# A comparative review of tone-mapping algorithms for high dynamic range video

G. Eilertsen[1],   R. K. Mantiuk[2],   J. Unger[1]

[1]Linköping University, Sweden    [2]Computer Laboratory, University of Cambridge, UK

**Abstract**

*Tone-mapping constitutes a key component within the field of high dynamic range (HDR) imaging. Its importance is manifested in the vast amount of tone-mapping methods that can be found in the literature, which are the result of an active development in the area for more than two decades. Although these can accommodate most requirements for display of HDR images, new challenges arose with the advent of HDR video, calling for additional considerations in the design of tone-mapping operators (TMOs). Today, a range of TMOs exist that do support video material. We are now reaching a point where most camera captured HDR videos can be prepared in high quality without visible artifacts, for the constraints of a standard display device.*

*In this report, we set out to summarize and categorize the research in tone-mapping as of today, distilling the most important trends and characteristics of the tone reproduction pipeline. While this gives a wide overview over the area, we then specifically focus on tone-mapping of HDR video and the problems this medium entails. First, we formulate the major challenges a video TMO needs to address. Then, we provide a description and categorization of each of the existing video TMOs. Finally, by constructing a set of quantitative measures, we evaluate the performance of a number of the operators, in order to give a hint on which can be expected to render the least amount of artifacts. This serves as a comprehensive reference, categorization and comparative assessment of the state-of-the-art in tone-mapping for HDR video.*

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

## 1. Introduction

High dynamic range (HDR) imaging has seen immense development the last two decades and is today an important aspect in a wide range of applications in computer graphics and image processing [RWP*10, BADC11]. HDR video acquisition is also maturing, where high quality HDR video sequences can be readily captured with versatile research prototypes [TKTS11, KGB*14]. Given this recent development, HDR video is now starting to establish itself in a number of applications [DCMM16], and constitutes an important topic for future research.

A vital component in the HDR imaging pipeline is tone-mapping. The purpose of tone-mapping is to adapt the final HDR content for viewing on a display device that is limited in terms of dynamic range and color gamut. Although HDR display systems do exist [SHS*04, RWP*10], and TVs with an extended dynamic range are currently establishing themselves in the commercial market, they have their limitations in terms of dynamic range and color gamut. Therefore, the material prepared for an HDR display will inevitably require some form of tone-mapping. Also, the absolute majority of displays are, and will most likely be, of a very limited dynamic range. Thus, the area of tone-mapping will continue to

be an important topic, and even more so with the rapid increase in availability of HDR video and solutions for distribution of such.

Methods for tone-mapping started to appear in the early 1990s, and have since evolved rapidly with a vast number of tone-mapping operators (TMOs) introduced in the literature. With the transition from static HDR images to video sequences, special considerations have to be made in the tone-mapping pipeline. Apart from the most evident challenge of maintaining a temporally coherent tone-mapping, the implications of the temporal dimension also include differences due to the capturing process. HDR videos are, for example, more likely to contain noise, motion blur and skin tones.

In this report, we provide a review of TMOs that explicitly account for the time domain. This is motivated by the challenges that tone-mapping of HDR video present. It is also a response to the recent and rapid development in HDR video capture. The review includes descriptions and categorization of the available methods, as well as a comparative assessment. The comparisons are based on a set of important properties, including temporal coherence, noise visibility, contrast and exposure. The result gives an indication of the operators ability to produce well-exposed and high contrast videos, without showing temporal artifacts and camera noise. Furthermore, to provide a starting point and context for our review, we

begin with an historical overview of tone-mapping in general. This is not intended to extensively cover the huge body of research in tone-mapping of images; instead, we focus on the evolution of the tone-mapping pipeline. The overview is followed by a general formulation of the tone-mapping problem, in order to clearly define the calculations involved.

The area of tone-mapping has already been covered to a certain extent in the literature. However, earlier attempts at summarizing the state-of-the-art are getting outdated [DCWP02]. Furthermore, more recent descriptions of TMOs are not explicitly focused on the problem of video tone-mapping, do not include many of the most recent contributions, and/or do not provide the same categorization and assessment we make [Hoe07, MMK08, RWP*10, MR11, BADC11, DCMM16]. Therefore, this report provides a novel contribution in the area. It serves as a comprehensive reference and comparison of the state-of-the-art in video tone-mapping, in order to facilitate future development in the area.

To summarize the content of the report, it is designed to provide the following:

1. A short literature review that provides an historical overview of the area of HDR and tone-mapping (Section 3).
2. A general formulation of a tone-curve, in order to clearly define the concept of tone-mapping and to facilitate a connection between different methods (Section 4).
3. A description of video tone-mapping, and the most important problems that need to be addressed specifically when operating on camera captured HDR video (Section 5).
4. A general suggestion for video TMO classification, using 5 distinct categories (Section 5.4). This is based on the concepts explained in the preceding sections.
5. A literature review over all TMOs we found that have an explicit algorithm for how to treat temporal aspects (Section 6).
6. A list of 26 classified video tone-mapping operators, according to the aforementioned categorization (Table 1).
7. A quantitative assessment on a set of the video tone-mapping algorithms, where we formulate measures to compare them in terms of a number of important properties (Section 7).

## 2. General background

Tone-mapping of HDR images involves transforming the HDR signal to a reduced dynamic range that can be presented on a standard display device. As such, the transformation typically involves a non-linear and compressive function. The complete tone-mapping pipeline can be constructed in many ways, but one of the most common designs is depicted in Figure 1. This transforms the input luminance to the log domain, where an edge-preserving filter is applied. The filtered image is then used to extract image details, which are kept separate from the compressive transformation performed by the tone-curve. Thus, local details are preserved while the overall dynamic range is reduced. Finally, the tone-mapped result is transformed back to the linear domain, colors are restored from the original input, and the material is prepared for a certain display device.

In this section we briefly describe the fundamentals of HDR and tone-mapping, and how the concepts in Figure 1 are motivated. A

more in-depth formulation of the actual mapping is then provided in Section 4. We also recognize that the motivation may be different based on the specific intent of the tone-mapping. This fact is later used to classify TMOs into three distinct groups.

### 2.1. Differences between LDR and HDR images

A *high dynamic range* image can describe a greater range of color and brightness than can be represented in a *low dynamic range* (LDR) image. While a standard LDR image typically uses 8 bits to store each color of a pixel, an HDR image is most often specified to floating point precision. The difference makes it possible to encode the entire range of colors and luminances visible to the human eye. However, the differences are not only in the representational capacity. An LDR image is almost exclusively meant to be sent to a display device. It is thus calibrated for this purpose, and said to be *display-referred*. For example, a gamma correction, $l = T^{1/\gamma}$, is already applied to the encoded image. Here, $T$ is intensity and $l$ is the final *luma* value that should be sent to the display. Typically, the gamma is in the range $\gamma \in [1.8, 2.8]$. While the correction originally was intended for CRT monitors, the non-linearity corresponds well with the human perception of brightness [MMS15]. Therefore, gamma correction can be viewed as a very simple form of tone-mapping. Furthermore, the gamma correction can be complemented to account for display and viewing environment, using a gamma-offset-gain model [Ber96, MMS15],

$$L_d(l) = l^\gamma \cdot (L_{max} - L_{black}) + L_{black} + L_{refl}. \tag{1}$$

$L_d$ describes the luminance, as measured coming from the display surface. $L_{max}$ and $L_{black}$ is the display peak luminance and black level, respectively. $L_{refl}$ is the ambient light that is reflected from the display surface, which is approximated from the ambient illuminance $E_{amb}$ and the reflectivity $k$ of the display,

$$L_{refl} = \frac{k}{\pi} E_{amb}. \tag{2}$$

In contrast to an LDR image, the HDR counterpart is typically *scene-referred*. It describes the physical properties of the lighting in a captured scene. Thus, HDR images can be used for several purposes where accurate physical measurement are needed, e.g. in image based lighting, automotive applications, medical imaging etc. In order to show an HDR image on a standard display it has to be transformed – or *tone-mapped* – to a display-referred format. Tone-mapping is also used for preparing a scene-referred HDR image for display on an HDR display, scaling and adapting it according to the certain characteristics of the display. Furthermore, tone-mapping can be involved in mapping between different display-referred formats. This can be accomplished e.g. by first transforming the image to a scene-referred format, followed by tone-mapping to the target display.

Strictly speaking tone-mapping is a mapping of image tones from one domain to another. However, today the by far most common use of the term is to describe a mapping from the linear scene-referred luminances in an HDR signal, to the restricted domain of
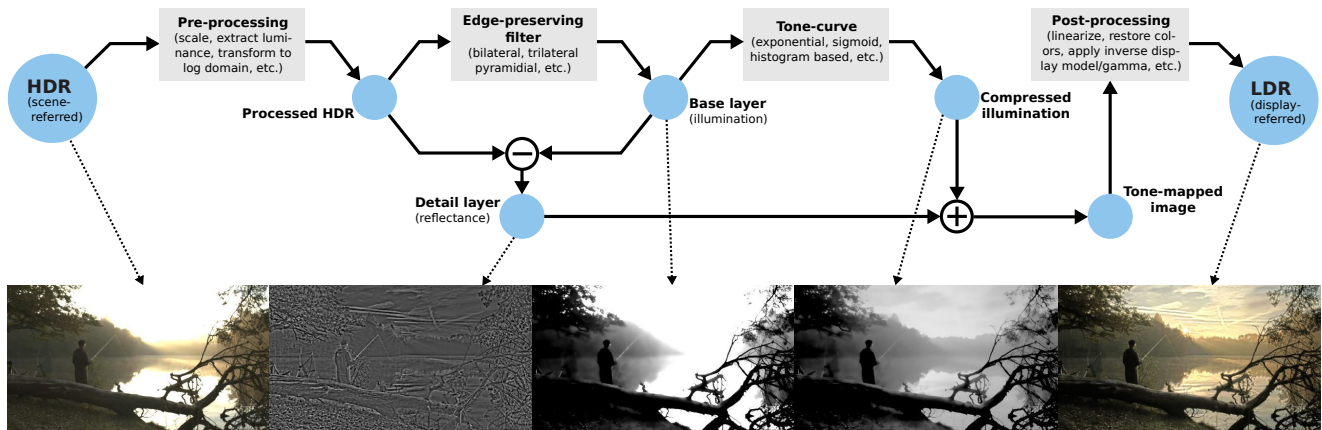
Figure 1: A typical pipeline setup for performing spatially varying (local) tone-mapping. The input scene-referred image is first taken through as set of pre-processing operations. The processed input is then filtered using an edge-preserving filter to create a base layer. The base layer is used to extract a detail layer. While the base layer is transformed using a tone-curve, the detail layer is kept untouched and added back after the mapping. Finally, the tone-mapped image with preserved details is prepared for display in a post-processing step. In the illustration intermediate images are shown as blue circles, and some are also displayed in the bottom of the figure.

a display device. Tone-mapping is now more or less synonymous with this definition, and it is in this form we use the term here. Usually the mapping goes to an intensity value $T$ that should be displayed. In order to finally create a display-referred LDR image, the tone-mapping is followed by clipping and gamma correction to adapt to the display device. Alternatively, the inverse of the display model in Equation 1 can be applied to compensate for display and viewing environment.

It should also be noted that tone-mapping is not only applicable to captured or rendered natural images. It can also be used on abstract representations, e.g. in volume visualization [YNCP05, YNCP06], where HDR can provide an extended dimension for conveying information [GTH05, KPB12]. Nevertheless, in most cases volume rendering is still done directly to a display-referred format, and by tweaking the transfer function a similar goal is achieved as with tone-mapping of a final HDR rendering [LKG*16].

## 2.2. HDR video capturing, display and distribution

Techniques for generating HDR images have been around since the early 1990s. These include e.g. physically based rendering engines [War94b], volume rendering techniques [GTH05], and reconstruction from camera exposure bracketing [MP95, DM97]. Although the rendering engines are capable of producing HDR video sequences, transitioning from static images to video for camera captured HDR is significantly more difficult. In order to record HDR video with a standard video camera, a temporal multiplexing can be used, that alternates between short and long exposures from frame to frame. This requires the video to be interpolated over time in order to reconstruct the HDR frames, e.g. using optical flow [KUWS03, MG10] or patch-based techniques [KSB*13]. Alternatively, HDR video can be captured with systems of multiple cameras for shooting the same scene [UGOJ04, WRA05, UG07]. However, high quality HDR video captured with dedicated versatile multi-sensor cameras did not appear until the 2010s [TKTS11,

KGB*14]. These have been demonstrated to exhibit a dynamic range of up to 20-24 *f-stops*. Recent development also brings forward professional HDR-camera systems such as Arri Alexa XT and Red Epic Dragon, with an extended dynamic range of up to 14-16.5 *f-stops*.

When it comes to HDR displays, some of the limitations of LCD panels can be overcome e.g. by a laser display [DB04]. However, the most common technique is to use dual modulation, increasing brightness by means of a back-lit LCD panel [SHS*04]. The HDR image is decomposed into a low resolution back-light image and a high resolution compensation image that modulates the intensity. In practice 8-bit images are used for back-light and compensation, effectively doubling the achievable bit-depth. The technique was further developed by Brightside, resulting in displays such as the DR37-P with a peak luminance of around 4000 cd/m$^2$. Later it was also used in the Dolby Pulsar monitor with a peak luminance of 4000 cd/m$^2$. Furthermore, Sim2 released monitors based on the technology in their HDR47 series, including the latest HDR47ES6MB capable of 6000 cd/m$^2$ peak luminance. Comparing to the professional HDR displays, a standard LCD monitor typically has a peak luminance in the vicinity of 200-300 cd/m$^2$, possibly exceeding 500 cd/m$^2$ for high fidelity displays. However, the consumer market is currently advancing rapidly, and a large number of HDR enabled TVs were released during 2016. While these monitors are far from achieving the same brightness as the professional devices, they are capable of an increased dynamic range and brightness, reaching – and for some TVs exceeding – a peak luminance of 1000 cd/m$^2$.

Although HDR displays can often match the dynamic range of the reproduced scenes, they do not eliminate the need for tone-mapping for several reasons. Firstly, HDR displays produce high contrast images only in the right viewing conditions, which is a dark room with no ambient light. If ambient light is present, their effective contrast is greatly reduced, thus introducing a need for

tone-mapping. Secondly, HDR displays vary greatly in their performance, so there is a need to adapt content to particular capabilities of a display. Tone-mapping is well suited for this task.

Distribution of HDR video can be accomplished using the power of existing video codecs, by transforming the luminance and encoding it at 10-12 bits [MKMS04]. Furthermore, there are techniques that allow for backward-compatibility [MEMS06, LK08]. In response to the recent development in HDR TV, there has also been many efforts from several organizations in standardization of the techniques involved in HDR video encoding. Already in the transition from high definition TV (HDTV) to ultra HDTV, a wider color gamut was standardized by moving from ITU-R Rec. BT.709 to BT.2020. As for HDR, and encoding of the wide dynamic range, a transformation of luminances was standardized with the perceptual quantizer (PQ) transfer function (SMPTE ST-2084) [MND13] and the Hybrid Log-Gamma (HLG). Both PQ and HLG are now included in the latest ITU-R recommendation, BT.2100. Many additional initiatives have also taken place, and we refer to the summaries by François et al. [FFH*16] and Schulte [SB16] for a thorough explanation. For distribution of HDR video the different initiatives have resulted in the Dolby Vision [Dol16] and MPEG HDR-10 [KYH16] codecs. Additionally, HDR video codecs can now be found implemented as open-source APIs [EMU16].

### 2.3. Perceptual motivation of tone-mapping

Tone-mapping is motivated by the properties of the human visual system (HVS), which simultaneously can register a dynamic range of about 3.7 log units in a scene [KR10, RAGB11]. While the individual neural units of the HVS are limited to respond in a range of about 1.5 log units [SW12], this capacity is achieved through the utilization of local adaptation mechanisms, such as spatially varying neural gain controls. The local processing makes it possible to restrict adaptation to an area of less than 0.5 visual degrees [VMGM15]. Furthermore, by adapting over time using e.g. pupil adjustments and pigment bleaching/regeneration processes, the HVS can function in luminances from around $10^{-6}$ cd/m$^2$ to $10^8$ cd/m$^2$. This makes for an effective dynamic range of about 14 log units. In tone-mapping the goal is to achieve the same – or superior – ability. Although a simulation of the HVS is not always the explicit objective, it is important to appreciate the intrinsic perceptual aspects. These play a fundamental role in the process of preparing HDR material for a display device.

As an example of the implications of tone-mapping being a perceptually motivated operation is the logarithmic space that often is used for performing the mapping. This stems from the close to logarithmic response of the HVS in large areas of the luminance range, according to the Weber-Fechner law [Fec65]. The law implies a logarithmic relationship between physical luminances and the perceived brightness. Thus, in many cases computations in the log domain allow for simpler formulations due to the improved perceptual linearity.

### 2.4. Detail preservation in tone-mapping

Already in early works of image processing, the idea of describing an image from intrinsic characteristics was introduced. The objective may be manifold, e.g. to provide some invariant description for the purpose of higher-level image understanding, or to be considered as an initial description of the early stages of the HVS. One such idea is to describe an image as separated into a product of illumination and reflectance [Hor74, BT78]. A human observer is able to determine the reflectance of objects in a large range of lighting situations, effectively describing a measure that is invariant to the illumination [GJ84]. The reflectance is one of the fundamental descriptions of a scene, and it is possible to specify using a very limited dynamic range (about 100:1). Thus, it makes intuitive sense to attempt to maintain this characteristic through the tone-mapping, while only compressing the dynamic range contained in the illuminance description. These principles are manifested in the construction of local tone-mapping operators, where the intent is to compress global contrasts while preserving local contrasts, or details. A common approach is to approximately separate an image into illuminance and reflectance using an edge-preserving filter, as illustrated in Figure 1.

### 2.5. Tone-mapping of colors

Although color appearance models sometimes are used in order to reproduce HDR colors [PFFG98, FJ04, AR06, KJF07, KWK09, RPK*12], for simplicity and computational efficiency tone-mapping typically only treats the luminance, and maintains colors from the original signal [Sch94]. Alternatively, the mapping can be performed separately on the color channels, in some suitable color space. However, tone-mapping on individual color channels tends to under-saturate colors, while separate treatment of a single luminance channel have an opposite effect. To account for this behavior, a heuristic de-saturation step can be performed when restoring colors from the original HDR image [THG99],

$$c = \left(\frac{C}{L}\right)^s T.$$ (3)

Here, $C$ and $c$ are the colors before and after tone-mapping, respectively, $L$ is the input luminance and $T$ is the tone-mapped value. The parameter $s$ steers the color saturation of the result, where a value $s < 1$ results in a de-saturation. Attempts have also been made to characterize the color saturation behavior, in order to derive the amount of saturation correction needed in a specific tone-mapping scenario [MMTH09, PAB*13].

### 2.6. Tone-mapping intent

Although a tone-mapping operator, as we define it here, is designed to map an HDR signal to the restricted space of a display, the objective may vary substantially – it is important to recognize that different TMOs try to achieve different goals [MR11]. Therefore, a natural way for classification of tone-mapping operators is their specific intent.

We use a classification with three different intents. This follows the distinction we used in previous work for evaluation of video TMOs [EWMU13] [DCMM16, Ch. 7]. The intents are illustrated in Figure 2, and can be described as follows:
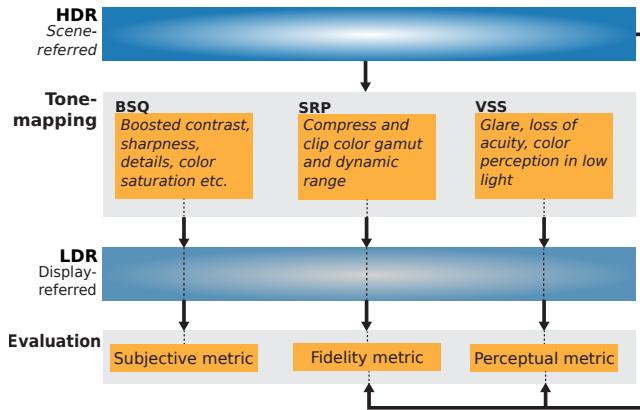
Figure 2: Tone-mapping using different types of intermediate processing depending on the intent. The motivation and design of an algorithm is based on assessing the quality of the tone reproduction with respect to its specific intent.

**Visual system simulator (VSS)** This category of operators try to simulate the properties of the human visual system (HVS), including its limitations. For example, a TMO can add glare, simulate the limitations of human night vision, or reduce colorfulness and contrast in dark scene regions. The goal is to achieve the minimal perceptual difference between viewing the real scene and its captured and tone-mapped counterpart on a certain display.

**Best subjective quality (BSQ)** In many circumstances a faithful reproduction of the depicted scene is not important, but rather to achieve a tone-mapping that is most preferred in terms of subjective quality. The subjective quality, however, is a wide term that is highly depended on the situation, and it can be established in closer terms, e.g. to be inline with a certain artistic goal. The category includes for example the creative aspects of grading feature films, where the result can be highly dependent on the particular viewing environment [CAS*15]. Comparing to the VSS category, the subjective preference is often to have a slightly exaggerated appearance, with increased contrasts, details and color saturation.

**Scene reproduction (SRP)** Here, the operators attempt to make the mapping $V(L)$ – to a display device with reduced capabilities – invariant to certain attributes. The goal is to preserve the original scene appearance as close as possible, measured in terms of the attributes. These could for example include the relative brightness, contrasts, colors, noise visibility, etc. However, only a restricted number of attributes can be captured simultaneously, and the tone-mapped image can still deviate substantially from the physical scene in many other respects.

## 3. Historical overview

Relating to the concepts in the previous section, we here give a short review of published work in the area of tone-mapping. This overview is by no means meant to be exhaustive. Since a large body of research has been devoted to the topic, with many specialized operators and refinements suggested through the years, we choose

to focus on tracking the evolution of tone-mapping. We then proceed to make a more rigorous review of video tone-mapping and algorithms for this purpose in Section 5 and 6, respectively.

Summarizing the number of operators we mention throughout this report, Figure 3 shows how many have been published each year, and how they are categorized. Although the plots do not include all possible TMOs, they give some insight into how the work of tone-mapping has been distributed.

### 3.1. Background

Techniques for compressing the dynamic range of a digital image signal was introduced already in the 1960s by Oppenheim et al. [OSS68]. Furthermore, in the 1980s the idea of matching perceptual attributes between real-world and display was elaborated on [MNM84, Ups85]. However, it was first in the early 1990s that the concept of tone-mapping was formally introduced in the computer graphics community [TR93]. Since then, tone-mapping as a research area has grown rapidly (see Figure 4). With techniques for reconstruction of natural HDR images from multi-exposure sequences, [MP95, DM97], it has become even more important to develop algorithms for display of both computer generated and camera captured HDR media. Furthermore, with the introduction of HDR video capturing [KUWS03, UGOJ04, WRA05, UG07, TKTS11, KGB*14], new problems were introduced that TMOs need to consider [EWMU13]. We return to the challenge of HDR video tone-mapping in Section 5.

To provide an illustration of the increasing popularity of HDR and tone-mapping, Figure 4 plots the occurrences of the terms as appearing in the database of Google Books used for the Google NGram viewer. The term tone-mapping appeared in a first wave in the context of linguistics during 1975-1990, describing a mapping of acoustic tones. A second wave came with the introduction of tone-mapping in computer graphics [TR93]. Hence, tone-mapping as we define it here – for display of HDR images – is a new and rapidly expanding area.

### 3.2. Tone-curves

The most trivial method for display of HDR images is by linear scaling, followed by clipping and a final gamma correction, effectively performing an exposure correction. As a first example, Ward uses a linear scaling, focusing on preserving the image contrasts [War94a]. The method is designed as a global scaling of luminances based on detection threshold experiment data produced by Blackwell [Bla81].

For a non-linear mapping of luminances, a linear scaling can be applied in the logarithmic domain, or equivalently as an exponential function in the linear domain. Such tone-curve was first used by Tumblin and Rushmeier [TR93], with the goal of preserving the perceived brightness during compression of the dynamic range, as described by Stevens and Stevens [SS60]. The method was later on improved, and used for a gaze-directed adaptation implementation [THG99]. Alternatively, instead of a tone-curve $V$ on the form $V(L;\alpha) = L^{\alpha}$, an exponential tone-curve can be formulated
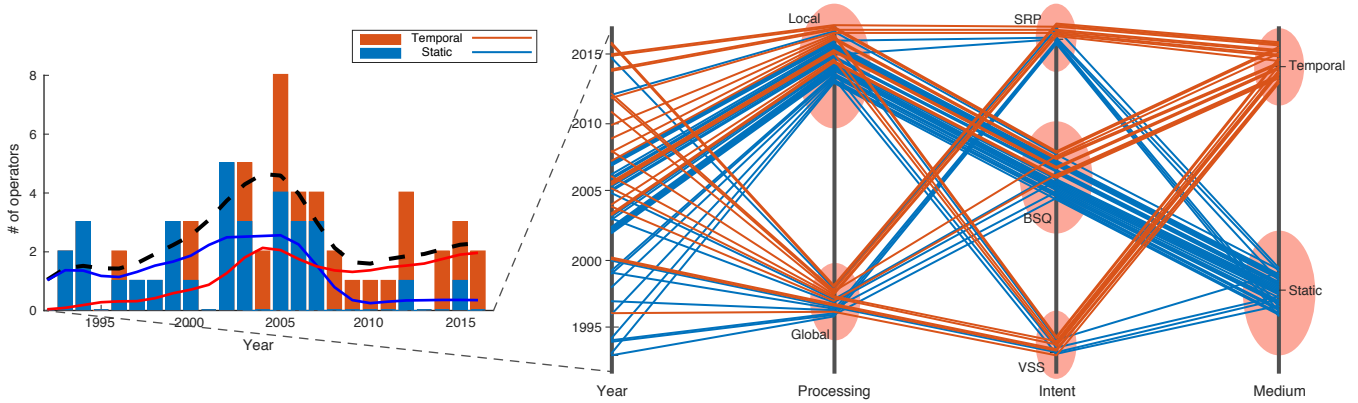
Figure 3: Histogram over time (left) of the number of published tone-mapping operators, and parallel coordinates plot of categorization (right). The plots are created using all TMOs that are referenced throughout this report, and do not include all of the existing TMOs. The lines in the TMO histogram are low-pass filtered from the static (blue), temporal (red) and all (black, dashed) TMO histograms, to show an indication on the progress through time.

as $V(L; \alpha) = 1 - e^{-\alpha L}$ [FTP94]. In the family of exponential tone-curves, one of the more frequently used in the literature is the logarithmic tone-curve originally proposed by Drago et al. [DMAC03].

Since the goal of tone-mapping is to map all input luminances to a predefined limited range, it makes intuitive sense to use a logistic, or sigmoid, function (see Equation 5). The function often occurs in natural processes. For example, it is referred to as the Michaelis-Menten equation in biochemistry [MM13] and the Naka-Rushton equation in the study of vision [NR66]. The sigmoid roughly corresponds to the response curve of the HVS, and is thus used in many of the VSS operators. Explicitly, the first use for tone-mapping was by Schlick [Sch94], although introduced with the motivation of being similar but computationally more economic as compared to exponential or logarithmic functions. Later on the sigmoid function

was used with the motivation of being similar to the photo-receptor response curve, e.g. by Pattanaik et al. [PFFG98].

Tone-curves that incorporate the mean or median value of the input are able to adapt to provide the best overall exposure of the output. However, using the image histogram a better distribution over luminances can be achieved. Formulating a tone-curve where the slope for each luminance $L_i$ is dependent on the corresponding probability $p_L(L_i)$ makes it possible to generate a result that shows better contrast and exposure in the more important areas of the image. The first method to use this technique was proposed by Ward et al. in [WLRP97]. It starts from a standard histogram equalization algorithm, and complements this with certain restrictions on the probabilities $p_L(L_i)$ in order to match the capabilities of the HVS. More recently, histogram based tone-curves have been used e.g. in order to provide minimal contrast distortions in the tone-mapping [MDK08, EMU15], to perform clustering based tone-mapping [Osk16], and to prevent noise visibility from increasing under the mapping [EMU15, LSRJ16].

Another important example of a family of tone-curves are those that are applied in the gradient domain. In [FLW02], Fattal et al. introduce this concept with a tone-curve that operates on image gradients by suppressing them proportional to their magnitudes.

### 3.3. Detail preservation

A Gaussian filter presents an efficient way of approximating an image's illuminance characteristics in order to perform local tone-mapping. In a first exploratory work, Chiu et al. make use of this idea. However, the approximation assumes smooth transitions in intensity. Close to large gradient edges the filter will be largely influenced by the different regions of intensity, which results in visible halos in the final tone-mapping. Gaussian filters where also used by Rahman et al. [RJW96] and Jobson et al. [JRW97], inspired by the center-surround operations used in the *retinex* theory [LM71].

A first experimental attempt to explicitly handle the problems caused by a standard low-pass filter can be found in [Sch94], where
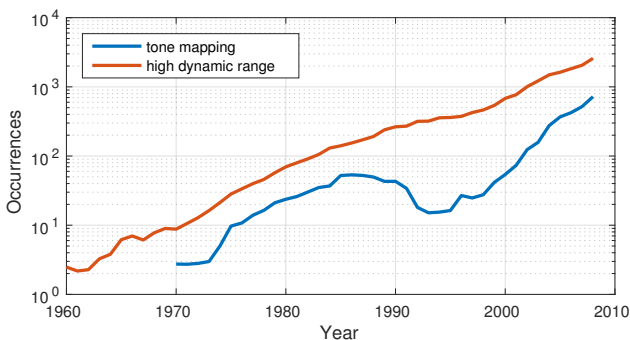


Figure 4: The number of occurrences of the phrases "tone mapping" and "high dynamic range" in the Google Books database of books used for the Google NGram Viewer. The plots reflect the popularity of the terms in the literature during the last 50 years (tone-mapping did not appear until during early seventies). Note that the y-axis is in log units, implying an exponential increase when the curves are linear.

local adaptation is suggested from the use of image segmentation on a simplified example. Although not practically applicable, the approach is a step towards the edge-preserving filters commonly used later on in tone-mapping. Later, a more general segmentation based approach for illuminance estimation was proposed by Yee and Pattanaik [YP03]. Segmentation has also been used to combine multiple TMOs in the same image [KMS05a,Čad07,BAS*12, SLF*15], for local interactive adjustments of the tone reproduction [LFUS06], and for combining different exposures directly from an exposure bracketing sequence without requiring an intermediate HDR reconstruction step [MKVR09].

Another way of alleviating the problems with a Gaussian filter is to perform filtering in an iterative manner, e.g. utilizing a Laplacian pyramid as described by Pattanaik et al. [PFFG98]. A more rigorous attempt at a hierarchical decomposition was proposed by Tumblin and Turk [TT99], using the LCIS filter. It draws inspiration from anisotropic diffusion [PM90] and represents a first serious attempt at using edge-preserving filtering in the context of tone-mapping. However, artifacts are still a problem, and the method is relatively expensive. Other early examples of pyramid filtering structures for local tone-mapping include the iterative Gaussian filter by Ashikhmin [Ash02], the Gaussian pyramid utilized by Reinhard et al. inspired by photographic practices [RSSF02], and the sub-band architecture proposed by Li et al. [LSA05].

A first explicit formulation of tone-mapping as a separation of illuminance and reflectance is given by Tumblin et al. [THG99]. The method uses a layering approach where illumination/reflectance pairs are extracted from a computer generated scene. It is, however, limited to computer generated images.

Extracting an approximation of the illuminance in a camera captured image, without the artifacts of a Gaussian filter, can be accomplished utilizing an edge-preserving filter. While the early filtering pyramid structures in effect are edge-preserving filters, the widely popular bilateral filter became a common method in the context of tone-mapping in the early 2000s. The idea of using the formulation of a bilateral filter kernel for local tone-mapping were first discussed in [DW00], although not really tested and confirmed. One of the first successful uses of a bilateral filter in tone-mapping was presented by Pattanaik and Yee [PY02]. The filter uses a uniform spatial kernel and an exponential kernel in the intensity domain, and is inspired from the SUSAN filter [SB97]. However, the by far most common formulation of a bilateral filter uses Gaussian kernels [AW95,TM98]. It was introduced in tone-mapping by Durand and Dorsey [DD02], and has since been used frequently in the tone-mapping literature.

While the bilateral filter provides a simple formulation which achieves good results in many circumstances, one of its main disadvantages is the computational complexity. For this reason, a larger body of work has been devoted to acceleration of the bilateral and related filtering formulations, [DD02, CPD07, AGDL09, ABD10, BJ10,YBY10,BCCS12,Yan12]. With many such schemes, running on modern hardware, real-time performance is fully possible. Another problem with the bilateral filter is that the reconstruction is biased at steep but smooth edges in intensity, which can result in banding artifacts in the detail layer [DD02, BPD06, EMU15]. At the cost of increased computational complexity, these problems can

be alleviated. Examples include smoothing at the edges [DD02], manipulating the gradients of the reconstruction [BPD06], and incorporating the image gradients in a trilateral filtering procedure [CT03]. A large number of modern multi-purpose filtering techniques have also been demonstrated to achieve high-quality edge-aware filtering suitable for detail preservation in tone-mapping, [KKEA06, TFM07, FFLS08, Fat09, SSD09, KS10, PHK11, GO12, Mil13,HST13].

In addition to general purpose edge-aware filters, a number of filters constructed especially for the purpose of tone-mapping have been proposed. These include the early iterative methods [PFFG98, TT99, Ash02, RSSF02], as well more recent filters [ASC*14, EMU15]. Fore example, the fast detail extraction diffusion in [EMU15] is designed from an iterative simplification of the bilateral filter, in order to efficiently overcome the problems of biased reconstruction at steep but smooth edges.

Local processing in order to preserve details does not necessarily have to be performed directly from the (possibly transformed) luminance values. Oppenheim et al. make use of a spatially varying gamma function, which depends on the local frequency content of the image [OSS68]. Fattal et al. present a method for gradient domain tone-mapping [FLW02]. The gradients are scaled at different levels using a Gaussian pyramid, with a scale factor described as a function of the gradient magnitudes. The manipulations are followed by solving a Poisson equation to return to the intensity domain. A similar approach was proposed by Mantiuk et al. [MMS05]. However, instead of using the image gradients directly, the method operates on a perceptually motivated formulation of image contrasts.

## 4. Definitions and taxonomy

In this section, we carefully construct a general formulation of the central calculations involved in performing tone-mapping. We start with the tone-curve, which is the principal element of a TMO. Subsequently, we complement the formulation to allow for spatially varying local calculations. The tone-curve and local processing are the main building blocks in the pipeline of a local TMO. Referring to Figure 1, it is the two central parts of the pipeline. These should be complemented with pre- and post-processing steps in order to construct a complete TMO.

### 4.1. Tone-curve

At the core of a tone-mapping algorithm is the function for mapping from one domain to another. In most cases this can be described as a continuous function – the tone-curve. However, it is also possible to have an implicit mapping that is not possible to describe in such manner.

The end objective of a TMO is to map from the HDR input $L \in \mathbb{R}^+$ to display-referred pixel values $l \in \mathbb{N}_d = \{0, \ldots, 255\}$, using a transformation $M : L \to l$. $L$ could be in arbitrary units, but for most tone-mapping algorithms it is expected to be linearly related to the captured luminance values. However, different methods tackle the tone-mapping differently. While some TMOs tone-map directly to the pixel values $\{0, \ldots, 255\}$, others map to display luminance $T$

and need a display model to transform them to pixel values. In what follows we describe the tone-curve as $V : L \to T$ to an output $T \in \mathbb{R}^+$. This should be complemented with a final transformation to a display value $l$ (the final step in Figure 1).

The tone-curve function $V$ may depend on a set of user-specified constants $\boldsymbol{\alpha} = \{\alpha_1, ..., \alpha_{K_\alpha}\}$, and also a set of image statistics $\boldsymbol{\phi}(L) = \{\phi_1(L_1, ..., L_N), ..., \phi_{K_\phi}(L_1, ..., L_N)\}$. The constants $\boldsymbol{\alpha}$ can for example be overall brightness, contrast, saturation etc. The image statistics can include measures such as the image mean $\phi = \mu = \frac{1}{N} \sum_s L_s$, some percentile such as the median, maximum or minimum value (0.5, 1 and 0-percentiles, respectively), or the image histogram $\phi_i = p_L(L_i)$.

Using $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$, a general formulation of a spatially invariant mapping at a pixel position $s$ can be written as

$$T_s = V(L_s; \boldsymbol{\phi}(L); \boldsymbol{\alpha}). \tag{4}$$

For example, a simple exposure compensation can be described as $T_s = V(L_s; \alpha) = \alpha L_s$. Another example is the sigmoid curve, which can be formulated to account for the mean value of the input, $\phi = \mu$, in order to adapt to the overall luminance,

$$T_s = V(L_s; \phi(L); \alpha) = \frac{L_s^\alpha}{L_s^\alpha + \phi^\alpha}. \tag{5}$$

## 4.2. Local adaptation

The tone-curve in Equation 4 can be formulated to compress the dynamic range, e.g. as in Equation 5, but this comes at the cost of compressed overall contrasts of the input. While the global contrast inevitably has to be reduced to fit a lower dynamic range, a common objective is to preserve local contrasts – or details – in order to mimic the abilities of the HVS, see Section 2.4.

In order to better reproduce local contrasts we can introduce a set of parameters $\boldsymbol{\xi}_s(L)$ that in the same way as $\boldsymbol{\phi}(L)$ depend on the information in $L$, but weighted in a local neighborhood $\Omega_s$ around a pixel $s$,

$$\boldsymbol{\xi}_s(L) = \boldsymbol{\xi}(L_{\Omega_s}) = \{\xi_1(L_{\Omega_s}), \cdots, \xi_{K_\xi}(L_{\Omega_s})\}. \tag{6}$$

By using $\boldsymbol{\xi}_s(L)$ to capture local properties around each pixel $s$, the tone-mapping can adapt spatially in order to maintain or enhance image details. The *local* tone-curve now takes the form

$$T_s = V(L_s; \boldsymbol{\xi}_s(L); \boldsymbol{\phi}(L); \boldsymbol{\alpha}). \tag{7}$$

A trivial example of a local adaptation measure is the weighted local mean, computed with a Gaussian kernel $G_\sigma$, $\boldsymbol{\xi}_s(L) = (G_\sigma * L)_s$. The spatially varying statistics can be used directly in the formulation of a tone-curve, for example in Equation 5 to adapt locally in the image instead of using the mean over all image pixels, $V(L_s; \phi(L); \alpha) \to V(L_s; \xi_s(L); \alpha)$. However, given a filter $\xi_s(L)$

that estimates the global changes in the image, a common formulation is to split the image into a base layer $B_s = \xi_s(L)$ and a detail layer $D_s = L_s - B_s$. The tone-curve is then applied on the base layer, while the detail layer is added back after the mapping,

$$
\begin{aligned}
T_s &= V(L_s; \xi_s(L); \boldsymbol{\phi}(L); \boldsymbol{\alpha}) \\
&= \tilde{V}(B_s; \boldsymbol{\phi}(L); \boldsymbol{\alpha}) + D_s \\
&= \tilde{V}(\xi_s(L); \boldsymbol{\phi}(L); \boldsymbol{\alpha}) + L_s - \xi_s(L).
\end{aligned} \tag{8}
$$

Here, the local tone-curve $V$ has been decomposed into a global mapping $\tilde{V}$ of the global image contrasts $B$, and details $D$ which are preserved separate from the mapping. It should also be noted that these operations generally are performed in the log domain, but this is not explicitly stated here to simplify the notation.

Finally, we can also change the mapping itself over the image, $\boldsymbol{V} = \{V_1, ..., V_n\}$, and introduce a set of spatially varying parameters $\boldsymbol{\theta}_s = \{\theta_1, \cdots, \theta_{k_\theta}\}$ that do not depend on the image content. These parameters could for example be formulated in terms of the vertical and horizontal pixel positions in the image. We now arrive at the most general formulation of a tone-curve,

$$T_s = V_s(L_s; \boldsymbol{\xi}_s(L); \boldsymbol{\theta}_s; \boldsymbol{\phi}(L); \boldsymbol{\alpha}). \tag{9}$$

## 5. Video tone-mapping

In this section we describe some of the main differences in tone-mapping of HDR video, as compared to static HDR images. The section is concluded with guidelines for classification of video TMOs. These are based on many of the concepts we have discussed throughout this report, and they are used to categorize the video TMOs described in Section 6.

Tone-mapping of a sequence of frames from an HDR video differs in many aspects from processing single static images. The most apparent difference is due to the correlation between frames that need to be maintained. Without careful consideration of temporal coherence, flickering and other artifacts can be introduced, and even the slightest disruption may be well visible due to the sensitive mechanisms of the HVS that respond to temporal changes [dLD58, Gre69]. However, temporal coherence is not the only problem faced in video tone-mapping. Another important aspect is computational efficiency, since a large number of frames need processing. Moreover, when dealing with camera captured HDR images, the material itself significantly differs due to the different capturing processes of single image and video sequences. These differences require separate consideration in order to achieve high quality tone-mapping in all possible situations.

### 5.1. Background

Although HDR video capturing did not fully mature until the early 2010s, the first tone-mapping algorithms that considered adaptation over time were presented more than 15 years earlier [FPSG96]. However, in many of the early methods with an explicit temporal model, the main purpose of the considerations in the time domain was to have a complete model of the adaptation mecha-

nisms of the HVS [FPSG96, PTYG00, LSC04, IFM05], or to provide means for interactive walk-through in computer generated scenes [DD00]. The lack of high quality camera captured HDR video has made it difficult to thoroughly assess the performance of new video tone-mapping operators. In many cases the only available material for testing has been created from either CGI, panning in static HDR panoramas, or from capturing static scenes with alternating lighting. A few examples also include custom built techniques and systems to record HDR video [KUWS03, WRA05]. However, first with the advent of dedicated HDR video cameras [TKTS11, KGB*14] it was possible to thoroughly test the existing algorithms [EWMU13]. From the survey and evaluation in our previous study [EWMU13], it was discovered that many of the assessed operators either suffer from artifacts, or lack in contrast and detail reproduction. Responding to the introduction of new HDR video cameras, during the 2010s research in video tone-mapping has been more focused to address the specific problems in tone-mapping of such material. The main objective is to achieve high quality tone-mapping despite the particular challenges the material may present [BBC*12, BCTB14b, ASC*14, EMU15, LSRJ16].

## 5.2. Temporal coherence

A categorization of possible temporal artifacts in tone-mapping was presented by Boitard et al. [BCTB14a]. This includes artifacts such as flickering, temporal contrast and object incoherence. We emphasize the global and local flickering and ghosting artifacts, as these are arguably the most prominent coherence problems. The other artifacts, concerning temporal contrast and object incoherence are more natural for a human observer; we adapt to the current lighting situation and thus accept changes in contrast and object brightness, as long as the changes are smooth. For a thorough treatment of the different incoherence definitions, we refer to the work by Boitard et al. [BBC*12, Boi14, BCTB14a] [DCMM16, Ch. 6].

If the mapping $V : L \rightarrow T$ in any way depends on the image content, $V = V(L_s; \boldsymbol{\phi}(L))$, it may potentially give rise to temporal artifacts. Many of the measures or image statistics $\boldsymbol{\phi}(L)$ used for the content-aware formulation of a tone-curve can potentially change rapidly from frame to frame, causing non-smooth changes of the tone-curve. The changes are perceived as flickering artifacts in the final tone-mapping.

In the case of global processing, using the same tone-curve for all pixels (Equation 4), smooth changes between frames can be achieved from a low-pass filtering of the image statistics over time using some kernel $G_\sigma$,

$$\tilde{\phi}(L_t) = \sum_{r=t-d}^{t+d} G_\sigma(\phi(L_r)). \tag{10}$$

However, this filter response is affected by future frames, which can yield inconsistencies in the timeline. More precisely, the tone-mapping can start to respond to a change in lighting, before it actually occurs. A more sensible choice is to only filter over past frames, e.g. using an exponentially decaying kernel, representing a leaky integrator. This can be formulated in its simplest form by blending between the the current and the previous samples in time, using a blend value $\alpha$,

$$\tilde{\phi}(L_t) = \alpha\phi(L_t) + (1-\alpha)\tilde{\phi}(L_{t-1}). \tag{11}$$

For local tone-mapping operators, Equation 8, the situation is more complicated. If the local processing, using some local image description $\boldsymbol{\xi}_s(L)$, is not carefully formulated with robust statistics it can give rise to quick changes locally in the image. Straightforward use of the temporal filtering procedures in Equation 10 and 11 on a per-pixel level will cause ghosting artifacts. An alternative is to low-pass filter the final tone-mapped values, $\tilde{T}_s = (G_\sigma * l)_s$, which can help on both temporal coherency and noise level at the cost of introducing some motion blur [vH06, BAHC09].

In order to perform per-pixel filtering over time without the side effects of motion blur, the local filter kernels can be restricted at large temporal gradients, using an edge-stopping filter [BM05], see e.g. Equation 17. However, this comes at the cost of limited or no filtering close to larger temporal gradients.

A natural extension to per-pixel filtering over time is to use a motion compensated temporal kernel, effectively filtering along motion paths in the time domain. If the motion estimation is correct, this makes it possible to filter in the time domain both in uniform regions and close to temporal edges and to do so without introducing motion blur. However, the motion field is not defined in all situations, e.g. for areas that are occluded between frames. Also, in other situations the estimation may fail. These problems leads either to artifacts, or the filtering have to be restricted in terms of the support in the temporal domain. Examples of motion compensated temporal filtering for tone-mapping include the method by Lee and Kim [LK07], utilizing block matching for acquiring motion vectors, and the framework described by Aydin et al. [ASC*14], which filters using an optical flow estimation. We refer to Section 6 for more information, and in particular Equation 18 and 21.

Using the above mentioned variants of temporal filtering, existing tone-mapping operators for static images can potentially be extended to work for HDR video. This is also a common approach when inspecting the literature, where a range of different static operators have been used. One of the most frequently used is the *Photographic TMO* [RSSF02], which has been complemented with temporal processing in many variations [GWWH03, RJIH04, KRTT12, BBC*12, BCTB14b].

Another approach for maintaining coherence over time is to consider temporal aspects a post-processing step, applicable to any static tone-mapping method. This means that the tone-mapping process is viewed as a black box, and the objective is to impose temporal coherence by some global or local pixel manipulations. To alleviate global flickering artifacts, one possibility is to analyze the mean value of each frame and constrain the changes over time [GKEE11], or compare it to the original HDR sequence [BBC*12]. However, such methods only work on a global level, by applying a separate scaling to each frame. The assumption is that all temporal incoherencies can be attributed to only changes in global image brightness. A more difficult situation would e.g. be temporally incoherent changes in contrasts, and local incoherencies.

A few methods for post-processing to achieve temporal coherence on a local level has also been presented [LWA*12, DBZY15, BTS*15], mostly relying on per-pixel motion fields from optical flow estimations. These are formulated for processing of a range of different video operations, and may be a good alternative for alleviating problems with coherence when the specific operation is unknown. However, as a viable solution for video tone-mapping it is rather counter-intuitive to generate temporally inchoherent results and then try to overcome these problems in a post-processing step. It will always be more efficient to maintain local coherence throughout the tone-mapping pipeline.

### 5.3. Noise visibility

One of the more serious implications of the HDR video capturing process is an increased amount of camera noise, especially in darker scenes. In order to capture dynamic scenes, the exposure time is limited. Thus, in order to incorporate shadows and dark areas in the range of luminances captured, the sensor gain needs to be set high, resulting in increased sensor noise. Although the noise may be invisible in the original video sequence, it can easily be revealed by the tone-mapping process, as demonstrated in Figure 5. This can be understood from the non-linearity that most tone-curves experience, which downscales the brightness of bright pixels while doing the opposite for dark pixels, resulting in an increased relative level of noise in dark areas.

A denoising algorithm can be applied in order to alleviate the problem with image noise. Some noise reduction also comes as a positive side-effect if a per-pixel filtering procedure is used for imposing temporal coherence in the tone-mapping, [BM05, vH06, BAHC09, ASC*14]. However, this comes at increased computational cost, and it is difficult to remove all noise without introducing artifacts. Also, if the tone-curve assigns a substantial increase in brightness of the dark pixels, some remaining noise or denoising artifacts will inevitably be revealed.

In order to make sure that the image noise is not made visible from the tone-mapping, the shape of the tone-curve can be explicitly controlled based on knowledge of the noise characteristics [EMU15, LSRJ16]. In this way the noise is concealed within the darker regions of the final tone-mapping. This approach can also be combined with denoising algorithms, in order to be able to allow for larger manipulations of darker image tones.

### 5.4. Categorization of video TMOs

We use the concepts described in the preceding sections to formulate a general categorization of video TMOs. This includes the type of tone-curve, processing, intent and temporal filter. Additionally, we complement with a categorization in terms of computational complexity. This is motivated by the increased amount of computations involved as compared to an image operation. Since a video sequence normally contains about 25-30 frames per second, interactive or real-time performance makes a large difference in terms of turnaround-time when tweaking parameters of the tone-mapping.
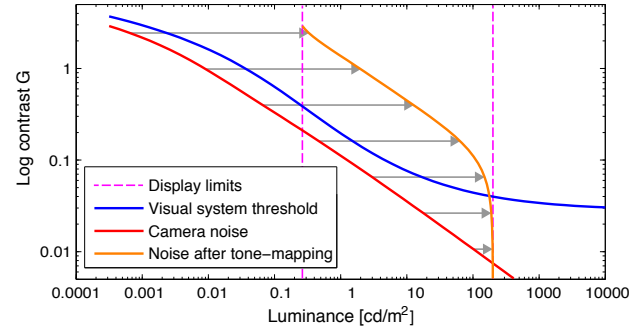
The categories are determined as follows:



Figure 5: Noise that is invisible to the human eye in the HDR input image, and thus below the threshold of the contrast visibility of the HVS, may be pushed above the threshold in the process of tone-mapping.

1. **Tone-curve**: The underlying transformation function for performing the tone-mapping. The reviewed TMOs in Section 6 use one of the following functions: *linear*, *logarithmic/exponential*, *sigmoid/s-shaped* or *histogram based*. Additionally, some operators perform the mapping in the *gradient domain*.

2. **Processing**: A *global* TMO uses the same tone-curve for all pixels of an image, while a *local* varies the function spatially.

3. **Intent**: The intent of the tone-mapping is according to: *visual system simulators* (VSS), *best subjective quality* (BSQ) operators and *scene reproduction* (SRP) operators.

4. **Temporal filter**: The temporal adaptation filter is either applied on a *global* level by filtering some of the tone-mapping parameters over time, or it is calculated per-pixel on a *local* level.

5. **Speed**: The computational speed of the TMO is capable of generating either *offline*, *interactive* or *real-time* performance. These have been roughly estimated from the complexity of the calculations performed by the specific TMO. Real-time performance is here defined as being above 20 frames/second, while interactive is when the TMO roughly is able to process more than a few frames/second. However, the frame-rate is highly dependent on the image resolution and the hardware used, so the categorization is approximate.

### 6. Video tone-mapping algorithms

In this section we describe existing tone-mapping operators for HDR video. All of the operators are listed in Table 1, using the classification in Section 5.4. Furthermore, for a subset of the operators the table also includes estimated performance in terms of four different criteria. These have been measured as described in Section 7.

In our review of the video tone-mapping literature, we found the following tone-mapping operators that have an exclusive component for how to treat temporal aspects:

***Visual adaptation TMO***, Ferwerda et al. [FPSG96]: As the first example of a tone-mapping operator with a time adaptation mechanism, Ferwerda et al. use data from a series of psycho-physical experiments to build a visual response model. The method is based on a *threshold-versus-intensity* (tvi) function, which is derived from

an experiment on detection thresholds over the full range of vision (scotopic and photopic). This is complemented with previously published data on changes in color appearance (spectral sensitivities), changes in visual acuity (ability to resolve spatial detail) over luminances, and adaptation over time (detection threshold over time during light and dark adaptation). A computational model is then built on the experimental data, in order to predict the behavior of the HVS in terms of visibility, color appearance and clarity of scene features. The prediction is made at a given background luminance and due to adaptation over time. The model uses the concept by Ward [War94a], where just notable differences (JNDs) are matched before and after tone-mapping. However, the matching is performed separately for rods and cones, and additional functionalities are added to account for visual acuity and adaptation over time.

***Interactive walk-through TMO***, Durand and Dorsey [DD00]: Durand design a method for interactive tone-mapping of computer generated walk-throughs. The TMO thus focuses on achieving interactivity. It uses a two-pass method in order to suit an online rendering engine. In the first step a down-sampled, low resolution, rendering is performed in the log domain to determine an adaptation level. This is used in the second step to perform rendering and appropriate tone-mapping. The operator is based on the work by Ferwerda et al. [FPSG96], but with extensions such as blue-shift for viewing night scenes and chromatic adaptation. The light adaptation is performed using an exponential filter based on data from Adelson [Ade82] and Hayhoe et al. [HBH87]. The chromatic adaptation is modeled by filtering per-channel adaptation levels over time. Finally, flares are simulated, and loss of acuity is modeled in the same manner as in the *Visual adaptation TMO* [FPSG96].

***Time-adaptation TMO***, Pattanaik et al. [PTYG00]: Pattanaik et al. present a model for efficient modeling of the time-adaptation mechanisms of the HVS. It is based on data from the psycho-physical measurements by Hunt [Hun95]. Using an adaptation model followed by an appearance model, the response of an image is simulated. Subsequently these are followed by the inverse appearance and adaptation models, in order to map the response to the display device. In the adaptation model, static responses are computed separately for cones and rods using sigmoid functions (Equation 5). Time adaptation is then modeled using exponential smoothing filters, with a decay fitted to data from Crawford [Cra37]. The data roughly agrees with the measurements by Adelson [Ade82] and Hayhoe et al. [HBH87]. Additionally, time-dependent pigment bleaching factors are included. In total four filters are used, for neural and photo-chemical mechanisms in both the rod and cone systems.

***Temporal exposure TMO***, Kang et al. [KUWS03]: Kang et al. propose a technique for capturing HDR video by alternating the exposure time temporally, from frame to frame. In order to reconstruct an HDR video, the frames are interpolated using optical flow, to provide two different exposures in each frame. For display of the reconstructed video a temporal filtering procedure is added to the *Photographic TMO* [RSSF02]. The tone-curve of the *Photo-*

*graphic TMO* is formulated using a sigmoid mapping on the following form:

$$T = \frac{L_s}{1 + L_s} \left( 1 + \frac{L_s}{L_{white}^2} \right), \qquad (12)$$

where $L_s = aL/\kappa$, with $\kappa$ being the geometric mean. The mean can potentially change quickly from frame to frame, causing flickering artifacts. To alleviate such problems, $\kappa$ is evaluated over a fixed number of frames, effectively performing a uniform filtering over time,

$$\kappa = \exp\left( \frac{1}{N} \sum_{s,t} \log(\varepsilon + L) \right). \qquad (13)$$

Here, $s$ runs over all spatial pixel positions, and $t$ goes over a set of frames. $\varepsilon$ is a small constant to avoid singular values.

***Time-dependent GPU TMO***, Goodnight et al. [GWWH03]: Addressing the problem of high computational complexity associated with local tone-mapping operators, Goodnight et al. describe a GPU acceleration of the *Photographic TMO* [RSSF02]. The method is able to achieve local tone-mapping at interactive framerates using the available hardware in the early 2000s, which means that today real-time performance should be no problem. As for the temporal behavior of the tone-mapping, a similar time-adaptation model is used as in the *Interactive walk-through TMO* [DD00], with an exponential low-pass filter.

***Adaptive temporal TMO***, Ramsey et al. [RJIH04]: Ramsey et al. also present a method to complement the *Photographic TMO* [RSSF02] with temporal adaptation. In the same way as in the *Temporal exposure TMO* [KUWS03], the geometric mean is calculated over time according to Equation 13. However, a dynamic range of frames are used, so that only the neighboring frames that have a similar mean value as the current frame are used. Effectively, this makes for an edge preserving temporal filtering that allows for fast adaptation when the input changes quickly.

***Local adaptation TMO***, Ledda et al. [LSC04]: Similar to the *Time-adaptation TMO* [PTYG00], Ledda et al. use retinal responses derived from experimental data. These are computed separately for the cones and rods using a sigmoid curve (Equation 5). However, in order to derive local responses, this is done separately for each pixel. To this end, a bilateral filter is used for deriving local adaptation levels which are used in the response functions. In order to model the time-course of local adaptation, the adaptation levels are filtered over time using exponential filters. Different filters are used for cones and rods, and for dark and light adaptation.

The operator was evaluated for changes in lighting in static scenes, but for dynamic scenes the local temporal filtering of adaptation levels often results in ghosting artifacts.

***Mal-adaptation TMO***, Irawan et al. [IFM05]: Irawan et al. present a VSS model for tone-mapping of HDR video which includes the state of mal-adaptation of the HVS. It attempts to match the scene visibility, while adapting over time using a model
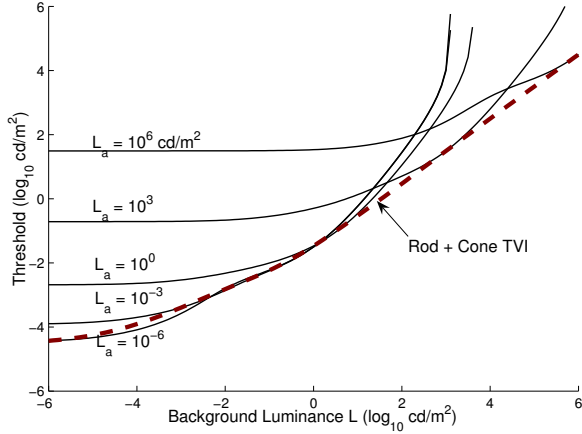
Figure 6: Threshold versus intensity and adaptation (*tvia*) function used in the *Mal-adaptation TMO*, [IFM05], for 5 different adaptation levels $L_a$. The *tvi* (dashed curve) approximately corresponds to the lower envelope of the *tvia* curves. (Image courtesy of Irawan et al. [IFM05]).

that accounts for both neural mechanisms and photo-pigment bleaching/regeneration. Furthermore, as it is able to model mal-adaptation, the TMO does not assume that the HVS is perfectly adapted to the current luminance level.

The visibility matching was taken from the method by Ward et al. [WLRP97]. It uses histogram adjustment to perform tone reproduction, but restricts the adjustments to the limitations in sensitivity of the eye with the help of a function for detection thresholds over luminances (*tvi*). Irawan et al. complement this tone-curve with the model of temporal adaptation from the *Time-adaptation TMO* [PTYG00]. Furthermore, they extend the threshold model to use a *tvia* (threshold versus intensity and adaptation) function, in order to model mal-adaptation. The *tvia* is derived numerically, from requiring that a change in luminance by the JND $\Delta L$, should yield a change $\Delta R$ in the response function. That is,

$$R(L + \Delta L(L, \sigma(L)), \sigma(L)) - R(L, \sigma(L)) = \Delta R. \quad (14)$$

The response curve $R$ is formulated with a sigmoid function, and $\Delta L(L, \sigma(L_a))$ is the *tvia* at adaptation level $L_a$. Some of the threshold curves of a mal-adapted HVS are shown in Figure 6. These were also confirmed to be inline with the probe-on-flash experiments by Walraven et al. [WECH*90] that measure threshold visibility of a mal-adapted HVS. The method also works as a low vision simulator, demonstrating the image as perceived by an HVS with visual impairments.

***Perceptual effects TMO***, Krawczyk et al. [KMS05b]: Similar to the *Time-dependent GPU TMO* [GWWH03], Krawczyk et al. construct a video TMO that is based on the *Photographic TMO* [RSSF02]. One of the main objectives is also to achieve local tone-mapping with real-time performance. However, the method augments the *Photographic TMO* with a set of perceptual effects. As the *Photographic TMO* requires building a Gaussian pyramid for deriving

local adaptation levels in the tone-mapping, additional effects that are based on this construction can be added with little computational overhead. Taking advantage of this observation, visual acuity and veiling luminance are included in the framework. Furthermore, the loss of color perception in the range of scotopic vision is also modeled.

Temporal adaptation is performed using an exponentially decaying function for filtering the adaptation level over time, in the same way as in the *Interactive walk-through TMO* [DD00]. However, instead of performing separate computations for the rods and cones, one single time-constant is used. This is interpolated from time-constants of the cones and rods based on the sensitivity of the rods. Finally, to achieve real-time performance the operator is implemented in graphics hardware.

***Gradient domain TMO***, Wang et al. [WRA05]: Wang et al. use a capturing device termed a *split aperture camera*. This is a rig with 3 CCD sensor cameras that simultanously can capture different ranges of luminances from the same scene. The radiance map for each frame is recovered with standard techniques, and the purpose is to use the resulting HDR video to exploit both spatial and temporal domains in the tone-mapping.

For tone-mapping, the gradient domain method of Fattal et al. is used [FLW02]. The method attenuates the image gradients $\nabla H_s^k$ at different scales $k$ at pixel position $s$, according to the scale factor $\varphi_s^k$,

$$\varphi_s^k = \frac{\alpha}{||\nabla H_s^k||} \left( \frac{||\nabla H_s^k||}{\alpha} \right)^\beta. \quad (15)$$

The modified gradient field $G_s = \Phi_s \nabla H_s$, where $\Phi$ is the combined attenuation factor over all $k$, is then used to solve for the tone-mapped values $T$ according to the Poisson equation,

$$\nabla^2 T = \text{div } G. \quad (16)$$

The difference compared to the original method though, is that the gradient field is in three dimensions, extending both spatially and over time. Solving the Poisson equation in 3D results in a consistent result over time in terms of scaling and shift, as opposed to a per-frame solution. For efficiency, the 3D Poisson equation is solved using a diagonal multi-grid technique. However, for the tone-mapping only the spatial gradients are attenuated since a temporal attenuation can cause motion blur.

***Virtual exposures TMO***, Bennett and McMillan [BM05]: The *virtual exposure camera* is designed as a filtering approach for video denoising. It performs a bilateral filtering both spatially and over time in order to reduce image noise. With this filter, Bennett and McMillan propose to integrate a video tone-mapping operator that uses the spatial bilateral filter for detail preservation [DD02]. To this end, each frame is decomposed into base and details layers. The base layer is tone-mapped using a tone-curve that draws inspirations from the method by Drago et al. [DMAC03], performing a logarithmic mapping. Subsequently the preserved details are added

back (see Equation 8). However, in order to reduce noise and help on increasing temporal coherency, a separate bilateral filter is employed in the time domain. For a pixel $s$ in frame $t$, the temporal filtering can be formulated as follows, where the filter support is $2d+1$ frames wide in the time domain:

$$\tilde{L}_{s,t} = \frac{1}{\bar{G}} \sum_{r=t-d}^{t+d} G_{\sigma_a} \left( ||L_{s,t} - L_{s,r}|| \right) G_{\sigma_b} \left( |t-r| \right) L_{s,r} . \qquad (17)$$

$G_\sigma$ denotes a Gaussian kernel, and $\bar{G}$ is a normalization factor of the bilateral kernels. The per-pixel temporal filtering stops at strong edges over time, in order to avoid problems with motion blur.

***Cone model TMO***, van Hateren [vH06]: As opposed to the previously published VSS tone-mapping methods, which use experimental data to model the HVS, van Hateren presents an approach that uses actual quantitative measurements. The operator is based on a system that dynamically models primate cones and horizontal cells [vH05]. This system is built upon quantitative measurements performed on cells in the macaque retina [SPLD01, LDSP03]. As the macaque retina operates in a nearly identical way as the human counterpart, the model is assumed to work as a good representation of the functionality of the human cones. Since the measurements are from the cones only, photopic vision is assumed and the influence of the rods is disregarded. Furthermore, a constant pupil size is assumed. In practice, the tone-mapping is perform with a sigmoid tone-curve (Equation 5).

The operator describes a dynamic response, modeling the cones over time. Motivated by the neural mechanisms, and especially integration of the cones over time, a temporal per-pixel low-pass filter is utilized. This also alleviates problems with image noise, at the cost of introducing motion blur for moving objects.

***Block matching TMO***, Lee and Kim [LK07]: Lee and Kim present a method for video tone-mapping based on the same idea as in the *Gradient domain TMO* [WRA05]. That is, the tone-mapping is performed in the gradient domain, with gradients extending both spatially and over time. The gradients are attenuated according to the gradient domain compression by Fattal et al. [FLW02]. However, as opposed to Wang et al., who only attenuate spatial gradients, Lee and Kim also modify the temporal gradients. In order to do so without introducing motion-blur, a block matching algorithm is used to allow for motion compensation. Given the estimated motion $\mathbf{v}_{s,t}$ of a pixel $s$, from frame $t-1$ to frame $t$, a cost function can be formulated as

$$C = \sum_s \left( L_{s,t} - L_{s-\mathbf{v}_{s,t},t-1} \right)^2 . \qquad (18)$$

The cost increases with dissimilar values of the pixels in two frames that should correspond to the same physical object point, which means that the motion estimation is potentially unreliable. The cost $C$ is then used in the Poisson equation to smooth gradients along the motion paths of the scene points. The result is an enforced temporal coherency and noise reduction. That is, not only is the method designed to not introduce any temporal artifacts, but
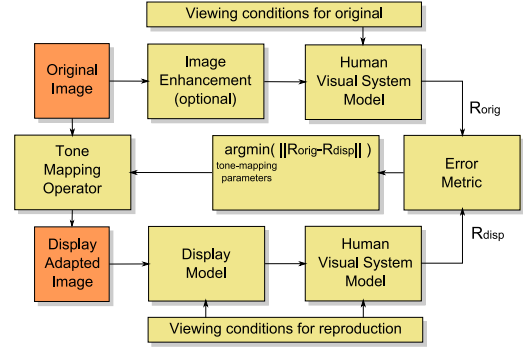


Figure 7: The formulation of tone-mapping as an optimization problem in the *Display adaptive TMO* [MDK08]. Image enhancements of the input HDR are also included, which could optionally be used in order to achieve sharper and more colorful images. (Image courtesy of Mantiuk et al. [MDK08]).

temporal flickering or incoherence problems of the input sequence are also alleviated.

***Display adaptive TMO***, Mantiuk et al. [MDK08]: Mantiuk et al. formulate tone-mapping as an optimization problem, for maximizing the similarity between the input HDR image and the tone-mapped image as viewed on a specific display device, see Figure 7. In order to estimate the luminance emitted from the display, the model in Equation 1 is used. With this information, a non-linear optimization problem is formulated. The objective is to minimize the squared difference between the contrast response of the original image and that after tone-mapping. That is, for the optimization the difference between the displayed image $L_d$ and the HDR input is measured in terms of their estimated contrast responses. These are computed from an HVS contrast perception model. However, rather than minimizing the per-pixel differences, the optimization uses the conditional probability density function of contrasts, estimated from histograms at different levels in a Laplacian pyramid. The optimization is subject to the tone-curve being an increasing function that maps luminances to the available dynamic range of the display. Finally, the problem is solved iteratively as a quadratic programming problem.

As for temporal adaptation, the nodes of the derived tone-curves are filtered over time using a low-pass filter.

***Retina model TMO***, Benoit et al. [BAHC09]: Benoit et al. design a VSS operator that models the early stages of the HVS. It describes operations in the cellular layers which process the response from the entry point at the photoreceptors and down to the ganglion cells output, complemented with spatio-temporal filters of the retina. The operator is developed from the models by Hérault and Durette [HD07] and Meylan et al. [MAS07]. These account for low level processes in the retina, including local adaptation properties.

First, local responses of the photo-receptors are modeled with sigmoid functions, where local adaptation levels come from spatial low-pass filtering of the image. The responses are then passed through a spatio-temporal filter, simulating the processing occurring in the Outer Plexiform Layer (OPL) [HD07]. This OPL fil-

ter can be considered as a difference between two low-pass spatio-temporal filters, and yields whitening of the image frequency spectrum. That is, it moves the spectrum of low frequencies towards higher frequencies and vice versa. The filter is argued to benefit noise reduction by low-pass filtering high frequencies, and to yield better temporal stability through filtering of the local adaptation levels over time. As a final step of the tone-mapping pipeline, the outputs from the OPL spatio-temporal filtering are fed to the ganglion cells response model. Similar to the initial photo-receptors, this is modeled using sigmoid functions, but utilizing different parameters.

As an additional operation, colors are processed by performing an initial spatial multiplexing – or mosaicing – of color images before the tone-mapping process. As a final step, colors are restored from a de-multiplexing, or de-mosaicing, step. The objective is to achieve a color constancy effect, motivated by the separate processing of the chromatic cone responses (L, M and S cones) in the retina [MAS07].

***Local mal-adaptation TMO***, Pajak et al. [PCA*10]: While the *Mal-adaptation TMO* simulates the effect of mal-adaptation in the HVS, this is done on a global level [IFM05]. The *Local adaptation TMO* on the other hand, models local adaptation levels but not the effect of mal-adaptation [LSC04]. With the *Local mal-adaptation TMO*, Pajak et al. present a model that accounts for spatially varying mal-adaptation. The model includes supra-threshold mechanisms such as visual masking. Furthermore, luminance adaptation and contrast sensitivity are derived as a function of spatial frequency. The spatial frequency sensitivity model uses the contrast sensitivity function (CSF) from the visible differences predictor (VDP) by Daly [Dal92]. Simulation of the time-course of adaptation is done similarly as in the *Mal-adaptation TMO*, but on a per-pixel level utilizing exponential filters. The simulation can be evaluated efficiently, as demonstrated with the use of a real-time GPU implementation.

***Flicker reduction TMO***, Guthier et al. [GKEE11]: Guthier et al. focus on reducing flickering artifacts in already tone-mapped video material, employing a post-processing procedure. In a tone-mapped sequence, flickering is detected when the difference between key values $\kappa_t$ (Equation 13) is above a certain threshold $H$ for two consecutive frames, $|\kappa_t - \kappa_{t-1}| > H$. The threshold is formulated from Stevens' power law [Ste62], $H = \Delta L = kL^{1/3}$, where $k$ is determined experimentally. For frames that are above the threshold, the brightness is changed iteratively until the key value falls into the range of acceptable brightness change, $\kappa_t \in [\kappa_{t-1} - H, \kappa_{t-1} + H]$. The method is applicable to any TMO which only suffers from global brightness incoherence, but is demonstrated on three previously published operators by Ward et al. [War94a, WLRP97] and Reinhard et al. [RSSF02].

***Real-time automatic TMO***, Kiser et al. [KRTT12]: Recognizing the inherent problems in tone-mapping of camera captured HDR video, Kiser et al. suggest two modifications of the *Photographic TMO* [RSSF02] with automated parameter estimation [Rei02]. First, based on the image histogram the black and white levels of the input image are clamped before the tone-mapping. This is in order to better utilize the available dynamic range of the display device, which otherwise potentially could be affected by extreme

values. Second, a leaky integrator is used for filtering parameters over time, which corresponds to an exponential averaging low-pass filter similar to the one used in the *Interactive walk-through TMO* [DD00]. The resulting operator allows for efficient implementation with real-time performance and automatically derived parameters.

***Color appearance TMO***, Reinhard et al. [RPK*12]: Reinhard et al. take inspiration from both color appearance modeling, tone-mapping and lightness perception, in order to construct an appearance calibration algorithm. The goal is to match image appearance in terms of photo-receptor responses, including local aspects of lightness perception. The matching is done with respect to room illumination as well as the display characteristics.

The appearance reproduction is performed from parameters describing the adapting luminance, maximum luminance, adapting white point, maximum white point and the degree of adaptation. These 5 parameters are kept separate for scene, display and viewing environment. The scene parameters are estimated from the input, while the others are tuned to specify the display environment for which to calibrate the input. Furthermore, the scene parameter estimation is localized to image regions calculated from the median cut algorithm [Deb08].

The final tone-mapping algorithm is formulated from a photo-receptor sigmoid response model, using the different environment parameters. For temporal stability a leaky integration is performed of the adaptation parameters, effectively performing an exponential filtering over time. Finally, the appearance reproduction is validated against a set of color datasets.

***Temporal coherence TMO***, Boitard et al. [BBC*12]: Apart from temporal flickering, Boitard et al. identify two additional potential problems in video tone-mapping. The first is termed *object inchoherency*, and concerns changes in the luminance of an object in the tone-mapped sequence; changes that do not correspond to the original HDR sequence, but occur due to varying image statistics. The second is called *brightness inchoherency*, and describes changes in overall brightness over time; the brightest frame in the HDR sequence do not always yield the brightest tone-mapped frame. In order to enforce a tone-mapping that preserves these properties, Boitard et al. propose to consider the entire HDR and tone-mapped sequences in a post-processing step, rather than operating locally in the time domain. For each frame $t$ the key values $\kappa_t^L$ and $\kappa_t^T$ are calculated (Equation 13), corresponding to the HDR and tone-mapped sequences, respectively. The key value of each frame of the tone-mapped sequence are then anchored to the maximum key value of the entire input HDR sequence, using a global scaling $\tau$ of the frames, $\tilde{T} = \tau T$, according to

$$\tau = \frac{\kappa_t^L \, \kappa_{max}^T}{\kappa_{max}^L \, \kappa_t^T} \, . \tag{19}$$

A blending value is also used on the scaling, in order to relax this anchoring and allowing smaller deviations from the input maximum key value,

$$\tau = \tau_{min} + (1 - \tau_{min})\tau . \tag{20}$$

The post-processing means that the algorithm is applicable to any TMO that only suffers from global incoherencies. However, Boitard et al. uses the tone-curve from the *Photographic TMO* [RSSF02], which allows for a slightly simplified formulation of the scaling. The post-processing adjustment using a brightness scaling is similar to the *Flicker reduction TMO* [GKEE11]. However, it considers the entire sequence in order to enforce object and brightness coherency, and does so by comparing to the input HDR sequence. This comes at the cost of reduced overall brightness over the video sequence, since most frames will be down-scaled in order to maintain a constant relative brightness as compared to the brightest input frame.

**Zonal coherence TMO**, Boitard et al. [BCTB14b]: In order to improve the results of the *Temporal coherence TMO*, in the *Zonal coherence TMO* Boitard et al. augment the method with additional features to perform the processing on separate segments in the luminance domain. Since preserving overall temporal brightness coherency results in reduced global brightness and contrast in all but the brightest frame, the objective of the new operator is to do this locally. And since a spatial segmentation would be difficult, especially to maintain coherent over the sequence, this is done in the luminance domain.

To segment into different luminance zones, for each frame the image histogram in the log domain is used. First the local maxima of the histogram is detected, and the local minima in between are defined to be the zone boundaries. In each of these zones key values are calculated separately. However, these zonal key values may change rapidly between frames, potentially causing flickering artifacts. To mitigate this problem, zones are calculated so that the boundaries are consistent throughout the sequence. This is done by calculating a sequence histogram from all the zonal key values in all frames, and once again perform the segmentation, arriving at a new set of zones. Finally, using the key values of these zones, the scaling adjustment from the *Temporal coherence TMO* (Equation 19) is performed separately for each of the luminance zones. To avoid problems at zone boundaries, the scaling ratios $\tau$ are blended between adjacent zones when close to the boundaries.

**Motion path filtering TMO**, Aydin et al. [ASC*14]: The approach taken by Aydin et al. is motivated by the problems common to video tone-mapping of camera captured material [EWMU13], and specifically the problem of reproducing local contrast while not introducing local temporal incoherence. In order to locally tone-map a sequence while enforcing local temporal coherency, the method is designed as a spatio-temporal filtering framework. The foundation of the framework is per-pixel motion estimation from an optical flow computation, that gives motion vectors $\boldsymbol{v}_{s,t \leftrightarrow r}$ for each pixel $s$, between frames $t$ and $r$. Using this vector field, a motion-compensated filtering over time can be performed for each pixel, over a $2d + 1$ window in time, $r \in [t-d, t+d]$, according to

$$\tilde{L}_{s,t} = \frac{1}{\bar{H}} \sum_{r=t-d}^{t+d} H\left(||L_{s,t} - L_{s-\boldsymbol{v}_{s,t \leftrightarrow r},r}||\right) L_{s-\boldsymbol{v}_{s,t \leftrightarrow r},r}. \quad (21)$$

For $t = r$ the vector field is $\boldsymbol{v}_{s,t \leftrightarrow t} = 0$. The kernel $H$, used for filtering along the motion paths, is described from a Lorentzian edge-stopping function. It uses the difference between the current pixel
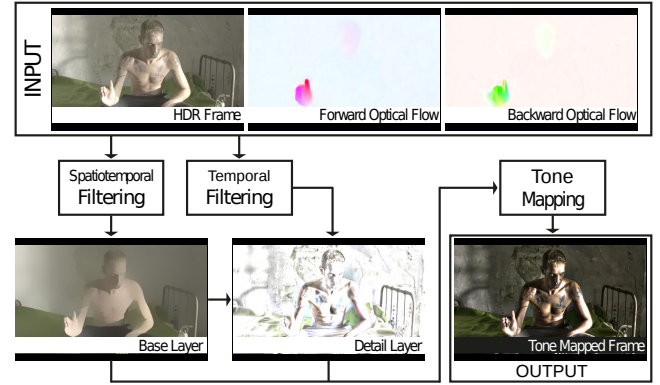


**Figure 8:** The pipeline of the *Motion path filtering TMO* [ASC*14]. Spatial and temporal filterings are used to decompose and tone-map the input sequence. (Image courtesy of Aydin et al. [ASC*14]).

in frame $t$ and the motion-compensated pixel from frame $r$, as a stopping criteria. Its purpose is to limit the filtering at strong temporal edges or were the optical flow fails.

As illustrated in Figure 8, the tone-mapping is performed by separating an input frame into a base and a detail layer, according to Equation 8. However, both the base layer and the original frame are filtered temporally using Equation 21, before extracting the detail layer. The edge-preserving filter used for estimating the base layer is formulated as an iterative smoothing procedure. Furthermore, compression of the base layer is done either by scaling in the log domain, or utilizing the tone-curve by Drago et al. [DMAC03].

The filtering procedure attempts at enforcing local temporal coherence of the tone-mapping, including temporal artifacts already present in the input sequence. A positive side effect is also a reduced level of camera noise.

**Hybrid TMO**, Shahid et al. [SLF*15]: Shahid et al. propose the use of a hybrid tone-mapping scheme, which utilizes two different TMOs in order to tone-map the same frame in an HDR sequence. The image is segmented into regions containing high-intensity changes, using a directional filter bank edge-detection algorithm [ATJ12]. These regions are subsequently tone-mapped using the iCAM06 TMO [KJF07], which is able to render well-preserved details due to local processing. In the remaining regions of the frame, the histogram adjustment from Ward et al. [WLRP97] is used. In order to handle temporal aspects, the tone-mapped frames are scaled to enforce the difference in average intensity between consecutive frames not to deviate above a predefined threshold, similar to the technique in the *Flicker reduction TMO* [GKEE11].

**Noise-aware TMO**, Eilertsen et al. [EMU15]: With a starting point in the authors' earlier survey and evaluation [EWMU13], Eilertsen et al. present a set of techniques for addressing the problems that it indicated as commonly appearing when tone-mapping camera captured HDR video material. The aim is to arrive at an efficient and temporally stable video tone-mapping method, that could handle large compression of dynamic range while maintaining local contrast and without increasing the visibility of camera noise.

For detail preservation, Equation 8 is used with a filter specif-

ically designed for the objective of tone-mapping. It is motivated from recognizing that standard edge-preserving filters, designed for purposes such as noise reduction, can reveal artifacts in some situations. The problem of detail extraction for tone-mapping is conceptually different, requiring a filter response at a large scale, and where the filtered output is not the final result. In order to efficiently extract details without artifacts, a simplification from a unified formulation of bilateral filtering and anisotropic diffusion is suggested, that can not show the type of reconstruction problems associated with anisotropic 0-order filters.

For the tone-mapping, a tone-curve is derived for preserving the contrasts of the original sequence as close as possible. This is formulated as an optimization problem, with the objective of minimizing the distortions of contrasts under the mapping of luminances. For an input contrast $G$ and the corresponding contrast $\tilde{G}$ after tone-mapping, the expected value of the squared distortions $(G - \tilde{G})^2$ is expressed from the distribution of contrasts $p(G|L)$ over log luminances $L$, as

$$E[||G - \tilde{G}||_2^2] = \int p(L) \int (G - \tilde{G})^2 p(G|L) \, dG \, dL. \qquad (22)$$

The equation represents an $L_2$ optimization problem that can be solved efficiently by parameterizing the tone-curve, and using the image histogram to describe $p(L)$. The general concept is similar to the contrast optimization in the *Display adaptive TMO* [MDK08]. However, the formulation allows for a significant reduction in complexity. To ensure smooth changes over time, the nodes of the tone-curves are filtered over time, either using a low-pass IIR or an edge-stop filter.

As illustrated in Figure 5, the mapping under a tone-curve can reveal noise that is invisible in the original sequence. To prevent this from happening, a concept of *noise-awareness* is introduced. The idea is to include a model of the image noise as a salience term in the optimization of the tone-curve in Equation 22. Furthermore, the extracted detail layer is also scaled according to the visibility of noise. The effect of the two mechanisms is to conceal the noise to make sure that it do not exceed the visibility threshold of the HVS.

As a final step of the tone-mapping process, the tone-mapped frames are passed through an inverse display model, Equation 1, in order to adapt the result to a certain display device.

***Noisy log sensor TMO***, Li et al. [LSRJ16]: Li et al. describe a histogram adjustment TMO, in the same manner as the work by Ward et al. [WLRP97]. The method uses the cumulative distribution function derived from the image histogram to formulate a tone-curve. While Ward et al. use the capabilities of the HVS as a threshold for clipping the image histogram before calculating a tone-curve, Li et al. instead use a noise model for the purpose of not increasing the visibility of image noise in the tone-mapping. The TMO is specifically tailored for processing of video from a logarithmic CMOS sensor in real-time. For temporal adaptation, the histogram bins are filtered over time, employing an exponential low-pass filter.

***K-means clustering TMO***, Oskarsson [Osk16]: Based on the author's previous work [Osk15], Oskarsson proposes to extend a clustering based TMO for the purpose of tone-mapping of HDR video.

Given an input HDR frame composed of a set of $N$ intensity levels, the objective is to assign $K < N$ output levels, minimizing the distance $D(N, K)$ as compared to the original image,

$$D(N, K) = \min_{T_1, ..., T_K} \sum_{i=1}^{N} p(L_i) \min_j |L_i - T_j|^2. \qquad (23)$$

Here, $L \in \{L_1, ..., L_N\}$ and $T \in \{T_1, ..., T_K\}$ are the input and output values, respectively, and $p(L)$ is the image histogram. The optimization minimizes the weighted distances between input values $L$ and the limited output levels $T$, yielding a set of $K$ representative clusters. That is, the tone-mapping is formulated as a one-dimensional weighted K-means clustering problem. Using a dynamic programming approach it is demonstrated that the global optima can be found iteratively with computational complexity proportional to $KN^2$. Typically, $K$ is set to 256 for a standard display, and the clustering is followed by a mapping of the $K$ levels to display values $l = \{0, ..., 255\}$.

To account for temporal changes, the tone-curve derived from the K-means clustering is calculated separately for some key-frames distributed throughout an HDR video sequence. The clustering used to derive tone-curves at the specific key-frames is based not only on the pixels of the key-frame itself, but from a combined set of frames around it. The tone-curves in-between the key-frames are then computed from linear interpolation of the tone-curves at the nearest key-frames.

## 7. Quantitative comparison

Assessing and comparing the quality of the many existing tone-mapping algorithms is problematic since the end result is judged by subjective measures. In order to provide a relative subjective quality rating experiments can be conducted, where the TMOs are compared by visual inspection from a set of participants. A number of such evaluations have been performed [DMMS02, YBMS05, LCTS05, KYL*07, AFR*07, ČWNA08, KHF10], and also including video TMOs [PM13, EWMU13, MBDC14]. We refer to our previous description [DCMM16, Ch. 7] for the details on evaluation of tone-mapping operators, and specifically those for HDR video. However, one of the problems with such evaluations is that the experiments always have to be restricted to a small number of TMOs and intents. Given the wide range of operators, it would be an overwhelming undertaking to assess the quality of all.

In this section we provide a brief comparative assessment of a set of the video TMOs in Section 6. However, in order to tabulate the strengths and weaknesses of the individual operators, we do this on a qualitative level using a set of objective measures. In this way we also avoid the difficulties in performing a subjective study. It should also be mentioned that there are examples of metrics for objective quality assessment of tone-mapped images [AMMS08, YW10, YW13] and video sequences [AvMS10, YWZ*16]. Although such quality indexes has shown to correlate with data from subjective evaluations, the heuristics involved in the quality prediction can not account for all the complex processes involved in a subjective comparison. Furthermore,

| Operator | Categorization | | | | | Estimated performance | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Tone-curve | Intent | Processing | Temporal filter | Speed | Temporal coherence | Contrast | Exposure | Noise treatment |
| *Visual adaptation TMO* [FPSG96] | Linear | VSS | Global | Global | Real-time | ★½☆☆☆ | ★½☆☆☆ | ★★☆☆☆ | ★★★★☆ |
| *Interactive walk-through TMO* [DD00] | Linear | VSS | Local | Global | Real-time | – | – | – | – |
| *Time-adaptation TMO* [PTYG00] | Sigmoid | VSS | Global | Global | Real-time | ★★★☆☆ | ★☆☆☆☆ | ★½☆☆☆ | ★★★★☆ |
| *Temporal exposure TMO* [KUWS03] | Sigmoid | BSQ | Local | Global | Interactive | – | – | – | – |
| *Time-dependent GPU TMO* [GWWH03] | Sigmoid | BSQ | Local | Global | Real-time | – | – | – | – |
| *Adaptive temporal TMO* [RJIH04] | Sigmoid | BSQ | Global | Global | Interactive | – | – | – | – |
| *Local adaptation TMO* [LSC04] | Sigmoid | VSS | Local | Local[1] | Offline | ★★½☆☆ | ★★☆☆☆ | ★½☆☆☆ | ★☆☆☆☆ |
| *Mal-adaptation TMO* [IFM05] | Hist. based[2] | VSS | Global | Global | Interactive | ★★★★½ | ★★★☆☆ | ★★★★½ | ★★★½☆ |
| *Perceptual effects TMO* [KMS05b] | Sigmoid | VSS | Local | Global | Real-time | – | – | – | – |
| *Gradient domain TMO* [WRA05] | Grad. domain[3] | BSQ | Local | Local[4] | Offline | – | – | – | – |
| *Virtual exposures TMO* [BM05] | Logarithmic | BSQ | Local | Local[5] | Offline | ★☆☆☆☆ | ★★★★★ | ★★★★½ | ★★★½☆ |
| *Cone model TMO* [vH06] | Sigmoid | VSS | Global | Local[1] | Offline | ★★½☆☆ | ★★☆☆☆ | ★☆☆☆☆ | ★★★☆☆ |
| *Block matching TMO* [LK07] | Grad. domain[3] | BSQ | Local | Local[4] | Offline | – | – | – | – |
| *Display adaptive TMO* [MDK08] | Hist. based[2] | SRP | Global | Global | Offline | ★★★★★ | ★★★½☆ | ★★★★½ | ★★★½☆ |
| *Retina model TMO* [BAHC09] | Sigmoid | VSS | Local | Local[1] | Offline | ★★☆☆☆ | ★½☆☆☆ | ★★½☆☆ | ★★☆☆☆ |
| *Local mal-adaptation TMO* [PCA*10] | Sigmoid | VSS | Local | Local | Real-time | – | – | – | – |
| *Flicker reduction TMO* [GKEE11] | –[6] | SRP | –[6] | Global[7] | Offline | – | – | – | – |
| *Real-time automatic TMO* [KRTT12] | Sigmoid | BSQ | Global | Global | Real-time | – | – | – | – |
| *Color appearance TMO* [RPK*12] | Sigmoid | SRP | Local | Global | Offline | ★☆☆☆☆ | ★★★★½ | ★☆☆☆☆ | ★★★★★ |
| *Temporal coherence TMO* [BBC*12] | –[6] | SRP | –[6] | Global[7] | Offline | ★★★★½ | ★★★½☆ | ★★★★★ | ★★★★☆ |
| *Zonal coherence TMO* [BCTB14b] | –[6] | SRP | Local | Local[7] | Offline | ★★★★½ | ★★★★½ | ★★★★★ | ★★★★☆ |
| *Motion path filtering TMO* [ASC*14] | Logarithmic | BSQ | Local | Local[8] | Offline | ★★★★½ | ★★★★☆ | ★★★★½ | ★★★½☆ |
| *Hybrid TMO* [SLF*15] | Sigm./hist. based | BSQ | Local | Global | Real-time | – | – | – | – |
| *Noise-aware TMO* [EMU15] | Hist. based[2] | SRP | Local | Global | Real-time | ★★★★½ | ★★★★★ | ★★★★★ | ★★★★★ |
| *Noisy log sensor TMO* [LSRJ16] | Hist. based[2] | SRP | Global | Global | Real-time | – | – | – | – |
| *K-means clustering TMO* [Osk16] | Hist. based[2] | SRP | Global | Global | Interactive | – | – | – | – |

1 Per-pixel low-pass filter
2 Content-adaptive tone-curve formulated using the image histogram
3 Tone-curve applied in the gradient domain, based on the gradient magnitudes
4 3D Poisson solver
5 Per-pixel bilateral filter
6 Applicable to any TMO, alleviating global temporal incoherence
7 Scaling in a post-processing step
8 Per-pixel filtering along motion paths

Table 1: List of HDR video tone-mapping operators, classified and with quantitatively measured expected performance provided for a subset of the operators. The performance measures have been estimated as described in Section 7. Please note that these only provide indicative evidence of the different properties; they are estimated from a limited amount of data, and are not guaranteed to generalize to all HDR video sequences. For a description of each of the TMOs, see Section 6.
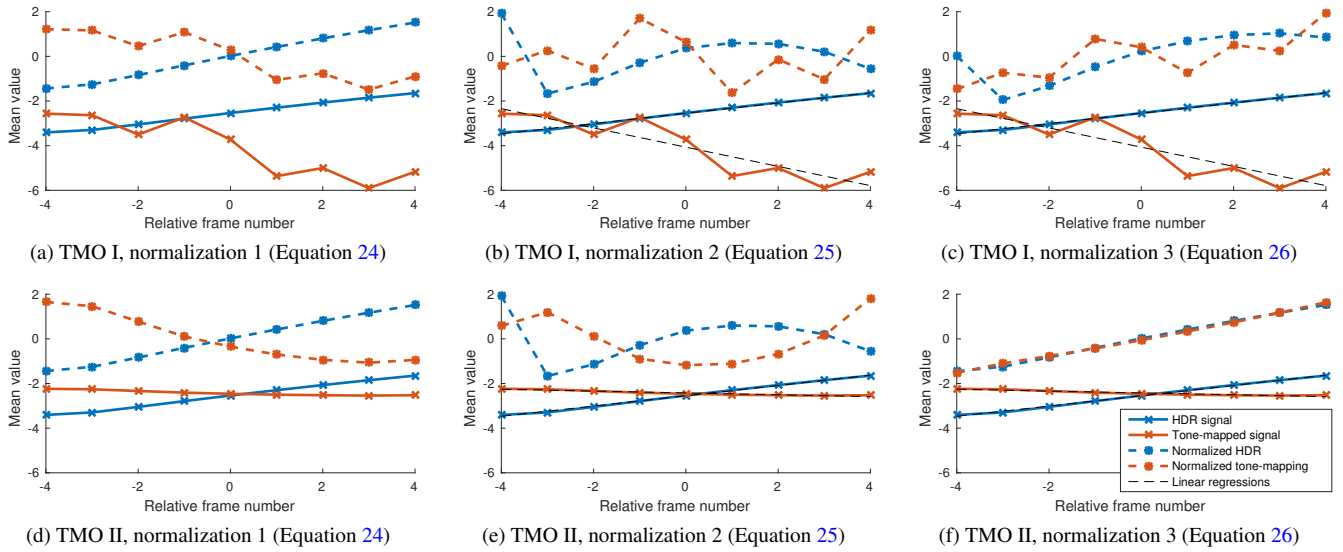
Figure 9: Normalization for calculating temporal coherence. The rows show two different TMOs, and the columns three different normalization techniques. Plotted are the mean log values of HDR and tone-mapped frames, for a local temporal neighborhood of 5 frames, and before and after normalization. The temporal coherence is estimated as the cross-correlation of the normalized signals. From a visual inspection, TMO II is temporally stable, while TMO I shows some flickering tendencies. However, it is only the normalization in Equation 26 that is capable of revealing this fact. The legend in (f) applies to all the sub-figures.

our objective is to evaluate some specific properties, not the overall subjective quality.

Our previous evaluation [EWMU13] concluded that the video TMOs under consideration either suffered from temporal incoherence (flickering, ghosting, etc.), lacked in local contrast, or increased the visibility of noise. We take these three properties as subjects for our qualitative evaluation, and complement with a fourth measurement to estimate the amount of over- and under-exposure. These are meant to indicate how well the TMOs can be expected to render a well-exposed tone-mapping with good local contrast, without showing temporal artifacts or revealing camera noise.

Our evaluation is similar to the approach by Čadík et al., where a set of image attributes were combined to create a quality measure [ČWNA06]. The attributes were measured in a subjective experiment, and different combinations were fitted to the estimated overall quality. The attributes included *brightness*, *contrast*, *details* and *colors*. In our evaluation, however, we design quantitative measures to estimate the attributes. The intention is not to generate a single quality score, but rather to evaluate the operators in terms of the different attributes. Furthermore, as we deal with video we include a very important attribute for temporal coherence.

Next, we describe the four measurements we use to estimate *temporal coherence*, *contrast*, *noise visibility* and *exposure*. We then present results of applying these on a set of the video TMOs.

### 7.1. Objective measures

**Temporal coherence** Evaluating the temporal coherence of a TMO is a matter of comparing the tone-mapped sequence to the

input HDR, deducing their correlation over time. The problem is how to account for transformations that the measure should be invariant to, such as scaling, exponentiation, temporal adaptation etc.

In order to quantitatively estimate temporal inconsistencies, a natural measure is the linear dependence of the HDR and tone-mapped signals, using their cross-correlation. For a normalized measure this can be calculated using the Pearson correlation coefficient,

$$\rho_1(L_t, T_t) = \frac{1}{2d} \sum_{r=t-d}^{t+d} \frac{L_r - \mu_{L_t}}{\sigma_{L_t}} \frac{T_r - \mu_{T_t}}{\sigma_{T_t}}. \qquad (24)$$

$L_t$ and $T_t$ denote the logarithmic HDR and tone-mapped values, respectively, taken from frame $t$. $T$ is evaluated from the final display value $l$ as $T = \log l^\gamma$. The linear dependence is evaluated over a local neighborhood in time, $r \in \{t-d, ..., t+d\}$. $\sigma_t$ and $\mu_t$ are used to normalize both signals to have the same mean and standard deviation within the temporal neighborhood $\{t-d, ..., t+d\}$. The measure can be applied per-pixel, $L_t = L_{t,s}$, or using the mean value in each frame, $L_t = \sum_s L_{t,s}/N$, in order to estimate the local or global temporal incoherencies, respectively.

One of the key properties of a video TMO is its ability to adapt to the input over time, based on statistics of the HDR sequence. This means that in practice the mean value, or a pixel value, can change over time in an opposite direction to the input HDR. This is not accounted for in Equation 24, as demonstrated in Figure 9(a) and (d). From this observation, we proceed to make the measure invariant to local linear changes over time. To this end, instead of

(a) *Time-adaptation TMO*, [PTYG00]



(b) *Local adaptation TMO*, [LSC04]



(c) *Virtual exposures TMO*, [BM05]



(d) *Display adaptive TMO*, [MDK08]



(e) *Temporal coherence TMO*, [BCTB14b]



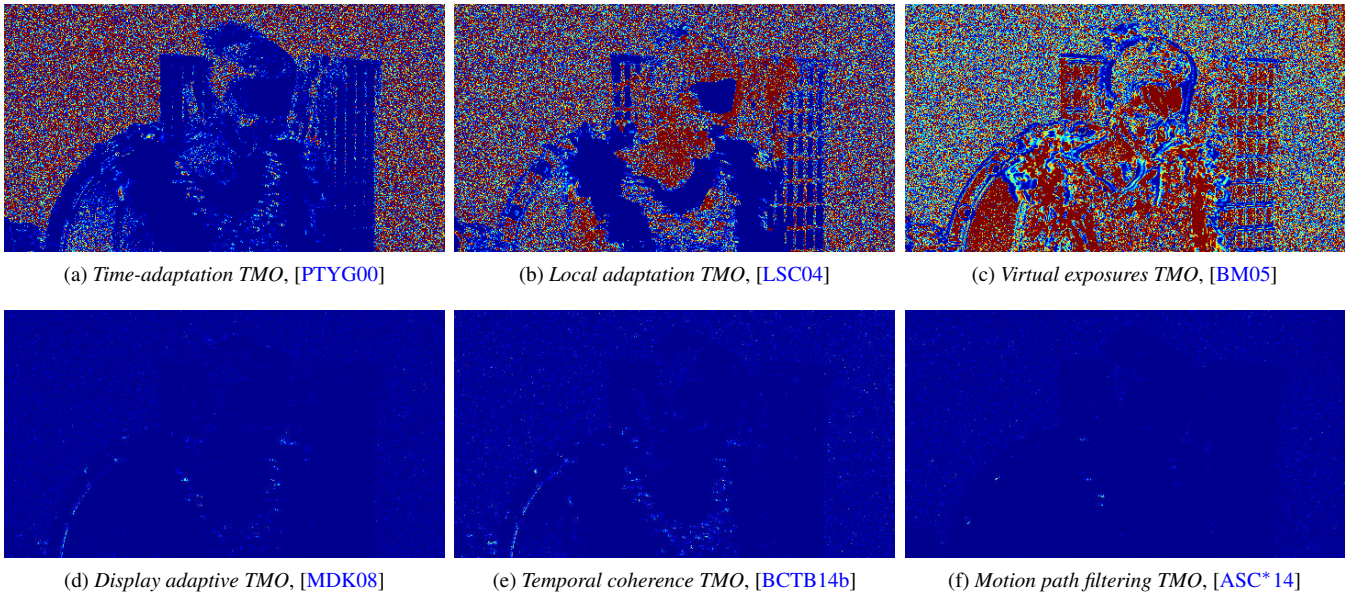(f) *Motion path filtering TMO*, [ASC*14]

Figure 10: Examples of measured local (per-pixel) temporal coherence, where blue and red corresponds to good and poor coherence, respectively. The examples demonstrate three TMOs that have problems with local temporal coherence (a)-(c), and three that show good coherence (d)-(f). Even though local incoherence is mostly associated with local TMOs (b)-(c), the global operator in (a) also suffers from incoherence due to amplified noise in the darker parts of the image. The frames are taken from measurements using Equation 26 on tone-mappings of the *Showgirl* HDR video sequence in Figure 19.

only using a 0-order normalization (the mean value $\mu$) in the measure, we perform a 1-order normalization from a linear regression, fitting a line segment $w(r-t)+\mu$ to the values of $L_r$ and $T_r$ at $r \in \{t-d,...,t+d\}$,

$$\rho_2(L_t, T_t) = \rho_1\left(\hat{L}_t, \hat{T}_t\right),$$
$$\hat{L}_r = L_r - w_{L_t}(r-t) - \mu_{L_t}, \tag{25}$$
$$\hat{T}_r = T_r - w_{T_t}(r-t) - \mu_{T_t}.$$

Here, $w_{L_t}$ and $w_{T_t}$ are the coefficients estimated from the linear regressions over $L_t$ and $T_t$, respectively. This formulation makes the coherence measure invariant to locally linear changes, see Figure 9. However, if the standard deviation $\sigma_{\hat{L}_t}$ of the corrected input $\hat{L}_t$ is small, the normalization in Equation 24 may amplify small-scale differences between the signals, as exemplified in Figure 9(e). For this reason, we scale $\hat{L}_t$ to have the same variance as $\hat{T}_t$. Then, the original slope $w_{L_t}(r-t)$ is put back to the signals,

$$\rho_3(L_t, T_t) = \rho_1\left(\hat{\hat{L}}_t, \hat{\hat{T}}_t\right),$$
$$\hat{\hat{L}}_r = \frac{\hat{L}_r}{\sigma_{\hat{L}_t}/\sigma_{\hat{T}_t}} + w_L(r-t), \tag{26}$$
$$\hat{\hat{T}}_r = \hat{T}_r + w_L(r-t).$$

That is, the normalization enforces $T_t$ to have the same slope as $L_t$, before the linear dependence is calculated. Inspecting the ex-

amples in Figure 9(c) and (f), this formulation makes it possible to distinguish temporal flickering even when a TMO is adapting over time. Further visualization of estimated coherence is exemplified in Figure 10, demonstrating per-pixel measurements.

Utilizing the measure in Equation 26, incoherence is estimated as $1 - \max(0, \rho_3(L_t, T_t))$. This means that negative correlations are clamped and considered equal to no correlation. The measure is then calculated for each frame $t$ in a sequence. To provide a representative value over a complete sequence, or multiple sequences, the average over all frames could be taken. However, coherence is often good except for shorter periods of the sequences which contain challenging transitions in intensity. Thus, the mean value may be underestimating the problem. Therefore, in order to provide an overall representative value we use the $95^{th}$ percentile of the measure. For the sequences we use in our results this corresponds to the best coherence of the worst 3 seconds of the total 1 min of sequences.

An alternative to the normalization approach above could be to use the second order temporal derivatives of the input and tone-mapped signals. However, experimenting with this measure gave unsatisfactory results, as it was difficult to tell the coherence of the TMOs apart. Some of the operators that were coherent from a visual inspection ended up with high estimated incoherence, and vice versa. Using the method described above this was not a problem.

**Contrast** The local contrast of a frame is estimated by extracting local contrasts – or details – using a bilateral filter. The contrast measure is taken as the mean absolute value of this detail layer, in

the log domain, and weighted by the input luminance in order to make it less influenced by noise,

$$C_{local}(T) = \frac{1}{N} \sum_s L_s \left| T_s - \sum_{p \in \Omega_s} G(T_p - T_s)G(p - s) \right|. \quad (27)$$

As for global contrast, this is estimated using a large Gaussian kernel,

$$C_{global}(T) = \frac{1}{N} \sum_s \sqrt{(G * T^2)_s - (G * T)_s^2}. \quad (28)$$

The two measures are applied to both the input HDR and the tone-mapped signal, in order to estimate the loss of contrast $c(L,T) = C(T) - C(L)$ in the tone-mapping.

**Noise visibility** As described in Section 5.3, camera noise can have a significant impact on the result of a tone-mapping method. As shown in Figure 5, the visibility of the noise can be altered under the transformation of a tone-curve. In order to quantify this change we use a computer generated HDR input $L$ without any noise. Modeling the camera noise as $N \sim \mathcal{N}(0, aI + b)$, a corrupted input is constructed according to $\hat{L} = L + N$. As a measure of the visibility of the added noise we use the HDR-VDP-2 (v2.2) quality predictor "Q" [MKRH11], $Q(\hat{L}, L)$. A high value means that the visual differences are small. The same procedure is then repeated for the tone-mapped frame, and finally the visibility before and after tone-mapping can be compared,

$$\begin{aligned} n(L,T) &= Q(\hat{L}, L) - Q(\hat{T}, T) \\ \hat{L} &= L + N \\ T &= V(L), \quad \hat{T} = V(\hat{L}) \\ N &\sim \mathcal{N}(0, aI + b). \end{aligned} \quad (29)$$

A positive difference in noise visibility, $n(L,T) > 0$, means that the noise is more visible after the tone-mapping, while the opposite describes a reduction in noise visibility.

**Exposure** To assess the ability of a TMO to produce a well-exposed tone-mapping, we estimate the number of under- and over-exposed pixels in the image. For over-exposure, the fraction of pixels $T \in [0, 1]$ with a value $\geq 0.95$ provides a measure of the amount of saturation, while for under-exposure we use the fraction of pixels $\leq 0.02$. A well-exposed tone-mapping that is capable of compressing the dynamic range sufficiently should have a low (but non-zero) value for both these measures. It should also be noted that for some scene characteristics a high number of saturated pixels is not necessarily a bad thing. The measure is meant to provide an indication on the TMOs ability to compress the dynamic range, when assessed over a number of different scenes.

### 7.2. Results

To evaluate the performance in terms of the measures above, we use the same set of operators that we used in our previous study

[EWMU13], plus three of the more recent methods that try to overcome the problems with temporal incoherence and lack of local contrast [BCTB14b,ASC*14,EMU15], for a total of 14 TMOs. The list of operators also includes a standard S-shaped camera curve, taken from a Canon 500D DSLR camera. For temporal processing, the exposure setting has been anchored to the mean luminance filtered over time using an exponential filter. This camera processing is named the *Camera TMO* in the results, and is provided as a simple reference.

The *temporal coherence*, *contrast* and *exposure* are calculated on 4 of the sequences captured by Fröhlich et al. [FGE*14]. Examples of tone-mapped frames and the properties of the sequences are shown in Figures 17-20. The sequences contain a set of challenging features, such as a high dynamic range of luminances, camera noise, rapid changes in lighting over time, complicated temporal transitions etc.

In Figure 11 the global and local incoherencies measured over the four sequences are shown. The global measure shows that in total five TMOs struggle to maintain coherence. Inspecting the sequences, three of these operators show some flickering in certain situations (*Virtual exposures TMO*, *Retina model TMO* and *Color appearance TMO*), possibly due to some parameters that are not correctly filtered over time. Two use per-pixel low-pass filters over time (*Cone model TMO* and *Retina model TMO*), which is causing motion blur that is punished by the coherence measure. Finally, the last TMO that shows global incoherence (*Visual adaptation TMO*) uses an adaptation mechanism that allows for rapid transitions over time that makes the mean of the tone-mapped frames not always comply with the mean of the input HDR. As for the local measure, it is generally TMOs that have local spatial and/or temporal processing that are prone to generate incoherence, with a value above 0.1. However, the *Time-adaptation TMO* only uses global calculations, and still shows local incoherencies. This could be explained from the operators tendency to amplify image noise in some situations, see Figure 10, 19 and 20 for examples.

The difference in contrast before and after tone-mapping is plotted in Figure 12. All operators inevitably have to reduce global contrast to match the limited output dynamic range, although to quite different amounts depending on the sequence. This is why the standard errors are large on the global contrast measures. However, when it comes to local contrast, only the local operators are able to maintain this property close to the original HDR. Of the global operators the *Display adaptive TMO* and *Temporal coherence TMO* show the least loss of local contrast.

Figure 13 shows the result of applying the exposure measures to the tone-mapped sequences. Many operators produce well-exposed results. A couple of TMOs give brighter results, such as the *Time-adaptation TMO* and *Local adaptation TMO*, which have saturated pixels for up to $10 - 15\%$ of the pixels. Under-exposure is most pronounced for the *Color appearance TMO*, with $> 20\%$ black pixels, and for the *Cone model TMO*. The *Camera TMO* shows a relatively large fraction of both dark and bright pixels, indicating that it is incapable of compressing the dynamic range enough. The results can also be confirmed in some of the example frames in Figures 17-20.

For the noise visibility difference measure demonstrated in Figure 14, 13 computer graphics generated HDR images have been
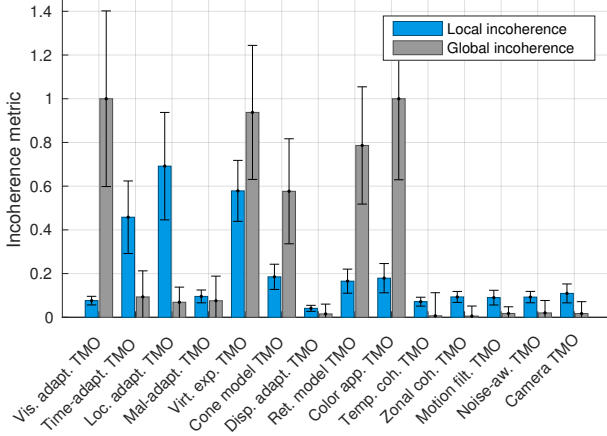
Figure 11: Temporal incoherence measure. Error bars show standard errors. A value of 0 means that the normalized HDR and tone-mapped signals (see Figure 9) are fully linearly correlated, while 1 means that they are either completely uncorrelated or negatively correlated. Since the bars are taken as the $95^{th}$ percentile, and the error bars are standard errors over the complete sequences, they extend above a value of 1 for some of the operators.
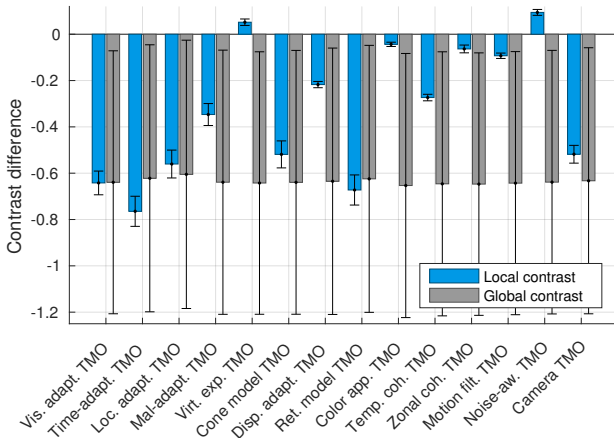


Figure 12: Measure of the loss of contrast from the tone-mapping. A negative value means that contrast is lost. Error bars show standard errors over all frames.
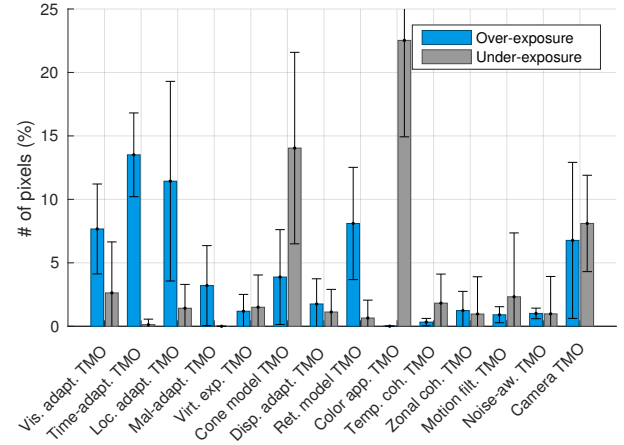


Figure 13: The fractions of estimated over- and under-exposed pixels averaged over all frames in the four sequences. Error bars show standard errors.
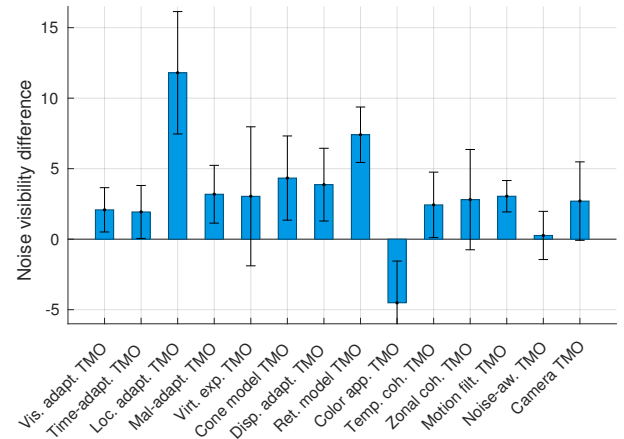


Figure 14: Noise visibility difference, comparing input HDR and tone-mapped sequences. Error bars show standard errors.

used. Figure 15 shows one example of such image. In order to make sure that each operator fully adapt to the input, each image has been used to compose a 1 second static HDR video sequence. Equation 29 has then been calculated for each of the TMOs, using $a = b = 10^{-4}$. In the result it is evident that most operators tend to increase the visibility of the noise. Exceptions are the *Color appearance TMO* and the *Noise-aware TMO*. Looking at Figure 13, the *Color appearance TMO* tends to under-expose, which explains why it also decreases the visibility of noise. The *Noise-aware TMO* on the other hand, is designed to tone-map without increasing the visibility of noise, while still providing an adequate exposure.

With the measurements in Figures 11-14, we can indicate which

of the operators that are expected to consistently produce results that are temporally coherent while reproducing contrasts without saturation, and how the noise visibility can be expected to change under the tone-mapping. To sum up these results, Figure 16 shows all the measures in one representative multidimensional visualization. The temporal coherence and local contrast are plotted along the x and y axes, respectively. To show the additional two dimensions of saturation and noise visibility, these are visualized with circles around each TMO. Finally, some of the categorization from Table 1 is also included as described in the legend.

Inspecting Figure 16, we see that three of the most recent operators are capable of preserving local contrast while maintaining temporal coherence and a small amount of over/under-exposed pixels. These are the *Zonal coherence TMO*, *Motion path filtering TMO* and *Noise-aware TMO*, which are designed by considering the particular challenges in tone-mapping of camera captured HDR video. The three operators represent quite different approaches to the
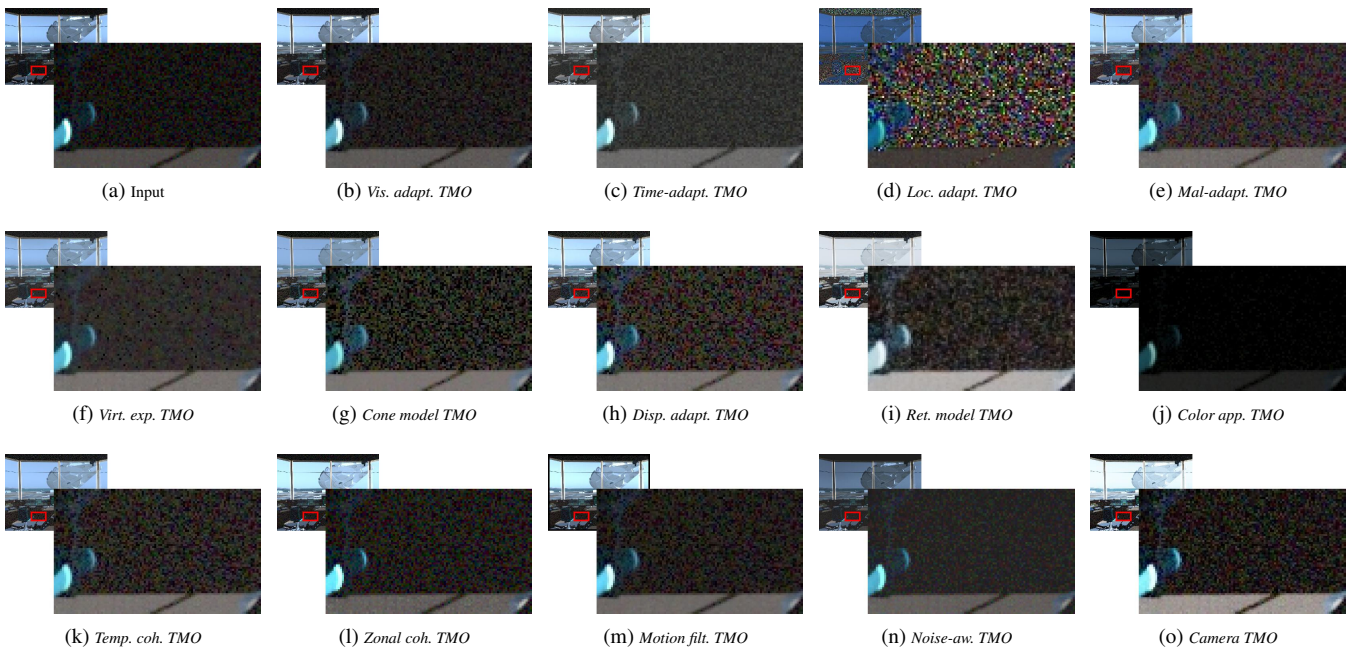
Figure 15: Example tone-mappings, (b)-(o), of the noisy input in (a). The operators show significantly different results when it comes to noise visibility, as highlighted by the enlarged dark parts of the images.

video tone-mapping problem. The *Zonal coherence TMO* considers the entire video sequence in order to locally remove both flickering and object incoherences in a post-processing step [BCTB14b]. The *Motion path filtering TMO* uses a temporal filtering along motion paths estimated from optical flow, both in order to enforce temporal coherence and to reduce noise [ASC*14]. Finally, the *Noise-aware TMO* is a more light-weight approach, with the goal of maintaining the properties of the input sequence as much as possible under the dynamic range compression, in terms of contrast, details and noise visibility [EMU15]. Apart from these methods, the evaluated video TMOs either have problem with local contrast reproduction or temporal coherence. This corresponds well to the results in our previous video TMO evaluation [EWMU13], which concluded that global TMOs often were preferred over their local counterparts, since these had problems with temporal coherence. The quantitatively measured results provided here indicate that this problem is resolved by later work in video tone-mapping.

Finally, the quantitative comparison is also used in Table 1, to complement the list of video TMOs with an indication on expected performance. However, it should be noted that the measures only have been calculated on a limited number of sequences, and they are not guaranteed to generalize to all possible HDR videos.

## 8. Conclusion

Methods for tone-mapping of HDR images have seen an immense development since their introduction almost 25 years ago. Operators designed with a model for temporal adaptation have also been around for almost as long time. However, with the introduction of camera captured HDR video, new problems arose; problems that

many of the as of then existing video operators could not properly address in all circumstances. With the most recent state-of-the-art methods, we are now in a position where high quality tone-mapping of this material can be consistently produced in most circumstances, taking into account the complicated temporal behavior and the specific properties of the HDR video streams.

In this work we have focused on algorithms for tone-mapping of HDR video, were we gave a comprehensive description of existing methods. A subset of these methods were subsequently evaluated in terms of *temporal coherence*, *contrast reproduction*, *exposure* and *noise visibility*, in order to give indication on their expected performance. From the results we conclude that, given the evidence provided by our quantitative measurements, some of the most recent TMOs are well-suited for tackling the problems associated with tone-mapping of camera captured HDR video.

Although the current state-of-the-art in video tone-mapping is capable of generating high quality results from existing HDR video material, some notable challenges still remain. For example, all TMOs have a set of parameters that can be tweaked for the optimal experience given a certain objective. The number of parameters can range from just a few to tens of possibilities to fine-tune the material, and even though a method may claim to be fully automatic there is always room to adjust contrasts, colors etc. The problem is how to find the optimal point in a multi-dimensional space of parameters, and to subjectively assess the result over time. While some experimental methods have been used for this purpose [YBMS05, EWMU13, EUWM14], there is no general methodology for how to fine tune a TMO. A second area that still needs attention is how to treat colors in the tone-mapping.
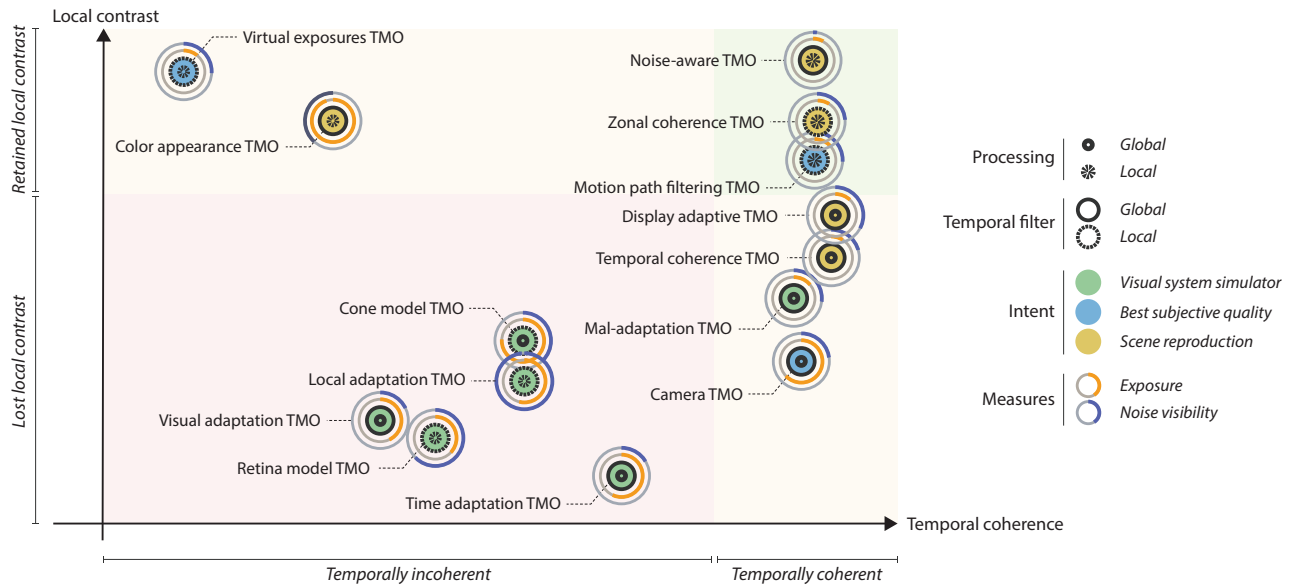
Figure 16: Expected performance of the evaluated TMOs, as suggested by the quantitative measurements according to Figures 11-14. The *temporal incoherence* has been taken as the sum of the local and global measures, and the *exposure* is the sum of under- and over-exposure. The categorization is according to Table 1.

Although color appearance models have been used for this purpose [PFFG98, FJ04, AR06, KJF07, RPK*12], most TMOs tend to only work on the luminance or separate color channels. Finally, since the quality of a tone-mapped sequence in the end is judged by subjective measures, evaluation of TMOs will always be a challenging problem. Unlike many other areas in image processing it is not possible to compare the result to a ground truth, and a subjective experiment is in most cases inevitable [DCMM16, Ch. 7].

## References

[ABD10] ADAMS A., BAEK J., DAVIS M.: Fast High-Dimensional Filtering Using the Permutohedral Lattice. *Computer Graphics Forum 29*, 2 (2010), 753–762. 7

[Ade82] ADELSON E. H.: Saturation and adaptation in the rod system. *Vision research 22*, 10 (1982), 1299–1312. 11

[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). 16

[AGDL09] ADAMS A., GELFAND N., DOLSON J., LEVOY M.: Gaussian kd-trees for fast high-dimensional filtering. *ACM Trans. Graphics 28*, 3 (2009), 21:1–21:12. 7

[AMMS08] AYDIN T. O., MANTIUK R., MYSZKOWSKI K., SEIDEL H.-P.: Dynamic range independent image quality assessment. *ACM Trans. Graph. 27*, 3 (Aug. 2008), 69:1–69:10. 16

[AR06] AKYÜZ A. O., REINHARD E.: Color appearance in high-dynamic-range imaging. *Journal of Electronic Imaging 15*, 3 (2006), 033001–033001. 4, 23

[ASC*14] AYDIN T. O., STEFANOSKI N., CROCI S., GROSS M., SMOLIC A.: Temporally coherent local tone mapping of HDR video. *ACM Trans. Graphics 33*, 6 (2014), 1–13. 7, 9, 10, 15, 17, 19, 20, 22

[Ash02] ASHIKHMIN M.: A tone mapping algorithm for high contrast images. In *Proceedings of the 13th Eurographics Workshop on Rendering* (Aire-la-Ville, Switzerland, Switzerland, 2002), EGRW '02, Eurographics Association, pp. 145–156. 7

[ATJ12] ANAND S., THIVYA T., JEEVA S.: Edge detection using directional filter bank. *International Journal of Applied Information Systems 1* (2012), 21–27. 15

[AvMS10] AYDIN T. O., ČADÍK M., MYSZKOWSKI K., SEIDEL H.-P.: Video quality assessment for computer graphics applications. *ACM Trans. Graph. 29*, 6 (Dec. 2010), 161:1–161:12. 16

[AW95] AURICH V., WEULE J.: Non-linear gaussian filters performing edge preserving diffusion. In *DAGM-Symposium* (1995), Sagerer G., Posch S., Kummert F., (Eds.), Informatik Aktuell, Springer, pp. 538–545. 7

[BADC11] BANTERLE F., ARTUSI A., DEBATTISTA K., CHALMERS A.: *Advanced High Dynamic Range Imaging: Theory and Practice*. AK Peters (CRC Press), Natick, MA, USA, 2011. 1, 2

[BAHC09] BENOIT A., ALLEYSSON D., HERAULT J., CALLET P. L.: Spatio-temporal tone mapping operator based on a retina model. In *Computational Color Imaging Workshop (CCIW'09)* (2009), pp. 12–22. 9, 10, 13, 17

[BAS*12] BANTERLE F., ARTUSI A., SIKUDOVA E., BASHFORD-ROGERS T., LEDDA P., BLOJ M., CHALMERS A.: Dynamic range compression by differential zone mapping based on psychophysical experiments. In *Proceedings of the ACM Symposium on Applied Perception* (New York, NY, USA, 2012), SAP '12, ACM, pp. 39–46. 7

[BBC*12] BOITARD R., BOUATOUCH K., COZOT R., THOREAU D., GRUSON A.: Temporal coherency for video tone mapping. In *Proc. SPIE 8499, Applications of Digital Image Processing XXXV* (San Diego, 2012), pp. 84990D–84990D–10. 9, 14, 17

[BCCS12] BANTERLE F., CORSINI M., CIGNONI P., SCOPIGNO R.: A low-memory, straightforward and fast bilateral filter through subsampling in spatial domain. *Computer Graphics Forum 31*, 1 (2012), 19–32. 7

[BCTB14a] BOITARD R., COZOT R., THOREAU D., BOUATOUCH K.: Survey of temporal brightness artifacts in video tone mapping. In *HDRi2014-Second International Conference and SME Workshop on HDR imaging* (2014). 9

[BCTB14b] BOITARD R., COZOT R., THOREAU D., BOUATOUCH K.: Zonal brightness coherency for video tone mapping. *Signal Processing: Image Communication 29*, 2 (2014), 229–246. 9, 15, 17, 19, 20, 22

[Ber96] BERNS R. S.: Methods for characterizing CRT displays. *Displays 16*, 4 (1996), 173–182. 2

[BJ10] BAEK J., JACOBS D. E.: Accelerating spatially varying gaussian filters. *ACM Trans. Graphics 29*, 6 (2010), 169:1–169:10. 7

[Bla81] BLACKWELL H.: An analytical model for describing the influence of lighting parameters upon visual performance. *Technical Foundations 1* (1981). 5

[BM05] BENNETT E. P., MCMILLAN L.: Video enhancement using per-pixel virtual exposures. *ACM Trans. Graphics 24*, 3 (2005), 845–852. 9, 10, 12, 17, 19

[Boi14] BOITARD R.: *Temporal Coherency in Video Tone Mapping*. PhD thesis, University of Rennes 1, 2014. 9

[BPD06] BAE S., PARIS S., DURAND F.: Two-scale tone management for photographic look. *ACM Trans. Graphics 25*, 3 (2006), 637–645. 7

[BT78] BARROW H., TENENBAUM J.: Recovering intrinsic scene characteristics from images. *Computer vision systems* (1978), 3–26. 4

[BTS*15] BONNEEL N., TOMPKIN J., SUNKAVALLI K., SUN D., PARIS S., PFISTER H.: Blind video temporal consistency. *ACM Trans. Graph. 34*, 6 (Oct. 2015), 196:1–196:9. 10

[Čad07] ČADÍK M.: Perception motivated hybrid approach to tone mapping. *WSCG* (2007), 129–136. 7

[CAS*15] CHAPIRO A., AYDIN T. O., STEFANOSKI N., CROCI S., SMOLIC A., GROSS M.: Art-directable continuous dynamic range video. *Computers & Graphics 53, Part A* (2015), 54 – 62. 40 years of Computer Graphics in Darmstadt. 5

[CPD07] CHEN J., PARIS S., DURAND F.: Real-time edge-aware image processing with the bilateral grid. *ACM Trans. Graphics 26*, 3 (2007), 103:1–103:9. 7

[Cra37] CRAWFORD B. H.: The change of visual sensitivity with time. *Proceedings of the Royal Society of London. Series B, Biological Sciences 123*, 830 (1937), 69–89. 11

[CT03] CHOUDHURY P., TUMBLIN J.: The trilateral filter for high contrast images and meshes. In *Proc. Eurographics workshop on Rendering 14* (2003), pp. 186–196. 7

[ČWNA06] ČADÍK M., WIMMER M., NEUMANN L., ARTUSI A.: Image attributes and quality for evaluation of tone mapping operators. In *National Taiwan University* (2006), Press, pp. 35–44. 18

[ČWNA08] ČADÍK M., WIMMER M., NEUMANN L., ARTUSI A.: Evaluation of HDR tone mapping methods using essential perceptual attributes. *Computers & Graphics 32*, 3 (2008). 16

[Dal92] DALY S. J.: Visible differences predictor: an algorithm for the assessment of image fidelity. *Proc. SPIE 1666* (1992), 2–15. 14

[DB04] DETER C., BIEHLIG W.: Scanning laser projection display and the possibilities of an extended color space. *Conference on Colour in Graphics, Imaging, and Vision 2004*, 1 (2004). 3

[DBZY15] DONG X., BONEV B., ZHU Y., YUILLE A. L.: Region-based temporally consistent video post-processing. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2015). 10

[DCMM16] DUFAUX F., CALLET P. L., MANTIUK R. K., MRAK M. (Eds.): *High Dynamic Range Video: From Acquisition, to Display and Applications*, vol. 1. Academic Press, 2016. 1, 2, 4, 9, 16, 23

[DCWP02] DEVLIN K., CHALMERS A., WILKIE A., PURGATHOFER W.: Tone Reproduction and Physically Based Spectral Rendering. In *Eurographics 2002 - STARs* (2002), Eurographics Association. 2

[DD00] DURAND F., DORSEY J.: *Rendering Techniques 2000: Proceedings of the Eurographics Workshop in Brno, Czech Republic, June 26–28, 2000*. Springer Vienna, Vienna, 2000, ch. Interactive Tone Mapping, pp. 219–230. 9, 11, 12, 14, 17

[DD02] DURAND F., DORSEY J.: Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. Graphics 21*, 3 (2002), 257–266. 7, 12

[Deb08] DEBEVEC P.: A median cut algorithm for light probe sampling. In *ACM SIGGRAPH 2008 Classes* (New York, NY, USA, 2008), SIGGRAPH '08, ACM, pp. 33:1–33:3. 14

[dLD58] DE LANGE DZN H.: Research into the dynamic nature of the human fovea – cortex systems with intermittent and modulated light. i. attenuation characteristics with white and colored light. *J. Opt. Soc. Am. 48*, 11 (Nov 1958), 777–784. 8

[DM97] DEBEVEC P. E., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1997), SIGGRAPH '97, ACM Press/Addison-Wesley Publishing Co., pp. 369–378. 3, 5

[DMAC03] DRAGO F., MYSZKOWSKI K., ANNEN T., CHIBA N.: Adaptive logarithmic mapping for displaying high contrast scenes. *Computer Graphics Forum 22* (2003), 419–426. 6, 12, 15

[DMMS02] DRAGO F., MARTENS W., MYSZKOWSKI K., SEIDEL H.-P.: *Perceptual evaluation of tone mapping operators with regard to similarity and preference*. Research Report MPI-I-2002-4-002, Max-Planck-Institut für Informatik, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany, August 2002. 16

[Dol16] DOLBY LABORATORIES: White paper: Dolby Vision for the home. https://www.dolby.com/us/en/technologies/dolby-vision/dolby-vision-white-paper.pdf, 2016. Accessed: 2017-01-24. 4

[DW00] DICARLO J. M., WANDELL B. A.: Rendering high dynamic range images, 2000. 7

[EMU15] EILERTSEN G., MANTIUK R. K., UNGER J.: Real-time noise-aware tone mapping. *ACM Trans. Graph. 34*, 6 (Oct. 2015), 198:1–198:15. 6, 7, 9, 10, 15, 17, 20, 22

[EMU16] EILERTSEN G., MANTIUK R., UNGER J.: A high dynamic range video codec optimized by large-scale testing. In *Proceedings of IEEE International Conference on Image Processing (ICIP)* (2016). 4

[EUWM14] EILERTSEN G., UNGER J., WANAT R., MANTIUK R.: Perceptually based parameter adjustments for video processing operations. In *ACM SIGGRAPH 2014 Talks* (New York, NY, USA, 2014), SIGGRAPH '14, ACM, pp. 74:1–74:1. 22

[EWMU13] EILERTSEN G., WANAT R., MANTIUK R. K., UNGER J.: Evaluation of Tone Mapping Operators for HDR-Video. *Computer Graphics Forum 32*, 7 (2013), 275–284. 4, 5, 9, 15, 16, 18, 20, 22

[Fat09] FATTAL R.: Edge-avoiding wavelets and their applications. *ACM Trans. Graphics 28*, 3 (2009), 1–10. 7

[Fec65] FECHNER G.: *Elements of psychophysics*. Holt, Rinehart & Winston, 1860/1965. 4

[FFH*16] FRANÇOIS E., FOGG C., HE Y., LI X., LUTHRA A., SEGALL A.: High dynamic range and wide color gamut video coding in hevc: Status and potential future enhancements. *IEEE Transactions on Circuits and Systems for Video Technology 26*, 1 (Jan 2016), 63–75. 4

[FFLS08] FARBMAN Z., FATTAL R., LISCHINSKI D., SZELISKI R.: Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graphics 27*, 3 (Aug. 2008), 67:1–67:10. 7

[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 9023, Digital Photography X* (2014), pp. 90230X–90230X–10. 20

[FJ04] FAIRCHILD M. D., JOHNSON G. M.: icam framework for image appearance, differences, and quality. *Journal of Electronic Imaging 13*, 1 (2004), 126–138. 4, 23

[FLW02] FATTAL R., LISCHINSKI D., WERMAN M.: Gradient domain high dynamic range compression. *ACM Trans. Graphics 21*, 3 (July 2002), 249–256. 6, 7, 12, 13

[FPSG96] FERWERDA J. A., PATTANAIK S. N., SHIRLEY P., GREEN-BERG D. P.: A model of visual adaptation for realistic image synthesis. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1996), SIGGRAPH '96, ACM, pp. 249–258. 8, 9, 10, 11, 17

[FTP94] FERSCHIN P., TASTL I., PURGATHOFER W.: A comparison of techniques for the transformation of radiosity values to monitor colors. In *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference* (Nov 1994), vol. 3, pp. 992–996 vol.3. 6

[GJ84] GILCHRIST A., JACOBSEN A.: Perception of lightness and illumination in a world of one reflectance. *Perception 13*, 1 (1984), 5–19. 4

[GKEE11] GUTHIER B., KOPF S., EBLE M., EFFELSBERG W.: Flicker reduction in tone mapped high dynamic range video. In *Proc. SPIE* (2011), vol. 7866, pp. 78660C–78660C–15. 9, 14, 15, 17

[GO12] GASTAL E. S. L., OLIVEIRA M. M.: Adaptive manifolds for real-time high-dimensional filtering. *ACM TOG 31*, 4 (2012), 33:1–33:13. Proceedings of SIGGRAPH 2012. 7

[Gre69] GREEN D. G.: Sinusoidal flicker characteristics of the color-sensitive mechanisms of the eye. *Vision Research 9*, 5 (1969), 591–601. 8

[GTH05] GHOSH A., TRENTACOSTE M., HEIDRICH W.: Volume rendering for high dynamic range displays. In *Fourth International Workshop on Volume Graphics, 2005.* (June 2005), pp. 91–231. 3

[GWWH03] GOODNIGHT N., WANG R., WOOLLEY C., HUMPHREYS G.: Interactive time-dependent tone mapping using programmable graphics hardware. In *Proceedings of the 14th Eurographics Workshop on Rendering* (Aire-la-Ville, Switzerland, Switzerland, 2003), EGRW '03, Eurographics Association, pp. 26–37. 9, 11, 12, 17

[HBH87] HAYHOE M., BENIMOFF N., HOOD D.: The time-course of multiplicative and subtractive adaptation process. *Vision Research 27*, 11 (1987), 1981–1996. 11

[HD07] HÉRAULT J., DURETTE B.: *Modeling Visual Perception for Image Processing.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 662–675. 13

[Hoe07] HOEFFLINGER B. (Ed.): *High-dynamic-range (HDR) vision.* Springer, 2007. 2

[Hor74] HORN B. K.: Determining lightness from an image. *Computer Graphics and Image Processing 3*, 4 (1974), 277–299. 4

[HST13] HE K., SUN J., TANG X.: Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence 35*, 6 (June 2013), 1397–1409. 7

[Hun95] HUNT R. W. G.: *The Reproduction of Colour.* Fountain Press, 1995. 11

[IFM05] IRAWAN P., FERWERDA J. A., MARSCHNER S. R.: Perceptually based tone mapping of high dynamic range image streams. In *Proc. Eurographics Conference on Rendering Techniques 16* (2005), pp. 231–242. 9, 11, 12, 14, 17

[JRW97] JOBSON D. J., RAHMAN Z.-U., WOODELL G. A.: A multi-scale retinex for bridging the gap between color images and the human observation of scenes. *Image Processing, IEEE Transactions on 6*, 7 (1997), 965–976. 6

[KGB*14] KRONANDER J., GUSTAVSON S., BONNET G., YNNERMAN A., UNGER J.: A unified framework for multi-sensor hdr video reconstruction. *Signal Processing : Image Communications 29*, 2 (2014), 203–215. 1, 3, 5, 9

[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). 16

[KJF07] KUANG J., JOHNSON G. M., FAIRCHILD M. D.: icam06: A refined image appearance model for {HDR} image rendering. *Journal of Visual Communication and Image Representation 18*, 5 (2007), 406–414. Special issue on High Dynamic Range Imaging. 4, 15, 23

[KKEA06] KATKOVNIK V., KATKOVNIK V., EGIAZARIAN K., AS-TOLA J.: *Local Approximation Techniques in Signal and Image Processing.* Press Monographs. SPIE Press, 2006. 7

[KMS05a] KRAWCZYK G., MYSZKOWSKI K., SEIDEL H.-P.: Lightness perception in tone reproduction for high dynamic range images. *Computer Graphics Forum 24*, 3 (2005), 635–645. 7

[KMS05b] KRAWCZYK G., MYSZKOWSKI K., SEIDEL H.-P.: Perceptual effects in real-time tone mapping. In *Proceedings of the 21st Spring Conference on Computer Graphics* (New York, NY, USA, 2005), SCCG '05, ACM, pp. 195–202. 12, 17

[KPB12] KROES T., POST F. H., BOTHA C. P.: Exposure render: An interactive photo-realistic volume rendering framework. *PloS one 7*, 7 (2012), e38586. 3

[KR10] KUNKEL T., REINHARD E.: A reassessment of the simultaneous dynamic range of the human visual system. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization* (New York, NY, USA, 2010), APGV '10, ACM, pp. 17–24. 4

[KRTT12] KISER C., REINHARD E., TOCCI M., TOCCI N.: Real time automated tone mapping system for HDR video. In *Proc. IEEE International Conference on Image Processing 19* (2012), pp. 2749–2752. 9, 14, 17

[KS10] KASS M., SOLOMON J.: Smoothed local histogram filters. *ACM Trans. Graphics 29* (July 2010), 100:1–100:10. 7

[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 (Nov. 2013), 202:1–202:8. 3

[KUWS03] KANG S. B., UYTTENDAELE M., WINDER S., SZELISKI R.: High dynamic range video. *ACM Trans. Graph. 22*, 3 (July 2003), 319–325. 3, 5, 9, 11, 17

[KWK09] KIM M. H., WEYRICH T., KAUTZ J.: Modeling human color perception under extended luminance levels. *ACM Trans. Graph. 28*, 3 (July 2009), 27:1–27:9. 4

[KYH16] KEROFSKY L., YE Y., HE Y.: Recent developments from mpeg in hdr video compression. In *2016 IEEE International Conference on Image Processing (ICIP)* (Sept 2016), pp. 879–883. 4

[KYL*07] KUANG J., YAMAGUCHI H., LIU C., JOHNSON G. M., FAIRCHILD M. D.: Evaluating hdr rendering algorithms. *ACM Trans. Appl. Percept. 4*, 2 (July 2007). 16

[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). 16

[LDSP03] LEE B. B., DACEY D. M., SMITH V. C., POKORNY J.: Dynamics of sensitivity regulation in primate outer retina: The horizontal cell network. *Journal of Vision 3*, 7 (2003), 5. 13

[LFUS06] LISCHINSKI D., FARBMAN Z., UYTTENDAELE M., SZELISKI R.: Interactive local adjustment of tonal values. *ACM Trans. Graph. 25*, 3 (July 2006), 646–653. 7

[LK07] LEE C., KIM C.-S.: Gradient domain tone mapping of high dynamic range videos. In *Proc. of International Conference on Image Processing* (2007). 9, 13, 17

[LK08] LEE C., KIM C. S.: Rate-distortion optimized compression of high dynamic range videos. In *Signal Processing Conference, 2008 16th European* (Aug 2008), pp. 1–5. 4

[LKG*16] LJUNG P., KRÜGER J., GROLLER E., HADWIGER M., HANSEN C. D., YNNERMAN A.: State of the art in transfer functions for direct volume rendering. *Computer Graphics Forum 35*, 3 (2016), 669–691. 3

[LM71] LAND E. H., MCCANN J. J.: Lightness and retinex theory. *J. Opt. Soc. Am. 61*, 1 (Jan 1971), 1–11. 6

[LSA05] LI Y., SHARAN L., ADELSON E. H.: Compressing and companding high dynamic range images with subband architectures. *ACM Trans. Graph. 24*, 3 (July 2005), 836–844. 7

[LSC04] LEDDA P., SANTOS L. P., CHALMERS A.: A local model of eye adaptation for high dynamic range images. In *Proc. International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa 3* (2004), pp. 151–160. 9, 11, 14, 17, 19

[LSRJ16] LI J., SKORKA O., RANAWEERA K., JOSEPH D.: Novel real-time tone-mapping operator for noisy logarithmic cmos image sensors. *Journal of Imaging Science and Technology 60*, 2 (2016), 1–13. 6, 9, 10, 16, 17

[LWA*12] LANG M., WANG O., AYDIN T., SMOLIC A., GROSS M.: Practical temporal consistency for image-based graphics applications. *ACM Trans. Graph. 31*, 4 (July 2012), 34:1–34:8. 10

[MAS07] MEYLAN L., ALLEYSSON D., SÜSSTRUNK S.: Model of retinal local adaptation for the tone mapping of color filter array images. *J. Opt. Soc. Am. A 24*, 9 (Sep 2007), 2807–2816. 13, 14

[MBDC14] MELO M., BESSA M., DEBATTISTA K., CHALMERS A.: Evaluation of {HDR} video tone mapping for mobile devices. *Signal Processing: Image Communication 29*, 2 (2014), 247 – 256. Special Issue on Advances in High Dynamic Range Video Research. 16

[MDK08] MANTIUK R., DALY S., KEROFSKY L.: Display adaptive tone mapping. *ACM Trans. Graphics 27*, 3 (2008), 68:1–68:10. 6, 13, 16, 17, 19

[MEMS06] MANTIUK R., EFREMOV A., MYSZKOWSKI K., SEIDEL H.-P.: Backward compatible high dynamic range mpeg video compression. *ACM Trans. Graph. 25*, 3 (July 2006), 713–723. 4

[MG10] MANGIAT S., GIBSON J.: High dynamic range video with ghost removal. *Proc. SPIE 7798* (2010), 779812–779812–8. 3

[Mil13] MILANFAR P.: A tour of modern image filtering: New insights and methods, both practical and theoretical. *IEEE Signal Processing Magazine 30*, 1 (2013), 106–128. 7

[MKMS04] MANTIUK R., KRAWCZYK G., MYSZKOWSKI K., SEIDEL H.: Perception-motivated high dynamic range video encoding. *ACM Transactions on Graphics (TOG)* (2004). 4

[MKRH11] MANTIUK R., KIM K. J., REMPEL A. G., HEIDRICH W.: Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graphics 30*, 4 (2011), 40:1–40:14. 20

[MKVR09] MERTENS T., KAUTZ J., VAN REETH F.: Exposure fusion: A simple and practical alternative to high dynamic range photography. *Computer Graphics Forum 28*, 1 (2009), 161–171. 7

[MM13] MICHAELIS L., MENTEN M. L.: Die kinetik der invertinwirkung. *Biochem. z 49*, 333-369 (1913), 352. 6

[MMK08] MYSZKOWSKI K., MANTIUK R., KRAWCZYK G.: High dynamic range video. *Synthesis Lectures on Computer Graphics and Animation 2*, 1 (2008), 1–158. 2

[MMS05] MANTIUK R., MYSZKOWSKI K., SEIDEL H.-P.: A perceptual framework for contrast processing of high dynamic range images. In *APGV '05: Proceedings of the 2nd symposium on Applied perception in graphics and visualization* (New York, NY, USA, 2005), ACM Press, pp. 87–94. 7

[MMS15] MANTIUK R. K., MYSZKOWSKI K., SEIDEL H.-P.: *High Dynamic Range Imaging*. John Wiley & Sons, Inc., 2015. 2

[MMTH09] MANTIUK R., MANTIUK R., TOMASZEWSKA A., HEIDRICH W.: Color correction for tone mapping. In *Computer Graphics Forum* (2009), vol. 28, Wiley Online Library, pp. 193–202. 4

[MND13] MILLER S., NEZAMABADI M., DALY S.: Perceptual signal coding for more efficient usage of bit codes. *SMPTE Motion Imaging Journal 122*, 4 (2013). 4

[MNM84] MILLER N. J., NGAI P. Y., MILLER D. D.: The application of computer graphics in lighting design. *Journal of the Illuminating Engineering Society 14*, 1 (1984), 6–26. 5

[MP95] MANN S., PICARD R.: On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of Society for Imaging Science and Technology's 48th Annual Conference* (1995), pp. 442–448. 3, 5

[MR11] MCCANN J. J., RIZZI A.: *The art and science of HDR imaging*, vol. 26. John Wiley & Sons, 2011. 2, 4

[NR66] NAKA K., RUSHTON W.: S-potentials from colour units in the retina of fish (cyprinidae). *The Journal of physiology 185*, 3 (1966), 536–555. 6

[Osk15] OSKARSSON M.: Democratic tone mapping using optimal k-means clustering. In *Proceedings of Scandinavian Conference on Image Analysis* (2015), pp. 354–365. 16

[Osk16] OSKARSSON M.: Temporally consistent tone mapping of images and video using optimal k-means clustering. *Journal of Mathematical Imaging and Vision* (2016), 1–14. 6, 16, 17

[OSS68] OPPENHEIM A., SCHAFER R., STOCKHAM T.: Nonlinear filtering of multiplied and convolved signals. *IEEE Transactions on Audio and Electroacoustics 16*, 3 (Sep 1968), 437–466. 5, 7

[PAB*13] POULI T., ARTUSI A., BANTERLE F., AKYÜZ A. O., SEIDEL H.-P., REINHARD E.: Color correction for tone reproduction. In *Color and Imaging Conference* (2013), vol. 2013, Society for Imaging Science and Technology, pp. 215–220. 4

[PCA*10] PAJAK D., CADÍK M., AYDIN T. O., MYSZKOWSKI K., SEIDEL H.-P.: Visual maladaptation in contrast domain. *Proc. SPIE 7527* (2010), 752710–752710–12. 14, 17

[PFFG98] PATTANAIK S. N., FERWERDA J. A., FAIRCHILD M. D., GREENBERG D. P.: A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques* (1998), ACM, pp. 287–298. 4, 6, 7, 23

[PHK11] PARIS S., HASINOFF S. W., KAUTZ J.: Local laplacian filters: edge-aware image processing with a laplacian pyramid. *ACM Trans. Graphics 30*, 4 (2011), 68:1–68:12. 7

[PM90] PERONA P., MALIK J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Analysis Machine Intelligence 12*, 7 (1990), 629–639. 7

[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). 16

[PTYG00] PATTANAIK S. N., TUMBLIN J., YEE H., GREENBERG D. P.: Time-dependent visual adaptation for fast realistic image display. In *Proc. SIGGRAPH 00* (2000), Annual Conference Series, pp. 47–54. 9, 11, 12, 17, 19

[PY02] PATTANAIK S., YEE H.: Adaptive gain control for high dynamic range image display. In *Proceedings of the 18th Spring Conference on Computer Graphics* (New York, NY, USA, 2002), SCCG '02, ACM, pp. 83–87. 7

[RAGB11] RADONJIĆ A., ALLRED S. R., GILCHRIST A. L., BRAINARD D. H.: The dynamic range of human lightness perception. *Current Biology 21*, 22 (2011), 1931–1936. 4

[Rei02] REINHARD E.: Parameter estimation for photographic tone reproduction. *Journal of Graphics Tools 7*, 1 (2002), 45–51. 14

[RJIH04] RAMSEY S. D., JOHNSON III J. T., HANSEN C.: Adaptive temporal tone mapping. In *Proceedings of the 7th IASTED International Conference on Computer Graphics and Imaging* (2004), Citeseer, pp. 124–128. 9, 11, 17

[RJW96] RAHMAN Z.-U., JOBSON D. J., WOODELL G. A.: A multiscale retinex for color rendition and dynamic range compression. In *SPIE International Symposium on Optical Science, Engineering and Instrumentation, Applications of Digital Image Processing XIX* (1996), vol. 2847, pp. 183–191. 6

[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, 14, 17, 23

[RSSF02] REINHARD E., STARK M., SHIRLEY P., FERWERDA J.: Photographic tone reproduction for digital images. *ACM Trans. Graph. 21*, 3 (July 2002), 267–276. 7, 9, 11, 12, 14, 15

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

[SB97] SMITH S. M., BRADY J. M.: Susan–a new approach to low level image processing. *Int. J. Comput. Vision 23*, 1 (May 1997), 45–78. 7

[SB16] SCHULTE T., BARSOTTI J.: *HDR Demystified: Emerging UHDTV Systems*. CreateSpace Independent Publishing Platform, 2016. 4

[Sch94] SCHLICK C.: Quantization techniques for visualization of high dynamic range pictures. Springer-Verlag, pp. 7–20. 4, 6

[SHS*04] SEETZEN H., HEIDRICH W., STUERZLINGER W., WARD G., WHITEHEAD L., TRENTACOSTE M., GHOSH A., VOROZCOVS A.: High dynamic range display systems. *ACM Trans. Graph. 23*, 3 (Aug. 2004), 760–768. 1, 3

[SLF*15] SHAHID H., LI D., FANASWALA A., POURAZAD M. T., NASIOPOULOS P.: A new hybrid tone mapping scheme for high dynamic range (hdr) videos. In *2015 IEEE International Conference on Consumer Electronics (ICCE)* (Jan 2015), pp. 351–352. 7, 15, 17

[SPLD01] SMITH V. C., POKORNY J., LEE B. B., DACEY D. M.: Primate horizontal cell dynamics: An analysis of sensitivity regulation in the outer retina. *Journal of Neurophysiology 85*, 2 (2001), 545–558. 13

[SS60] STEVENS S. S., STEVENS J.: Brightness function-parametric effects of adaptation and contrast. In *Journal of the Optical Society of America* (1960), vol. 50, pp. 1139–1139. 5

[SSD09] SUBR K., SOLER C., DURAND F.: Edge-preserving multiscale image decomposition based on local extrema. *ACM Trans. Graphics 28*, 5 (Dec. 2009), 147:1–147:9. 7

[Ste62] STEVENS S.: The surprising simplicity of sensory metrics. *American psychologist 17*, 1 (1962), 29. 14

[SW12] SPILLMANN L., WERNER J. S.: *Visual perception: The neurophysiological foundations*. Elsevier, 2012. 4

[TFM07] TAKEDA H., FARSIU S., MILANFAR P.: Kernel regression for image processing and reconstruction. *IEEE Trans. Image Processing 16*, 2 (2007), 349–366. 7

[THG99] TUMBLIN J., HODGINS J. K., GUENTER B. K.: Two methods for display of high contrast images. *ACM Transactions on Graphics (TOG) 18*, 1 (1999), 56–94. 4, 5, 7

[TKTS11] TOCCI M. D., KISER C., TOCCI N., SEN P.: A versatile hdr video production system. *ACM Trans. Graphics 30*, 4 (2011), 41:1–41:10. 1, 3, 5, 9

[TM98] TOMASI C., MANDUCHI R.: Bilateral filtering for gray and color images. In *Proc. International Conference on Computer Vision 6* (1998), pp. 839–846. 7

[TR93] TUMBLIN J., RUSHMEIER H.: Tone reproduction for realistic images. *Computer Graphics and Applications, IEEE 13*, 6 (1993), 42–48. 5

[TT99] TUMBLIN J., TURK G.: Lcis: a boundary hierarchy for detail-preserving contrast reduction. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1999), SIGGRAPH '99, ACM Press/Addison-Wesley Publishing Co., pp. 83–90. 7

[UG07] UNGER J., GUSTAVSON S.: High-dynamic-range video for photometric measurement of illumination. In *Proc. SPIE* (2007), vol. 6501, pp. 65010E–65010E–10. 3, 5

[UGOJ04] UNGER J., GUSTAVSON S., OLLILA M., JOHANNESSON M.: A real time light probe. In *Proceedings of the 25th Eurographics Annual Conference* (2004), vol. Short Papers and Interactive Demos, pp. 17–21. 3, 5

[Ups85] UPSTILL S. D.: *The Realistic Presentation of Synthetic Images: Image Processing in Computer Graphics*. PhD thesis, 1985. AAI8610255. 5

[vH05] VAN HATEREN J. H.: A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells. *Journal of Vision 5*, 4 (2005), 331–347. 13

[vH06] VAN HATEREN J. H.: Encoding of high dynamic range video with a model of human cones. *ACM Trans. Graphics 25* (2006), 1380–1399. 9, 10, 13, 17

[VMGM15] VANGORP P., MYSZKOWSKI K., GRAF E. W., MANTIUK R. K.: A model of local adaptation. *ACM Trans. Graph. 34*, 6 (Oct. 2015), 166:1–166:13. 4

[War94a] WARD G.: A contrast-based scalefactor for luminance display. *Graphics gems IV* (1994), 415–421. 5, 11, 14

[War94b] WARD G. J.: The radiance lighting simulation and rendering system. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1994), SIGGRAPH '94, ACM, pp. 459–472. 3

[WECH*90] WALRAVEN J., ENROTH-CUGELL C., HOOD D. C., MACLEOD D. I., SCHNAPF J. L.: *The control of visual sensitivity: Receptoral and postreceptoral processes*. Academic Press, 1990, pp. 53–101. 12

[WLRP97] WARD LARSON G., RUSHMEIER H., PIATKO C.: A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Trans. Visualization and Computer Graphics 3*, 4 (1997), 291–306. 6, 12, 14, 15, 16

[WRA05] WANG H., RASKAR R., AHUJA N.: High dynamic range video using split aperture camera. In *IEEE 6th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, Washington, DC, USA* (2005), Citeseer. 3, 5, 9, 12, 13, 17

[Yan12] YANG Q.: Recursive bilateral filtering. In *Proc. European Conference on Computer Vision 12* (2012), pp. 399–413. 7

[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. 16, 22

[YBY10] YOSHIZAWA S., BELYAEV A. G., YOKOTA H.: Fast gauss bilateral filtering. *Computer Graphics Forum 29*, 1 (2010), 60–74. 7

[YNCP05] YUAN X., NGUYEN M. Z., CHEN B., PORTER D. H.: High dynamic range volume visualization. In *VIS 05. IEEE Visualization, 2005.* (Oct 2005), pp. 327–334. 3

[YNCP06] YUAN X., NGUYEN M. X., CHEN B., PORTER D. H.: Hdr volvis: high dynamic range volume visualization. *IEEE Transactions on Visualization and Computer Graphics 12*, 4 (July 2006), 433–445. 3

[YP03] YEE H. Y., PATTANAIK S.: Segmentation and adaptive assimilation for detail-preserving display of high-dynamic range images. *The Visual Computer 19*, 7 (2003), 457–466. 7

[YW10] YEGANEH H., WANG Z.: Objective assessment of tone mapping algorithms. In *2010 IEEE International Conference on Image Processing* (Sept 2010), pp. 2477–2480. 16

[YW13] YEGANEH H., WANG Z.: Objective quality assessment of tone-mapped images. *IEEE Transactions on Image Processing 22*, 2 (Feb 2013), 657–667. 16

[YWZ*16] YEGANEH H., WANG S., ZENG K., EISAPOUR M., WANG Z.: Objective quality assessment of tone-mapped videos. In *2016 IEEE International Conference on Image Processing (ICIP)* (Sept 2016), pp. 899–903. 16

(a) *Vis. adapt. TMO*    (b) *Time-adapt. TMO*    (c) *Loc. adapt. TMO*    (d) *Mal-adapt. TMO*    (e) *Virt. exp. TMO*    (f) *Cone model TMO*    (g) *Disp. adapt. TMO*

(h) *Ret. model TMO*    (i) *Color app. TMO*    (j) *Temp. coh. TMO*    (k) *Zonal coh. TMO*    (l) *Motion filt. TMO*    (m) *Noise-aw. TMO*    (n) *Camera TMO*
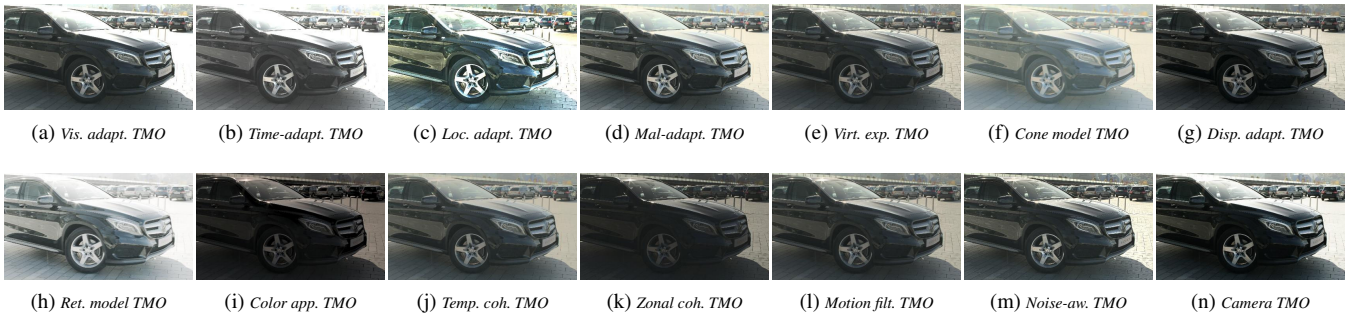
Figure 17: Tone-mapped frames from the *Cars fullshot* HDR video sequence. The video shows a daylight scene with a limited amount of camera noise. The scene is static while the camera performs a panning movement.



(a) *Vis. adapt. TMO*    (b) *Time-adapt. TMO*    (c) *Loc. adapt. TMO*    (d) *Mal-adapt. TMO*    (e) *Virt. exp. TMO*    (f) *Cone model TMO*    (g) *Disp. adapt. TMO*

(h) *Ret. model TMO*    (i) *Color app. TMO*    (j) *Temp. coh. TMO*    (k) *Zonal coh. TMO*    (l) *Motion filt. TMO*    (m) *Noise-aw. TMO*    (n) *Camera TMO*
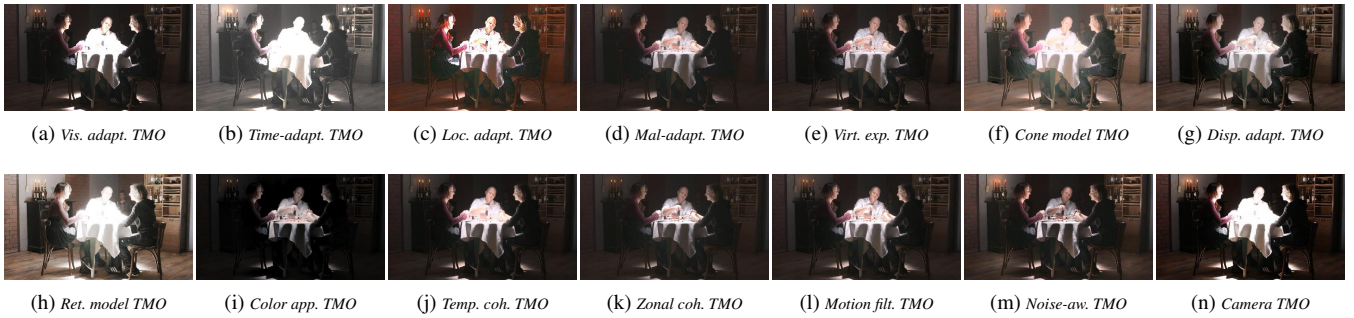
Figure 18: Tone-mapped frames from the *Poker fullshot* HDR video sequence. The video shows an indoor scene with a moderate amount of camera noise. The camera is static while the actors perform a moderate amount of movements.



(a) *Vis. adapt. TMO*    (b) *Time-adapt. TMO*    (c) *Loc. adapt. TMO*    (d) *Mal-adapt. TMO*    (e) *Virt. exp. TMO*    (f) *Cone model TMO*    (g) *Disp. adapt. TMO*

(h) *Ret. model TMO*    (i) *Color app. TMO*    (j) *Temp. coh. TMO*    (k) *Zonal coh. TMO*    (l) *Motion filt. TMO*    (m) *Noise-aw. TMO*    (n) *Camera TMO*
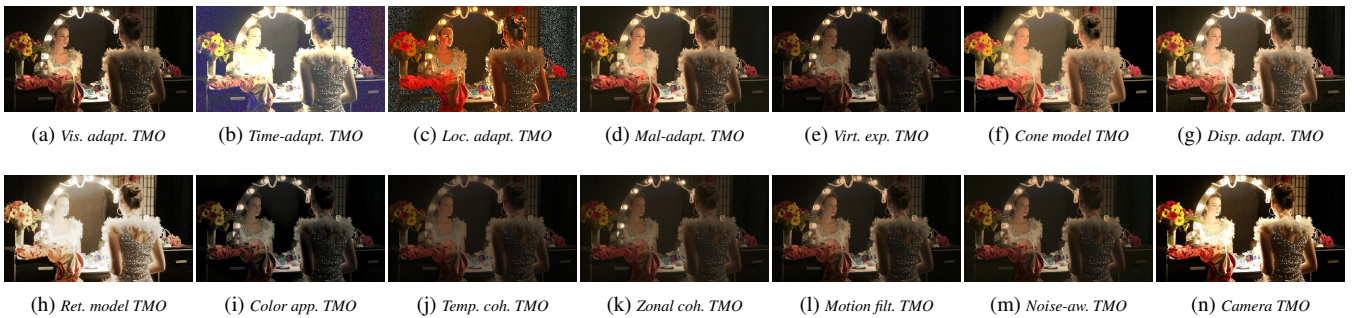
Figure 19: Tone-mapped frames from the *Showgirl* HDR video sequence. The video shows an indoor scene with a high amount of camera noise. The sequence contains a sharp transition in lighting when the actor stands up and a bright light source is switched on.



(a) *Vis. adapt. TMO*    (b) *Time-adapt. TMO*    (c) *Loc. adapt. TMO*    (d) *Mal-adapt. TMO*    (e) *Virt. exp. TMO*    (f) *Cone model TMO*    (g) *Disp. adapt. TMO*

(h) *Ret. model TMO*    (i) *Color app. TMO*    (j) *Temp. coh. TMO*    (k) *Zonal coh. TMO*    (l) *Motion filt. TMO*    (m) *Noise-aw. TMO*    (n) *Camera TMO*
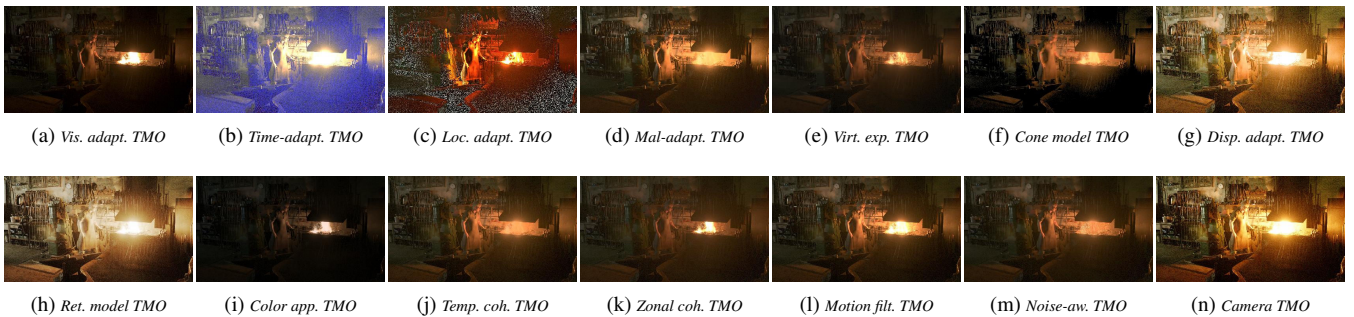
Figure 20: Tone-mapped frames from the *Smith hammering* HDR video sequence. The video shows an indoor scene with a high amount of camera noise. The camera is slowly zooming while the actor is moving.