input location with low temporal variation (red), that is mainly dependent
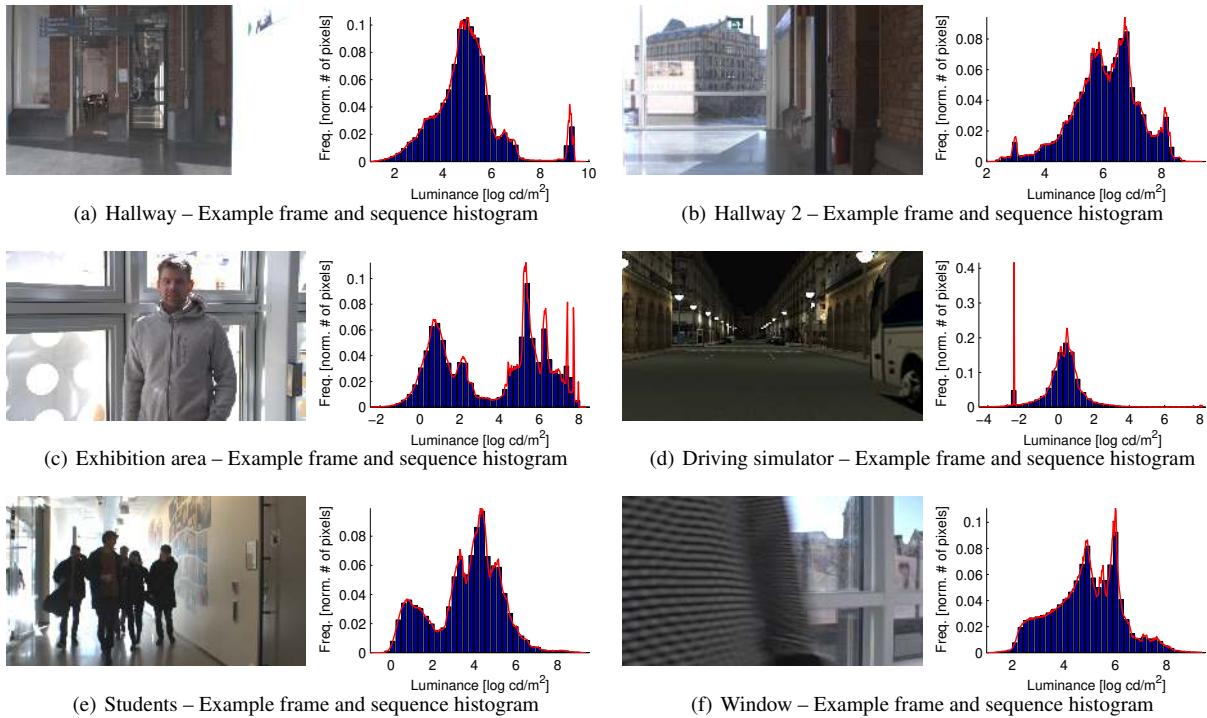
(a) Hallway – Example frame and sequence histogram

(b) Hallway 2 – Example frame and sequence histogram

(c) Exhibition area – Example frame and sequence histogram

(d) Driving simulator – Example frame and sequence histogram

(e) Students – Example frame and sequence histogram

(f) Window – Example frame and sequence histogram

Figure 2: *Example frames from the HDR-video sequences used in the experiments. The images are linearly scaled and gamma mapped for display. The histograms are computed over all frames in each sequence to show the dynamic range in each scene.*

on global effects of the TMO, and 3e - 3g show the response of a location with a high temporal variation (green). From Figures 3d and 3g it is evident that the *Virtual exposures TMO* (blue) and the *Local adaptation TMO* (black) introduce flickering and overshoots respectively.

Since TMOs are typically evaluated by comparing them to each other, it is necessary to identify and remove problematic TMOs due to the fact that: a) Comparing TMOs with severe and often unacceptable artifacts to each other is very difficult. If both results are unacceptable, the judgement (which one is better) does not provide much useful information. b) Pair-wise comparison gives only ranking (or rating after scaling) of operators without proper understanding of why one operator is better from the other. It is therefore difficult to find what the particular problems with operators are from a comparison study alone.

Method. The qualitative evaluation was carried out as a rating experiment where six video clips, see Figure 2, were tone-mapped with all operators listed in Table 1. Five expert observers viewed each clip in a random order and provided categorical rating of the following attributes: overall *brightness*, overall *contrast*, overall color *saturation*, temporal color *consistency* (objects should retain the same hue, chroma and brightness), temporal *flickering*, *ghosting* and excessive *noise*. The attributes were selected to capture the most common problems in video sequences and represent all of quality feature groups presented in [WP02]: based on spatial gradi-

ents, based on chrominance information, based on contrast information and based on absolute temporal information. In addition to categorical rating, the observers could also leave comments for each attribute and an overall comment for a particular sequence.

Results. The rating results are shown in Figure 4 and are also exemplified in the supplementary video. The two most salient problems were flickering and ghosting. Either of the artifacts rendered results of the operators not fit to be used in practice. For that reason we eliminated from further analysis all operators for which either artifact was visible in at least three scenes: *Virtual exposures TMO*, *Retina model TMO*, *Local adaptation TMO* and *Color appearance TMO*. Several operators revealed excessive amount of noise in the clips but we found noisy clips much less objectionable than those with ghosting and flickering. In terms of color reproduction, some operators produced results consistently too bright (*Retina model TMO*, *Visual adaptation TMO*, *Time-adaptation TMO*, *Camera TMO*), or too dark (*Virtual exposures TMO*, *Color appearance TMO*, *Temporal coherence TMO*). That, however, was not as disturbing as the excessive color saturation in *Cone model TMO* and *Local adaptation TMO*.

Table 2 presents a summary of the comments made by the observers and is color coded to give an overview of the results from the rating experiment. Based on the comments it is evident that temporal artifacts such as *flickering* and *ghosting*

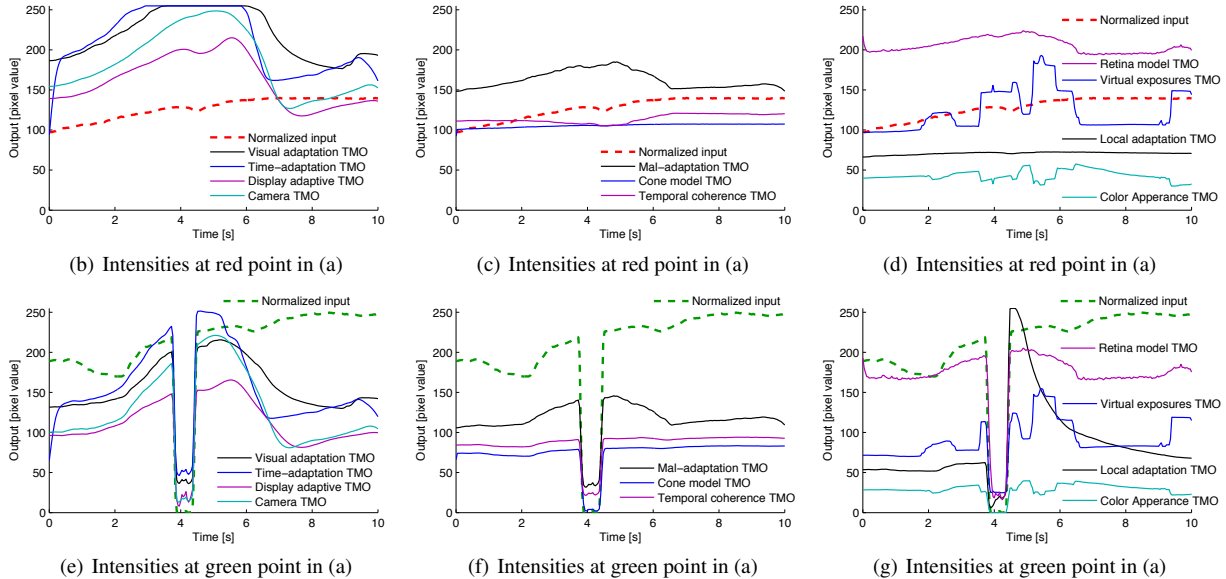(a) Frames of input sequence, with two measurement points indicated

(b) Intensities at red point in (a)

(c) Intensities at red point in (a)

(d) Intensities at red point in (a)

(e) Intensities at green point in (a)

(f) Intensities at green point in (a)

(g) Intensities at green point in (a)

Figure 3: *(a) shows a set of frames from a 10 sec. HDR sequence, where two positions are marked with red and green. The tone mapped intensity values of the different TMOs at these points are shown in (b)–(d) and (e)–(g), respectively, to illustrate the TMOs temporal properties. The operators are roughly grouped in different plots according to their behavior over time.*

are unacceptable. All artifacts are clearly exemplified in the supplementary video and the full set of comments from the observers can be found in the supplementary material.

7. Pairwise comparison experiment

The qualitative rating experiment delivered a number of useful observations. However, such a direct rating method is not sensitive to small differences between the results of two operators. We also would like to know which operator produces the best results in terms of the VSS intent. Therefore, we conducted a pair-wise comparison experiment for the 7 short-listed operators.

Method. In each trial the observers were shown two videos of the same HDR scene tone-mapped using different operators and asked which one of them looked more similar to what they imagined the scene would have looked like in real life. The full pairwise design, in which each pair of TMOs is compared, would require $5 \times \frac{1}{2} \times 7 \times 6 = 105$ comparisons, making it too time-consuming. Instead, we used a reduced design in which the Quicksort algorithm was used to reduce the number of compared pairs [SF01, MTM12]. 18 observers completed the experiment.

Results. Our result analysis is focused on practical significance, which estimates what portion of the population will

select one operator as a better reproduction of a memorized scene than the other. A convenient method to obtain such information for pair-wise comparison data is JND scaling. When 75% of observers select one condition over another, we assume that the quality difference between them is 1 JND. To scale the results in JND units we used the Bayesian method of Silverstein and Farrell [SF01]. In brief, the method maximizes the probability that the collected data explains the experiment under the Thurstone Case V assumptions. The optimization procedure finds a quality value for each image that maximizes the probability, which is modeled by the binomial distribution. Unlike standard scaling procedures, the Bayesian approach is robust to unanimous answers, which are common when a large number of conditions are compared.

The results of the pairwise-comparison experiment are shown in Figure 5. The results are reported per scene rather than averaged, because JND is a relative (interval) scale with different absolute values per scene. The results differ between scenes but there is also an observable pattern: The most frequently preferred operators are *Mal-adaptation TMO*, *Display adaptive TMO* and *Camera TMO*. *Cone model TMO* and *Time-adaptation TMO* were most often rejected. The practical difference between the best performing operators is not very large for some clips, for example the three best operators differ by at most 0.5 JND units for the *Window* clip. Interestingly, *Camera TMO* was selected as the best in the *Hallway*
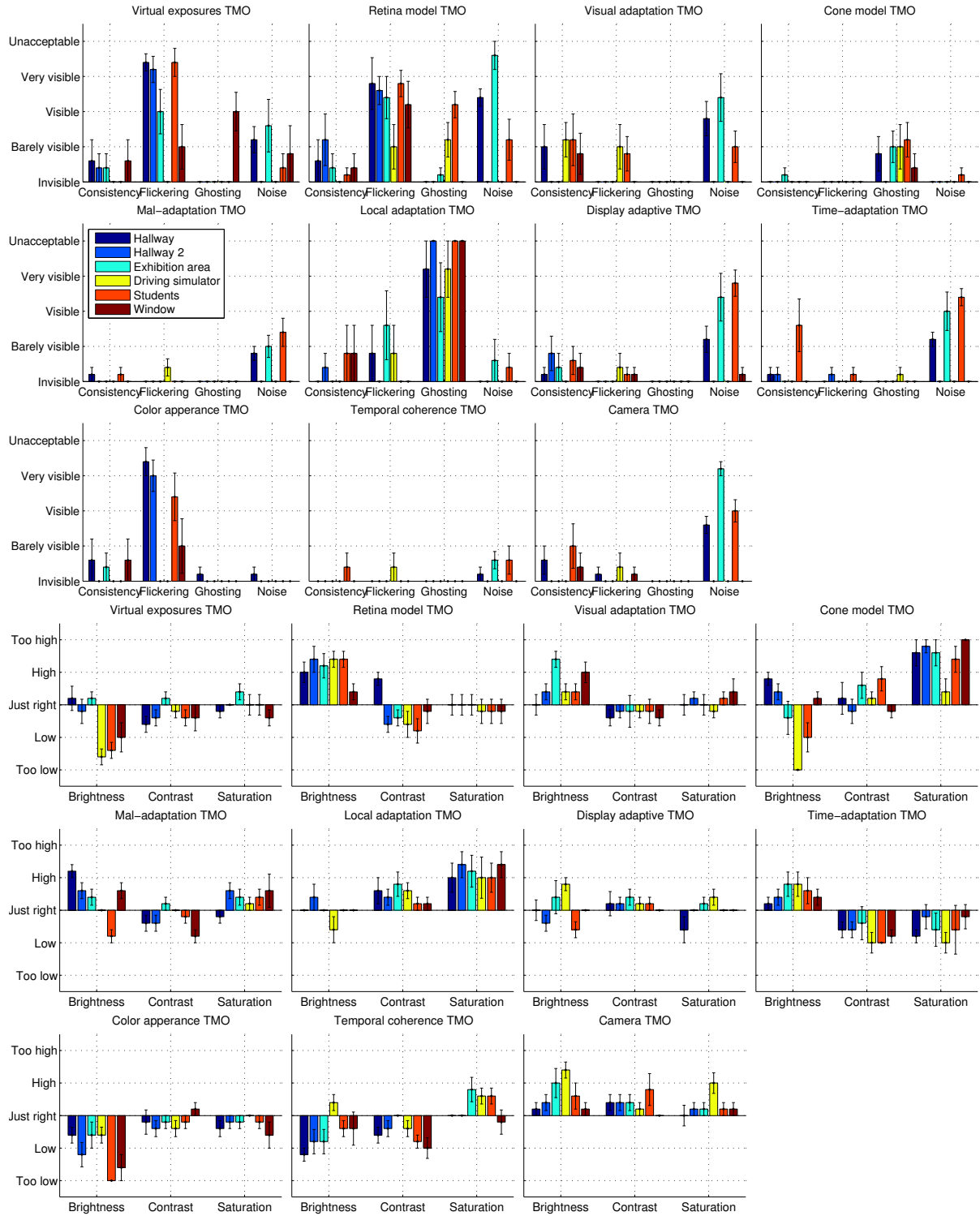
Figure 4: *Rating of the artifacts (top) and color-rendition problems (bottom) in tone-mapped clips (colors) for each operator (separate plots). The results are averaged over observers. The errors bars denote standard errors. The color codes are the same for all plots (refer to the legend).*

Operator	Brightness	Contrast	Color saturat.	Color consist.	Flickering	Ghosting	Noise
Visual adaptation TMO	Over-exp. when adapting to dark environment.	Dynamic range compression limited.		Some inconsistent brightening.	Adaptation can be perceived as flickering.		At overexposed locations.
Time-adaptation TMO	Over-exposure problems.	Low contrast. Dynamic range compr. limited.	Consistently de-saturated.	Some inconsistent changes in *Students* seq.			Noise incr. due to boost of low intensity regions.
Local adaptation TMO		Somewhat high.	Significant saturation.	Ghosting may be read as inconsistency.	Ghosting may be read as flickering.	Severe ghosting due to local adaptation.	Visible in ghosting artifacts.
Mal-adaptation TMO	Both low/high depending on sequence.	Somewhat too low for most sequences.	Tendency to over-saturate.				Boost of low intensity regions.
Virtual exposures TMO	Under-exposed under some conditions.	Somewhat too low for most sequences.			Flickering in intensity transitions.	In *Window* seq. cause of filtering problems.	Boost of low intensity regions.
Cone model TMO	Both low/high depending on sequence.	Somewhat too high for some sequences.	Significant saturation.			Motion blur.	
Display adaptive TMO	Both low/high depending on sequence.	Dynamic range compression limited.					Boost of low intensity regions.
Retina model TMO	Consistently too bright.	Both low/high depending on sequence.		Some block artifacts.	High-frequency flickering.	Motion blur.	Much to visible due to exposure problems.
Color appearance TMO	Under-exposure problems.			Some inconsistent adaptations.	Flickering in intensity transitions.		
Temporal coherence TMO	Under-exposed for most sequences.	Somewhat too low for most sequences.	Somewhat too high for some sequences.				
Camera TMO	Over-exposed under some conditions.	Somewhat too high.		Small saturation changes over time.			Boost of low intensity regions.

Table 2: *Summary of the problems recognized in the qualitative evaluation. From the experiment result we map to the following clarifications:* **red**; *critical problems that to a large extent affect the perceived visual quality of the tone reproduction.* **yellow**; *issues of less obvious character, but which add to a weaker outcome of the operator.* **green**; *no visible artifacts or weaknesses.*

clip (though only by 0.6 JND difference) and was one of the best operators in other clips, even though this operator does not attempt to simulate any perceptual effects. Both *Mal-adaptation TMO* and *Display adaptive TMO*, which scored highly, are histogram-based operators. This demonstrates that content-adaptive tone curve gives an advantage in the scenes with difficult lighting conditions, such as *Exhibition area* and *Students* clips.

Note that the results are not meant to provide a ranking of the operators as there is no guarantee that the results for five clips generalize for the entire population of possible video clips. The confidence intervals account for random effect of the group of observers, but they cannot predict variation in JND scores for different video clips

8. Discussion

The availability of HDR-video content has been limited until recently due to lack of HDR video cameras. As a result most

of the TMOs have not been tested on genuinely challenging video material and thus not properly validated. Most test footage was limited to rendered material and high quality still HDR images, lacking important features such as camera noise, rapid local luminance variations, skin tones, transitions from bright to dark scenes etc. By including such features in the test material, we were able to identify which aspects of current TMOs work satisfactorily and which must be improved. Below we discuss these aspects in order to arrive at a number of research problems that need to be solved in order to enable robust and high quality tone mapping.

Temporal artifacts. From the qualitative experiments, we have seen that even very small amounts of ghosting or flickering make a TMO unacceptable to use in practice. An interesting observation is that "simpler" TMOs with *global* processing are in general significantly more robust compared to *local* operators. In particular, all TMOs that were excluded in Section 6 due to temporal artifacts rely on local processing.
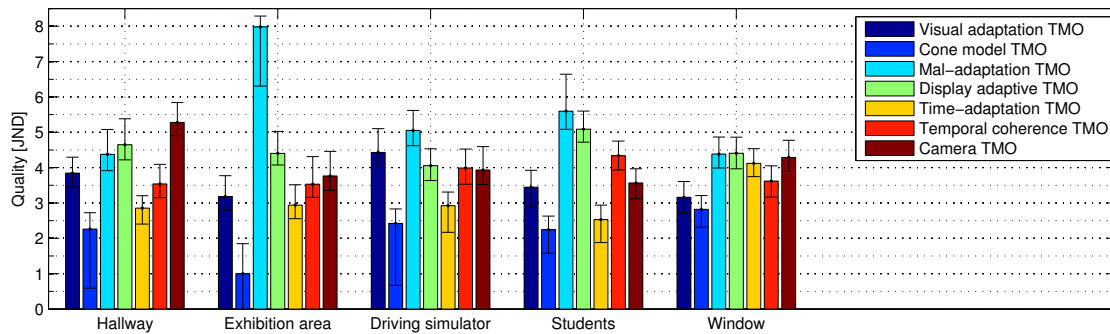
Figure 5: *Results of the pairwise-comparison experiment scaled in JND units (the higher, the better) under Thurstone Case V assumptions, where 1 JND corresponds to 75% discrimination threshold. Note that absolute JND values are arbitrary and only relative differences are meaningful. The error bars denote 95% confidence intervals computed by bootstrapping.*

Another interesting observation is that temporal processing that makes the adaptation time too short or too long (for rapidly changing luminance levels) is often perceived as being incorrect as in e.g. *Camera TMO* and *Time-adaptation TMO*. Examples of this adaptation behavior are shown in Figure 3b and 3e, where the color of a pixel that should remain constant changes significantly over time.

Contrast and brightness. Many operators suffer from low contrast. This is a common problem for global operators, which often reduce contrast to compress the dynamic range. TMOs that perform local processing are able to retain details and local contrast while compressing the dynamic range but are prone to temporal artifacts as discussed above.

Noise. Noise has been largely ignored in case of TMOs for still images. For video content, however, noise becomes a significant problem. From all evaluated TMOs, three addressed the issue of noise: *Virtual exposures TMO*, *Cone model TMO* and *Retina model TMO*. The treatment of noise ranges from naïve filtering over time, to more advanced methods that account for different noise characteristic in low intensity areas (*Virtual exposures TMO*). The simpler methods, however, have the side-effect of introducing ghosting and motion blur.

Implications. Given the results in the artifact rating experiment and the final evaluation, some correlations are of interest. From a quick look at Figure 4 we see that the cumulated importance of the different artifacts and color rendition problems roughly correspond to their performance in the evaluation result in Figure 5. We also note that some operators generating well visible amounts of noise in certain sequences still come out on top in the evaluation experiment. This can be caused by noise being of lower perceptual importance in the evaluations, or since these operators outperform other TMOs in other aspects. Investigating such connections and their implications further would be an interesting addition in future work.

High quality tone mapping for HDR-video. From this discussion we draw the conclusion that the problem of tone mapping of HDR-video sequences is far from being solved.

In order to do so, the ideal TMO should have the following properties:

- Temporal model free from artifacts such as flickering, ghosting and disturbing (too noticeable) temporal color changes.
- Local processing to achieve sufficient dynamic range compression in all circumstances while maintaining a good level of detail and contrast.
- Efficient algorithms, since large amount of data need processing, and turnaround times should be kept as short as possible.
- No need for parameter tuning.
- Calibration of input data should be kept to a minimum, *e.g.* without the need of considering scaling of data.
- Capability of generating high quality results for a wide range of video inputs with highly different characteristics.
- Explicit treatment of noise and color.

9. Conclusion and Future Work

This paper presented the first systematic survey and evaluation of tone mapping operators for HDR-video content. Eleven representative state-of-the-art TMOs were investigated in a series of experiments. First, the optimum parameters were found, then artifacts and color renditions problems were rated and commented, and finally, the TMOs were compared to each other. Based on the results from the evaluation we arrive at a list of challenges that need to be solved in order to develop a robust TMO that can produce visually pleasing results for general input data. As future work, we will take the main results from this evaluation as a starting point in the development of computationally efficient TMOs that handles temporal variations and image/video noise in a robust way.

formation Technology (CENIIT), and COST Action IC1005 on HDR video.

References

[AFR*07] AKYÜZ A. O., FLEMING R., RIECKE B. E., REINHARD E., BULTHOFF H. H.: Do HDR displays support LDR content? A psychophysical evaluation. *ACM Trans. on Graph. 26*, 3 (2007). 2

[AG06] ASHIKHMIN M., GOYAL J.: A reality check for tone mapping operators. *ACM Transactions on Applied Perception 3*, 4 (2006). 2

[BAHC09] BENOIT A., ALLEYSSON D., HERAULT J., CALLET P.: *Computational Color Imaging*. Springer-Verlag, Berlin, Heidelberg, 2009, ch. Spatio-temporal Tone Mapping Operator Based on a Retina Model. 4

[BBC*12] BOITARD R., BOUATOUCH K., COZOT R., THOREAU D., GRUSON A.: Temporal coherency for video tone mapping. In *Proc. of SPIE 8499, Applications of Digital Image Processing XXXV* (October 2012). 4

[BM05] BENNETT E. P., MCMILLAN L.: Video enhancement using per-pixel virtual exposures. *ACM Trans. on Graph. 24*, 3 (July 2005). 4

[ČWNA08] ČADÍK M., WIMMER M., NEUMANN L., ARTUSI A.: Evaluation of HDR tone mapping methods using essential perceptual attributes. *Computers & Graphics 32*, 3 (2008). 2

[DZB05] DELAHUNT P. B., ZHANG X., BRAINARD D. H.: Perceptual image quality: Effects of tone characteristics. *Journal of Electronic Imaging 14*, 2 (2005). 2

[EUWM13] EILERTSEN G., UNGER J., WANAT R., MANTIUK R.: Survey and evaluation of tone mapping operators for HDR video. In *ACM SIGGRAPH 2013 Talks* (New York, USA, 2013), SIGGRAPH '13, ACM. 4

[FPSG96] FERWERDA J. A., PATTANAIK S. N., SHIRLEY P., GREENBERG D. P.: A model of visual adaptation for realistic image synthesis. In *Proc. of SIGGRAPH '96* (New York, USA, 1996), ACM. 4

[Hun95] HUNT R. W. G.: *The Reproduction of Colour*. Fountain Press, 1995. 4

[IFM05] IRAWAN P., FERWERDA J. A., MARSCHNER S. R.: Perceptually based tone mapping of high dynamic range image streams. In *Proc. of Eurographics Symposium on Rendering* (June 2005), Eurographics Association. 4

[KGBU13] KRONANDER J., GUSTAVSON S., BONNET G., UNGER J.: Unified HDR reconstruction from raw CFA data. In *Proceedings of the IEEE International Conference on Computational Photography* (2013). 1, 3

[KHF10] KUANG J., HECKAMAN R., FAIRCHILD M. D.: Evaluation of HDR tone-mapping algorithms using a high-dynamic-range display to emulate real scenes. *Journal of the Society for Information Display 18*, 7 (2010). 2

[KYJF04] KUANG J., YAMAGUCHI H., JOHNSON G. M., FAIRCHILD M. D.: Testing HDR image rendering algorithms. In *Proc. IS&T/SID 12th Color Imaging Conference* (Scotsdale, Arizona, 2004), pp. 315–320. 2

[LCTS05] LEDDA P., CHALMERS A., TROSCIANKO T., SEETZEN H.: Evaluation of tone mapping operators using a high dynamic range display. *ACM Trans. on Graph. 24*, 3 (2005). 2

[LSC04] LEDDA P., SANTOS L. P., CHALMERS A.: A local model of eye adaptation for high dynamic range images. In *Proc. of AFRIGRAPH '04* (New York, NY, USA, 2004), ACM. 4

[MDK08] MANTIUK R., DALY S., KEROFSKY L.: Display adaptive tone mapping. *ACM Trans. on Graph. 27*, 3 (2008). 4

[MR12] MCCANN J. J., RIZZI A.: *The Art and Science of HDR Imaging*. Wiley, Chichester, West Sussex, UK, 2012. 2, 3

[MTM12] MANTIUK R. K., TOMASZEWSKA A., MANTIUK R.: Comparison of four subjective methods for image quality assessment. *Computer Graphics Forum 31*, 8 (2012). 6

[PM13] PETIT J., MANTIUK R. K.: Assessment of video tone-mapping : Are cameras' S-shaped tone-curves good enough? *Journal of Visual Communication and Image Representation 24* (2013). 2, 3

[Pow64] POWELL M. J. D.: An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal 7*, 2 (Jan. 1964). 4

[PTYG00] PATTANAIK S. N., TUMBLIN J., YEE H., GREENBERG D. P.: Time-dependent visual adaptation for fast realistic image display. In *Proc. of SIGGRAPH '00* (New York, USA, 2000), ACM Press/Addison-Wesley Publishing. 4

[RHP*10] REINHARD E., HEIDRICH W., PATTANAIK S., DEBEVEC P., WARD G., MYSZKOWSKI K.: *High dynamic range imaging: acquisition, display, and image-based lighting*, 2nd ed. Morgan Kaufmann, 2010. 1

[RPK*12] REINHARD E., POULI T., KUNKEL T., LONG B., BALLESTAD A., DAMBERG G.: Calibrated image appearance reproduction. *ACM Trans. on Graph. 31*, 6 (November 2012). 4

[RSSF02] REINHARD E., STARK M., SHIRLEY P., FERWERDA J.: Photographic tone reproduction for digital images. In *Proc. of SIGGRAPH '02* (2002), ACM. 4

[SF01] SILVERSTEIN D., FARRELL J.: Efficient method for paired comparison. *Journal of Electronic Imaging 10* (2001). 6

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

[UG07] UNGER J., GUSTAVSON S.: High-dynamic-range video for photometric measurement of illumination. *Proc. of SPIE 6501* (2007), 65010E. 1

[VH06] VAN HATEREN J. H.: Encoding of high dynamic range video with a model of human cones. *ACM Trans. on Graph. 25* (October 2006). 4

[VL10] VILLA C., LABAYRADE R.: Psychovisual assessment of tone-mapping operators for global appearance and colour reproduction. In *Proc. of Colour in Graphics Imaging and Vision 2010* (Joensuu, Finland, 2010). 2

[War94] WARD G.: A contrast-based scalefactor for luminance display. In *Graphics gems IV*, Heckbert P. S., (Ed.). Academic Press Professional, Inc., San Diego, USA, 1994. 4

[WLRP97] WARD LARSON G., RUSHMEIER H., PIATKO C.: A visibility matching tone reproduction operator for high dynamic range scenes. In *ACM SIGGRAPH 97 Visual Proceedings: The art and interdisciplinary programs of SIGGRAPH '97* (New York, USA, 1997), SIGGRAPH '97, ACM. 4

[WP02] WOLF S., PINSON M.: Video quality measurement techniques. *2002.* (2002), 47–54. 5

[YBMS05] YOSHIDA A., BLANZ V., MYSZKOWSKI K., SEIDEL H.-P.: Perceptual evaluation of tone mapping operators with real world scenes. In *Proc. of SPIE Human Vision and Electronic Imaging X* (San Jose, CA, 2005), vol. 5666. 2

[YMMS06] YOSHIDA A., MANTIUK R., MYSZKOWSKI K., SEIDEL H.-P.: Analysis of reproducing real-world appearance on displays of varying dynamic range. *Computer Graphics Forum 25*, 3 (2006). 2