Linköping Studies in Science and Technology. Dissertations No. 1939

The high dynamic range imaging pipeline

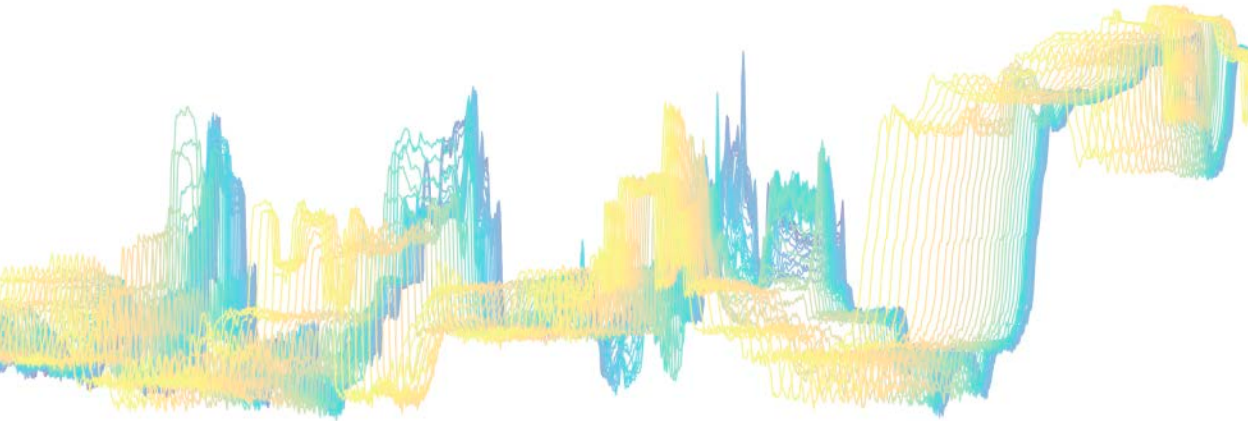
Tone-mapping, distribution, and single-exposure reconstruction

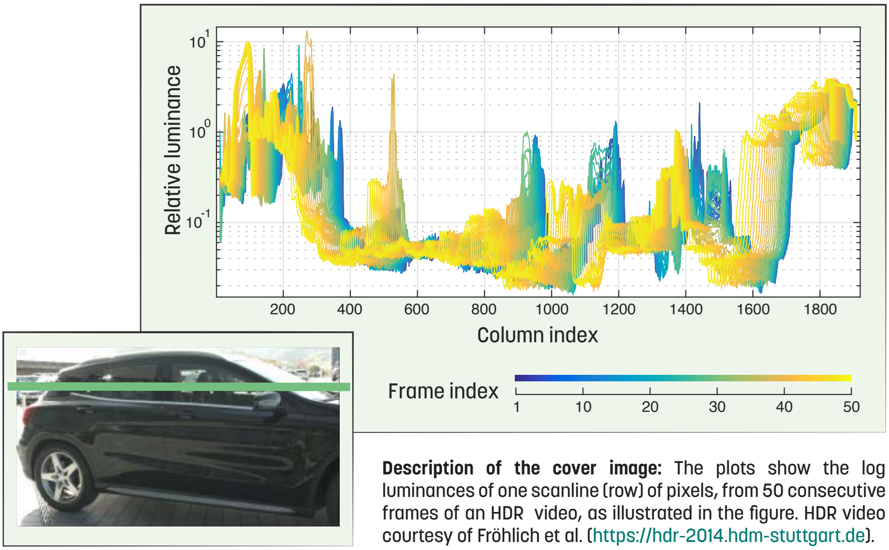
Gabriel Eilertsen

LINKÖPING
UNIVERSITY

Division of Media and Information Technology
Department of Science and Technology

Linköping University
SE-601 74 Norrköping, Sweden
Norrköping, June 2018





Description of the cover image: The plots show the log luminances of one scanline (row) of pixels, from 50 consecutive frames of an HDR video, as illustrated in the figure. HDR video courtesy of Fröhlich et al. (https://hdr-2014.hdm-stuttgart.de).

The high dynamic range imaging pipeline: tone-mapping, distribution, and single-exposure reconstruction

Division of Media and Information Technology
Department of Science and Technology

Campus Norrköping, Linköping University
SE-601 74 Norrköping, Sweden

ISBN: 978-91-7685-302-3
ISSN: 0345-7524

Printed in Sweden by LiU-Tryck, Linköping, 2018

Det heter inte improvisera,
det heter forska när man
inte vet vad man gör.

Magnus och Brasse

Abstract

Techniques for *high dynamic range* (HDR) imaging make it possible to capture and store an increased range of luminances and colors as compared to what can be achieved with a conventional camera. This high amount of image information can be used in a wide range of applications, such as HDR displays, image-based lighting, tone-mapping, computer vision, and post-processing operations. HDR imaging has been an important concept in research and development for many years. Within the last couple of years it has also reached the consumer market, e.g. with TV displays that are capable of reproducing an increased dynamic range and peak luminance.

This thesis presents a set of technical contributions within the field of HDR imaging. First, the area of HDR video tone-mapping is thoroughly reviewed, evaluated and developed upon. A subjective comparison experiment of existing methods is performed, followed by the development of novel techniques that overcome many of the problems evidenced by the evaluation. Second, a large-scale objective comparison is presented, which evaluates existing techniques that are involved in HDR video distribution. From the results, a first open-source HDR video codec solution, *Luma HDRv*, is built using the best performing techniques. Third, a machine learning method is proposed for the purpose of reconstructing an HDR image from one single-exposure *low dynamic range* (LDR) image. The method is trained on a large set of HDR images, using recent advances in deep learning, and the results increase the quality and performance significantly as compared to existing algorithms.

The areas for which contributions are presented can be closely inter-linked in the HDR imaging pipeline. Here, the thesis work helps in promoting efficient and high-quality HDR video distribution and display, as well as robust HDR image reconstruction from a single conventional LDR image.

Keywords: high dynamic range imaging, tone-mapping, video tone-mapping, HDR video encoding, HDR image reconstruction, inverse tone-mapping, machine learning, deep learning

Populärvetenskaplig sammanfattning

Utvecklingen av kameror har gått mycket snabbt de senaste årtiondena, och de utnyttjas idag för en stor mängd ändamål. Till exempel är kameran ett viktigt verktyg inom produktkontroll och övervakning, för att inte tala om inom filmindustrin som är en av de allra största i världen. Kameran utgör också en naturlig del i privatpersonens liv, för att dokumentera familj, resor och vardag. Det genomslag kameran har haft kan ses på den mängd kameror vi omger oss med, som separata enheter eller integrerade i datorer och telefoner. Men kameran har sina tydliga begränsningar. Vi har nog alla upplevt situationer där vi tvingas kompromissa i hur en bild ska exponeras när det finns både mörka skuggor och ljusa högdagrar i den miljö som ska fotograferas. Även om en betraktare samtidigt kan urskilja detaljer i både skuggor och ljusa delar, så klarar inte kameran av att registrera all information. Antingen avbildas de ljusa delarna som helt vita, eller så försvinner detaljer i de mörka delarna av bilden. Detta beror på att en konventionell kamera är begränsad i hur stora skillnader i ljus som kan registreras i en och samma bild. Jämför man med det mänskliga ögat, så har det en mycket bättre förmåga att uppfatta detaljer i ett stort omfång av ljusintensiteter.

Med hjälp av tekniker för att fotografera i ett utökat spann av ljusintensiteter kan en bild med stort dynamiskt omfång (HDR, från engelskans High Dynamic Range) infångas, exempelvis genom att kombinera flera bilder med olika exponering. Inom forskning och produktion har HDR-formatet använts i många år. Då bilderna kan representera en fysikaliskt korrekt mätning av det omgivande ljuset kan de t.ex. användas för att ljussätta datorgenererade fotorealistiska bilder, och i en uppsättning av efterbehandlingsapplikationer. De senaste åren har HDR-format också etablerat sig på konsumentmarknaden, exempelvis med TV-apparater som kan visa ett utökat dynamiskt omfång och en högre ljusintensitet. Också för konventionella skärmar och TV-apparater kan HDR-bilder tillhandahålla en förbättrad tittarupplevelse. Genom metoder för s.k. tonmappning kan bildinnehållet komprimeras till ett lägre dynamiskt omfång, medan detaljer bibehålls i mörka och ljusa bildregioner, och resultatet efterliknar på så sätt hur det mänskliga ögat uppfattar den fotograferade scenen. Andra målsättningar för tonmappning är också möjliga, t.ex. att försöka skapa en bild med den subjektivt bästa kvalitén, eller en bild som så bra som möjligt återger en specifik bildegenskap.

Denna avhandling presenterar ett antal tekniska forskningsbidrag inom HDR-fotografi och video. De första bidragen är inom tonmappning av HDR-video. Först presenteras en studie där existerande metoder för tonmappning av HDR-video utvärderas. Resultaten visar på problem som ännu var olösta vid tidpunkten för studien. I ett efterföljande projekt fokuserar vi på att lösa dessa problem i en ny metod för videotonmappning. Vi visar hur metoden kan åstadkomma hög bildkvalité med snabba beräkningar, medan detaljnivån bibehålls och bildbrus undertrycks.

För att spara och distribuera HDR-video kan inte existerande format för standardvideo användas utan modifikation. Det krävs nya strategier för att uppnå tillräckligt hög precision och färgåterbildning. I och med att HDR-video etablerar sig inom TV-industrin har en standardisering av tekniker för detta ändamål påbörjats. Avhandlingen presenterar en utvärdering av olika teknikerna involverade i att distribuera HDR-video, samt utveckling av ett ramverk för kodning och avkodning av HDR-video som använder de bäst presterande teknikerna. Den resulterande mjukvaran, *Luma HDRv*, publiceras med öppen källkod, och erbjuder på så sätt ett första fritt tillgängligt alternativ för distribution av HDR-video.

Ett problem med HDR-fotografi är att det krävs dyra, begränsade eller tidskrävande tekniker för att fotografera ett stort dynamiskt omfång. Den absoluta majoriteten av existerande bilder är dessutom fotograferade med konventionella metoder, och för att kunna använda dessa i HDR-applikationer behöver det dynamiska omfånget utökas. Ett av de viktigaste och svåraste problemen med detta är att försöka återskapa detaljer och information i bildens ljusa delar, och inga metoder har tidigare lyckats göra det på ett övertygande sätt. I det sista projektet som presenteras i avhandlingen använder vi de senaste framstegen inom deep learning (maskininlärning med “djupa”, mycket kraftfulla, modeller) för att återbilda ljusintensitet, färg och detaljer i bildens ljusa delar. Metoden lär sig från en stor uppsättning av HDR-bilder, och resultaten visar en stor förbättring jämfört med tidigare existerande metoder.

Tillämpningarna av de olika forskningsbidragen är tätt sammankopplade i den kedja/pipeline av tekniker som behövs för att infånga och visa HDR-bilder. Här bidrar de olika metoderna som avhandlingen presenterar till att lättare och mer effektivt skapa, distribuera och visa HDR-material. Givet den senaste utvecklingen och populariteten inom HDR-TV, så förväntas också att tekniker för HDR-fotografi bara kommer att bli viktigare framöver. Framtiden för HDR-bilder ser ljus ut!

Acknowledgments

In the same manner as the *human visual system* has a non-linear response, where the log luminance is closer to describing the perceived brightness, this is also true for time perception. In order to describe the perceived elapsed time as a function of age, a logarithmic relationship is probably also a decent generalization. However, the experience of time is also heavily affected by other parameters. For example, time tends to fly by when you are occupied with a lot of things to do, and when you really enjoy something. Children are also one of the most profound accelerators of the perceived time. Given all these considerations, it is not surprising that my years as a Ph.D. student are the shortest years I have experienced. It feels as if it was yesterday I started my journey towards the disputation. At the same time, considering the things I have learned and the ways in which I have grown as a researcher and as a human, it also feels as if it was far away in the distant past. The relative nature of perception, and therefore also life, is truly remarkable.

Over the course of my years as a Ph.D. student, I have met many extraordinary individuals. I would like to take this opportunity to express my gratitude to the *high dynamic range* of people that have, in one way or the other, contributed to the thesis.

The work that the thesis is built on would not be there without the support of my supervisors. First and foremost I would like to thank my main supervisor Jonas Unger. It has been a privilege to work under your supervision. With your skills, you have provided an excellent balance of guidance and encouragement, which have helped me develop and gain confidence as a researcher. I am also very grateful for all the help from my co-supervisor Rafał Mantiuk. Your expertise, through suggestions for possible directions to explore and with all the insightful feedback, has had a significant impact on the focus and quality of the thesis work. Thank you also for having me as a visiting researcher at Bangor University and in the Computer Laboratory at the University of Cambridge. I hope our collaboration can continue in the future. Furthermore, I would like to thank my co-supervisor Anders Ynnerman. I truly appreciate the research environment that has been made available through your efforts. You have also been an inspiration since my last years as an undergraduate student and one of the contributing reasons that I decided to pursue research.

Despite only my name appearing on the thesis, the work that it presents is truly a collaborative effort. I would like to thank all the co-authors for their work on the thesis papers: Jonas Unger, Rafał Mantiuk, Robert Wanat, Joel Kronander, and Gyorgy Denes. In terms of the more practical matters, thank you to Per Larsson for all the help with hardware and software. Your skills have also helped in resolving a number of disagreements between me and my computer. Thank you also to Eva Skärblom for all the support with administrative concerns and for the help with the practicalities related to the thesis. Your knowledge and patience are much appreciated.

Working in the Computer Graphics and Image Processing group has been a much greater experience thanks to my fellow Ph.D. students. Thank you to Ehsan Miandji, Saghi Hajisharif, Apostolia Tsirikoglou, and Tanaboon Tongbuasirilai for all the discussions, sharing experiences and knowledge about work, courses, and completely different matters. You have taught me a lot of things and provided me with much-needed company in the sometimes solitary work of a Ph.D. student. I would also like to thank my previous colleagues in the Computer Graphics and Image Processing group. Joel Kronander, thank you for sharing your knowledge with such enthusiasm. Andrew Gardner, thank you for all the discussions, advice, and company. Reiner Lenz, thank you for interesting conversations and perspectives.

Finally, this really goes without saying, but saying it a million times is not enough – thank you to my beloved family. Jenny Eilertsen, you are the love of my life, my best friend, my comfort. With your amazing ability of reasoning and clear thinking, you always support me with invaluable advice. “*True love is a big deal*”. During my years as a Ph.D. student, I have also had the honor of becoming the father of two. Ebba Eilertsen and Olle Eilertsen, you are my never-ending source of reality and a constant reminder of what is important. I love you more than I can ever put into words.

Gabriel Eilertsen
Norrköping, May 2018

Publications

The work presented in the thesis is built on the following publications:

Paper A: G. Eilertsen, R. K. Mantiuk, and J. Unger. A comparative review of tone-mapping algorithms for high dynamic range video. *Computer Graphics Forum (Proceedings of Eurographics 2017)*, 36(2):565–592, 2017.

Paper B: G. Eilertsen, R. Wanat, R. K. Mantiuk, and J. Unger. Evaluation of tone mapping operators for HDR-video. *Computer Graphics Forum (Proceedings of Pacific Graphics 2013)*, 32(7):275–284, 2013.

Paper C: G. Eilertsen, R. K. Mantiuk, and J. Unger. Real-time noise-aware tone mapping. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2015)*, 34(6):198:1–198:15, 2015.

Paper D: G. Eilertsen, R. K. Mantiuk, and J. Unger. A high dynamic range video codec optimized by large-scale testing. In *Proceedings of IEEE International Conference on Image Processing (ICIP 2016)*, pages 1379–1383, 2016.

Paper E: G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger. HDR image reconstruction from a single exposure using deep CNNs. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2017)*, 36 (6):178:1–178:15, 2017.

A number of additional publications were also part of the work leading up to the dissertation, but not included in the thesis. These are listed here in reverse chronological order:

1. G. Eilertsen, P.-E. Forssén, and J. Unger. BriefMatch: Dense binary feature matching for real-time optical flow estimation. In *Proceedings of Scandinavian Conference on Image Analysis (SCIA 2017)*, pages 221–233, 2017.

2. G. Eilertsen, R. K. Mantiuk, and J. Unger. Real-time noise-aware tone-mapping and its use in luminance retargeting. In *Proceedings of IEEE International Conference on Image Processing (ICIP 2016)*, pages 894–898, 2016.

3. G. Eilertsen, R. K. Mantiuk, and J. Unger. Luma HDRv: an open source high dynamic range video codec optimized by large-scale testing. In *ACM SIGGRAPH 2016 Talks*, pages 17:1–17:2, 2016.

4. J. Unger, F. Banterle, G. Eilertsen, and R. K. Mantiuk. The HDR-video pipeline - from capture and image reconstruction to compression and tone mapping. In *Eurographics 2016 Tutorials*, 2016.

5. G. Eilertsen, J. Unger, and R. K. Mantiuk. Evaluation of tone mapping operators for HDR video. In F. Dufaux, P. L. Callet, R. K. Mantiuk, and M. Mrak, editors, *High Dynamic Range Video: From Acquisition, to Display and Applications*, chapter 7, pages 185–207. Academic Press, 2016.

6. G. Eilertsen, J. Unger, R. Wanat, and R. K. Mantiuk. Perceptually based parameter adjustments for video processing operations. In *ACM SIGGRAPH 2014 Talks*, pages 74:1–74:1, 2014.

7. G. Eilertsen, J. Unger, R. Wanat, and R. K. Mantiuk. Survey and evaluation of tone mapping operators for HDR video. In *ACM SIGGRAPH 2013 Talks*, pages 11:1–11:1, 2013.

Contributions

The thesis provides a set of contributions to the field of *high dynamic range* (HDR) imaging. The main focus is on tone-mapping of HDR video, for compressing the dynamic range to be displayed on a conventional display device (Paper **A**, **B**, **C**). However, there are also important contributions related to the reverse problem of reconstructing an HDR image given a *low dynamic range* (LDR) input image (Paper **E**), as well as HDR video encoding (Paper **D**).

Paper A provides a review that serves as a comprehensive reference, categorization, and comparative assessment of the state-of-the-art in tone-mapping for HDR video. It constitutes a complementary part of the background for the tone-mapping work presented in this thesis, as it describes the foundations in HDR imaging and tone-mapping. The report includes a literature overview of tone-mapping in general, as well as a categorization and description of all, at the time, existing tone-mapping algorithms for HDR video. Finally, a quantitative analysis is performed in order to tabulate the strength and weaknesses of a set of representative video tone-mapping operators.

The publication was presented as a state-of-the-art report (STAR) at Eurographics 2017 in Lyon, France [84].

Paper B presents the results of a subjective evaluation of tone-mapping operators for HDR video. This constitutes the foundation of the video tone-mapping contributions in this thesis, and was one of the first tone-mapping evaluations that considered the temporal domain. The results show that even though tone-mapping is a well-researched area, there are still a number of unsolved challenges related to tone-mapping for HDR video. This laid the ground for the subsequent work on overcoming the challenges in a novel video tone-mapping operator (Paper **C**).

The paper was presented at Pacific Graphics 2013 in Singapore [75]. A pilot study that preceded the work was also described in a talk at Siggraph 2013 in Anaheim, USA [74]. The technique used in order to calibrate the different tone-mapping operators was presented in a talk at Siggraph 2014 in Vancouver, Canada [76]. Finally, a more general text on strategies and existing work within HDR video evaluation was included as a chapter [81] in the book "*High Dynamic Range Video: From Acquisition, to Display and Applications*" [71].

Paper C introduces a novel tone-mapping operator for HDR video, which overcomes a number of the problems of the, at the time, existing

methods. It is temporally stable, while operating locally on the image with minimal artifacts around edges. It considers the noise characteristics of the input HDR video in order to not make noise visible in the tone-mapped version. It compresses the dynamic range to a specified display device while minimizing distortion of image contrasts. All calculations run in real-time so that interactive adjustments of all the parameters are possible.
The paper was presented at Siggraph Asia 2015 in Kobe, Japan [77].

Paper D presents an HDR video codec that is released as an open-source library and *application programming interface* (API) named *Luma HDRv*. The HDR video encoding is built by first performing a large-scale evaluation on a high-performance computer cluster, and measuring differences using a perceptual image quality index. The evaluation considers a set of existing techniques for color encoding, luminance transformation, and compression of the final bit-stream. By choosing the highest performing combination, the final codec pipeline allows for the best compression performance given the techniques examined. The paper was presented at the International Conference on Image Processing (ICIP) 2016 in Phoenix, USA [79]. The work was also described in a talk at Siggraph 2016 in Anaheim, USA [80]. The HDR video codec is available on GitHub: `https://github.com/gabrieleilertsen/lumahdrv`.

Paper E demonstrates how recent advances in deep learning can be applied to the reverse problem of tone-mapping; that is, to expand the dynamic range in order to reconstruct an HDR image from an input LDR image. The method can robustly predict high quality HDR image information given a standard 8 bit single-exposed image. It uses a *convolutional neural network* (CNN) in an auto-encoder design, together with HDR specific transfer-learning, skip-connections, color space, and loss function. The proposed method demonstrates a steep improvement in the quality of reconstruction as compared to the, at the time, existing methods for expanding LDR into HDR images. The quality of the reconstructions is further confirmed in a subjective evaluation on an HDR display, which shows that the perceived naturalness of the reconstructed images are in most cases on par with the ground truth HDR images.
The paper was presented at Siggraph Asia 2017 in Bangkok, Thailand [83]. Code for inference and training with the HDR reconstruction CNN is available on GitHub: `https://github.com/gabrieleilertsen/hdrcnn`.

Contents

Chapter 1

Introduction

A camera is designed for a similar task as the *human visual system* (HVS) – to capture the surrounding environment in order to provide information for higher level processing. Given this similarity, a naïve conception would be that a physical scene captured by a camera and viewed on a display device should invoke the exact same response as observing the scene directly. However, this is very seldom the case, for a number of reasons. For example, there are insufficient depth cues in the captured image and there are differences in color and brightness. Also, one of the most prominent differences in many scenes is a mismatch in *dynamic range*. The camera and the display are unable to cover the wide range of luminances that the HVS can detect simultaneously, which means that there is more visual information available in the scene than what can be captured and reproduced. For example, when attempting to capture an object in a dark indoor environment in front of a bright window, one has to choose between properly exposed background or foreground, while the other information is lost in dark or saturated image areas, respectively. However, it is usually not a problem for the human eye to simultaneously register both foreground and background. The limitations of the camera as compared to the HVS becomes evident. With techniques for *high dynamic range* (HDR) imaging information can be captured in both dark and bright image regions, matching or outperforming the dynamic range of the HVS.

The thesis presents a number of technical research contributions within the HDR imaging pipeline. This chapter first gives a brief introduction to the concept of high dynamic range and the HDR image format. Next, the thesis contributions are briefly described and put in a context. Finally, the structure of the thesis is outlined.

1.1 High dynamic range

The difference in the dynamic range of the HVS as compared to conventional cameras/displays gives a natural motivation for developing techniques that can capture and display HDR images, which can better match the sensation of watching the real scene. Since a camera sensor is limited in the range of luminances that can be captured, the most common technique for generating HDR images is to combine a set of images that have been captured with different exposure times, as demonstrated in Figure 1.1. With long exposures, the details in dark image areas are captured while information in bright image areas disappears due to sensor saturation. With short exposures, the bright image features can be registered while the darker parts are lost in noise and quantization. Combining different exposures means that both dark and bright image features, which are outside the range of a conventional sensor, can be represented and thereby providing a large increase in captured information and dynamic range.

1.1.1 Definition

The incident light from the surrounding environment onto a specific point on a surface in a scene – the *illuminance* – is reflected based on the properties of the surface material. The integrated outgoing light over an area in a certain direction is the *luminance*, and this is what we measure when registering the light as it falls on the area of a pixel in a camera sensor. The SI unit for measuring the luminance in a scene or on a screen is *candela per square meter* (cd/m^2). In the TV/display manufacturing industry, the same unit is also commonly referred to as *nit* (1 nit = 1 cd/m^2). In Figure 1.2a, the typical luminances for some objects are illustrated to give a reference for the range of observable values.

The dynamic range is the ratio between the smallest and largest value registered by an imaging sensor or depicted on a display. For the HVS, it is between the smallest and the largest observable luminance of a scene. For a camera sensor, it is between the smallest detectable luminance above the noise floor and the largest measurable luminance before the sensor saturates. For a display, it is between the smallest and largest pixel luminances that can be rendered simultaneously on the screen. For example, if the lowest and largest values are 0.001 and $1,000$ cd/m^2, respectively, the dynamic range is $1,000,000$:1, or 6 $\log_{10}$ units. In photography, the dynamic range is often measured in *stops/f-stops*, which uses $\log_2$ units. Alternatively, the dynamic range can also be specified with the *signal-to-noise ratio* (SNR), usually specified in decibels, where SNR = $20\log_{10}(I_{ceil}/I_{noise})$ dB. For a camera sensor I_{ceil} is the saturation point and I_{noise} is the noise floor. For the previous example, we thus have a dynamic range $1,000,000$:1 = 6 $\log_{10}$ units = 19.93 stops = 120 dB.

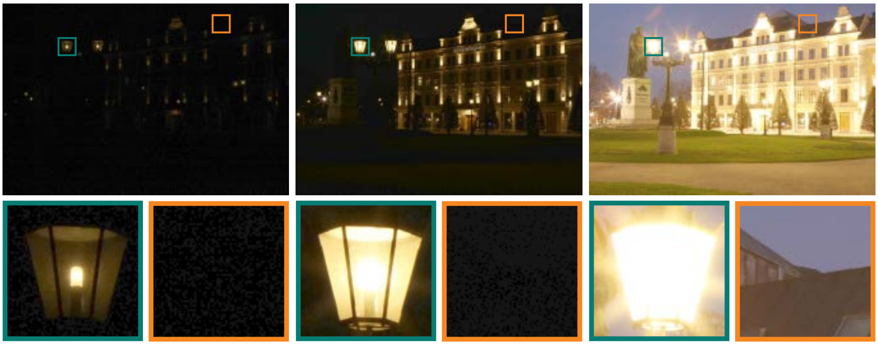

(a) Exp.: 1/180s, -5.8 stops
(b) Exp.: 0.3s, ±0 stops
(c) Exp.: 20s, +6.1 stops

Figure 1.1: An HDR image can capture the full range of luminances in the scene. The top row shows 3 of the in total 7 exposure bracketed images used to create the HDR image in Figure 1.3. The bottom row shows enlarged bright and dark image areas. The numbers specify absolute exposure times, as well as the relative exposures in relation to **(b)**. The example demonstrates that a very large difference in exposure is required in order to capture both highlights **(a)** and details of shadowed image regions **(c)**, and there are still some saturated pixels in the brightest highlights of the darkest image.

From the literature in HDR imaging, it is not exactly clear what the definition of high dynamic range is and it may vary depending on the application. The term is generally used for anything that has larger dynamic range than the conventional cameras/displays. In some cases this may be misleading though, where an HDR image actually can have a rather limited dynamic range. To denote images that are not HDR, the terms *low dynamic range* (LDR) or *standard dynamic range* (SDR) are used interchangeably.

1.1.2 The dynamic range of the HVS

Figure 1.2 shows typical dynamic ranges in order to compare the capabilities of the HVS to different capturing and display techniques. The HVS can observe a very large range of luminances, from around 10^{-6} cd/m^2 up to 10^8 cd/m^2, for a total dynamic range of ≈14 $\log_{10}$ units [93]. However, in order to do so the eye needs to adapt to the different lighting situations. This is achieved partly by changing pupil size, but mostly from bleaching and regeneration processes in the photoreceptors. The processes can take considerable time, especially for regeneration of photopigment when adapting to a dark environment. This is evident for example when transitioning from a bright outdoor environment into a dark room – it takes several minutes before details can be discerned, and up to

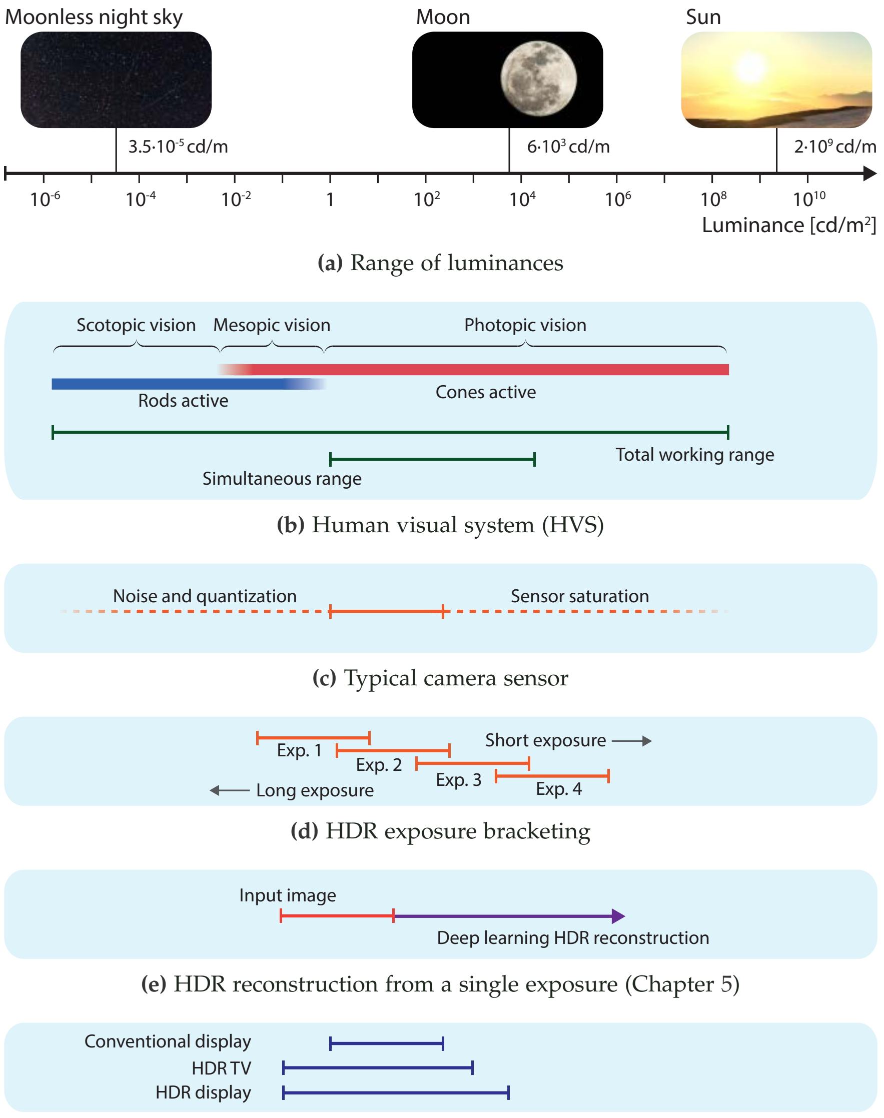
30 minutes for complete dark adaptation. There are two types of photoreceptors on the retina, which are active in different ranges of luminances. The *rods* are more sensitive, but provide poor acuity and no color vision, while the *cones* are active in brighter environments and give colors and higher resolution. The working ranges of the different photoreceptors are illustrated in Figure 1.2b. The range over which only rods are active is termed the *scotopic* vision, and when the rods have saturated only the cones are responsible for the *photopic* vision. There is a significant overlap in the working ranges, where both rods and cones contribute, which is the *mesopic* vision.

The simultaneous dynamic range of the eye, which also is illustrated in Figure 1.2b, is difficult to quantify due to the complexity of how the HVS operates. The response range of the individual neural units is limited to around 1.5 $\log_{10}$ units [232]. However, adaptation can be restricted to an area of less than 0.5 visual degrees [251], so that the effective dynamic range over the observed scene is larger, around 3.7 $\log_{10}$ units [141, 207]. Moreover, we constantly use saccadic eye movements, and adapt to the lighting close to the focal point both in focus and exposure. This means that the perceived dynamic range can be much larger than the actual simultaneous dynamic range of the retinal image.

1.1.3 Camera and display dynamic range

The dynamic range of a camera sensor can vary greatly, from just over 2 $\log_{10}$ units in compact digital cameras, above 4 $\log_{10}$ units for high-end *digital single-lens reflex* (DSLR) cameras, and up to 5 $\log_{10}$ units for professional HDR capable cinematographic video cameras. Figure 1.2c illustrates the dynamic range for a typical consumer level camera sensor. Luminances above the highest measurable value for the current exposure time cannot be registered since the sensor has saturated. Information below the lowest detectable value is lost due to noise and quantization. This means that the dynamic range can actually extend to a lower point on the luminance axis, but these values only contain noise and do not carry any information. The difference in dynamic range between sensors is mainly due to the ability to handle noise, where e.g. a large sensor with low resolution can reduce the noise level by integrating over the larger pixel areas. The noise floor of a sensor can be measured in different ways, and the numbers reported by manufacturers tend to be very optimistic. This means that the dynamic ranges specified above, with up to 5 $\log_{10}$ units, can be difficult to achieve in practice.

In order to capture an HDR image, a set of different exposures can be combined into one image using methods for *HDR reconstruction*. Figure 1.2d illustrates how the dynamic range can be extended in this way. Another strategy for extending the dynamic range is illustrated in Figure 1.2e. It relies on only one



(a) Range of luminances

(b) Human visual system (HVS)

(c) Typical camera sensor

(d) HDR exposure bracketing

(e) HDR reconstruction from a single exposure (Chapter 5)

(f) Different display devices

Figure 1.2: Dynamic ranges of different capturing and display techniques. The axis in **(a)** shows a range of luminances, together with some example scenes for reference. **(b)**-**(f)** show typical dynamic ranges in relation to the axis in **(a)**.

single exposure, and the bright image areas are reconstructed by means of deep learning techniques. This is the topic of Chapter 5.

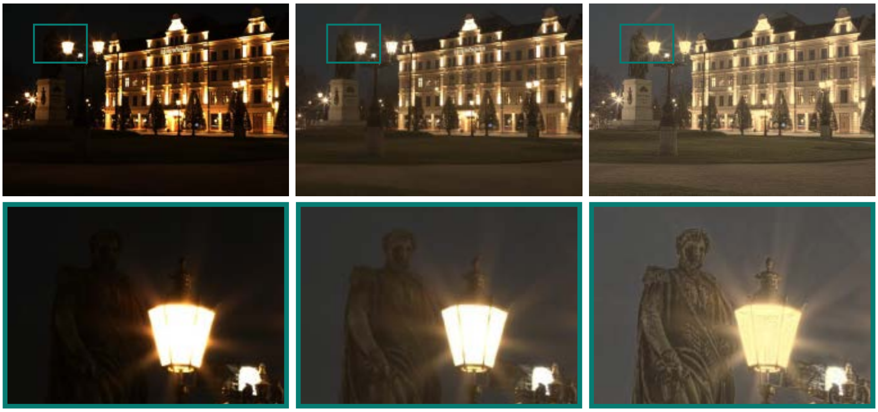
Finally, Figure 1.2f illustrates the typical dynamic ranges of some display devices. For a conventional *liquid-crystal display* (LCD) it is around 2.3-2.7 $\log_{10}$ units, which approximately matches the dynamic range of a consumer level camera sensor, Figure 1.2c. However, when the dynamic range of the image is much higher than the display device, image details are lost in shadows or highlights when displayed. By applying methods for *tone-mapping*, using *tone-mapping operators* (TMOs), the dynamic range of the image can be compressed to match the display while retaining most of the details. An example of the differences between directly displaying an HDR image and by applying a TMO is shown in Figure 1.3. Tone-mapping is not only applicable for the purpose of mapping an HDR image to a conventional display. It can also be used to account for smaller differences in dynamic range and color capabilities of cameras and displays.

For displays, the dynamic range is not the only important feature for supporting HDR material. For example, an *organic light emitting diode* (OLED) screen can have a very large dynamic range even though the peak luminance is equivalent or less than in a conventional LCD device. This is possible due to the very low black level, which in principle can be 0. However, if HDR content is scaled to fit within this range, a large portion of the luminance range will be in the dark image regions, and even in the rod-mediated scotopic vision range. This results in a loss in acuity and color vision in the perceived image. It is probably also not true to nature, so that the displayed luminance is substantially lower than in the captured scene and thus not intended for scotopic vision. Moreover, the display is very sensitive to ambient lighting, so that the dynamic range is drastically decreased as soon as some light is reflected on the screen.

1.1.4 Calibration

Most of the existing digital images are stored using 8-bit integer values, providing $2^8 = 256$ different levels for representing the intensity of each color channel in a pixel. HDR images, on the other hand, are typically stored using a floating point representation, allowing for greater precision and representational power, with a substantial increase in the range of possible brightnesses and colors. However, the differences in dynamic range and precision between HDR and LDR images are not the only aspects when comparing the formats. There is also a fundamental difference in how the formats are calibrated.

Since a conventional digital LDR image almost exclusively is meant to be displayed in one way or the other (monitor, projector, printed paper, etc.), it is calibrated for this purpose. We refer to this format as *display-referred* images. Typically, the calibration includes a *gamma correction*, $l = L^{1/\gamma}$, which performs a



(a) Linear **(b)** Gamma corrected **(c)** Tone-mapped

Figure 1.3: Difference between scene-referred linear values **(a)**, gamma corrected display-referred pixels with $\gamma = 2.2$ **(b)**, and a locally tone-mapped image **(c)**, using the method from Paper **C**. The tone-mapping can compress the dynamic range considerably, while retaining local contrast by means of local processing.

non-linear correction of the linear luminance L in order to generate the final *luma* value l that should be encoded and sent to the display. The gamma value is usually in the range $\gamma \in [1.8, 2.8]$, performing a compression of the dynamic range. Originally, this correction was intended to compensate for the non-linearity of *cathode ray tube* (CRT) displays, but it is also used for modern displays by simulating the non-linearity. This is because the correction also compensates for a similar non-linearity of the HVS within the range of LDR image intensities, so that the range of encoded values is closer to linear from a perceptual standpoint. This means that when encoding an image at the limited precision provided from 8 bits, the *quantization errors* due to rounding off to the nearest representable value, will be perceived as equally large across the range of pixel values. From applying the correction before encoding, and undoing it on the display side, the 256 values are in general enough to make the quantization errors invisible, i.e. it is not possible to distinguish between pixel value l and $l + 1/255$ for any value $l \in [0, 1]$. As the gamma correction in this way relates to perceived brightness, it may be considered a simple form of tone-mapping for LDR images.

The gamma correction operation can also be extended to account for the display and viewing environment, with the gamma-offset-gain model [34, 175],

$$L_d(l) = l^{\gamma} \cdot (L_{max} - L_{black}) + L_{black} + L_{refl} . \tag{1.1}$$

It models the final luminance L_d emitted from the display surface, as a function of the luma value $l \in [0,1]$, taking into account the display characteristics and the ambient lighting of the surrounding environment where the display is used. The display is characterized by its minimum and maximum luminance; the black level L_{black} and the peak luminance L_{max}, respectively. The ambient lighting affects L_d as it is reflected off the display surface, L_{refl}. This term can be approximated given the measured ambient lighting E_{amb} (in lux) and the reflectivity k of the display,

$$L_{refl} = \frac{k}{\pi} E_{amb} . \tag{1.2}$$

By inverting the gamma-offset-gain model, a display-referred calibration that accounts for the particular display and viewing environment can be made.

For digital cameras, the captured image is usually calibrated in-camera, before encoding. Depending on camera brand and model, the non-linear calibration, or *camera response function* (CRF), may have different shapes and accomplishes different calibration/tone-mapping results. For example, one camera can apply a larger compression of the dynamic range in order to reveal more of the RAW pixels captured by the sensor, while another accomplishes better contrast reproduction. In order to allow for more flexibility, most modern DSLR cameras provide an option to directly access the linear RAW sensor read-out, so that it can be prepared for display in post-processing. The RAW image is stored at an increased bit-depth, typically 12-14 bits, and can contain a wider dynamic range as compared to the display-referred 8-bit image.

In contrast to the LDR image format, HDR images are not meant to be sent directly to a display device. Instead, the calibration is *scene-referred*, so that pixel values relate to the physical lighting in the captured scene, by measuring the linear relative luminance. Apart from the high dynamic range and precision provided, the linearity of pixel values is the most essential attribute of HDR images.

In techniques for generating HDR images from conventional cameras, either the linear RAW images can be used, or the non-linear transformation applied by the CRF needs to be estimated and inverted. An absolute calibration of the pixels, though, is more difficult to achieve. It depends on a large set of camera parameters, including exposure time, aperture, gain, etc., as well as the imaging sensor itself. One option for providing absolute calibration is to use a luminance meter for measuring a reference point within the captured scene, and subsequently scale the relative luminances of the HDR image in order to correspond with the measurement.

Given the different domains of display and scene calibrated images, the process of preparing an HDR image for display – or tone-mapping – involves not only compression of the dynamic range, but also a transformation from a scene-referred to a display-referred format. The effect of using gamma correction in order to transform to a display-referred format is demonstrated in Figure 1.3. The correction compresses the dynamic range so that more of both shadows and highlights can be displayed. Even more of the image information can be made visible by also using a tone-mapping operator, which provides a result that is closer to how the HVS would perceive the real scene.

1.1.5 Applications

In addition to improving the direct viewing experience, on HDR displays or by means of tone-mapping, HDR imaging is useful in a number of other applications. As HDR techniques can capture the full range of luminances in a scene, an HDR image can represent a photometric measurement of the physical lighting incident on the camera plane. This information is important for example in *image-based lighting* (IBL) [60, 247], where an HDR panorama is used as lighting when synthesizing photo-realistic images in *computer-generated imagery* (CGI). IBL is often used within the *visual effects* (VFX) industry, where an HDR panorama can be captured at a position in a filmed shot and subsequently used to insert computer graphics generated image content that complies with the lighting in the shot.

In general, HDR imaging can be used whenever accurate physical measurements, or information across a larger range of luminances, are needed for processing or information visualization. This can be the case in automotive applications and other computer vision tasks, medical imaging, simulations, virtual reality, surveillance, to name a few.

Although HDR imaging has been used frequently for many years in research and industry/production, within the last couple of years it has also reached major applications for the consumer market. In the TV industry, HDR is the latest buzzword, and an abundance of HDR capable TVs are now available from a number of manufacturers. Although these devices cannot match the dynamic range of previous research prototypes [223], they offer a significantly extended range of luminances and higher peak luminance, as compared to earlier TV models. The introduction of HDR TV has also pushed forward techniques for distribution of HDR video, and a standardization process is currently ongoing [94]. Major online streaming services (Netflix, Youtube, Vimeo, Amazon Prime Video, etc.) have also started to introduce HDR video in order to provide material for the HDR TVs. Considering this recent development, the topics within this thesis are ever so important, and contributions are presented for both generation, distribution, and display of HDR images and video.

1.2 Context

Clearly, the increasing applicability of HDR images and video will make for higher demands on robust techniques for creation, distribution, and display of the format in the future. This thesis contributes to the field of HDR imaging in three different areas. These are the software components of the HDR imaging pipeline; reconstruction, distribution, and tone-mapping, as illustrated in Figure 1.4. The papers that the thesis is built on are listed on page vii in the preface and their individual contributions on page ix. In order to give a clear motivation for the thesis within the HDR imaging pipeline, in what follows are brief descriptions of the papers in the context of the three aforementioned areas:

- **Tone-mapping (Paper A, B, C):** This is the largest area of contribution, with three papers that help in advancing techniques for tone-mapping of HDR video material. The work started with Paper B, which demonstrates an evaluation of the, at the time, existing methods for tone-mapping of HDR video. The evaluation reveals a number of issues with the TMOs, such as loss in local contrast or temporal artifacts and increased visibility of noise. Paper B is used as a starting point for the techniques presented in Paper C. This paper proposes a novel real-time tone-mapping operator that can achieve high local contrast with a minimal amount of spatial and temporal artifacts. It also considers the noise characteristics of the input HDR video in order to make sure that the noise level of the tone-mapped video is below what can be discriminated by the HVS. Finally, in Paper A we recognize that existing literature that describes the area of tone-mapping is getting outdated, and do not cover the recent developments related to video tone-mapping. The paper presents a thorough literature review on tone-mapping in general, and especially focusing on HDR video. It provides descriptions and categorization of the state-of-the-art in video tone-mapping, as well as a quantitative evaluation of their expected performances. The assessment indicates that many of the problems found in the evaluation in Paper B have been resolved in the most recent TMOs, including the method in Paper C.
- **Distribution (Paper D):** HDR video can be stored with existing techniques for LDR video compression, by encoding at a higher bit-depth. In order to do so, the HDR pixels need to be mapped to the available bit-depth. A number of techniques for this mapping have been proposed, but lack in comparison. Paper D makes a large-scale comparison of such techniques, as well as different color spaces used for encoding. The paper also presents *Luma HDRv*, which is the first open-source library for HDR video encoding and decoding. The library is accompanied with applications for encoding and decoding, as well as an application programming interface (API) for easy integration in software development.

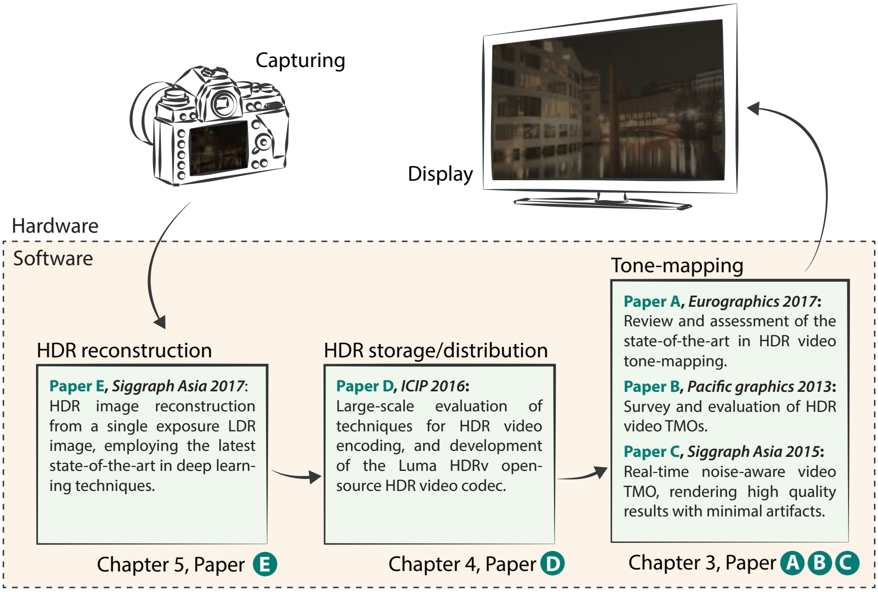

Capturing
Display
Hardware
Software
Tone-mapping
Paper A, Eurographics 2017: Review and assessment of the state-of-the-art in HDR video tone-mapping.
Paper B, Pacific graphics 2013: Survey and evaluation of HDR video TMOs.
Paper C, Siggraph Asia 2015: Real-time noise-aware video TMO, rendering high quality results with minimal artifacts.
HDR reconstruction
Paper E, Siggraph Asia 2017: HDR image reconstruction from a single exposure LDR image, employing the latest state-of-the-art in deep learning techniques.
HDR storage/distribution
Paper D, ICIP 2016: Large-scale evaluation of techniques for HDR video encoding, and development of the Luma HDRv open-source HDR video codec.
Chapter 5, Paper E
Chapter 4, Paper D
Chapter 3, Paper A B C

Figure 1.4: Brief summary of the thesis contributions, where the individual papers are listed in context of the HDR imaging pipeline. Contributions are made in each of the software components of the pipeline. A more general illustration of the pipeline is provided in Figure 2.1 in Chapter 2.

- **Reconstruction (Paper E):** With increasing popularity of HDR image applications, but limited availability of HDR image material, an interesting topic is how to enable using LDR images in these applications. A number of methods for this purpose have been presented, labeled *inverse tone-mapping operators* (iTMOs). However, these are very limited as they boost the dynamic range without really reconstructing the missing information in the LDR images. In Paper E we present an HDR reconstruction method that uses recent advancements in deep learning in order to reconstruct saturated regions of an LDR image. The method shows a substantial improvement over existing techniques and makes it possible to use LDR images in a wider range of HDR applications than was previously possible.

Although the thesis work considers three different aspects of HDR images, in the HDR imaging pipeline these are closely inter-linked, as demonstrated in Figure 1.4. A possible scenario for using the contributions in connection could, for example, be to enable compatibility with existing LDR image material in HDR streaming. First, the single exposure method in Paper E can be used to transform the LDR material into HDR. The HDR video stream is then possible to distribute with the Luma HDRv codec in Paper **D**, which allows for open-source development. Finally, the techniques in Paper **C** can adapt the HDR stream to a certain HDR display, or compress the dynamic range in a fast and robust manner to be displayed in high-quality on a conventional LDR monitor.

1.3 Author's contributions

The work that is presented in this thesis has been performed in collaboration with a number of co-authors. In order to clarify the individual contributions from the author of the thesis, in what follows are brief descriptions of the author's work related to each of the papers:

- **Paper A**: The report is an individual work and literature study, written in a first draft by the author. The final publication has the same content, but was complemented, rearranged, and rephrased to a smaller extent after feedback from the co-authors.
- **Paper B**: The author implemented a number of methods for evaluation and conducted major parts of the experiments. The author took part in analyzing the outcome of the experiments, and in extracting general problems with existing methods for tone-mapping. The paper was written in a collaborative effort with the co-authors.
- **Paper C**: The author implemented the complete tone-mapping operator for execution on the GPU and together with a graphical user interface. The filtering method described in the paper was formulated by the author, while ideas and initial implementations of the tone-curve were provided by a co-author. The author conducted the comparison study and produced the results. For the paper, the author wrote most of the filtering and result sections, and helped in writing other parts.
- **Paper D**: The author implemented the Luma HDRv codec library and API. The author conducted the testing on a large-scale computer cluster, with guidelines and functions for making comparisons provided by a co-author. The results were put together by the author. The paper was written by the author, followed by feedback and complementing text by co-authors.
- **Paper E**: The author was responsible for the idea, design, implementation, training, putting together results, and writing of the paper. Co-authors helped in coming up with suitable deep learning architectures and training strategies, some initial implementation, and evaluation of the results on an HDR display. The author did most of the paper writing, and co-authors complemented the text and wrote the section on evaluation using an HDR display.

1.4 Disposition

This introductory chapter intended to introduce, define and motivate the field of HDR imaging. It also briefly described and contextualized the contributions provided in the thesis. The upcoming chapters will provide a more thorough background on HDR imaging and discuss the work presented in the different thesis papers. These chapters constitute the first part of the thesis. The second part is composed of the five selected papers that have been published within the scope of the thesis work.

A general background and related work of the field of HDR imaging is provided in Chapter 2, in the context of the HDR imaging pipeline. To this end, the different components of the pipeline are discussed in turn; capturing, reconstruction, distribution, tone-mapping, and display.

In Chapter 3, the context, content, and contributions of the papers considering tone-mapping are described. This work makes specific considerations for HDR video and the implications of tone-mapping of temporally varying data. First, in Section 3.2 a subjective evaluation of different methods for video tone-mapping is described (Paper **B**). In Section 3.3 this is followed by a presentation of a video TMO that uses a set of novelties in order to enable robust and high-quality tone-mapping (Paper **C**). In Section 3.4, a set of quantitative experiments are explained, which intend to point to which video TMOs can be expected to render a good level of exposure and contrast, with the least amount of artifacts (Paper **A**). For this part of the thesis, Paper **A** should also be considered a background description and a literature review, which categorizes and describes the state-of-the-art in tone-mapping for HDR video.

Chapter 4 treats storage and distribution of HDR video. It describes a large-scale objective evaluation of the techniques involved in preparing HDR video for encoding (Paper **D**). It also presents the Luma HDRv codec, which is built taking into consideration the results of the evaluation.

Chapter 5 deals with the problem of reconstructing HDR image information from a single-exposed LDR image. A method that uses deep learning techniques in order to predict the HDR values of saturated pixels is described and discussed (Paper **E**). It makes use of a convolutional neural network that is designed and trained with special consideration of the challenges in predicting HDR pixels.

Finally, Chapter 6 provides a unified summary of the work and contributions. The chapter, and the thesis in its whole, is then wrapped up by an outlook towards the future of HDR imaging, with possible directions for research and development.

Chapter 2

Background

The HDR imaging pipeline, from capturing to display, is illustrated in Figure 2.1. The physical scene can be exposed onto one or more imaging sensors, followed by processing the captured information using techniques for HDR reconstruction (Section 2.2). Alternatively, an HDR camera can be used in order to directly infer an HDR image, either with a sensor that can cover a large dynamic range or with a multi-exposure system (Section 2.1). The captured HDR image or video sequence is then stored using some HDR capable format, where a variety of different solutions have been proposed for both static images and video (Section 2.3). The next step in the pipeline is to prepare the HDR image for display, using a tone-mapping algorithm (Section 2.4). The objective is to compress the dynamic range to the constrained range of the display while retaining visual image information, and to transform the image to a display-referred format. The final component in the pipeline is the actual display of the tone-mapped image, either on an HDR capable display (Section 2.5) or on a conventional monitor.

This chapter will discuss the five components of the HDR imaging pipeline in Figure 2.1: capturing, reconstruction, distribution, tone-mapping, and display. The presentation attempts to cover the most important techniques and literature within these individual areas, in order to give a background on research and development in HDR imaging. It also places the individual thesis papers in relation to previous work, demonstrating how they contribute to the area. For a wider description of HDR imaging and its applications, the reader is referred to recent books on the topic, treating HDR imaging in general [28, 175, 211] and specializing on HDR video [49, 71].

2.1 Capturing with HDR cameras

When it comes to HDR cameras, we discern two different techniques for covering a large range of luminances; either with multi-exposure camera systems, or with a single exposure using a sensor that, through some mechanism, has the capability of capturing a much higher dynamic range as compared to conventional sensors.

Strictly speaking, the HDR reconstruction step also takes place when using multi-exposure HDR camera systems, in the same way as for exposure bracketed images when capturing with a conventional camera. However, these systems are dedicated HDR capturing devices where the reconstruction potentially could take place live onboard the camera, as opposed to using a conventional camera where this is an explicit post-processing operation. Consequently, we categorize the versatile multi-exposure systems as HDR cameras that directly output HDR images.

2.1.1 Single-exposure HDR cameras

The most capable single-exposure cameras, in terms of the specified dynamic range, can be found in the film industry. The increased dynamic range of a high-end cinematographic camera can partly be attributed to the large size and production quality of the sensor, which makes for a reduction in the noise floor of the captured image. There may also be additional techniques used in order to boost the dynamic range, for example by employing dual gain readouts. However, these details of the camera construction and capturing techniques are not always specified for commercial cameras.

The camera manufacturing company RED has probably had the most impact during the last decade, starting with their first model RED ONE in 2007. In 2013 they released the RED Epic Dragon, with at that time incredible specifications and a dynamic range that was claimed to be more than 16.5 stops ($\approx$5 $\log_{10}$ units). A major impact has also been from manufacturer ARRI with their Alexa model. The camera features a dual gain architecture (DGA), which makes use of two gain readouts from each pixel on the sensor in order to boost the achievable dynamic range, for a total of 14 stops according to the manufacturer.

There has also been a large development in cinematographic cameras within the last years, possibly spurred by increasing demands with the establishment of HDR TVs. RED introduced the Helium 8K sensor in 2016 and the Monstro 8K large-format sensor in 2017 (although only slightly larger area than a traditional full-format sensor), which is claimed to have a dynamic range of above 17 stops. Together with the recent camera body called Weapon, the latest flagship from RED is the Weapon Monstro 8K VV. A recently upcoming contender –

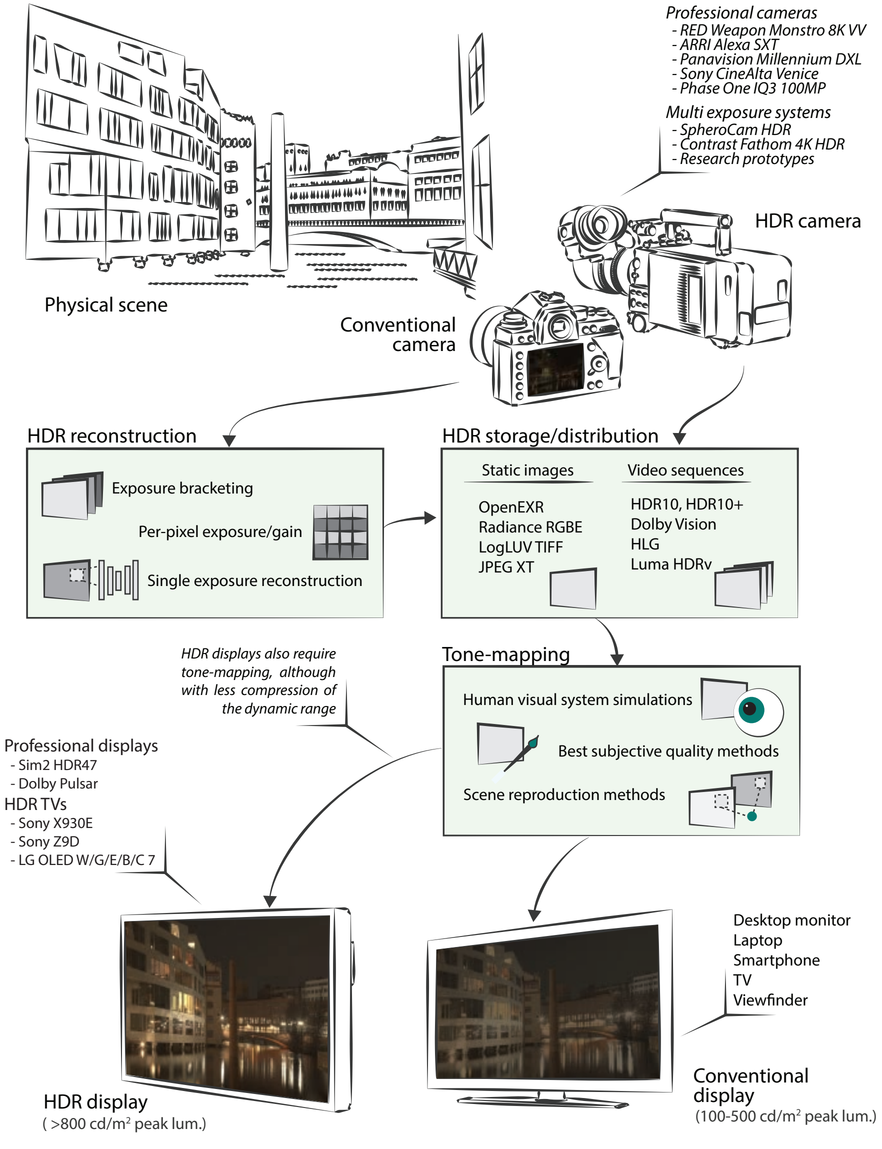


Figure 2.1: The HDR imaging pipeline, from capturing to display. The three intermediate blocks represent the software section of the pipeline.

and allegedly a superior camera in terms of many technical aspects for the production environment – is a joint effort by Panavision, RED, and Light Iron in order to create the top-of-the-line cinematographic camera Panavision Millennium DXL. This device also features an 8K large-format sensor, which is specified to have a dynamic range of 15 stops. Sony has also recently announced a top-segment cinematographic camera; the Sony CineAlta Venice, which is the manufacturer's next flagship after the F65 model. The camera is scheduled for release in early 2018. It is equipped with a 6K full-frame sensor, with a 15 stop dynamic range according to the specifications.

In addition to the high-end cinematographic cameras, there has also been a segment of more affordable alternatives presented within the last couple of years. These include, but are certainly not limited to, the Grass Valley LDX 82, the Kinefinity KineMAX, and the Blackmagic Ursa. The dynamic capabilities are specified in the range of 15-16 stops according to the manufacturers.

In common for the cinematographic video cameras are specified dynamic ranges between 14-17 stops, which is significantly higher than in conventional cameras. However, the measured dynamic range is highly dependent on the specific measurement procedure, and the manufacturers' numbers tend to be in optimal conditions. This means that the specified dynamic ranges can be difficult to reproduce in practice.

The high-end segment of DSLR cameras is also expected to be close to the cinematographic devices in terms of dynamic range. There is a trade-off between pixel size and dynamic range, as larger pixels allow for lower noise level and traditionally DSLRs have had higher resolution than cinema cameras. However, this is not always the case anymore, with cinema cameras supporting 8K (≈35 megapixels). Among the abundance of high-end DSLRs, two notable examples are the Sony α7R III and the Phase One IQ3 100MP. Sony α7R III uses a full-format sensor and is known for its good noise characteristics. The large-format sensor in Phase One IQ3 should definitely be in the same category as the high-end cinema cameras considering the larger sensor (53.7 x 40.4mm) and its high resolution (101 megapixels). According to the manufacturers, both these cameras are able to capture a dynamic range of 15 stops. However, in the tests carried out by Photons to Photos the Sony and the Phase One cameras were measured to have dynamic ranges of 11.65 and 13.06 stops, respectively [202]. This highlights the problem of reproducibility of manufacturers' dynamic range specifications.

There are also alternative sensor techniques that enable coverage of a significantly larger dynamic range, but which impose other forms of limitations. For example, *log sensors* are able to extend the range of captured luminances by having a logarithmic dependence between the light incident on a pixel and the photo-voltage induced by the photons. However, these have limited resolution and weak low-light performance, with high levels of *fixed pattern noise* (FPN) [128]. As such, log sensors are typically used for machine vision and surveillance applications, but are too limited for e.g. feature film. One example is the Photonfocus HD1-D1312, with a 1.4 megapixel CMOS sensor that features a logarithmic capturing mode that can achieve a dynamic range of around 120 dB (≈20 stops). There are also examples of sensors that use locally adaptive exposures in order to capture a high dynamic range of linear values. In the Silicon Vision LARS III (Lokal-AutoadaptiveR Sensor) [158], the integration time of each pixel is individually and automatically controlled. If a pixel exceeds a certain reference voltage the integration is terminated, preventing saturation of the pixel. The sensor technology alleviates the problems with FPN, but the resolution is limited to 0.37 megapixels. Another type of special purpose sensor is used in so-called *event-based cameras* [155]. These capture the temporal derivatives, with pixels that trigger based on relative changes in intensity, and which are read as an asynchronous stream. HDR images can then be produced from integration over time, but as with log sensors the limitations mean that the main applications are within computer vision.

In summary, there exists a multitude of both cinematographic cameras and DSLR cameras that qualify into the category of single-sensor HDR – or extended dynamic range – capturing devices, which extends up to approximately 17 stops of dynamic range. This is enough to cover the dynamic range needed for e.g. HDR TV devices, and makes extensive post-processing possible. Alternative sensor techniques, on the other hand, can capture a larger dynamic range of around 20 stops, but are limited to e.g. computer vision applications.

2.1.2 Multi-exposure HDR camera systems

In order to capture a dynamic range of ≥20 stops at high resolution and quality, multi-exposure techniques are still required. This large range of luminances is for example often needed for IBL, and in other applications that demand accurate photometric measurements.

There are a number of special purpose HDR cameras commercially available, which can capture static scenes with a very high dynamic range and resolution, in order to provide accurate measurement for e.g. IBL. These include devices such as Spheron SpheroCam HDR, Weiss AG Civetta, and Panoscan MK-3. For example, the SpheroCam HDR can capture a dynamic range of 26 stops at a horizontal resolution of up to 100K pixels. The device rotates and captures vertical scanlines with different exposures, which are combined into a final HDR panorama.

Also, many conventional cameras now have specific multi-exposure HDR capturing modes implemented. This goes both for more expensive DSLRs and low-end cameras such as in smartphone devices. While the HDR capturing techniques can vary, the typical approach is to complement with some additional exposures, both shorter and longer than the current exposure. After capture, and onboard the device, the exposures are aligned and fused to an HDR image. Alternatively, a burst of images with short exposure times can be combined to improve noise level and dynamic range, such as in Google's HDR+ software [112]. With state-of-the-art techniques in image registration, deghosting, and machine learning, these methods can achieve good results in a variety of situations, including scenes with moderate amounts of motion. However, for video sequences or scenes with fast motions, alternative techniques are required.

The most challenging scenario is capturing of HDR video in high resolution and quality using multiple exposures. A number of techniques have been demonstrated for this purpose [95, 126, 127, 163, 245, 246, 254]. These will be closer examined in Section 2.2. However, only a few truly versatile multi-exposure HDR video camera systems have been built. One example is the prototype developed in collaboration between SpheronVR and the University of Warwick [48]. It uses a single lens and partitions the incoming light onto multiple sensors by means of a beam splitter arrangement. The system captures 30 frames per second at 1920×1080 pixels resolution and a dynamic range of around 20 stops. Contrast Optical's amp HDR prototype, presented by Tocci et al. [236], also splits the incoming light onto 1920×1080 pixel resolution sensors. A common approach with this technique is to place *neutral-density* (ND) filters in front of the sensors in order to absorb light and thus simulate different exposures. This means that not all incoming light contributes to the final image. However, the amp HDR system is able to make use of 99.96% of the incoming light, exposed on 3 sensors, by reusing the majority of the light that is transmitted and reflected by the beam splitters. The dynamic range of the prototype was measured to 17 stops. Recently, the technology has been incorporated in the commercialized Fathom 4K HDR camera, specified to have a dynamic range of 13 stops and 4912×3684 pixels resolution [53]. Another example prototype, shown in Figure 2.2, was developed in collaboration between Linköping University and SpheronVR [135, 136]. It utilizes 4 sensors, differently exposed through the same lens using beam splitters and ND filters. The device can capture a dynamic range of 24 stops at 2336×1752 pixels resolution. For HDR reconstruction from the sensor data, a unified approach is proposed, which considers debayering, denoising, alignment, and exposure fusion as a single operation, in order to improve quality and to enable real-time performance.

Finally, in addition to Contrast's Fathom HDR camera, there are already a number of devices commercially available that employ multiple sensors, but which combine the sensory data for other purposes than HDR. For example, the Light L16 camera has in total 16 individual sensors and lenses. The different

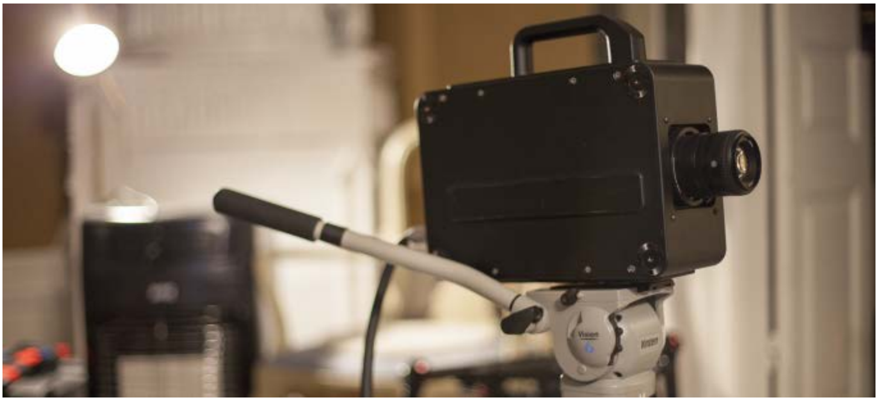


Figure 2.2: Multi-sensor HDR video camera developed in collaboration between Linköping University and SpheronVR [136], capable of capturing a simultaneous dynamic range of up to 24 stops.

images are combined in order to enable a higher resolution and quality than is possible with the individual sensors, and to provide options for changing the focal length without using moving optical elements. The camera could potentially also be modified to use different exposures, in order to enable HDR capturing. Furthermore, multi-sensor cameras are also popular in surveillance and virtual reality, for the purpose of capturing panoramic images. For example, the Axis Q3708-PVE uses 3 sensors for covering a 180 degrees field of view in video surveillance. Notably, this camera also has a feature termed "Forensic WDR", which employs a dual gain setup for increasing the dynamic range [16]. For multi-sensor HDR capture, however, the different lenses have to be adjusted to a common image plane. Given that commercial multi-sensor devices are increasing in number, alternatives for HDR video capturing using such techniques will most likely become common in the near future.

2.2 HDR reconstruction from conventional sensors

Techniques for combining multiple exposures from conventional sensors, in order to infer an HDR image, have been around for well over 20 years [61, 160, 164]. A large number of methods have been proposed, both for capturing different exposures and for how to combine these. We distinguish between the ones that use altering exposures over time and those that perform the multiplexing in the spatial domain. Additionally, there are also techniques that only consider one single exposure, in order to transform a conventional LDR image for use in HDR applications.

While multiple exposures can be captured and registered in many different ways, the problem of optimal fusion of the final pixel values – or HDR reconstruction – is similar in most methods. The problem is generally composed of two distinct steps. First, the *camera response function* (CRF) needs to be estimated and inverted in order to derive pixel values that are linearly dependent on captured luminances [61, 186, 217]. In modern DSLR cameras, however, it is possible to directly access the linear RAW sensor read-out, which is stored at an increased bit-depth (usually 12-14 bits). Second, the set of differently exposed linear pixel values should be combined, possibly accounting for the specific characteristics of the sensor. Trivial methods for HDR fusion include picking one single exposure per pixel [160] or using a simple triangular filter [61]. Other methods extend to use a weighting based on the response function derivative, in order to avoid quantization errors [164], or assuming certain noise behaviors [186, 217]. Tsin et al. [240] developed on the influence of noise in the reconstruction, with a weighting that is based on the standard deviation from a camera noise model. Later methods combine exposures using variance estimates from more comprehensive sensor noise models, such as the weighting proposed by Granados et al. [103]. More recent methods also perform the different image reconstruction steps (demosaicing, denoising, alignment, exposure fusion) as a unified single operation [109, 114, 135, 136].

2.2.1 Temporally multiplexed exposures

The classical technique for HDR image acquisition is to capture a set of differently exposed images one after the other [61, 160, 164]. Without additional processing, however, the capturing is limited to static scenes, and results in ghosting artifacts if this is not the case. In order to handle small amounts of camera shake/motion, the exposures need to be globally registered [165, 238, 259]. For dynamic scenes the problem is more difficult, requiring registration on a local level. There is a large body of work on HDR image registration and deghosting, facilitating HDR exposure bracketing of dynamic scenes. For example, for per-pixel registration optical flow can be used [108, 125, 280], or patch-based approaches [118, 224]. For a thorough survey and categorization, we refer to the state-of-the-art report by Tursun et al. [244].

A number of attempts have been made for reconstructing HDR video using temporal exposure multiplexing. The typical scenario is to use two different exposure times and alternate between these for every frame. Subsequently, the inter-frame correspondences are estimated on a local level, so that information from multiple exposures can be provided for reconstruction in each frame. The first method to exploit this scheme was proposed by Kang et al. [127], where optical flow is employed for registration of the different exposures. More recently, Mangiat and Gibson [163] demonstrated improved reconstruction performance by using block-based motion estimation followed by motion refinement and cross-bilateral filtering. Kalantari et al. [126] combined optical flow with a patch-based matching strategy, which improves reconstruction in regions of fast motion as compared to the previous methods.

2.2.2 Spatially multiplexed exposures

In order to overcome the inherent problems with a time multiplexed HDR capturing method, different exposures can be captured in the same shot by varying the exposure spatially. All such techniques can potentially be used for HDR video capturing, since dynamic scenes can be recorded without the need for complicated local registration of the different exposures.

Multiple exposures can be captured simultaneously in three levels of spatial separation. First, multiple separate camera devices can be used to capture the same scene. Second, multiple sensors can capture the scene through the same single lens. Third, different exposures can be interleaved, or spatially encoded in some other arrangement, on the same single sensor. A number of techniques within these categories are described next.

Multi-camera methods: Combining images from multiple separate cameras provides a relatively inexpensive alternative for spatial exposure multiplexing. For example, there are a number of methods that exploit stereo camera capturing rigs, where the two cameras are set to capture different exposures [31, 156, 235, 239]. The images require reliable stereo matching in order to align the separated views of the cameras. This problem can be overcome by aligning the camera views through an external beam-splitter. Fröhlich et al. [95] captured a wide variety of HDR videos with such setup using two Arri Alexa cameras, achieving a dynamic range of up to 18 stops. A more general framework was presented by McGuire et al. [180], which can capture a unified view with multiple cameras using an optical splitting tree.

Multi-sensor methods: Having multiple cameras with separate optics may be difficult in terms of calibration, where all lenses need to be synchronized for equal view, focus, etc. Moreover, the systems tend to be very bulky and difficult to maneuver. In order to alleviate these problems, the beam-splitter can be placed behind a single lens in a single camera body, where the light is split onto multiple sensors [111]. By restricting light using different ND filters, this setup can capture a stack of exposure bracketed images for HDR image reconstruction [2, 3, 254]. In more recent work, the multi-sensor concept has been extended to provide versatile HDR video camera prototypes [48, 136, 236], as explained in Section 2.1.

Single-sensor methods: While a multi-sensor HDR camera presents an effective alternative for capturing multiple exposures in a single shot, the custom-built systems are expensive. For this reason, a high diversity of techniques have been explored for capturing multiple exposures simultaneously on a single sensor. Some of the techniques require custom-built sensor add-ons, while others can be implemented only with modification in camera software. In common with the methods is that they can potentially be used by existing conventional cameras. However, since multi-exposure information is captured within the same sensor, the increased dynamic range comes at the price of lower resolution, i.e. there is a trade-off between dynamic range and spatial resolution.

The first method for capturing spatially varying exposures in a conventional sensor was presented by Nayar and Mitsunaga [190]. The spatial variations in exposure are accomplished by means of an ND filter mask, where 4 different levels of transmittance are interleaved in a regular pattern over the sensor. The method was later extended to also include a color filter array for HDR color acquisition [189], and by optimizing the particular exposure/color filter layout [270]. Furthermore, there are examples where the layout of the filter mask has been changed to non-regular patterns [4, 221], in order to alleviate problems with interpolation aliasing artifacts. Serrano et al. [225] approached the problem from a different standpoint. Instead of interpolating between spatially varying exposures, the method uses a learned convolutional filter bank that can decode exposure patterns with techniques in convolutional sparse coding. Furthermore, an alternative to the per-pixel ND filter array arrangements is to use a beam-splitter for partitioning incoming light onto different regions of the same sensor [162]. The technique can be realized by an optical element that is inserted between the lens and the camera body of a conventional DSLR camera, and by using a different ND filter for each of the regions of the sensor.

All the above single-sensor techniques rely on ND filters, which inevitably restrict some of the incident light on the sensor. Other techniques for accomplishing spatially varying exposures include per-row modification of the sensor readout. This can be done in order to get rows of different exposure [104] or gain [109, 110, 245] within the same shot. There are also more unconventional techniques for encoding highlight information in an LDR image. Rouf et al. [218] proposed a significantly different form of spatial encoding as compared to per-pixel exposure or gain. The method uses a star filter for capturing, which scatters highlights as one-dimensional streaks in a sparse set of directions. This means that 1D techniques can be applied for decoding the information into one LDR image without the scattered light and one image with recovered highlights. The two images are subsequently combined into an HDR image.

2.2.3 Single-exposure techniques

Single-exposure techniques attempt to extend the dynamic range without requiring information from multiple exposures, nor special equipment or capturing techniques. Hence, methods in this category can be applied to the vast number of existing LDR images and video, facilitating their use in HDR applications. Single-exposure reconstruction can be separated into three distinct sub-problems; decontouring, tone expansion, and reconstruction of under/over-exposed image areas. Additionally, noise is also a highly relevant problem, deteriorating information in the dark areas of the image. However, denoising is a classical and well-researched image processing task [43, 44, 54], and not specific to the single-exposure HDR imaging problem.

Decontouring: LDR pixels are almost exclusively encoded at 8 bits per color channel. When expanding the dynamic range the quantization can potentially reveal visible banding artifacts, viewed on an HDR display or by means of tone-mapping. One method for alleviating the problem is to use a dithering based method, which applies noise in order to conceal the artifacts. The dithering can be performed either before [55] or after [5, 35] the quantization. These methods are intended to conceal false contours at the same bit-depth as the input image. In order to actually increase the bit-depth, there are a number of filtering based methods [56, 150, 159, 229]. For example, the method proposed by Daly and Feng [56] filters the image followed by quantization at the input bit-depth. The difference between filtered and quantized image represents false contours and is subtracted from the input image. Although bit-depth extension methods are limited, they can increase the precision by around 1-2 bits.

Tone expansion: In order to map an LDR image to HDR, the *camera response function* (CRF) needs to be inverted, expanding the dynamic range and mapping the image tones to the linear domain. However, the most common goal for single-exposure HDR techniques is to display LDR images on HDR monitors. Given that the result of the tone expansion E is assessed on an HDR display, it describes a composite mapping $E = V \circ f^{-1}$, where f is the CRF and V represents a tone-mapping operation for the particular HDR display. Furthermore, since it is difficult to reconstruct highlights convincingly, the optimal mapping E may be different than it would be if this information was available. A second common goal is to use the LDR image in IBL. If highlight information is missing, a global boost in brightness generally yields an IBL rendering that is preferred over the otherwise too dark result. Consequently, tone expansion is, in general, a different matter than the inversion of a CRF, and the optimal end result may be very different from the true underlying HDR image.

A method for expanding the dynamic range of LDR images is commonly referred to as an *inverse tone-mapping operator* (iTMO), as introduced by Banterle et al. [24, 25]. However, dynamic range expansion can be traced back to a simple trick presented by Landis [143], for the purpose of using LDR images in IBL. For display of LDR images on HDR displays, a number of perceptual studies have pointed to the fact that a global mapping may be preferred, either using a gamma function [36, 177, 178] or a linear scaling [8, 226].

Under/over-exposure reconstruction: The most difficult problem in inferring an HDR image from a single-exposed image is how to recover lost information in under- and over-exposed areas. Generally, over-exposure is the most significant problem, as the majority of HDR applications require the bright image information but not the dark. A number of iTMOs attempt to alleviate the problem by applying separate expansion to pixels that are classified as saturated. For example, Meylan et al. [182, 183] applied different linear functions in saturated and non-saturated image areas. Banterle et al. [24, 25] used the median cut algorithm in order to derive an *expand map* for boosting highlights. The method was also extended for video processing and with cross-bilateral filtering of the expand map [26]. Another expand map method was presented by Rempel et al. [214], which simplifies the estimation using a Gaussian filter for real-time performance. It was later modified by Kuo et al. for improved robustness [142]. A more recent similar method was described by Kovaleski and Oliveira [132], using a cross-bilateral expand map [131]. The method aims at operating in a wider range of exposures than previous iTMOs. A different approach was proposed by Didyk et al. [64], where a semi-manual classifier separates the image into diffuse, reflections, and light sources. The diffuse part is left untouched, while the other layers are expanded to a wider dynamic range. As compared to the global iTMOs, which expands the dynamic range without explicit consideration of saturated regions, these highlight boosting methods are expected to generate results that more closely resembles the true HDR image. This was also confirmed in a pair-wise comparison experiment performed by Banterle et al. on an HDR display [27]. However, the boosting is a very crude approximation of luminance, and it cannot reconstruct details and colors in saturated image regions.

A second category of methods for correcting over-exposure aims at reconstructing colors and details given statistics of nearby non-saturated pixels. Zhang and Brainard [278] applied Bayesian estimation in order to infer the values of 1-2 saturated color channels of a pixel, given information of the non-saturated channel(s) of the same pixel. Masood et al. [179] extended to use color channel ratios in a neighborhood of the pixel being reconstructed. Furthermore, Guo et al. [105] and Xu et al. [267] also considered reconstruction of pixels with all color channels saturated, and the methods can handle larger areas of missing information. However, all these exposure correction methods are limited in that the dynamic range is extended only by a small amount. High-intensity highlights are not considered, which are essential for HDR reconstruction.

Finally, there are some methods that aim at reconstructing both high intensities, colors, and details of saturated image regions. The method proposed by Wang et al. [255] separates the input image into a high-frequency texture/reflectance layer and a low-frequency illumination layer. Saturated regions in the texture layer are reconstructed by transferring – or inpainting – from similar textured areas in the image. The illumination is approximated by fitting Gaussian lobes to the saturated areas, similar to how highlight boosting is performed in iTMOs. While convincing results can be achieved, the method is limited to textured areas and it requires some manual interaction. More recently, a number of methods employ deep learning strategies for single-exposure HDR image reconstruction [85, 151, 176, 276], including the method of Paper **E** [83]. The paper is discussed in Chapter 5 and related to the other deep learning reconstruction methods in Section 5.2. In summary, the method from Paper **E** can predict high-quality high intensities, colors, and details in a large range of situations, and in a completely automatic fashion. It uses a *convolutional neural network* (CNN) that has been specifically designed considering the characteristics of HDR data, and which is trained on a large augmented database of HDR images. The reconstructions show a substantial improvement in quality over earlier methods and enables the use of LDR images in a wider range of HDR applications than was previously possible.

2.3 HDR distribution

In order to store and distribute HDR images and video, either custom encoding schemes need to be applied or the display-referred HDR pixels can be adapted for encoding with existing algorithms for LDR images/video. When it comes to static images, there are a few floating point pixel formats that have been developed particularly for HDR data. Inter-frame encoding of HDR video, on the other hand, as well as backward-compatible encoding schemes for static images, rely on the use of modifications or extensions of existing codecs for LDR data.

2.3.1 Floating point HDR pixel formats

A natural goal for an HDR image format is to store the linear pixel values with floating point precision, e.g. in the RGB color space. However, assuming 32-bit floating numbers, this means that 96 bits per pixel (bpp) have to be used in order to encode colors. For a 10 megapixel image, this amounts to a file size of 120 MB with no compression applied, which in many situations is unfeasible. For this reason, floating point HDR image formats use reduced pixel descriptions. The two most widely used formats are Radiance RGBE and OpenEXR.

The HDR pixel format used by the Radiance renderer [262] employs the RGBE pixel description introduced by Ward [257]. It stores RGB values with 32 bits; 8 bits mantissa for each color channel, plus an 8-bit common exponent. The common exponent makes the format limited in terms of color saturation, i.e. when there are large differences between color channels. This means that highly saturated colors outside the sRGB color gamut cannot be represented. In order to alleviate the problem, there is also an option to use the XYZE pixel description, which employs the same coding scheme but in the CIE XYZ color space. The final bit stream is stored uncompressed or by means of run-length encoding, which means that the format is lossless up to the particular precision of the pixel representation.

The OpenEXR (EXtended Range) HDR image format [37] was developed by Industrial Light & Magic (ILM). It was released as an open source library in 2003. The format has gained a widespread use, where it for example often is employed in the visual effects industry and in commercial software. Pixels are typically stored with "half" floats, which use 16 bits for each color channel. The bits are allocated for 1 sign bit, 5 exponent bits, and 10 mantissa bits. There are also options for 32-bit floats and 32-bit integers. The pixels can be encoded both by lossy and lossless compression schemes. For example, with ILM's PIZ format, there is a lossless compression to around 35-55% of the uncompressed size, employing Huffman encoding of a wavelet transformed image [124].

As an example of the performances of the two formats, we compute the mean bit-rate of encoding the entire Stuttgart HDR video dataset captured by Fröhlich et al. [95]. This represents a diverse set of scenes in 33 HDR video sequences, with various amount of noise. Thus, it is a good representation of HDR images in general. With RGBE and run-length encoding, the mean bit-rate is 26.52 bpp. OpenEXR achieves a bit-rate of 23.78 bpp, employing the PIZ wavelet encoding. For this example, OpenEXR reduces the size to 49.5% as compared to uncompressed pixels. This means that although the pixel format of OpenEXR is larger (48 bits) than RGBE (32 bits), the encoding scheme allows for better compression performance.

2.3.2 HDR encoding using LDR formats

While the floating point formats can distribute high-quality HDR pixels, the file size is still large compared to common LDR formats. This is especially problematic for video sequences, as these HDR formats do not explore inter-frame correlations. For example, with OpenEXR a 1-minute sequence at full HD 2K resolution (1920×1080 pixels) and 24 frames/second (fps) would require around 8.8 GB with the PIZ encoding. While this can be accepted in the industry, where quality is a high priority, it is not feasible e.g. for HDR TV streaming. In order to provide viable solutions for lossy encoding of HDR images and video, a number of different techniques have been suggested for encoding HDR data using existing LDR codecs. There are several benefits to this strategy. First, LDR codecs have evolved for a long time and are today very efficient. Second, by employing an LDR codec it is easy to enable support of HDR material in existing software, and also to allow for backward-compatibility. Moreover, LDR codecs rely on integer pixel representations, which allow for better compression properties as compared to floating points.

Single-layer encoding: The most straightforward approach for adapting scene-referred floating point HDR pixels for integer encoding is to transform the luminance to a perceptually linear domain, using a so-called *perceptual transfer function* (PTF) or *electro-optical transfer function* (EOTF). A subsequent rounding operation to the particular bit-depth of the LDR codec will then result in quantization artifacts that are approximately perceptually uniformly distributed across different luminance levels. The concept is related to gamma correction for LDR images, which achieves a similar goal. However, gamma correction is only a good approximation of the HVS response for a very limited range of luminances. Stretching the gamma correction over a wide dynamic range will result in that quantization artifacts are perceived as larger for lower luminance levels. Another alternative is a logarithmic transformation, but this is only a reasonable approximation for luminance levels within the photopic vision, approximately above 1 cd/m^2 (see Figure 1.2b), and will spend too many bits on low luminance levels. For this reason, a number of PTFs have been proposed that rely on psychophysical experiments. These functions have shapes that are somewhere in-between the gamma and logarithmic mappings, see Figure 2.3.

The first example of HDR image encoding using an existing image file format was presented by Ward [263]. This is referred to as LogLuv and it is implemented as an extension to the TIFF (Tagged Image File Format) library. The pixel format is described with log-transformed luminance and CIE u'v' perceptually linear chromaticity coordinates. It uses 8 bits for each chroma channel, 15 bits for log luminance, and 1 sign bit, for a total of 32 bits. While TIFF describes a number of different encoding schemes, the LogLuv format is primarily intended for lossless encoding, as described in the baseline TIFF specification. A similar method for the JPEG 2000 coding scheme was proposed by Xu et al. [268]. However, this transforms RGB values to the log domain before

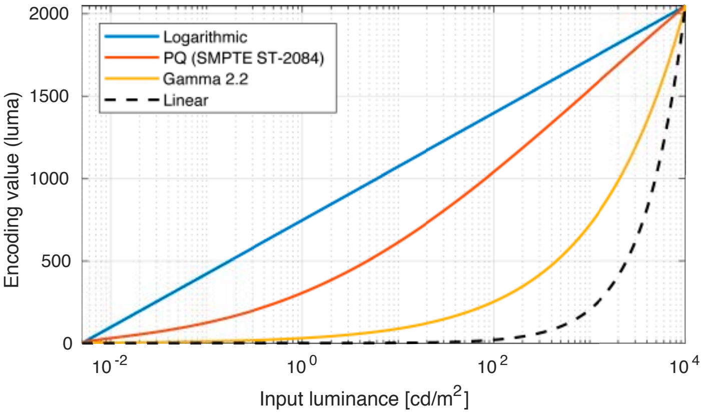


Figure 2.3: The SMPTE ST-2084 standard perceptual transfer function [185], compared to log, gamma, and linear mappings. The functions map physical luminance in the range 0.005 - 10,000 cd/m^2 to integer luma values for encoding at 11 bits. The gamma function has been stretched to cover the same range of luminance.

encoding each channel with the 16-bit integer format provided by JPEG 2000. Compared to other methods, the performance of HDR-JPEG 2000 shows advantages for lossy encoding at low bit-rates. In 2009, JPEG XR was made available, which similarly to JPEG 2000 provides a range of different pixel formats and bit-depths that can facilitate HDR image encoding [70]. However, compared to JPEG 2000 the new standard allows for lower computational complexity and better flexibility.

Mantiuk et al. [166] demonstrated the first method for inter-frame encoding of HDR video. It is also the first to derive a PTF based on experiments on the sensitivity of the HVS. The PTF is formulated to ensure that quantization errors are below the visibility threshold, given the experimental data from Ferwerda et al. [93]. The pixel format uses the u'v' color space, storing chroma at 8 bits/channel, while luminance is mapped to 11 bits. This is enough to encode the full range of perceivable luminances without visible quantization artifacts. The encoding is implemented by modifying the Xvid MPEG-4 codec so that it can encode at a higher bit-depth. Additionally, a modification is made for encoding HDR edge information separately, in order to avoid artifacts around high contrast edges in synthetic HDR video sequences.

With the introduction and rapid growth in popularity of HDR TV displays, during the last couple of years there has been a lot of activity around HDR video encoding. Already with the transition from high definition TV (HDTV) to ultra HDTV, the ITU-R recommendation BT.2020 was introduced for describing a wider color gamut than sRGB (ITU-R BT.709). For HDR, the focus has been on techniques for single-layer encoding, where PTFs/EOTFs have been standardized through the perceptual quantizer (PQ) function (SMPTE ST-2084) and the Hybrid Log-Gamma (HLG). These are now part of the ITU-R recommendation BT.2100, which specifically concerns HDR video distribution. The PQ function [185] is derived in a similar way as the PTF by Mantiuk et al. [166], but using contrast sensitivity data by Barten [30]. It is fitted to an analytic function and describes a mapping for luminance values up to 10,000 cd/m^2. It has also been verified that PQ results in good perceptual uniformity [41] and encoding performance [79]. The HLG function is a combination of a gamma function and a log mapping. For low luminance values the gamma function is a good representation of perceptual linearity, similar to gamma correction for LDR images, and for larger values, in the photopic vision, the log is representative according to Weber-Fechner's law [90]. With the gamma correction in the range of LDR luminances, encoding with HLG makes it possible to directly display the LDR range on a standard LDR monitor without depending on metadata.

The initiatives in HDR video encoding have resulted in a set of HDR video formats that have gained widespread support by HDR TVs and streaming services. The HDR10 format specifies luminance encoding using PQ (SMTP2084), and C_bC_r color primaries according to recommendation ITU-R BT.2020. Both luminance and color channels are encoded at 10 bits. The format from DolbyVision specifies encoding luminance at 12 bits, in order to support levels up to 10,000 cd/m^2. Additionally, DolbyVision stores "dynamic" metadata that can be used to adapt to a certain display device on a per-frame basis. HDR10 has also been updated in order to support dynamic metadata, in the recent HDR10+ format. Furthermore, HLG has also been introduced as an independent specification, which is similar to HDR10 but using the HLG transfer function for better compatibility with LDR displays.

While HDR10 and HDR10+ are open standard specifications, implementations rely on proprietary codecs, e.g. employing the High Efficiency Video Coding (HEVC) compression scheme. That is, HDR video distribution has not been available on open source terms. In Paper **D** [79] a first open source HDR video codec, *Luma HDRv*, is presented. It uses the PQ PTF and u'v' chromaticity coordinates, together with Google's VP9 codec. These components were demonstrated to give the best performance in a large-scale objective evaluation. However, the software also supports other PTFs and color spaces, so that e.g. HDR10 can be encoded/decoded. The evaluation and the codec are further explained in Chapter 4.

Multi-layer encoding: Backward-compatibility for HDR image/video distribution can be achieved by having two disjoint image streams; one with HDR data and one with its LDR counterpart. However, since these are highly correlated, a large reduction in file size can be achieved by encoding the streams together, so that the HDR data is decoded from the LDR component by incorporating a residual layer. For the encoding, the LDR stream can be provided separately, or it can be computed within the encoding scheme using a tone-mapping algorithm.

The first example of multi-layer image encoding for extending the dynamic range was proposed by Spaulding [231], separating the HDR image into a tone-mapped image and a residual layer. A readily available implementation capable of a much higher dynamic range was provided by Ward and Simmons [260, 261], with the JPEG-HDR extension to the JPEG coding scheme. The method stores a tone-mapped image as a standard 8-bit JPEG, which is backward-compatible with any JPEG decoder. However, a ratio image is provided in the JPEG metadata tag, so that the original HDR image can be restored when the two layers are multiplied. In a more recent effort, the JPEG XT standard has been announced, with the intention of providing HDR encoding with JPEG in a completely backward-compatible manner, using a two-layer layout [10].

For backward-compatible HDR video encoding, the first method was presented by Mantiuk et al. [168]. It does not put any restrictions on how the LDR stream is constructed, as LDR and HDR streams are provided separately to the encoder. The two streams are then de-correlated by attempting to find a reconstruction function that can predict the HDR pixels from the LDR counterparts. This means that the residual of HDR and reconstructed LDR streams is kept to a minimum. LDR and residual data are subsequently encoded using MPEG-4 and give approximately a 30% increase in file size as compared to only encoding the LDR data.

A number of succeeding methods attempt to improve on the layered HDR image encoding strategy in various ways. For example, Okuda and Adami [191] used an analytic function for reconstructing HDR from the LDR stream before computing the residual, where parameters are chosen based on image content. Lee and Kim [149] explored motion information between frames in tone-mapping for the LDR stream. The LDR and residual streams are encoded at different quality levels in order to improve the compression performance. Based on a statistical model, Mai et al. [161] derived a tone-curve for the LDR stream that is optimized for the best quality of the reconstructed HDR data.

While backward-compatibility is an important feature in transitioning to better support for HDR data in general, the single-layer encoding approaches tend to provide better rate-distortion performance [21, 187]. That is, single-layer HDR encoding can provide higher quality for a given bit-rate.

2.4 Tone-mapping

Methods for tone-mapping can reduce the dynamic range of HDR images, for the purpose of display on a medium that is limited in its dynamic range, including computer monitor, TV, smartphone, and printed paper. Strictly speaking, tone-mapping can describe any transformation of image tones, but the term is almost exclusively referring to a mapping from scene-referred HDR tones to display-referred LDR pixels. The tone compression generally aims at revealing information over a larger range of luminances than what is possible with conventional LDR images, similarly to how the HVS operates.

Techniques for compressing the dynamic range of an image signal dates back to the 1960s [192]. In the 1980s there were attempts at matching the appearance between a real-world scene and an image displayed on a screen [184, 249]. In the early 1990s, the problem was formally introduced in the computer graphics community [241], for the purpose of displaying images generated by physically based rendering methods. Subsequently, during the last 25 years tone-mapping has been an active area of research, resulting in the development of many hundreds of different methods.

2.4.1 Categorization

In order to distinguish between the large number of existing TMOs, they are commonly grouped in different ways. The most general distinction is to classify TMOs as either *global* or *local* operators. A global TMO applies the same operation for all pixels, while a local can change the transformation spatially as a function of a local neighborhood of pixels. Local TMOs can better preserve local contrasts of the HDR image, but are generally computationally more expensive and more prone to generate artifacts.

Another distinction can be made between TMOs that are only designed to process static images and those that also are applicable for HDR video sequences. The video TMOs use mechanisms for adapting the tone processing over time, in order to avoid temporal artifacts such as ghosting and flickering.

Furthermore, a third categorization of TMOs considers the specific intent. Although the tone-mapping algorithms take an HDR signal and compresses the dynamic range to the limited range of a display device, the objective, or intent, of this mapping may vary. The intent is decided upon how the quality of the final tone-mapping should be evaluated. Following the categorization introduced in Paper **B**, a natural differentiation can be made using three major intents: visual system simulators (VSS), best subjective quality (BSQ) operators, and scene reproduction (SRP) operators.

Visual system simulator (VSS): One of the most natural objectives of a TMO is to attempt to mimic the capabilities of the HVS. Since the HVS can register a higher dynamic range than a conventional camera, this means that an increased amount of visual information is made visible as compared to a typical LDR image. It also means that the deficiencies of the HVS should be simulated, including loss of acuity, glare, and decreased color saturation in low light conditions. The optimal result of a VSS is the image that minimizes the perceived difference when comparing the tone-mapped image to the original captured scene. However, there may also be features for simulating different vision impairments, such as age-dependent factors and color blindness, which do not improve the perceptual similarity, but which can demonstrate how the image may be perceived by an HVS with disabilities.

One of the first VSS tone-mapping algorithms was presented by Ferwerda et al. [93]. It models the adaptation mechanisms of the HVS, based on a series of psychophysical experiments. Pattanaik et al. proposed one of the most comprehensive perceptual models for tone-mapping [198]. It uses a multiscale representation of luminance, detail, and color processing of the HVS, and it accounts for both threshold and supra-threshold perception. In subsequent work, Pattanaik et al. combined adaptation and appearance models in order to simulate the response of the captured HDR scene [199]. By inverting the models, the response can be mapped to an LDR display device. A similar concept was used by Ledda et al., but on a local – per-pixel – level of the image [145]. Furthermore, Irawan et al. extended the adaptation modeling to also include the state of mal-adaptation of the HVS [121], thus not assuming that the HVS is perfectly adapted to the background luminance level. The concept of mal-adaptation for tone-mapping was further extended by Pajak et al. [193], in order to work on a local level.

VSS methods are most often based on data from psychophysical experiments, but there are also examples where actual quantitative measurements are used. For example, van Hateren employed a model that is built from measurements performed on the retina of macaques [250]. Moreover, there are also VSS methods that model the actual HVS components instead of its high-level behavior. One example is the TMO proposed by Meylan et al. [183], which makes use of a model that accounts for low-level processing in the retina. The method was extended for HDR video tone-mapping by Benoit et al. [33].

Best subjective quality (BSQ) operator: A common objective for tone-mapping is to generate the image that is most preferred upon visual inspection. That is, the image with highest subjective quality without comparing to the reference HDR image. Compared to the VSS category, this often means that abilities which are superior to the HVS are favored, such as increased contrast, sharpness, details, and color saturation, as well as a larger compression of the dynamic range. However, depending on the particular application, and the individual that is judging the result, the tone-mapping may be very different. The intent can be established in closer terms, e.g. to better comply with a specific artistic goal.

There is a large range of TMOs that qualify as BSQ operators, from the first work within tone-mapping [51, 91, 208, 220], to some of the most frequently appearing operators in the literature [68, 73, 89, 210]. Also, BSQ tone-mapping examples often appear in connection to presentations of novel edge-aware filtering techniques [14, 52, 73, 87, 113, 195], in order to provide a common application for demonstrating the filtering performance.

Scene reproduction (SRP) operator: A numerical comparison of the overall perceptual differences between a tone-mapped image and the reference input HDR image is complex. Instead, a TMO can focus on minimizing the difference in terms of an isolated image attribute. That is, an SRP operator attempts to make the tone-mapping invariant to this certain attribute, in order to preserve its original appearance. The attribute can, for example, be the relative brightness, contrast, color, or temporal behavior. However, while optimizing for one particular attribute, the final image may still deviate substantially from the reference HDR image in terms of other attributes.

With the introduction of tone-mapping to the computer graphics community, Tumblin and Rushmeier proposed a method for preserving the apparent, or perceived, brightness of the HDR image [241]. Ward attempted to preserve the contrasts from the HDR image [258], using a global scaling factor. However, as this method in essence performs an automatic exposure adjustment, it also means that much of the visual information is lost in dark and saturated image areas. Another approach is to aim at minimizing the changes in contrasts, given that the dynamic range is compressed to a certain display device [77, 170]. Other SRP goals include, for example, preservation of visibility [264], perceived lightness [133], color appearance [138, 212], and temporal consistency [38, 107].

2.4.2 Tone-mapping pipeline

A tone-mapping method can be designed in many ways. However, the typical procedure is displayed in Figure 2.4. There are four distinct steps involved, which can be altered to accomplish different intents:

1. Pre-processing: The scene-referred HDR image is first transformed into a format that is suitable for the tone compression. The transformation may vary depending on how the TMO is constructed. For example, there are examples of

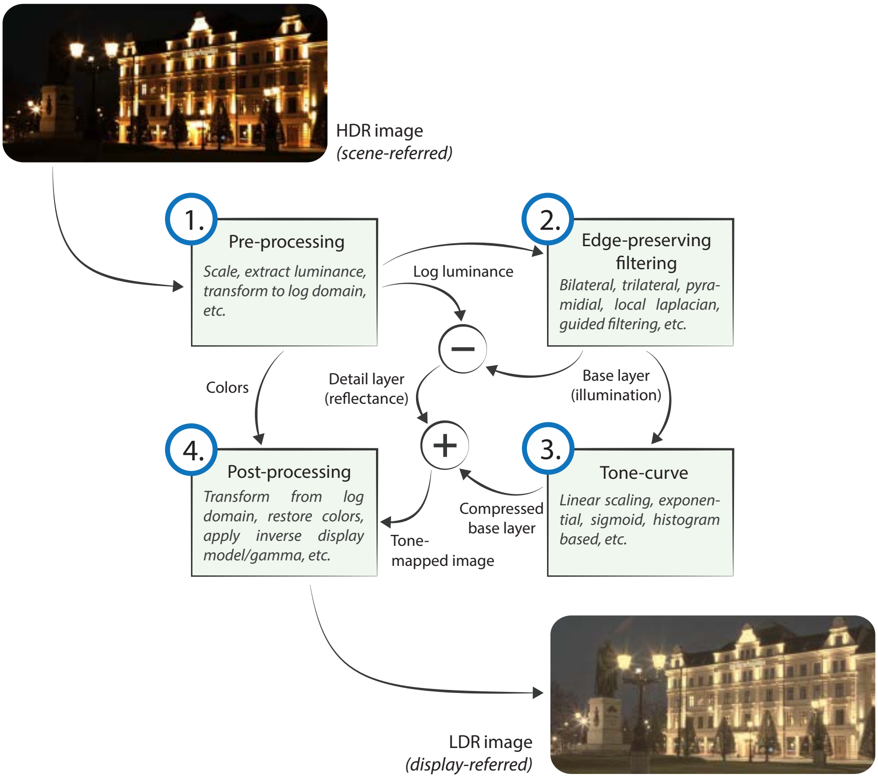

HDR image
(scene-referred)
1.
Pre-processing
Scale, extract luminance, transform to log domain, etc.
Log luminance
2.
Edge-preserving filtering
Bilateral, trilateral, pyramidial, local laplacian, guided filtering, etc.
Colors
Detail layer
(reflectance)
Base layer
(illumination)
4.
Post-processing
Transform from log domain, restore colors, apply inverse display model/gamma, etc.
3.
Tone-curve
Linear scaling, exponential, sigmoid, histogram based, etc.
Compressed base layer
Tone-mapped image
LDR image
(display-referred)

Figure 2.4: Typical pipeline for performing tone-mapping. While the edge-preserving filtering enables local processing, the pipeline can also describe a global tone-mapping by substituting this with the identity mapping.

methods that perform the compression in the gradient [89, 148, 254] or contrast [167] domain. Also, a number of methods attempt to model the appearance of colors [6, 86, 129, 138, 198, 212]. However, the most common method is to only consider luminance, and restore colors after this has been compressed [220]. Furthermore, in most cases tone-mapping is not performed on linear luminances, but in the log domain. The reason is that in a large range of luminances the HVS has a close to logarithmic response, according to the Weber-Fechner law [90]. Therefore, operating on log luminances often makes for a simpler problem description due to the increased perceptual linearity.

2. Edge-preserving filtering: Local processing makes it possible to achieve similar or superior capabilities as the local adaptation mechanisms of the HVS. However, instead of performing a per-pixel tone-mapping depending on a local neighborhood, the processing is usually decomposed by means of a low-pass filter. The filtered image represents the *base layer B*, which then is used in order to extract a *detail layer D* from the HDR luminance L. In the log domain the details are separated from subtracting the base layer, $D = L - B$, as illustrated in Figure 2.4. While the dynamic range of the base layer is compressed, the detail layer is by-passed this step and added back after the tone compression. In this way, local contrast and details are preserved. This methodology is analogous to separating the image into a product of illumination and reflectance layers [29, 116], which is similar to how the HVS processes a scene. It can discriminate reflectance over a wide range of luminances while disregarding the illumination [98]. The reflectance is of low dynamic range and contains image details and textures, while the illumination is responsible for the high dynamic range and describes global variations within the scene. Therefore, it makes intuitive sense to maintain the reflectance unmodified while only compressing the illumination.

For decomposing the image into base and detail layers, the choice of the specific filter used is critical in order to avoid visible artifacts. For example, some first attempts at local tone-mapping make use of Gaussian low-pass filters [51, 123, 208], which assumes that there are no sharp boundaries within the scene. If this is not the case, there will be haloing artifacts around the sharp image features. For this reason, a range of different edge-aware filters has been demonstrated in connection to tone-mapping. Some have been presented solely for the purpose of tone-mapping, e.g. early attempts that employ multi-scale structures [12, 198, 210, 242], as well as more recent techniques [20, 77]. There are also many multi-purpose edge-aware filters that have been used in local tone-mapping [14, 87, 88, 113, 195, 234]. One of the most frequently appearing filters in the tone-mapping literature is the bilateral filter [15, 237]. The idea of local tone-mapping using a bilateral kernel was first discussed by DiCarlo and Wandell [63], and later independently demonstrated in different formulations by Durand and Dorsey [73] and Pattanaik and Yee [197]. The filter allows for a simple formulation, and can also be accelerated in different ways for real-time performance [1, 50, 194, 265, 269]. However, on smooth high-contrast edges the anisotropic filter kernels are biased towards one side of the edge. This can generate gradient reversals in the extracted detail layer, which cause visible banding artifacts [22, 73, 77]. The problem can be alleviated at the expense of added computational complexity [22, 52]. For the TMO presented in Paper **C** [77], we introduce an iterative and isotropic simplification of the bilateral filter. The technique is both fast and it overcomes the problems with banding artifacts. The filter is further explained in Section 3.3.

3. Tone-curve: A tone-curve V describes a mapping $V : L \rightarrow T$ that takes the input relative HDR luminance L and transforms it to a compressed domain of LDR luminance values T, as shown in Figure 2.5. To avoid inconsistencies in the output luminance levels, V should be a monotonic nondecreasing function. The simplest form is a linear function, performing a scaling – or exposure correction – of luminance levels [258]. A linear scaling in the log domain corresponds to an exponential function in the linear domain, which can be used to compress the dynamic range [68, 91, 241]. Moreover, one of the most frequently occurring functions in the tone-mapping literature is a logistic, or sigmoidal, function,

$$V(L) = \frac{L^n}{L^n + \sigma^n}. \tag{2.1}$$

The parameter n can be used to control the slope of the function, and σ shifts it along the horizontal axis. The sigmoid transforms all luminance levels to the range $[0, 1]$, and it performs a similar compressive mapping as is done by biological visual systems [188]. The first use within tone-mapping can be found in the method by Schlick [220], and a few years later Pattanaik et al. introduced the function to describe an approximation of the photo-receptor response curve [198].

Image statistics are often accounted for in the aforementioned tone-curves, in order to adapt to the overall luminance level. For example, σ in Equation 2.1 can be formulated using the image mean or median. However, for an improved distribution of tone-mapped values the shape of the tone-curve can be controlled by means of the image histogram [69, 170, 203, 264], similarly as in histogram equalization. For the TMO presented in Paper **C** [77], we use the image histogram in order to minimize the differences in contrasts between input and tone-mapped images. The tone-curve is further explained in Section 3.3.

4. Post-processing: As a final step in the tone-mapping pipeline, a number of post-processing operations can be applied. For example, the colors can be restored from the original image. However, the re-coloring can make for a visible increase in color saturation, especially when the tone-mapping performs a very large compression of the dynamic range. The problem can be alleviated by incorporating a heuristic desaturation operation in the re-coloring step [243],

$$c = \left(\frac{C}{L}\right)^s T. \tag{2.2}$$

Here, T is the tone-mapped luminance, $T = V(L)$, while c and C are tone-mapped and input color channels, respectively. The amount of color saturation is specified by the exponent s, where a value $s < 1$ accomplishes a desaturation.

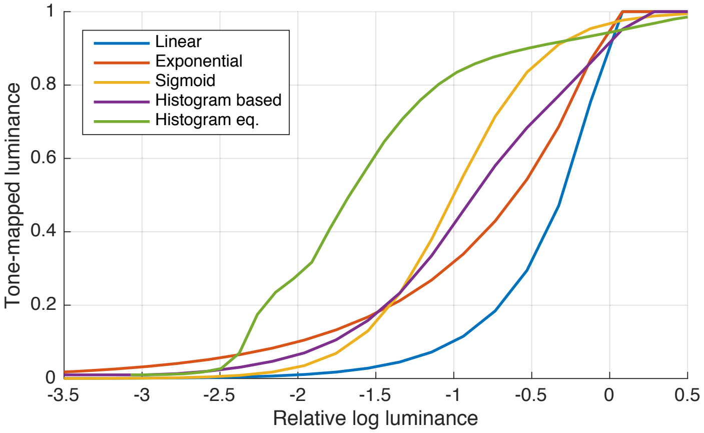


Figure 2.5: Different types of tone-curves. These map linear scene-referred luminances to a limited dynamic range. The tone-mapping should be followed by a display adaptation for mapping to a display-referred format, e.g. by means of gamma correction. The histogram based tone-curve is derived using the method in Paper **C**. A tone-curve that performs histogram equalization in the log domain is also included for reference. The histogram used is from the input HDR image in Figure 2.4.

There are also attempts at characterizing the behavior of the color saturation in tone-mapping [11, 171, 204], for automatic calibration of the required desaturation.

After color processing, the final step is to prepare the tone-mapped image for display, mapping it to a display-referred format. This can be accomplished e.g. by means of a gamma correction. Display characteristics and viewing environment can also be accounted for using the display model in Equation 1.1.

2.4.3 Temporal aspects

For HDR video sequences there are some critical differences that need to be accounted for by a tone-mapping algorithm, as compared to tone-mapping of static HDR images. The most prominent difference is that temporal coherence needs to be maintained, both globally and locally. Computational complexity also becomes an important aspect, as large amounts of data need processing. Moreover, there are differences caused by the different capturing techniques. HDR video is, for example, more prone to carry visible amounts of image noise.

Non-trivial TMOs rely on image statistics, so that the tone-curve $V(L_p, S(L))$ applied to a pixel at position $\boldsymbol{p}$ depends both on the pixel value L_p and on some measure $S(L)$ over the whole image. Many different statistics can be used, such as the image mean, median, or histogram. In an HDR video sequence, these can change rapidly from frame to frame, which can be perceived as flickering artifacts in the final tone-mapped output. To prevent this from happening, a low-pass filter can be applied to $S(L)$ over time. This can be done e.g. using an exponentially decaying filter kernel over a set of past frames [72, 102, 127, 209], or equivalently using a leaky integrator. In order to prevent flickering artifacts in a local tone-mapping algorithm, local image statistics can be filtered over time [145]. However, this may lead to visible ghosting artifacts. Another alternative is to filter the final pixel values [33, 250], which promotes temporal coherency at the cost of introducing motion blur. The per-pixel filtering can also employ an edge-preserving filter [32], so that filtering is restricted at large temporal gradients. Finally, the local filtering can be performed over motion compensated temporal pixels, using block matching [148] or optical flow [20].

There are also techniques for imposing temporal coherence in tone-mapping as a post-processing step, alleviating flickering artifacts of arbitrary global TMOs [38, 39, 107]. For local coherence, motion estimation by means of e.g. optical flow can be utilized, as suggested in methods for imposing temporal coherence of different types of video processing operations [42, 66, 144].

Classically, HDR images have been produced from either CGI or exposure bracketing with little restriction on the exposure time. Consequently, image noise has not been a major problem, especially when noise is considered in the HDR reconstruction [7, 103, 135]. For HDR video, on the other hand, the capturing methods are more susceptible to generating noise. Since TMOs use non-linear mappings, increasing the intensities of dark pixels while doing the opposite for bright pixels, the visibility of the noise can be amplified. With denoising methods [44, 54], or per-pixel filtering for temporal coherence [20, 32], the amount of noise can be reduced. However, this may be expensive, and it is difficult to remove all noise without introducing artifacts. Another approach is to control the shape of the tone-curve based on an estimation of the image noise, in order to not reveal the noise in the tone-mapping [77, 154]. This concept was introduced in Paper **C** [77] and will be explained in Section 3.3.

A thorough review of tone-mapping for HDR video is provided in Paper **A** [84], including brief descriptions and categorizations of 26 video TMOs from the literature.

2.4.4 Evaluation

For many image processing operations, assessment of the result can be made from an averaged direct pixel-wise comparison to a reference image, e.g. by means of the *root-mean-square error* (RMSE) or the *peak signal-to-noise ratio* (PSNR). Although such measures are not expected to be linearly correlated with the perceived differences, they provide direct insight into the obtained performance. There are also measures that better agree with perceived visual quality, such as the *multi-scale structural similarity* (MS-SSIM) index [256] or the *HDR visual difference predictor* (HDR-VDP-2) [172]. For tone-mapping, however, these measures cannot be used directly, as a reference image is not available. Therefore, an important aspect within tone-mapping is strategies for quality evaluation, in order to enable comparisons between different TMOs.

There are some methods that have been developed for objective quality assessment of TMOs, comparing the tone-mapped image to the HDR source [17, 46, 271], or making a similar comparison with video sequences [19, 272]. However, while these measures have been demonstrated to correlate with subjective evaluations, the heuristics employed cannot completely replace a human's high-level judgments, which are based on both long-term memory and low-level visual information processed by the HVS.

A number of studies have been conducted in order to evaluate the subjective quality of tone-mapping, attempting to compare different TMOs against each other. In performing such study, there are a couple of possible strategies for reference/non-reference comparison of the tone-mapping results, as illustrated in Figure 2.6:

1. The most straightforward strategy is to evaluate by only displaying tone-mapped images, in a non-reference setup. This is probably also most true to how the images are to be viewed in the end. For these reasons, non-reference evaluations have been employed most often, both for tone-mapped images [8, 13, 47, 62, 67, 137, 274] and video sequences [75, 201].
2. Another strategy is to make comparisons to the real-world scene. This is a natural setup and it directly tests one of the main intents of tone mapping, namely fidelity with reality. However, it is challenging to execute. The images differ not only in dynamic range, but also in depth cues, field of view, colors, etc. Despite these differences, a number of studies use this setup [13, 252, 273, 274], and some have also demonstrated correlations between reference and non-reference evaluations [47, 139].
3. A third strategy is to compare the tone-mapped image to a reference displayed on an HDR monitor [140, 146, 181]. Although the HDR display also has restrictions as compared to the real-world scene, it provides a more well-controlled reference.
4. Finally, comparison of isolated perceptual attributes is also possible. It can be realized e.g. by means of magnitude estimation methods [233], where subjects judge the magnitude of a certain stimulus. However, more complex attributes may be difficult to compare, and an overall match in image appearance is not guaranteed from a limited set of measurements.

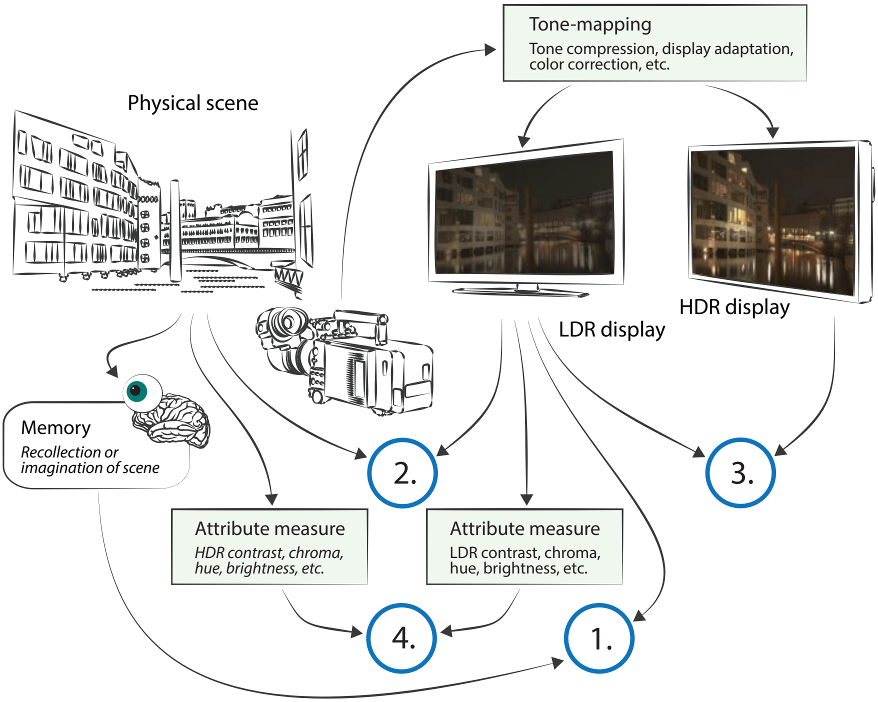

Tone-mapping
Tone compression, display adaptation, color correction, etc.
Physical scene
LDR display
HDR display
Memory
Recollection or imagination of scene
Attribute measure
HDR contrast, chroma, hue, brightness, etc.
Attribute measure
LDR contrast, chroma, hue, brightness, etc.
1.
2.
3.
4.

Figure 2.6: Different methods for evaluating the quality of tone-mapping. **1.)** Non-reference perceptual comparison, assessing fidelity with memorized scene or the subjective quality. **2.)** Direct perceptual comparison to the physical scene, assessing fidelity with reality. **3.)** Perceptual comparison with HDR display, assessing fidelity with HDR reproduction. **4.)** Perceptual comparison in terms of isolated image attributes, assessing the appearance match.

The outcome of a comparison study is also dependent on a range of additional parameters. In a non-reference setup, a critical parameter is the particular criteria specified for making a comparison. For example, different results can be expected if asking to assess the subjective quality as compared to fidelity with memorized scene. Another deciding factor is how parameters of the TMOs are calibrated. For the sake of simplicity, a common solution is to use default parameters. However, with tweaking the tone-mapping quality can potentially be improved to better agree with a certain intent [76, 273]. Moreover, how subjects perform the judgments may also affect the result. Assessing quality by means of e.g. rating, ranking or pair-wise comparisons can impact on precision and length of the experiment. Nevertheless, the results of the different strategies are expected to be correlated [47, 139, 174].

While most evaluations concern assessment of tone-mapped static images, a few more recent studies particularly focus on video [75, 181, 201]. In Paper **B** [75] we report on one of the first evaluations of video TMOs. This reveals a number of unsolved challenges specific to tone-mapping of video. Furthermore, in Paper **A** [84] a quantitative assessment is performed for a number of video TMOs. This is not intended to deduce which operator is "best" or most preferred. Instead, it tabulates the individual strengths and weaknesses in terms of a number of important attributes, indicating which TMO can be expected to show the least amount of artifacts. The evaluations are further explained in Chapter 3.

2.5 HDR displays

A number of research prototypes have been presented for supporting display of HDR images and video, with little compromise of the dynamic range. These have led to a smaller set of professional devices with very high peak brightness. Furthermore, the last couple of years has seen a rapid growth in the number of commercial TV devices with increased brightness and dynamic range, which can support HDR encoded material. There will always be restrictions of displays though, in terms of e.g. brightness, black level, and color gamut, so that tone-mapping is required in order to map the HDR stream to the specific display. However, with HDR displays a much smaller compression of the dynamic range is required.

2.5.1 Professional HDR display devices

The most common technique for achieving the brightness required for display of HDR imagery is by means of dual modulation. A *liquid crystal display* (LCD) panel is back-lit by a high-intensity light array that can be spatially controlled. The LCD display modulates the high intensities by displaying a compensation image, effectively performing an optical multiplication of the independent images. Typically 8-bit precision is used for both images, which means that the total bit-depth is doubled. The technique was originally proposed by Seetzen et al. [222, 223], with two different prototypes. One uses a digital light processing (DLP) projector back-light and achieves a peak luminance of 2,700 cd/m^2, while the other demonstrates a more versatile solution using a low-resolution array of individually controllable *light emitting diodes* (LEDs). The LED-based display can deliver a maximum intensity of 8,500 cd/m^2 and a dynamic range above 200,000:1, but it requires more extensive pre-processing for decomposing the HDR image into back-light and compensation images. Since the low-resolution LED array depicts a smoothly varying image, the compensation image needs to account for this around sharp features, in order to avoid bleeding/blooming artifacts in the displayed HDR image.

For research purposes, a number of prototypes have been built following the technique introduced by Seetzen et al. Most use the projector based setup [92, 253], which is more straightforward to build. An HDR projector has also been demonstrated, by Damber et al. [57], which utilizes dual modulation within the projector in order to lower the black level substantially as compared to a conventional projector. However, the peak brightness is still limited. In more recent work, Damberg et al. presented a light steering projection system [58], which can steer light away from dark to bright image areas. This means that although the brightness of a full white image is still limited, when the image only contains smaller highlights these can be boosted to a large extent. Since natural images often have this property, with an intensity distribution that has small values towards high luminances, the steering projection could potentially be an important concept in future HDR projector systems.

The dual modulation prototypes by Seetzen et al. were developed on and realized in 2005 by Brightside (formerly SunnyBrook Technologies). For example, the DR37-P is back-lit by an LED array and can reach a brightness of 4,000 cd/m^2, and the SBT1.3 uses a projector for back-lighting and has a peak luminance of 2,700 cd/m^2. In 2007 Brightside was acquired by Dolby Laboratories, and production of these devices was terminated. The technology was later used in the Dolby Pulsar reference monitor, also with a peak brightness of 4,000 cd/m^2, and in the HDR47 series by Italian electronics company Sim2. The latest model, HDR47ES6MB, is specified with 6,000 cd/m^2 peak luminance. The most recent news to the top-performing segment of HDR displays is Sony's prototype showcased at CES 2018 [230]. The 85-inch device features 8K resolution and allegedly it can reach a peak brightness of 10,000 cd/m^2.

2.5.2 HDR TVs

We are today in a position where ultra HD is the norm within the consumer TV market, with most TV devices specified for 4K resolution. Now, 8K resolution is expected to appear in a very near future. From this previous trend in maximizing spatial information, the current focus is on expanding in the intensity domain. TVs with HDR support is a new segment in the TV industry, which has seen a large development in the last few years. The development focuses on increasing peak brightness, and improving techniques for local dimming to achieve better dynamic range. Moreover, a standardization on the HDR format is currently ongoing, see Section 2.3.

Most HDR TVs use the same principle as the professional – high performance – HDR displays, with back-light modulation for local dimming. However, the back-lighting is less bright and not as precise. The most common technique is to utilize LCD modulation with back-light provided from LEDs mounted on the edges of the display panel. This allows for cheaper and thinner construction. The light from the LEDs is reflected from the rear by means of a set of guides. A rough local control can be achieved for spatially varying dimming in order to increase the dynamic range, but blooming can be a problem. There are also more high-end devices with rear-mounted LED arrays, which can achieve better local control. However, in contrast to the professional HDR displays, the LEDs are in general not possible to control on a per-unit level, but instead local dimming is provided through a set of different zones of LEDs. Currently, in terms of peak brightness the highest performing LCD HDR TVs can approach – or even exceed – 1,500 cd/m^2, such as the edge-lit Sony X930E or the full-array Sony Z9D [219].

Another promising technique is *organic* LED (OLED) display panels, which do not require back-lighting. Instead, each pixel in an electro-luminescent layer is individually controllable in terms of emitted light, and can be switched off to achieve a 0 black level. Although this makes for a very high dynamic range, OLED displays cannot yet match the LCD based displays in terms of brightness. Because of this, the dynamic range is very sensitive to ambient lighting. However, the technique is progressing, for example with the 5 OLED TVs revealed by LG at CES 2017. These provide 4K resolution and increased brightness as compared to previous OLED displays, peaking around 700 cd/m^2 [219]. Furthermore, there are other single-modulation techniques emerging, such as *micro* LED (mLED or μLED) display panels. These use individually controllable micro LEDs for each pixel, which potentially can allow for higher brightness than OLED while still having 0 black level.

Clearly, on the consumer market there is an ongoing transition towards HDR material and HDR displays. The future will see improved techniques for back-lighting and local dimming, as well as single-modulation solutions. This means that the dynamic range and brightness capabilities of current professional devices may soon be surpassed by some HDR TVs, and at a higher resolution. The future of HDR displays is looking *bright*!

Chapter 3

Tone-mapping of HDR video

With the plethora of existing tone-mapping techniques, one can argue that there are not many more avenues to explore within the area. However, the absolute majority of existing work only considers tone-mapping of static images. Tone-mapping for HDR video sequences introduces a number of problems that either do not appear, or are not as prevalent in tone-mapping of static images. This thesis presents a first systematic survey and evaluation of existing methods for video tone-mapping, in which a set of problems were identified. Problems that had not been properly accounted for at the time of the study. These problems formed the basis for the development of the new TMO presented in this thesis and in Paper **C**.

This chapter discusses the work and contributions of Papers **A**, **B**, and **C**, which focus on tone-mapping of HDR video sequences. Following a short motivation of the work in Section 3.1, the survey and evaluation from Paper **B** are described in Section 3.2. From the findings of the evaluation, the algorithms introduced by Paper **C**, which are the topic of Section 3.3, are developed specifically considering the problems faced in tone-mapping for video. In Section 3.4, the quantitative evaluation from Paper **A** is discussed. The evaluation includes some of the most recent TMOs, and indicates that many of the problems found in Paper **B** have been addressed in the most recent work for video tone-mapping. Finally, in Section 3.5 the chapter wraps up the contributions of the papers and discusses some of the limitations and possible directions for future work.

For a thorough background on tone-mapping of HDR images and video sequences, the thesis provides a literature study of the area in Paper **A**. This gives a historical overview of tone-mapping, discusses the particular challenges in tone-mapping of video sequences, and lists brief descriptions and categorizations of all TMOs with explicit temporal processing that could be found.

3.1 Motivation

As discussed in Section 2.4.3, the most evident problem faced by video TMOs is maintaining the temporal coherence. Global problems with coherence can cause extensive flickering, while local problems can be manifested as e.g. ghosting artifacts. It is also possible for a local TMO to cause flickering artifacts on a local level. For example, imagine a spatial artifact caused by the local tone processing, and which is barely visible in a static image. If the artifact changes quickly over time due to local variations in image content, it may be perceived as a significantly more salient degradation in image quality. For these reasons, explicit consideration of temporal coherence is important in tone-mapping for video, and especially for local TMOs.

Temporal aspects in tone-mapping were considered already more than 20 years ago [93] and many VSS methods attempt to model the temporal adaptation mechanisms of the HVS [33, 93, 121, 199, 250]. There are also examples of other methods that consider video tone-mapping [32, 102, 170, 209]. However, all these TMOs were developed when there was an insufficient number of HDR videos available to allow for thorough testing of the tone-mapping quality. Consequently, most have only been demonstrated on artificial HDR videos, such as CGI, panning in HDR panoramas, or from capturing static scenes with alternating lighting. A few examples also include custom-built techniques and systems to record HDR video [127, 254]. With the advent of versatile HDR video camera systems [48, 136, 236] and professional cinematographic cameras with extended dynamic range, a number of new challenges were introduced. For example, HDR videos are more likely to contain image noise, which may be revealed by tone-mapping. The videos can also present challenging transitions in intensity, and certain dynamic objects that are not common in static HDR images. One such example is skin tones, which is important to render with the appropriate hue and saturation.

The lack in testing of existing video TMOs with diverse HDR video data, motivates the study carried out in Paper **B** [75]. Furthermore, the problems that were established by this study motivates the development of new techniques for HDR video tone-mapping, as presented in Paper **C** [77]. Finally, with the recent development in tone-mapping for HDR video, partially due to the findings in Paper **B**, the work in Paper **A** [84] contributes with an up-to-date reference, categorization, and assessment of the state-of-the-art in tone-mapping for HDR video.

3.2 Evaluation of TMOs

The ultimate question when inspecting the multitude of existing TMOs is: which produces the best results? This question is impossible to answer as it depends on many individual factors, such as the specific intent (Section 2.4.1), the particular viewer, and the viewing condition. However, by conducting perceptual comparison experiments some important insight can be gained given the certain experimental setup that is used [81].

For the evaluation presented in Paper **B**, the motivation is not only to provide a relative ranking of existing methods for video tone-mapping. An important part of this work is to identify major problems and challenges that need to be addressed by video TMOs, highlighting the differences as compared to tone-mapping of static images. For this reason, a number of challenging HDR video sequences were used in the experiments, captured using a multi-sensor HDR video camera system [135, 136], as well as a RED EPIC cinematographic camera, and a computer-generated sequence. This provided a wide variety of content and genuinely challenging conditions, which the TMOs under consideration had not been tested for.

The TMOs that were included in the study are listed in Table 3.1. These were chosen with the criterion of having explicit treatment, or model, of temporal aspects in the tone-mapping. The study considered VSS methods, but other operators were also included since these may yield competitive performance although the intent differ.

3.2.1 Parameter calibration

A major difficulty in staging an evaluation experiment with image processing operations is that the operations may require parameter calibration in order to achieve optimal results. This is complicated mainly due to two reasons:

1. Computationally expensive operations cannot be tweaked with real-time feedback of the result, which is essential in order to make a calibration experiment feasible. The problem is even more pronounced in evaluation of video operations – the parameters may affect temporal aspects, which require the result of a particular calibration to be assessed on video sequences. Thus, it may take many minutes, or even hours, to process the large number of frames needed for assessment of one single parameter calibration.
2. In general, the operators have many parameters that can be tweaked. How can we find the perceptually most optimal point in a high dimensional space of parameters?

Because of these difficulties, most previous studies of TMOs use the default parameters that were suggested by the authors of the different methods. However,

Name	Processing	Intent
Visual adaptation TMO [93]	Global	VSS
Time-adaptation TMO [199]	Global	VSS
Local adaptation TMO [145]	Local	VSS
Mal-adaptation TMO [121]	Global	VSS
Virtual exposures TMO [32]	Local	BSQ
Cone model TMO [250]	Global	VSS
Display adaptive TMO [170]	Global	SRP
Retina model TMO [33]	Local	VSS
Color appearance TMO [212]	Local	SRP
Temporal coherence TMO [38]	Global	SRP
Camera TMO (see Paper **B**)	Global	BSQ

Table 3.1: List of video TMOs included in the study of Paper **B**. See Section 2.4.1 for a description on the different categorizations. The bottom TMO uses a conventional camera curve, measured from a Canon 500D DSLR camera, with the exposure setting filtered over time.

default parameters are not always available, or they can produce unacceptable results in certain situations. Another strategy was reported by Yoshida et al. [273], where a parameter adjustment experiment was conducted prior to the evaluation. In this experiment, a number of observers, experienced in imaging, were to choose between a limited set of different parameter calibrations. We generalize this idea and suggest a method for perceptual optimization of parameters, which potentially can explore the complete multi-dimensional space of parameters. The method was used in Paper **B**, but was described in closer details in subsequent work [76].

Interpolated calibrations: In order to solve the first of the above mentioned problems, enabling tweaking of computationally expensive video TMOs with real-time feedback, we suggest to interpolate between a sparse set of pre-computed parameter calibrations. However, linear changes in parameter values may result in highly non-linear changes in image content. This means that at certain locations in the parameter space, the interpolated video can deviate substantially from the ground truth calibration. The differential $\partial L_{\mathbf{\Theta}}/\partial \theta_k$ caused by a change in parameter value θ_k can be used to quantify changes in image content, e.g. by means of the RMSE,

$$E(\theta_k) = \sqrt{\frac{1}{N}\sum_{\boldsymbol{p}}\left|\frac{\partial L_{\boldsymbol{p},\boldsymbol{\Theta}}}{\partial\theta_k}\right|^2}, \tag{3.1}$$

where the image L is calibrated with the K-dimensional parameter vector $\boldsymbol{\Theta} = \{\theta_1, ..., \theta_K\}$. The sum is taken over all N pixels $\boldsymbol{p}$ in the image. The measure $E(\theta_k)$ may change non-linearly across the range of the parameter. In order to make the changes uniform, the normalized inverse of the integrated parameter changes describes a transformation to a linearized domain,

$$\Lambda(\theta_k) = \int_{\theta_{k,min}}^{\theta_k} E(\phi)d\phi, \tag{3.2a}$$

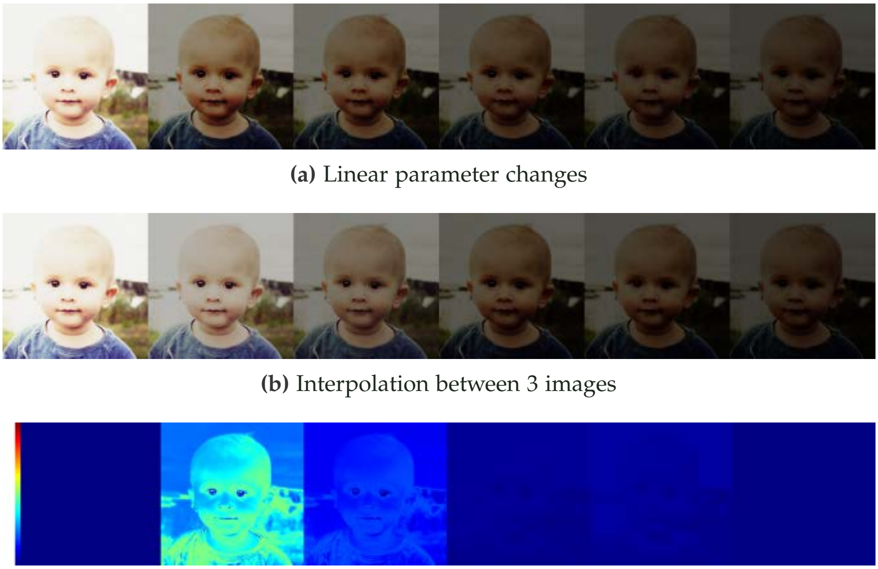
$$\hat{\theta}_k = \Gamma(\theta_k) = \frac{\Lambda^{-1}(\theta_k)}{\int_{\theta_{k,min}}^{\theta_{k,max}} \Lambda^{-1}(\phi)d\phi}. \tag{3.2b}$$

Here, $\Lambda(\theta_k)$ integrates the image changes between the minimum and the current parameter value. With a sparse uniform sampling of the transformed parameters $\hat{\theta}_k$, the RMSE interpolation error is kept to a minimum over the range of the parameter. For a simple demonstration, Figure 3.1 shows the images for a uniform sampling of the parameter σ in Equation 2.1, between 0.05 and 3. That is, $\theta_k = \sigma \in [0.05, 3]$. Using three calibrations for interpolation, $\theta_k = \{0.05, 1.525, 3\}$, the error is large close to small values. By instead sampling the parameter in the transformed domain, $\hat{\theta}_k = \Gamma(\theta_k)$, the error is significantly reduced as exemplified in Figure 3.2.

The transformed parameter calibration not only increases interpolation quality; it also improves the perceptual linearity of parameter changes. This means that parameter adjustments are more intuitive and easier to control.

In practice, the linearization transformations are calculated over a set of video sequences, in order to find a function that generalizes better to different situations. However, the linearization of a particular parameter θ_k at a certain point $\boldsymbol{\Theta}_a$ in the parameter space is not guaranteed to be valid at a different point $\boldsymbol{\Theta}_b$. A more general approach should not consider each parameter individually. Furthermore, more sophisticated metrics could also be used, to allow for minimal interpolation error in terms of perceived differences. These considerations could be topics for future work in calibration for subjective evaluation. For our purpose, the simple method described above was found to work well in the parameter adjustment experiment.

Parameter optimization: With the interpolation strategy, a very limited number of sampling points can be used for interactive exploration of the parameter



(a) Linear parameter changes

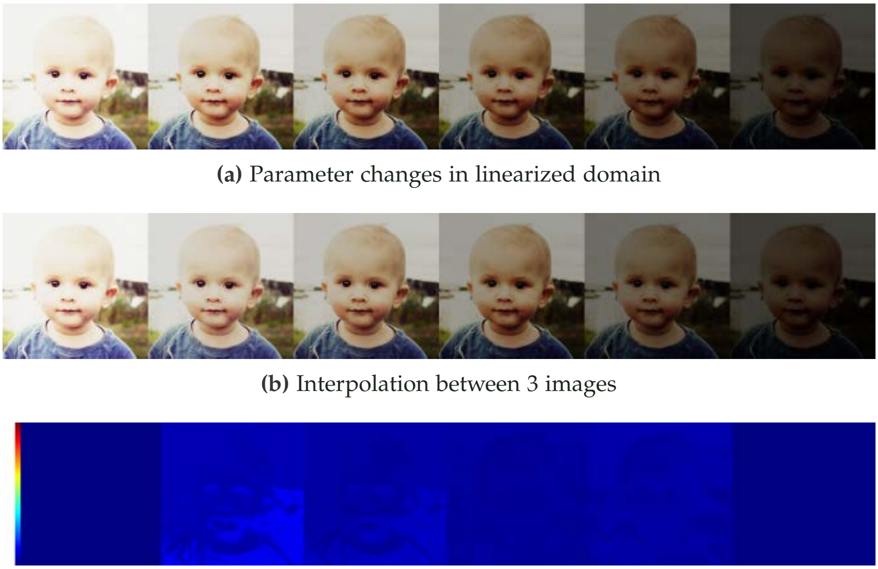
(b) Interpolation between 3 images

(c) Absolute error between **(a)** and **(b)**

Figure 3.1: Linear changes of the tone-mapping parameter σ in Equation 2.1, in the range $\sigma \in [0.05, 3]$. Using 3 interpolation images, at $\sigma = \{0.05, 1.525, 3\}$, there are large errors when image content is changing rapidly.

space. In the example in Figure 3.2, 3 points generate approximations with small errors, but to generalize to more complicated situations we use 5 points in the parameter adjustment experiment. However, even though this is a small number of sampling points, for a large number of dimensions, K, sampling the entire parameter space is impractical or even impossible. Moreover, it is also a very difficult problem to find the optimal point in such high dimensional space. To overcome these problems, we employ a conjugate gradient search, as proposed by Powell [205]. The search strategy allows for finding the local optimum of a non-differentiable function, from searching along conjugate gradient directions. The method is also robust to the high variance that is expected to be present in perceptual measurements. For an example, Figure 3.3a shows how the conjugate directions are explored for finding the optimal point in a 2D parameter space, using a few linear searches. Figure 3.3b shows the same example, but where errors are introduced in the searches. The optimal point can still be found by complementing with a few additional searches.

Given the search and interpolation strategies, a perceptual parameter optimization is performed by interpolating between 5 videos along one direction of the



(a) Parameter changes in linearized domain

(b) Interpolation between 3 images

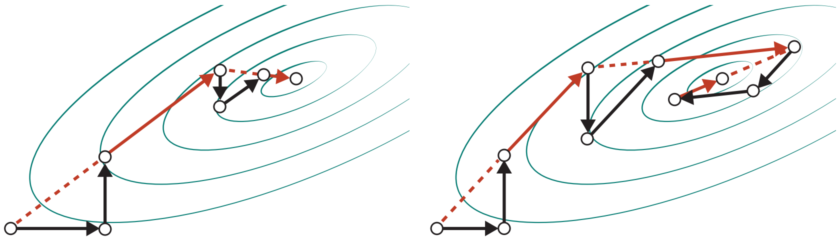
(c) Absolute error between **(a)** and **(b)**

Figure 3.2: Mapping the parameter σ to a domain where uniform changes in the parameter value yields approximately uniform changes in image content. This means that interpolation errors are smaller and better distributed across the parameter range. The 3 images used for the interpolation are located at parameter settings $\sigma = \{0.05, 0.4, 3\}$.

parameter space. The user is presented with a slider for selecting the optimal position along the direction. When this is found, 5 new videos are generated so that the search can continue along the next direction. This procedure is repeated, choosing directions according to Powell's method, in at least two full iterations, i.e. along $\geq 2K$ directions given K parameters. For the results in Paper **B**, four TMOs were selected for parameter optimization. These were the ones that did not offer default values or were deemed to generate unacceptable results with the default parameters. Four expert users performed the experiment on three different HDR video sequences, and the average optimum was used as final calibration.

3.2.2 Qualitative evaluation experiment

From initial experiments, presented in a pilot study [74], it was revealed that many of the existing methods for tone-mapping of HDR video produce unacceptable temporal artifacts. In order to identify and estimate the magnitude of



(a) Perfect judgments

(b) Noise applied

Figure 3.3: Parameter optimization in a 2D space by means of conjugate gradient updates, where the conjugate directions are illustrated in red. The examples show how exploration is performed with perfect measurements **(a)**, and with noise caused by non-consistent perceptual judgments **(b)**. The method is robust to the noise and can still find a good local optimum with additional iterations.

the different problems, a qualitative analysis was performed prior to the main subjective study and quality comparison.

Five expert users provided ratings of six tone-mapped HDR video clips. The ratings were made both in order to assess generated artifacts (flickering, ghosting, consistency, and noise level) and for measuring the appearance reproduction in terms of individual image attributes (brightness, contrast, and color saturation). The experiment provides valuable insights into common problems in video tone-mapping. Also, based on the results, four TMOs were excluded from the final pair-wise comparison experiment due to excessive flickering or ghosting artifacts. Since these artifacts are visually very prominent, it would not make sense to attempt making comparisons. With flickering or ghosting as the most salient feature in a tone-mapped video, it would potentially mask out comparisons in terms of other features.

In order to draw high-level conclusions from the qualitative experiment, the result presented in Paper **B** have been distilled in Figure 3.4. To this end, we provide an overall objective score of expected artifacts for each TMO, estimated by averaging over all different artifacts and across all the six video sequences. However, the ratings for noise level have been excluded. This is due to the observation that noise is a less objectionable image artifact, which can be accepted to a larger extent compared to other artifacts. The attribute ratings have been averaged in a similar fashion, using the absolute value of the scores. The errors provided in Figure 3.4 have also been averaged in the same way as the ratings. Thus, error bars represent the average standard errors for all individual sequences and categories. Calculating the standard errors across all the sequences and rating categories would be less informative, resulting in very large values.

The different plots in Figure 3.4 facilitate a direct comparison between the qualitative ratings and the subjective preference results from the pair-wise comparison experiment. The conclusions will be discussed in the next section.

3.2.3 Pair-wise comparison experiment

The final pair-wise comparison experiment was performed using the non-reference method, see Figure 2.6, asking for the video that appears most true to nature, or conception of the true scene. Although this task can be considered vaguer than a reference comparison, the setup is closer to how videos are viewed in real-life situations.

In total 18 observers conducted the experiment, comparing 7 TMOs in 5 HDR video sequences. The results are summarized in Figure 3.4, together with the averaged results from the rating experiment. The detailed results are provided in Paper B, reported individually for the 5 video sequences. The results are scaled in *just-noticeable difference* (JND) units [200], providing relative per-sequence quality differences. That is, the absolute level may differ between the sequences. In order to approximate an overall single quality indication for each TMO, we need to average the results across different sequences. To do so, while accounting for the different absolute levels, the per-sequence average is subtracted prior to averaging across sequences,

$$Q_t = \frac{1}{N_s} \sum_{s \in S} q_{t,s} - \mu_s, \tag{3.3a}$$

$$\mu_s = \frac{1}{N_t} \sum_{t \in T} q_{t,s}. \tag{3.3b}$$

Here, S and T are the set of sequences and TMOs, respectively. There are in total N_t TMOs and N_s sequences. $q_{t,s}$ is the quality level of a certain TMO t and sequence s. The measure Q_t should only be regarded as an indicator of the overall quality of the TMO t over the set of evaluated sequences, since the JNDs have been estimated per-sequence. The error bars in Figure 2.6 have also been calculated by averaging and thus represent mean 95% confidence intervals across the sequences.

First, Figure 3.4 demonstrates the overall artifact levels and attribute rendition problems of the four most problematic TMOs, which experience excessive flickering or ghosting artifacts. These are all local TMOs, and were excluded from the subjective evaluation. The remaining TMOs can all be regarded as global operators and seem to be significantly more robust in the temporal domain. This highlights the problems in retaining temporal coherence in advanced methods for local tone-mapping. The conclusion is not that global TMOs are preferred

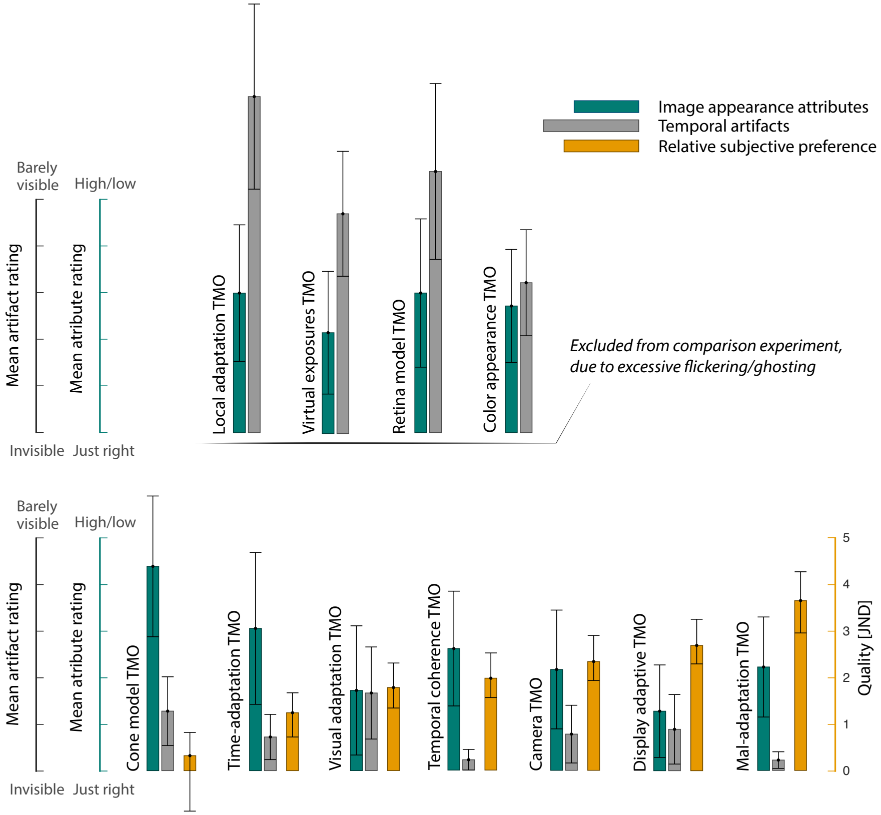


Figure 3.4: Results of qualitative rating experiment (image attributes and artifacts) and the pair-wise comparison experiment (subjective preference). The attribute and artifact ratings have been averaged across all the different attributes and artifacts, respectively, and across sequences. The relative subjective preferences have been averaged over mean subtracted per-sequence results. Error bars represent mean standard errors for the ratings, across artifacts/attributes and sequences. For the subjective preferences, error bars represent average sequence 95% confidence intervals.

over local TMOs, but rather that at the time of the evaluation there was a lack of temporally robust local methods for tone-mapping. Therefore, one of the major problems in video tone-mapping is to achieve a good level of detail and local contrasts, without introducing visible artifacts over time. These two goals are contradicting to a certain extent, where local processing increases the risk of generating spatial and temporal artifacts.

Among the global methods, the *Camera TMO*, using a simple camera curve, shows competitive performance in some of the sequences. However, for more complicated scenes the histogram-based methods can better adapt to the specific scene content and produce a higher level of overall contrast while compressing the dynamic range. This observation also agrees with other TMO evaluations performed on video sequences [201].

Figure 3.4 also reveals a connection between the two experiments performed in Paper **B**. There is a high negative correlation between the qualitative judgments and the end subjective quality. That is, with an increase in artifact and color rendition ratings, there is a decrease in subjective quality. This points to the importance of producing results with little artifacts and with well-balanced brightness, contrast, and colors. The ratings are apparently very good indicators of the performance in terms of subjective preference, where a weighted sum of the different ratings can be used as an accurate prediction of the subjective quality in the pair-wise comparison experiment. This also agrees with previous studies in quality of tone-mapping for static images, which show a correlation between separated image attributes and subjective quality [46, 47, 139]. For video material, however, the ratings on temporal artifacts are clearly of high importance, where high artifact ratings provide substantial evidence of a low subjective quality.

The work in Paper **B** concludes with a list of problems that were considered important to address in future development within tone-mapping for HDR video. This includes the challenge of maintaining a good level of detail and contrast, while at the same time not introducing temporal artifacts. Treatment of noise is also recognized as an important aspect of HDR video tone-mapping, which has received little attention in the literature. Moreover, efficient algorithms should be promoted, due to the large increase in data to be processed. These problems were subsequently dealt with in the work presented in Paper **C**.

3.3 New algorithms

The goal of the work in Paper **C** is to cater for high-quality local tone-mapping in real-time, without introducing visible temporal or spatial artifacts. To achieve this, we present a set of novel techniques for a) how to perform local tone-mapping, b) how to formulate a tone-curve for dynamic range compression, and c) taking into account the noise characteristics of the HDR input.

Relating to the categorization in terms of intent, as explained in Section 2.4.1, the TMO in Paper **C** most closely resembles an SRP method. The tone-curves attempt to preserve contrasts from the HDR scene as close as possible given the limitations of a certain target display. However, the technique for detail preservation also allows for strong enhancement without introducing visible spatial or temporal artifacts, which allows for artistic freedom inline with the BSQ tone-mapping intent.

3.3.1 Filtering for tone-mapping

As described in Section 2.4.2, detail-layer separation is usually performed by means of edge-preserving filtering. There exists a large number of different multi-purpose low-pass filtering techniques, which adapt to the edges within an image. One of the most common applications of such filters is noise reduction. However, this application differs from detail extraction in two important aspects. First, details are often significantly larger features than image noise. This means that an increased filter support is needed, both in the spatial and in the intensity domain. Second, while the filtered image is the end result in the case of noise reduction, for tone-mapping it is used to extract a detail layer from the input image. This detail layer is highly sensitive to how the filter accounts for image edges and easily reveals artifacts due to bias within anisotropic filter kernels.

One of the most commonly used multi-purpose edge-preserving filters in tone-mapping is the bilateral filter [15, 73, 237]. It allows for a simple formulation and can also be accelerated in different ways for relatively fast evaluation. Given the pixel value $L_{\boldsymbol{p}}$, at position $\boldsymbol{p}$ within the image L, the filtered pixel, $\hat{L}_{\boldsymbol{p}}$, is computed as

$$\hat{L}_{\boldsymbol{p}} = \sum_{\boldsymbol{q} \in \Omega_{\boldsymbol{p}}} \omega_s(\|\boldsymbol{q} - \boldsymbol{p}\|) \omega_r \left(\|L_{\boldsymbol{q}} - L_{\boldsymbol{p}}\| \right) L_{\boldsymbol{q}}. \tag{3.4}$$

The point $\boldsymbol{q}$ runs over a local neighborhood $\Omega_{\boldsymbol{p}}$ surrounding point $\boldsymbol{p}$. The bilateral weights ω_s and ω_r are usually formulated with Gaussian kernels, decaying with increasing spatial distance and intensity difference, respectively. Thus, filtering is suppressed both at large spatial distances from $\boldsymbol{p}$ and across large differences in intensity (edges). Since ω_r modulates the individual filter weights, the bilateral kernel makes for an anisotropic filtering close to edges, as visualized in Figure 3.5. This means that the filter can be biased towards the side of an edge that is closer in the intensity domain. The bias is manifested in a sharpening effect in the filtered image. For most applications, this is not a problem and not visually prominent for small filter kernels. However, in detail extraction for tone-mapping, the bias can create visible banding/ringing artifacts, where the image gradients are reversed as compared to the input image. For an example of the sharpening effect and banding artifacts in the detail layer, Figure 3.5 demonstrates detail extraction by means of the bilateral filter.

The artifacts in the detail layer are especially problematic for the BSQ intent, which often favors an exaggerated level of local contrasts. The same problem can also be found with the majority of the classical edge-preserving filters. Moreover, in video sequences the banding artifacts generally show an incoherent behavior from frame to frame, increasing their visual prominence.

In order to prevent banding artifacts, and enable robust tone-mapping in BSQ operators and video sequences, the filter from Paper **C** is based on isotropic filter kernels. In Equation 3.4, if the bilateral weight ω_r is removed, what remains is a standard Gaussian low-pass filter, $\hat{L}_{\boldsymbol{p}} = \left(G_\sigma * L_{\boldsymbol{p}}\right)$. Instead, by weighting different Gaussian convolutions, it is possible to adapt spatially to the image content and avoid filtering across edges,

$$L_{\boldsymbol{p}}^k = (1 - \omega_r)\, L_{\boldsymbol{p}}^{k-1} + \omega_r \left(G_{\sigma_{k-1}} * L_{\boldsymbol{p}}^{k-1}\right). \tag{3.5}$$

The filtering is performed iteratively, $k = \{1, 2, ..., K-1, K\}, L^0 = L, \hat{L} = L^K$, similarly to a diffusion process. In each step, the filtered image is weighted using the edge-stop function $\omega_r\left(||\nabla L_{\boldsymbol{p}}^{k-1}||\right)$, based on the gradient $\nabla L_{\boldsymbol{p}}^{k-1}$ at the point $\boldsymbol{p}$. This means that in uniform areas the end result is equivalent to filtering with a large Gaussian kernel, while the filter support is smaller close to edges. On the edges, there is no filtering, which is preferred over risking introducing artifacts. Also, since the edge itself is the most salient feature, it masks the reduced amount of extracted details in these regions. An example of the spatially varying isotropic filter kernels is visualized in Figure 3.5, together with the filtered image and extracted details. The details are effectively extracted using the isotropic filtering technique, without creating the banding artifacts.

Since the isotropic detail extraction filtering strategy is based on a sequence of separable Gaussian filters, it allows for efficient execution. The filter is implemented for hardware acceleration, computed by consecutively convolving the image with 1D filter kernels stored in the constant memory of the GPU.

3.3.2 Tone-curve

A tone-curve controls how the dynamic range of an HDR image should be compressed over the range of different luminances. Inevitably, since the dynamic range of a display device is limited, this means that contrasts have to be

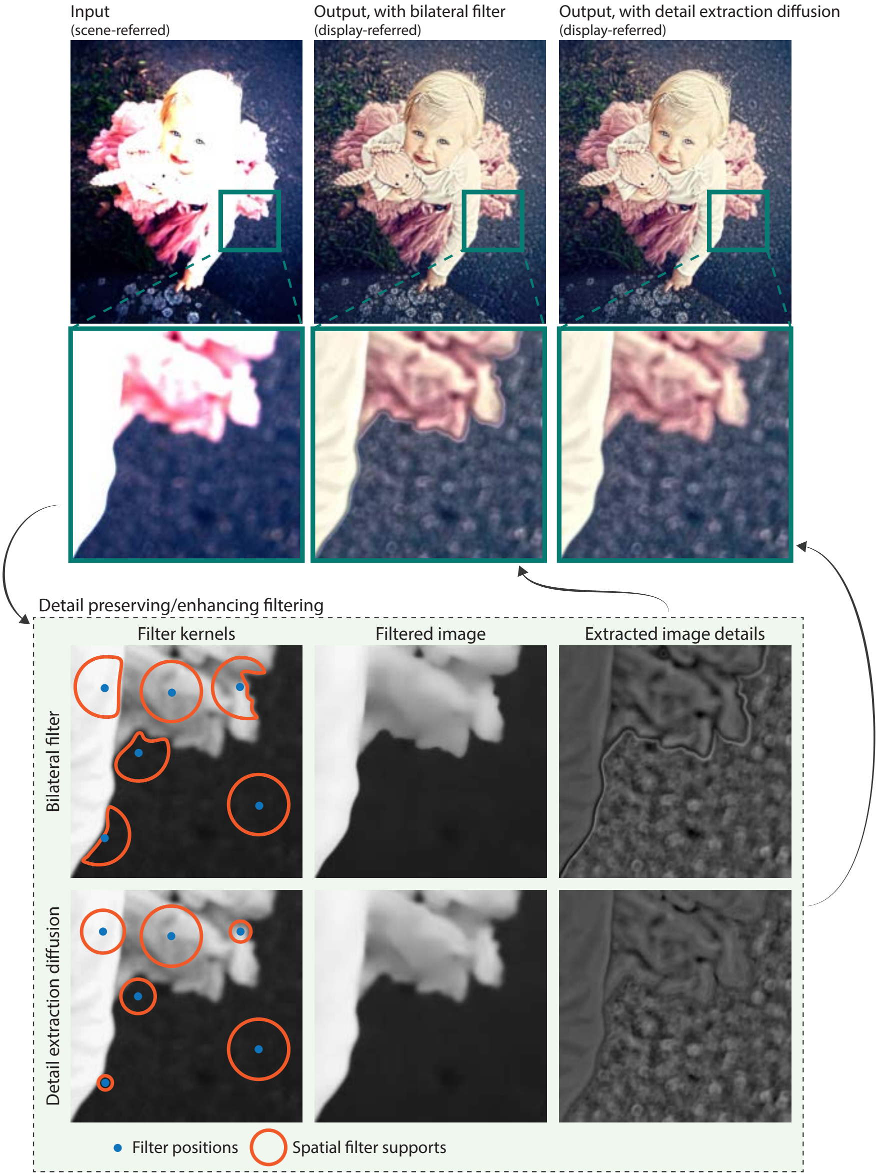


Figure 3.5: Example of detail extraction for tone-mapping. The input image is locally tone-mapped, with an enhanced level of details. This is accomplished with the detail extraction diffusion from Paper **C** and compared to using bilateral filtering. The detail extraction diffusion employs isotropic filtering kernels, which prevent the banding artifacts that are common to the anisotropic kernels of the bilateral filter.

compressed, or distorted. Traditionally, the tone-curve is S-shaped to preserve contrasts in middle tones at the cost of higher distortions in low and high tones, see Figure 2.5. In Paper C, the derivation of a tone-curved is posed as an optimization problem, with the objective of minimizing the distortions in contrasts.

Given a tone-curve $V : L \to T$, which maps the HDR luminance L to a compressed tone value T, in broad terms this amounts to the optimization problem

$$\arg\min_{V} ||\Pi(L) - \Pi\left(V(L)\right)||, \tag{3.6}$$

where $\Pi(L)$ is the contrast of L. This is subject to V mapping to the dynamic range of the target display device. By parameterizing the tone-curve as a piece-wise linear and monotonically increasing function, the slopes of each segment can be optimized given the image histogram for representing the probability distribution of contrasts over different luminance levels. An analytic solution can be derived and solved for very efficiently.

Examples of the minimum contrast distortion tone-curves are demonstrated in Figure 2.5 and Figure 3.6. These use the same input HDR image, but are plotted in linear and log domains, respectively. Compared to histogram equalization, the slope is constant for bin probabilities above a certain threshold (viewed in the log domain, Figure 3.6). This is in order not to increase contrasts from the tone-mapping. Contrasts should only be preserved to the extent possible and, therefore, the slope needed to achieve this should not be exceeded.

The content-adaptive nature of the tone-curve allows for minimal contrast distortions in different situations. Thus, a good overall distribution of contrasts in the tone-mapped image can be achieved. However, in order to better maintain local image contrasts, the tone-curves are computed over a set of local image regions. In order to avoid discontinuities due to widely different local image content, the tone-curves are computed by blending the local histograms with a small amount of the global image histogram. The mapping is then performed for each pixel using a per-pixel tone-curve interpolated from neighboring tone-curves.

The spatially varying tone-curves mean that the compound TMO in Paper C has two mechanisms for local adaptation. The local regions used for the tone-curves are on a relatively large spatial extent, in the vicinity of 5 visual degrees, while the detail separation filtering preserves or enhances the more local image features (approximately operating around 1 visual degree).

In order to maintain coherence over time, the nodes of the local tone-curves are temporally filtered, either using a low-pass IIR or an edge- stop filter. Due to the nature of the detail extraction filter that is employed, which only uses

Gaussian filters, there are no visible temporal artifacts related to the details. Hence, the tone-curve filtering is enough to ensure temporally coherent local tone-mapping of video sequences.

The tone-mapping pipeline follows the steps in Figure 2.4, using the special-purpose detail extraction filter and the minimum contrast distortions tone-curves. As a final step, the tone-mapped image is passed through the inverse of the display model in Equation 1.1. This transforms the image to a display-referred format, accounting for display dynamic range and ambient lighting.

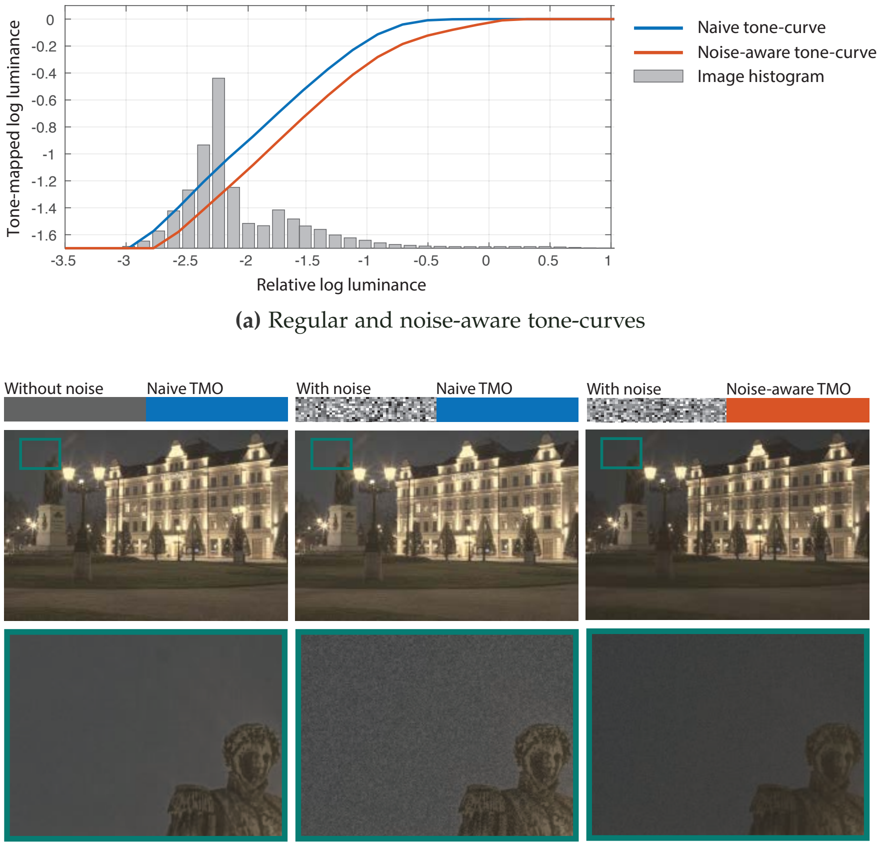
3.3.3 Noise-awareness

The presence of image noise has largely been disregarded in tone-mapping of static images, as this has not been a major problem. For video sequences, on the other hand, noise is an important aspect due to the difference in capturing techniques as compared to static images. While noise reduction has been researched for a long time, and also accounted for in HDR reconstruction, there is little work on the problem of noise specifically for tone-mapping.

Since a tone-curve generally compresses the dynamic range while attempting to preserve image content, the dark image areas are often boosted in intensity. This means that the noise is amplified and that noise not visible in the original image is revealed. The TMO presented in Paper **C** uses a noise-aware tone-mapping strategy for controlling the shape of the tone-curve, in order to make sure that noise is kept hidden in the dark image areas of the image. Based on measured or estimated noise characteristics, this is done by adding a saliency term that scales the bin probabilities of the histogram before optimizing for minimum contrast distortion. That is, preserving contrasts is not the only objective in the optimization – it should not be at the cost of revealed noise. Furthermore, since noise can also be retained in the detail layer, this is scaled according to the noise characteristics when it is added back after the tone compression.

Given knowledge about the viewing conditions and the noise level, the proposed method can ensure that noise is kept below the visibility threshold of the HVS. This noise-aware tone-mapping strategy presents a light-weight technique that is complementary to classical denoising methods. For example, if there is a substantial amount of noise, the end result can come out darker than intended, in order to conceal the noise. On the other hand, with denoising algorithms it is difficult to remove all the noise without introducing filtering artifacts, and the artifacts and/or remaining noise can be revealed by the tone-mapping. By combining denoising and noise-awareness the best compromise can be made between both removing and concealing noise.

An example of the impact of a noise-aware tone-mapping is shown in Figure 3.6. The image is captured with exposure bracketing, so that it contains only a very



(a) Regular and noise-aware tone-curves

(b) Noise-free HDR image **(c)** Artificial noise added **(d)** Noise accounted for

Figure 3.6: Demonstration of the noise-aware techniques from Paper **C**. In the images in **(c)** and **(d)**, artificial noise has been added. The tone-mapped images in **(b)** and **(c)** use the same tone-curve and detail level, disregarding noise. The tone-mapping in **(d)** uses the noise-aware processing (tone-curve and detail scaling). The tone-curves are shown in **(a)**, where the noise-aware version is computed by taking into account the amount of noise that has been added to the noisy images. The differences in tone-mapping are best viewed in the electronic version of the thesis. However, it should be noted that since the result may be viewed in different conditions (dynamic range, viewing distance, etc.), it cannot be guaranteed that noise is not visible in the noise-aware tone-mapping.

Name	Processing	Intent
Zonal coherence TMO [39]	Local	SRP
Motion-path filtering TMO [20]	Local	BSQ
Noise-aware TMO, Paper **C** [77]	Local	SRP

Table 3.2: List of video TMOs included in the comparisons in Paper **A**, in addition to the TMOs listed in Table 3.1. Thus, in total 14 TMOs were considered.

small amount of noise, as shown in the tone-mapping in Figure 3.6b. Next, noise has been artificially added followed by tone-mapping with the same tone-curve and detail level. The result reveals clearly visible amounts of noise, in Figure 3.6c. By using the noise-aware mechanisms, the added noise can be concealed in darker image areas and by reducing the level of details, as demonstrated in the tone-mapping in Figure 3.6d. Comparing the naive and the noise-aware tone-curves, in Figure 3.6a, the latter has a decreased slope for the dark parts of the image which contain most noise. In this way, the image noise is not boosted from the tone-mapping.

3.4 Recent developments

While Paper **A** provides an introduction and overview of tone-mapping, and particularly for video TMOs, it also attempts to assess the latest progress in video tone-mapping. To this end, an objective evaluation is performed, which is similar to the one in Paper **B**. This includes all the 11 TMOs (Table 3.1) that were considered in Paper **B**, plus three more recently published TMOs. These are listed in Table 3.2 and include the TMO from Paper **C**. They have all been developed specifically considering the challenges in tone-mapping of HDR video.

As opposed to the perceptual qualitative experiment of Paper **B**, the evaluation made in Paper **A** makes use of a set of quantitative measures for assessing temporal artifacts and image attributes:

1. **Temporal incoherence:** The temporal coherence at a pixel $\boldsymbol{p}$ in frame t, is measured using the cross-correlation $\rho(L_{p,t}, T_{p,t})$ between the HDR luminance L and tone-mapped luminance T. The correlation is measured in a window over K frames in time, from $t - \lfloor K/2 \rfloor$ to $t + \lfloor K/2 \rfloor$. It is formulated to account for different types of adaptation that can take place in the tone-mapped video. For example, when adapting to a new lighting situation a pixel of the tone-mapped video can potentially make a transition in intensity that is opposite in direction as compared to the same pixel in the HDR. This

should not be directly penalized by the correlation measure. As an example, this situation can occur if the scene contains a light source that is switched on in the HDR video sequence. The light can affect the background to show an increase in luminance, while at the same time the tone-mapping has to lower the overall brightness in order to fit the light source into the limited dynamic range of the display.

The measure ρ can be used for evaluating both the global correlation,

$$\Phi_{global} = \rho\left(\frac{1}{N}\sum_{p} L_{p,t}, \frac{1}{N}\sum_{p} T_{p,t}\right), \tag{3.7}$$

and the mean local correlation,

$$\Phi_{local} = \frac{1}{N}\sum_{p} \rho\left(L_{p}, T_{p}\right). \tag{3.8}$$

Here, N is the number of pixels in each frame. The final measure for incoherence is then formulated as $1 - \max(0, \Phi)$, disregarding negative correlations.

2. **Details:** The level of detail preservation in the tone-mapped images is estimated by extracting detail layers from both original HDR and tone-mapped images. The mean absolute values of the detail layers in the log domain represent the total amount of details within the images. Then, by comparing the measures between HDR and tone-mapped images, the decrease/increase in the amount of details after tone-mapping can be deduced.

3. **Exposure:** The amount of over- and under-exposure of a tone-mapped image are measured as the fractions of pixels that are bright and dark, respectively. This is different from measuring absolute brightness, but can better indicate if the tone-mapped image retains information in dark and bright image areas.

4. **Noise visibility:** In order to measure how much the visibility of noise has been increased or reduced by the tone-mapping, a set of computer-generated images are used. These are noise-free or contain very low levels of noise. After adding artificial noise to the HDR images, the perceptual difference compared to the noise-free image is measured using HDR-VDP-2 (v2.2) [172]. By tone-mapping both the original and the noisy images, the perceptual difference can then be measured also after the tone-mapping. Then, comparing the visibility of the noise, before and after the tone-mapping, the difference indicates if the noise-visibility is reduced, retained or boosted by the TMO.

The image attributes (details and exposure) are different than the ones used in Paper **B** (brightness, contrast and color saturation). Furthermore, the noise measure is also different, comparing the difference in noise-visibility as opposed

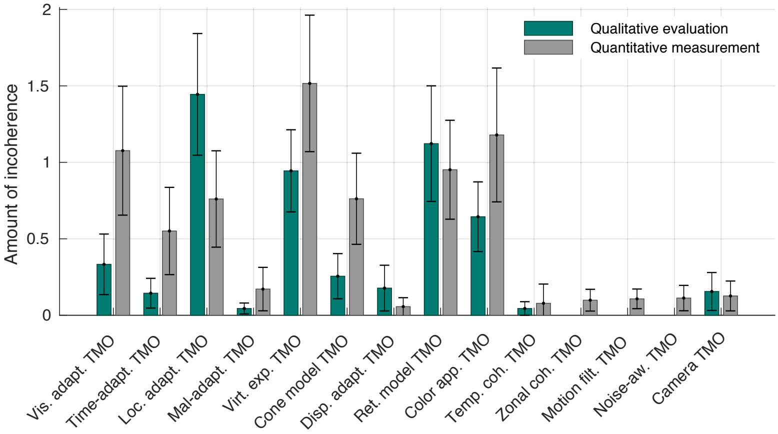


Figure 3.7: Temporal incoherence for the 11 TMOs in Table 3.1, measured from qualitative ratings in Paper **B**, and for the 14 TMOs in Tables 3.1 and 3.2 using the quantitative measures described in Paper **A**. The ratings and the measures have been evaluated over different sets of HDR video. Still, there is a good overall correlation between the two. Error bars show mean standard errors.

to the absolute perceived level of noise. However, the measure for temporal coherence is expected to be similar between the two evaluations. This is also confirmed in Figure 3.7, where the temporal artifact ratings from Figure 3.4 are plotted next to the sum of the global and local incoherence measures (using Equations 3.7-3.8). The three most recent TMOs lack the rating results, but for the other TMOs there is a high correlation between the perceptual ratings and the quantitative measures. The two evaluations are also performed with different sets of HDR videos, demonstrating that the correlation generalizes to different HDR video sequences.

While Figure 3.7 shows a general agreement between perceptual and quantitative measurements, for some of the TMOs the differences are larger. For example, the *Cone model TMO* performs a per-pixel filtering that is punished by the quantitative measure. Perceptually, the problem is not as prominent, since motion-blur is much less objectionable than e.g. flickering. Also, the *Visual adaptation TMO* measures much higher using the quantitative approach, presumably due to the way adaptation is handled by the method, allowing for rapid changes in intensity.

One of the central problems discovered in Paper **B**, which is one of the main focuses in Paper **C**, is the difficulty in performing local tone-mapping with a good level of local contrasts, while at the same time retaining a good temporal coherence. From the quantitative measurements in Paper **A**, we can show some evidence of this problem being addressed by the more recent TMOs. In Figure 3.8 the measured difference in details between HDR and tone-mapped images is plotted against the estimated coherence. This is taken as the negative sum of local and global incoherence, so that a higher value means better coherence. The figure shows that among the TMOs used in the evaluation in Paper **B**, only the global methods can retain a good temporal coherence. However, these cannot preserve the level of image details that is present in the original HDR images. The TMO that comes closest is the *Display adaptive TMO*, presumably due to its content-adaptive tone-curve, which renders better local contrast than a simpler tone-curve (e.g. in the *Camera TMO*). With three of the more recent TMOs (Table 3.2), which focus on HDR video tone-mapping, temporal coherence can be retained without sacrificing image details.

All in all, the results of the quantitative measurements in Paper **A** indicate which TMOs can be expected to render tone-mapped videos with good temporal coherence, details, exposure and low noise visibility. Given the discussion in Section 3.2.3, on how the different ratings provided for the qualitative evaluation in Paper **B** correlates with the end subjective preference, we can also expect that this is true for the different measures provided in Paper **A**. That is, the TMOs that provide the best result in terms of the different measures, can also be expected to provide competitive performance in a subjective comparison. Especially the measure of temporal coherence is central for tone-mapping of video, and this shows a general agreement with the perceptual ratings. In light of these observations, we can further confirm that the TMO from Paper **C** is capable of generating high-quality results with minimal amounts of artifacts.

3.5 Summary

The recent availability of high-resolution HDR video with a wide variety of content has made it possible to test TMOs against challenging dynamic scenes. The work presented in Paper **B** is the first to do so, and the results point to several deficiencies with the, at the time, existing TMOs for HDR video. The method in Paper **C** follows up on the work and proposes techniques for alleviating the specific problems that were pointed out by the study. Finally, in Paper **A** a quantitative analysis is performed, which shows that the method indeed can be expected to produce good local tone-compression while maintaining temporal coherence and without revealing image noise. Thus, the papers follow a natural chain of motivations, from uncovering existing problems, followed by developing techniques for addressing these, and finally verifying that tone-mapping with good performance can be expected. Moreover, a broad background and up-to-date reference on tone-mapping for HDR video is provided through the state-of-the-art report in Paper **A**.

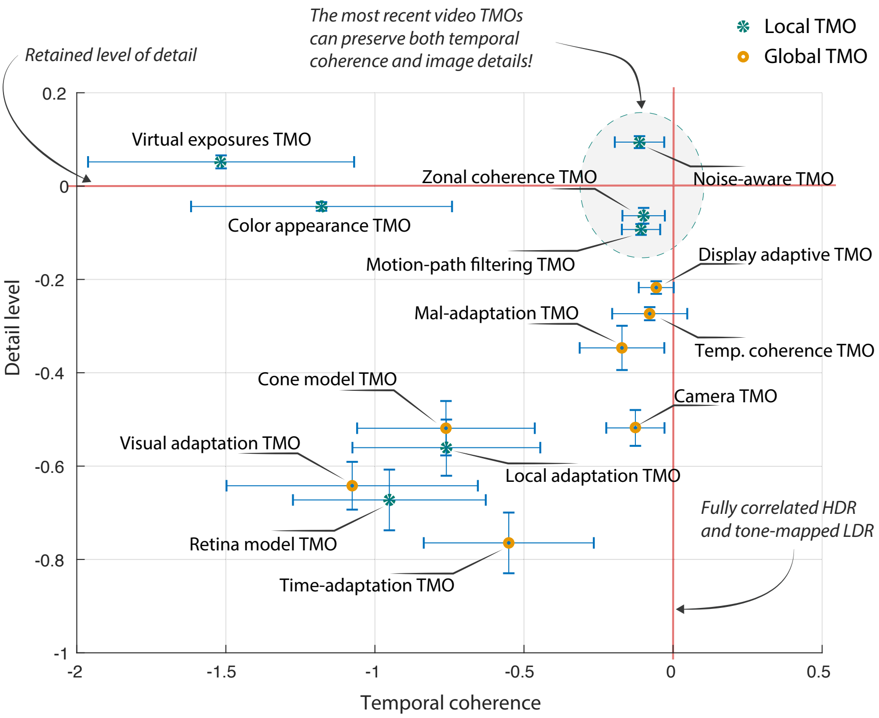


Figure 3.8: Temporal coherence versus local contrast/details for the 14 TMOs in Tables 3.1 and 3.2. The measurements are from Paper **A**, and indicate that more recent video TMOs can achieve a good temporal coherence and preserved image details. Vertical and horizontal error bars represent standard errors for detail and coherence measures, respectively.

3.5.1 Limitations and future work

While the experiments in Paper **B** provide a number of important insights to the problem of tone-mapping for HDR video, there are also several difficulties in evaluating the performance of TMOs. It cannot be emphasized enough that TMO evaluation is a very difficult task, and insights hereof were also gained through the evaluation work. The results of a study are highly dependent on the particular experimental setup. For example:

- The *Mal-adaptation TMO*, which is ranked as the best performing TMO in Figure 3.4, was also included in the parameter adjustment experiment. Other TMOs may potentially also gain in visual performance with optimized parameters.

- Details regarding interpretation, implementation, and usage of methods, or bugs in the code for that matter, can potentially affect the results. It is sometimes difficult to entirely conform to how the original authors of a method intended for it to be implemented and used. The work in this thesis applies the best efforts in order to stay true to the different methods that were evaluated.
- Generalization of performance from a limited set of sequences cannot be guaranteed.

With the rapid increase in HDR video and methods for tone-mapping of such, there are ample opportunities and motivation for conducting more studies. With differences in e.g. material, TMO selection, and experimental setup, further insight can be gained, facilitating future development.

While the techniques in Paper **C** make for high-quality video tone-mapping with minimal amounts of artifacts, there are also some situations which are more difficult to tackle. For example, the detail extraction filter has problems with detecting thin images features. For large amounts of detail enhancement, this can result in visible halo artifacts upon close inspection of such features in the final tone-mapping. The problem could potentially be resolved by exploring a better edge-stop criterion. Moreover, the tone-curve compresses the dynamic range of the HDR input to entirely fit the dynamic range of the display. This means that highlights, such as fire and light sources, in some situations may look artificial, see e.g. the lamp in Figure 1.3c. Special considerations could be made in high-intensity image regions, in order to allow for some clipping. Finally, the local tone-curves can in certain situations reveal visible borders between regions, despite the interpolation. Future work could explore how to better blend tone-curves, how to evaluate tone-curves based on content-dependent local regions, or how to employ local tone-curves at multiple spatial scales.

In addition to improving the different techniques of the method in Paper **C**, future work could also explore other aspects that are not included in the TMO. For example, it could be complemented with dedicated color appearance modeling. It would also be of interest to investigate the different parameters, in order to facilitate easier or automatic calibration that depends on the situation.

Another interesting avenue for future exploration is to investigate applications of the different quantitative artifact and attribute measures from Paper **A**. For example, these could potentially be combined in order to create a subjective quality index. This would be specially tailored for evaluation of HDR video tone-mapping, using the temporal incoherence measure as an important component.

Chapter 4

Distribution of HDR video

HDR imaging has for many years constituted an important component in computer graphics applications within research and production. The last decade has also shown a steady increase in research interest in HDR video. Moreover, within the last couple of years HDR has been introduced to the consumer market and, spurred by latest developments in HDR TV displays, it is rapidly gaining in popularity. Hence, there is a lot of activity around HDR video for commercial purposes and standardization has been ongoing for quite some time. However, the concept of HDR video for the consumer market is still in its infancy. There is a long way to go before hardware and software have fully adapted to this new format. One of the most central aspects of the transition towards HDR support is how to encode the HDR video content, providing viable options for distribution in different situations.

With the developments around HDR video distribution, there is a need for comparing and evaluating the techniques that have been proposed for the different components of the HDR video encoding pipeline. This chapter discusses the work and contributions of Paper **D**, which aims at assessing a number of such techniques. The work also recognizes the lack in availability of non-proprietary solutions for HDR video encoding, by presenting the *Luma HDRv* software, which is released under open source terms. In Section 4.1, a brief context and motivation is provided, followed by a discussion of the evaluation from Paper **D** in Section 4.2. The Luma HDRv codec is described in Section 4.3, and the chapter is summarized in Section 4.4, together with a discussion on limitations and possible directions for future work.

4.1 Motivation

As described in Section 2.3.2, the most straightforward, convenient, and efficient strategy for encoding of HDR video, is to make use of existing video codecs that are intended for LDR data [166]. This requires the floating point, scene-referred, pixels to be transformed to an integer format that is better suited for encoding. In the same manner as gamma correction and the sRGB color space (BT.709) make quantization errors spread approximately perceptually uniformly across the range of LDR values, this transformation should accomplish a similar goal for HDR values. Moreover, the encoding of the transformed luminance needs to be performed at an increased precision (usually 10-12 bits) as compared to LDR data. Despite active development of different techniques for how to transform HDR luminances and colors, these lack a comprehensive comparison. One previous comparison was performed by Boitard et al. [41]. They conducted a perceptual study for estimating the minimum bit-depth required for encoding HDR data without visible distortions, and the perceptual uniformity of different color and luminance encodings. This is accomplished by evaluating differences between gradient patches, which are encoded with the different techniques. Compared to this work, the evaluation in Paper **D** aims at assessing the final overall quality of different encoding schemes when applied to a wide variety of natural HDR videos.

Another observation that motivated the work of Paper **D** was that, while standards had been specified for the purpose of HDR video distribution, at the time there were no solutions available for HDR video encoding on open source terms. The work that is described in Paper **D** presents both a comparison of different pixel encodings and an open source HDR video codec solution. For each of the components in the encoding pipeline, the codec is designed by choosing the technique that indicates the best performance in the comparisons.

4.2 Evaluation

The individual steps that are involved in preparing an HDR video for integer encoding with a video codec are illustrated in Figure 4.1. Assuming that the input HDR pixels are specified by RGB colors, these are transformed to decorrelated luminance and chrominance channels. Next, the luminance is mapped with the PTF, to a domain of increased perceptual linearity. If the color separation is omitted, preserving the RGB coordinates, all channels need to be transformed by the PTF. Following a quantization to the target bit-depth, the final bit-stream is then compressed with a conventional video codec.

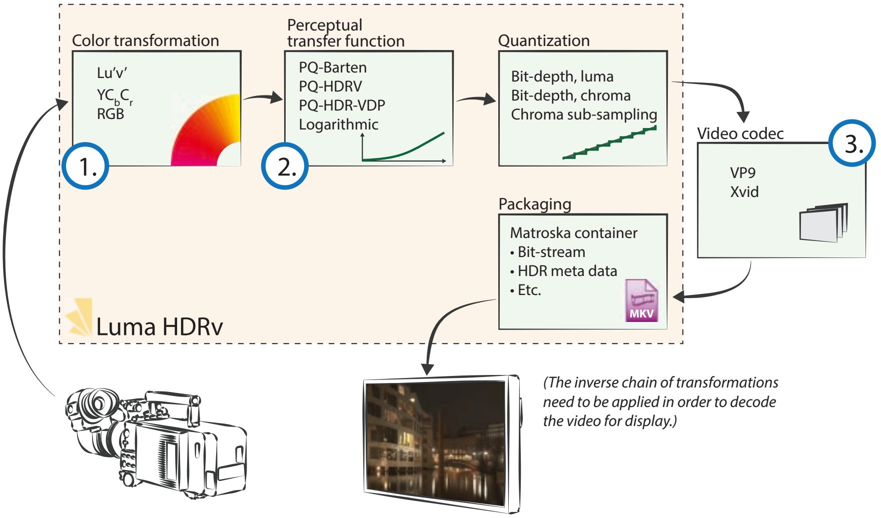


Figure 4.1: The pipeline for preparing HDR video for integer encoding with a conventional video codec. The numbers 1-3 are components for which the different settings are compared in the evaluation in Paper **D**. As indicated, the Luma HDRv software provides a layer of HDR specific features that can adapt a codec for HDR video. While any video codec can be used, that provides encoding at 10-12 bits, Luma HDRv is currently bundled with VP9.

4.2.1 Setup

The evaluation in Paper **D** considers 33 HDR video sequences, encoded at 15 different quality levels and using 9 different settings of the encoding pipeline. Due to the excessive amount of data (4,455 videos in total), a subjective evaluation was not an option. Moreover, many of the compared conditions experience only sub-threshold differences, so that it would be difficult to differentiate between the videos in a subjective comparison. For these reasons, the perceptual similarity comparing the input videos and the encoded-decoded counterparts were instead measured with two perceptual objective measures: HDR-VDP-2 [172] and PU-MSSIM; the multi-scale structural similarity index (MSSIM) [256] applied after perceptual uniform (PU) encoding [18]. Both measures have been demonstrated to correlate well with subjective comparisons. For a closer description of the practicalities involved in quality prediction of HDR content, we refer to the explanations by Mantiuk [173].

Although the comparisons were made by means of computational measures, the large amount of data was still a problem, requiring encoding of the 4,455 videos, and with close to 0.5M image to image comparisons. As the objective measures are computationally expensive, the total time of the evaluation would be several months running on a single multi-core machine. Instead, the computer cluster at High Performance Computing Wales[1] (HPC Wales) was employed, which made it possible to run all comparisons in a matter of a few days.

4.2.2 Results

The evaluation in Paper D considers different settings for three of the components in the HDR video encoding pipeline, as illustrated by the numbers 1-3 in Figure 4.1. Each component is treated separately, by varying its settings while keeping the rest of the pipeline unchanged. Following, the different settings and the results are briefly described:

1. **Color transformation:** The YC_bC_r color difference encoding is commonly used for video material. In the ITU-R Recommendation BT.2020, this was extended to a wider gamut, as compared to the previous BT.709, in order to accommodate HDR content. However, even the updated specification cannot represent the full gamut of visible colors. In Paper D we compare YC_bC_r to the wider gamut of Lu'v' [263], and also include RGB as a reference.
   The results in Paper D show a clear advantage of Lu'v' over YC_bC_r. This also agrees with the results presented by Boitard et al. [41], demonstrating that Lu'v' is better at separating the information between luminance and chrominance, thus decreasing the inter-channel correlations. Finally, comparing Lu'v' and YC_bC_r to encoding directly in RGB space shows, as expected, how the latter is clearly inferior.
2. **Perceptual transfer function:** Three different perceptual luminance encodings (PTFs) are included in the evaluation, and compared to a logarithmic mapping. The PTFs are a) PQ-Barten [185], b) PQ-HDRV [166], and c) PQ-HDR-VDP [172]. These are derived in a similar fashion, but using different psychophysical measurements. PQ-Barten is commonly referred to as PQ (perceptual quantizer) and is employed e.g. in the standards of HDR10 and Dolby Vision. This function is plotted in Figure 2.3 together with the log transform. PQ-HDRV and PQ-HDR-VDP show some variations, but have similar shapes as PQ-Barten.
   The results in Paper D show that the simple log transform clearly gives inferior performance. This is expected, as the log transform only is a good approximation of perceptual linearity for larger luminances (photopic vision). However, from the measurements it is difficult to differentiate between the three perceptual encodings. All three are most likely good options for luminance encoding in HDR video.

[1] http://www.supercomputing.wales

3. **Video codec:** Since the work in Paper D aims at providing an open source solution for HDR video encoding, the underlying codec itself needs to be released on similar terms. At the time of the evaluation there were not many such choices that were able to encode at an increased bit-depth. The choice fell on Google's VP9, which has been demonstrated to have a similar performance as the widely used H.264/AVC standard [216]. The older MPEG-4 Part 2 encoding, provided through the XVID codec, is also included for comparison. This was used in the seminal HDR encoding work by Mantiuk et al. [166], modified to provide a higher bit-depth.
   As expected, the results in Paper D show a substantial improvement using the more recent VP9 codec. It is able to provide the same HDR-VDP-2 quality prediction as XVID, but at around half the bit-rate.
   With the transition from H.264/AVC to H.265/HEVC, a significant improvement in encoding performance can be achieved [216]. Although there were no open source implementations of HEVC at the time of the evaluation, the situation is different today. The possibilities in improving encoding performance using more recent codecs will be discussed in Section 4.4.

4.2.3 Comparison to HDR10

For all of the comparisons discussed above, the range of encoded luminance is between 0.005 cd/m^2 and 10,000 cd/m^2. Furthermore, the transformed luminance, or luma, is encoded at 11 bits, while the final chrominance channels, or chroma, are encoded at 8 bits. These bit-depths have been demonstrated to be the minimum required in order to make sure that quantization artifacts are kept below the visibility threshold for the particular range of luminances when encoding using the Lu'v' color space [41, 166]. The YC$_b$C$_r$ encoding, on the other hand, requires > 8 bits per chroma channel, as it is not as effective in decorrelating information between luminance and chrominance. This was e.g. shown in the experiments by Boitard et al. [41]. The same experiments also indicated that the final luma channel using the YC$_b$C$_r$ color space may require a slightly less number of bits than for the Lu'v' transformation. 10 bits luma was also demonstrated to be enough by Miller et al. [185]. That is, it seems that YC$_b$C$_r$ and Lu'v' provide different distributions of information between luminance and chrominance. While 10 bits for both luma and chroma channels is the minimal requirement for YC$_b$C$_r$, 11 and 8 bits are better suited for luma and chroma, respectively, of the Lu'v' color space.

While Lu'v' luma/chroma at 11/8 bits was determined to be the best choice from the evaluation in Paper D, and is used as default settings for the Luma HDRv codec, the evaluation do not include a comparison to YC$_b$C$_r$ luma/chroma at 10/10 bits. The latter option corresponds to the most widely used HDR encoding standard, HDR10. In order to give additional insight into the differences

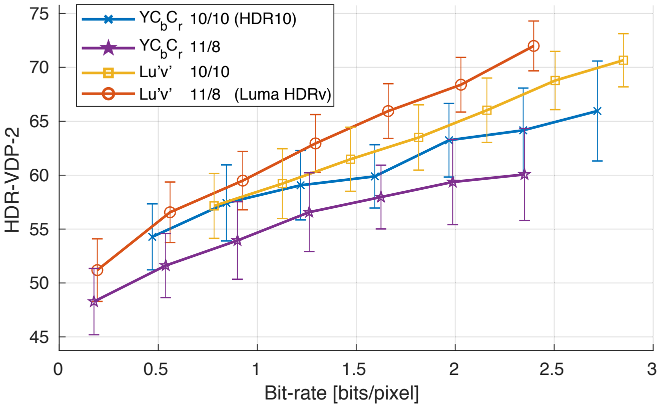
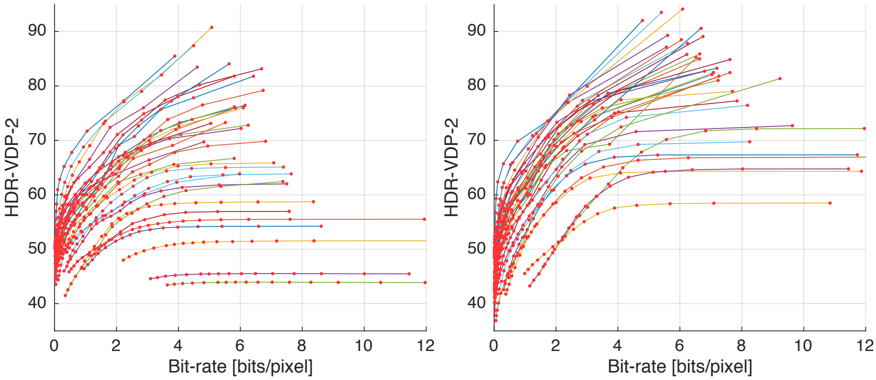


Figure 4.2: Rate-distortion plots, comparing the Lu'v' color space at 11/8 bits luma/chroma (Luma HDRv) and the YC$_b$C$_r$ color space at 10/10 bits luma/chroma (HDR10). These bit-depths indicate the best performance for the respective color spaces, as compared to encoding YC$_b$C$_r$ at 11/8 and Lu'v' at 10/10. Error bars denote standard errors.

between these two settings, Figure 4.2 complements Paper **D** with an additional comparison. This has been estimated in the same manner as the comparisons in Paper **D**, using the same 33 HDR video sequences provided by Fröhlich et al. [95], and encoding with VP9. Also, PQ-Barten is employed by both Luma HDRv and HDR10. However, the comparison is made only in terms of HDR-VDP-2 and with only 1 second from each video instead of 5 seconds as was used in the original experiments. The results for Luma HDRv and HDR10 in Figure 4.2 have been estimated from the per-sequence results in Figure 4.3, averaging across equal bit-rates of the sequences at 7 different sampling points. For each sequence, the qualities at these specific bit-rate sampling points have been computed from interpolation between neighboring measured bit-rates.

For comparison, the results in Figure 4.2 also include YC$_b$C$_r$ encoded at 11/8 luma/chroma and Lu'v' encoded at 10/10 luma/chroma. Thus, the rate-distortion plots that use 11/8 luma/chroma are the same as in Paper **D**, comparing these to 10/10 luma/chroma. However, there may be some smaller differences compared to the results in Paper **D**, due to different sampling and filtering, and since a newer version of VP9 is used for encoding. The results show that, as expected, the YC$_b$C$_r$ color encoding benefits from 10/10 luma/chroma, as in HDR10, and that the opposite is true for Lu'v'. However, there still seem to be clear advantages of using the Lu'v' color space, especially at higher bit-rates.



(a) HDR10 (Y$C_b$$C_r$, 10/10 bits) **(b)** Luma HDRv (Lu'v', 11/8 bits)

Figure 4.3: Per-sequence rate-distortion plots of the 33 different HDR video sequences evaluated, using the settings of HDR10 **(a)** and Luma HDRv **(b)**.

4.3 Luma HDRv

The software that was implemented and released together with Paper **D**, Luma HDRv[2], provides C++ libraries/API and applications for HDR video encoding and decoding, as well as playback of HDR video:

- **libluma_encoder** and **libluma_decoder:** Libraries for preparing, encoding, packaging, and decoding HDR video. In the current version, Luma HDRv is bundled with VP9 for the final encoding of luma and chroma. Packaging is provided through the Matroska[3] container, which makes it easy to add support for Luma HDRv in existing applications.
- **lumaenc** and **lumadec:** Applications that can encode and decode HDR videos with a range of settings using the Luma HDRv libraries. The default settings employ the best performing techniques from the evaluation, Section 4.2. However, there are options to change the settings for each of the components in Figure 4.1. The input/output HDR video can either be stored as OpenEXR[4] frames [37], or it can be piped from/to the PFSTools[5] HDR processing applications [169] to allow for extended compatibility with HDR formats.
- **lumaplay:** Minimal HDR video player that decodes and displays the video with OpenGL/GLSL. The simple GUI of the player provides options for changing the exposure of the video, in order to reveal the full dynamic range.

[2] http://lumahdrv.org/

[3] http://www.matroska.org

[4] http://www.openexr.com

[5] http://pfstools.sourceforge.net

Luma HDRv provides a light-weight layer on top of any high bit-depth codec, as illustrated in Figure 4.1. It is currently released bundled with Google's VP9 codec, but it would not require a big update to allow support for other codecs. Thus, Luma HDRv can be thought of as a video codec abstraction layer, which makes it possible to add the processing required for encoding and decoding of HDR video.

With the different settings provided by Luma HDRv, it is possible to encode according to existing HDR video standards. For example, by encoding colors from the YC_bC_r color space, with 10 bits luma and chroma, the result complies with the most widespread standard, HDR10. In the latest release of Luma HDRv, the packaging has also been updated in order to store the correct metadata associated with HDR10. This makes the video compatible with applications that support HDR10 and can decode VP9 video stored in a Matroska container. For example, it has been verified that HDR10 encoded with Luma HDRv can be uploaded to Youtube and Vimeo, where it is correctly recognized and processed as HDR video.

4.4 Summary

The area of HDR video distribution is becoming increasingly more important with the ongoing transition to HDR content, especially within the TV industry. The work discussed in this chapter contribute to the area by providing insights to compression efficiency for a number of different pixel encoding schemes, and by making the Luma HDRv open source HDR video coding software available. In order to complement this work, which is presented in Paper D, we have also recognized that the YC_bC_r and Lu'v' spaces may require different proportions of bit-depth between luma and chroma, to allow for optimal encoding performance. This is not considered in Paper D, which uses 11 bits luma and 8 bits chroma for all of the luminance-chrominance separated conditions. In Section 4.2 we have complemented with an additional comparison, which indicates a significant improvement of the 11/8 bits Lu'v' luma/chroma used as default settings in Luma HDRv, as compared to the 10/10 bits YC_bC_r luma/chroma specified by HDR10.

4.4.1 Limitations and future work

With the activity around HDR video, we expect that the there will be an increasing number of solutions available for HDR video distribution, and also open source alternatives. For example, BBC's Turing codec provides an open source implementation of the HEVC compression scheme. Turing supports HDR video encoding integrated with the codec, using the PQ-Barten and

Hybrid Log-Gamma (HLG) transfer functions. There are also a number of additional royalty-free alternatives to HEVC that, similarly to the Turing codec, have appeared within the last year or so. As with the Turing codec, it will likely become more common that such implementations provide direct HDR support by integrating the transformations needed with the codec, and providing a specific encoding option for HDR content.

An interesting development is also the initiative from the Alliance for Open Media (AOM), where a group of the leading internet companies have joined forces, including Apple, Amazon, Cisco, Google, Intel, and Netflix. The first objective of AOM is to create the next generation video codec, AV1, with increased performance as compared to HEVC. AV1 will be released as an open and royalty-free video codec, in order to avoid the patenting problems associated with existing video codecs such as AVC and HEVC. AV1 will be built by considering a number of elements from existing open source video codec initiatives, such as Google's VP10, Mozilla's Daala, and Cisco's Thor.

Given this direction of development, with open source high performing video codecs that potentially can support HDR content, the Luma HDRv software may soon be considered obsolete. However, we believe that Luma HDRv still can provide a useful HDR abstraction layer and packaging application, which gives easy and flexible control over the HDR specific settings of the encoding. This can be achieved by making a more explicit disconnection of the codec used under the hood, as illustrated in Figure 4.1, so that Luma HDRv easily can be compiled with a number of different codecs. This would allow for a joint and easy to use interface, where a range of different HDR specific encoding settings could be controlled regardless of the particular underlying codec. Moreover, Luma HDRv also provides an API that makes HDR video encoding and decoding easy to integrate in software development.

Another possible direction for future work is to extend both the evaluation in Paper **D**, as well as Luma HDRv, with alternative solutions for color encoding, luminance encoding, and video codec. For example, it would be interesting to see how the HLG function compares to the other PTFs. Also, following the discussion above, comparisons could be made in order to see how a HEVC implementation compares to VP9 for HDR video encoding, and what improvements can be expected from the AV1 codec when this becomes available. Finally, a perceptual study could be performed on a selected number of videos and conditions, in order to connect the objective results to expected subjective preference.

Chapter 5

Single-exposure HDR image reconstruction

Throughout the thesis, we have emphasized the benefits of HDR images. A wide variety of applications can take advantage of the extra information that the format provides. However, due to the inherent limitations of conventional camera sensors, HDR image capturing is still expensive and/or time-consuming. Moreover, the absolute majority of existing image and video material do not provide HDR information. Therefore, there would be great benefits in being able to estimate the extra information from a single-exposure image, so that HDR images can be provided from an unmodified conventional camera, or derived from the vast amount of existing images.

In this chapter, the method from Paper **E** is discussed, which takes a machine learning approach to the problem of HDR reconstruction from a single-exposure image. The method can provide convincing HDR images in a wide range of scenes, provided that areas with all color channels saturated are limited in size. The reconstruction is possible by making use of the recent progress in deep learning, and from careful considerations in terms of data augmentation of a gathered set of HDR images that are used for training. First, the single-exposure reconstruction problem is closer defined and motivated in Section 5.1. Then, the recent trend of applying deep learning strategies in HDR imaging is discussed in Section 5.2. In Section 5.3 the techniques used in Paper **E**, and extensions thereof, are described. Finally, Section 5.4 summarizes the chapter and provides a discussion on the limitations of the work, as well possible directions for future work.

5.1 Motivation

5.1.1 Relation to inverse tone-mapping

In Section 2.2.3, the problem of inferring an HDR image from a single exposure was divided into the three sub-problems of 1) decontouring, 2) tone expansion, and 3) reconstruction of under/over-exposed image areas. Most inverse tone-mapping operators fall into the second category and the objective is usually not to reconstruct the HDR image as close as possible. Instead, they attempt at achieving the best visual performance in end applications such as HDR display or IBL. For display on an HDR capable device this means that global mappings are generally preferred [177, 226], due to problems in reconstructing colors and details in over-exposed areas. Thus, saturated image regions remain saturated on the HDR display. Since we are accustomed to viewing images with saturated pixels, this may work well. For IBL, on the other hand, the iTMOs that attempt to boost highlights using expand maps or similar can be expected to provide better reproduction of the lighting from high-intensity light sources. However, due to problems in estimating the saturated image regions, the rendering quality can generally benefit from a global boost in image intensity of the IBL panorama.

The iTMOs can give substantial improvements when using LDR images in HDR applications. However, due to the different problems, the end HDR image can actually deviate more from the ground truth HDR image than the input LDR image. The problem of actually reconstructing, or approximating, the missing information as close as possible, similar to HDR reconstruction from multiple exposures, is conceptually different from the iTMO approaches. Although there are some previous methods that consider the reconstruction of colors and details in saturated pixels [105, 179, 267, 278], these only work for smaller corrections of over-exposure, or for textured highlights by requiring some manual interaction [255]. In Paper E we show successful reconstruction of both colors, details, and high intensities, which have not been demonstrated to be possible before. With the reconstructed information, LDR images can be used in a much wider range of HDR applications than previously possible, such as exposure correction, tone-mapping, and glare simulation. And for other applications, such as HDR display and IBL, the result can be much closer to what a real HDR image would yield.

5.1.2 Where is the dynamic range?

Comparing LDR to HDR images, in general the most significant difference is due to lost information in over-exposed image areas. This can be explained by inspecting the image histograms of natural HDR images. An example is plotted

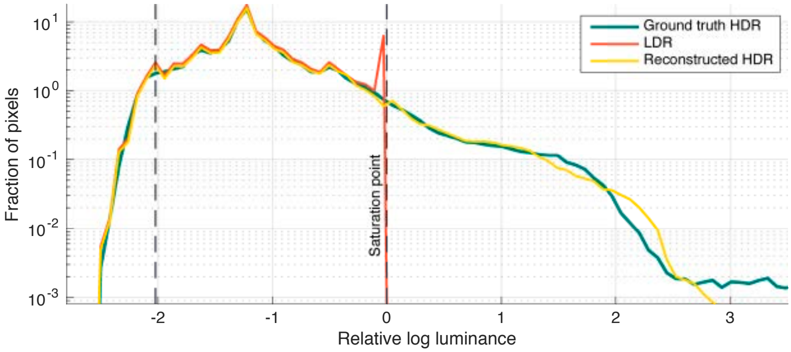


Figure 5.1: The histograms of the images from Figure 5.5, calculated in the log domain. The vertical lines show the 5% and 95% percentiles. However, due to log axis, it may seem like more than 5% of the pixels are located above the saturation point. The log axis helps in showing the tail of decaying high intensities of the distribution, which contains a small fraction of pixels but a large fraction of the dynamic range.

in Figure 5.1, which shows the distribution of pixel values in the log domain for the HDR image in Figure 5.5. The left and right vertical lines show the 5% and 95% percentiles, respectively. That is, the histogram values outside the left and right lines contain 5% of the darkest and brightest pixels, respectively. The distance between the lines is ≈2 $\log_{10}$ units, which is in the order of the dynamic range of a conventional camera sensor. This means that, for this example, such sensor can capture around 90% of the image information. The information lost in dark image regions shows less than 0.5 $\log_{10}$ units of additional dynamic range, and does not contribute very much to the final image. The information that is lost due to saturation of the sensor, on the other hand, contains an additional >3 $\log_{10}$ units of dynamic range. Figure 5.1 also shows how the information above the saturation point (right line) has been clipped, in order to simulate an LDR image. Consequently, the rightmost histogram bin contains about 5% of the pixels. By mapping the clipped image through a camera curve, quantizing to 8 bits, and reconstructing the lost information using the method in Paper E, the histogram of the reconstructed HDR image shows that most of the dynamic range has been recovered. Thus, by providing reconstruction of only 5% of the saturated pixels, the dynamic range is boosted by several $\log_{10}$ units.

Although the above example uses a night scene, with a very skewed distribution of pixels, natural images, in general, show similar statistics, where a small number of bright pixels store a large amount of the dynamic range. This is illustrated in Figure 5.2, which is computed from averaging over the database of >3,700 HDR images used in Paper E. The blue and red histograms have

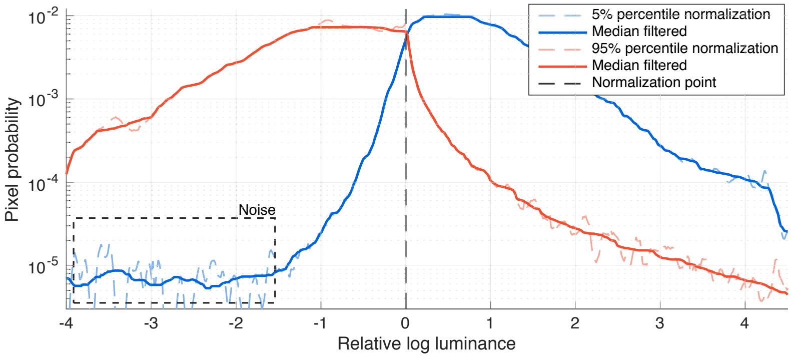


Figure 5.2: Averaged histograms in the log domain, using the 3.7K HDR images from Paper E. The two histograms have been computed after normalization by anchoring the 5% and 95% percentiles to 1, respectively, for the two different averages. For the 5% normalization the averaged histogram is dominated by noise at low values. Median filtered histograms are provided to better illustrate the shapes of the histograms.

been calculated after each image has been normalized using the 5% and 95% percentiles, respectively. The distribution of the 5% darkest pixels decreases fast and is soon dominated by image noise. In the 5% brightest areas, the slope of the histogram is less steep and there is information available for many $\log_{10}$ units. Although there are low probabilities for the brightest pixels, these are often important in HDR applications and capture the very essence of high dynamic range.

5.1.3 Focusing on the important

In Paper E we consider the problem of reconstructing pixels that have been clipped due to sensor saturation. As discussed above, this is the most important problem when transforming an LDR image to HDR.

The darkest pixels are only revealed in less common situations, e.g. when an image is captured with an overall too short exposure, or in extreme tone-mapping situations. Nevertheless, reconstruction of under-exposed pixels would probably also be well-suited for the same method as presented in Paper E. However, the most prominent feature of dark pixels is noise, which makes the problem considerably different from reconstruction of saturated pixels. Also, noise in the ground truth data will be a significant issue, as seen in Figure 5.2, making training problematic.

The problem of decontouring would likely also be well-suited for deep learning. In the same manner as for predicting under-exposed pixels, there are very

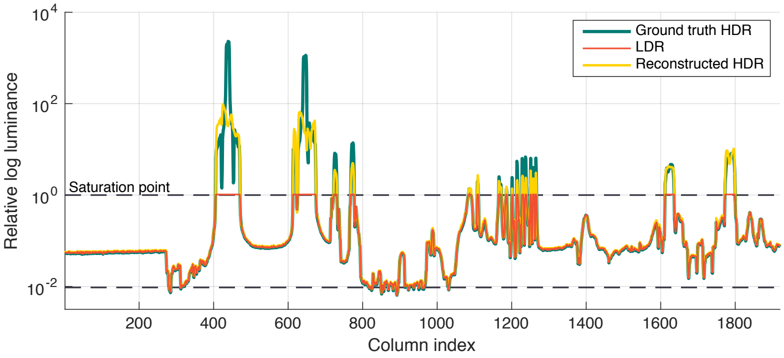


Figure 5.3: Luminances across one row of pixels from Figure 5.5. The row is indicated with the respective colors in the images in Figure 5.5. The horizontal lines correspond to the vertical lines in Figures 5.1.

different characteristics of this problem, where more local information needs to be reconstructed in order to undo the quantization.

Attempting within the same trained network to do solve all the three different problems of over-exposure, under-exposure, and quantization can be thought of as trying to do in-painting, denoising, and super-resolution in a single operation. It is unlikely that this would give satisfying results. The sub-problems of inferring HDR from a single-exposed 8-bit image have equally distinct differences as these three applications. Therefore, in order to allow for higher quality reconstruction of the most important pixels, it is a sensible strategy to treat over-exposure as a separate problem.

Reconstruction of over-exposed image regions is in many situations also a tractable problem, which is facilitated by the spatial arrangement of the saturated pixels. A common property of the most intense parts of a scene, e.g. specular highlights, light sources, sun, and moon, is a limited spatial extent. This can be seen in Figure 5.3, which shows one row of pixel values, as illustrated in Figure 5.5. The pixels with high luminance are often represented by sharp peaks, which only extend short distances in the image space. This makes it easier to approximate the clipped pixels from neighboring information. The figure also demonstrates the luminances of the row from the LDR image, as well as from the HDR image reconstructed using the method from Paper E.

In the LDR image in Figures 5.1 and 5.3, there are still pixels below the line that indicates where under-exposure starts. This is because a display-referred image can express a relatively large contrast. If the linear RAW image captured by a camera sensor would be noise-free, the dynamic range between the smallest and largest representable pixel value would be in the range 3.6 – 4.2 $\log_{10}$ units, for bit-depths between 12 – 14 bits. By compressing the dynamic range using a non-linear CRF, it can be fitted to the 8 bits provided in a conventional image. Thus, the 8-bit display-referred image can represent a captured dynamic range that is much larger than would be possible if it was scene-referred. However, as mentioned above, the dark pixels are usually deteriorated from noise and quantization, which makes the effective dynamic range much lower.

5.2 Deep learning for HDR imaging

Deep learning has shown a great success in a wide range of computer vision and image processing tasks, especially using *convolutional neural networks* (CNNs). A CNN learns the weights of a number of filter kernels in each layer of a deep neural network, where the layers represent different abstraction levels. By convolving the image with the learned filter weights, features can be extracted at different spatial locations using the same kernel. CNNs have shown unprecedented performance in applications such as image classification [134, 227], object detection [215], semantic segmentation [157], colorization [119, 277], style transfer [97], super-resolution [65], and many more.

The field of deep learning has seen a tremendous progress and gain in popularity over the last decade. Combined with the recent increase in HDR image data, e.g. due to a number of publicly available HDR video datasets [23, 40, 95, 136], deep learning for HDR imaging is now an interesting topic to be explored. There are only very few examples of deep learning for HDR imaging published before the year 2017, including estimation of reflectance maps from images [213]. In 2017, however, a number of publications appeared. These make use of CNNs for a variety of problems related to HDR imaging, demonstrating various degrees of improvement over previous work. For example, there are CNNs for HDR reconstruction from multiple exposures in separate images [106, 125, 266] and from single-shot, spatially varying, exposures [9]. Other techniques attempt to estimate outdoor [115] and indoor [96] illumination maps from conventional LDR images. Furthermore, there are examples of HDR image quality assessment [122], estimation of the camera response function from a single LDR image [152], and tone-mapping of HDR images [117].

The idea of deep learning HDR image reconstruction from a single exposure, was first introduced by Zhang and Lalonde [276], who used a CNN to predict HDR panoramas from LDR images for the purpose of IBL. However, the method is limited to low-resolution outdoor images, where the sun is assumed to be located at a certain azimuthal position. There are also more general methods, which were developed concurrently or after the work in Paper E was published. The methods by Endo et al. [85] and Lee et al. [151] predict a set of LDR images with different exposures, which subsequently are combined into an HDR image.

Marnerides et al. [176] proposed to use a multi-level CNN, which processes local and global information in separate branches. These methods are to some extent complementary to Paper **E**, as they consider the compound problem of transforming an LDR image to HDR. The method by Marnerides et al. can also in certain situations produce better approximation for large saturated areas, with less tiling artifacts due to the multi-level network. In comparison, the method in Paper **E** only attempts at recovering information in over-exposed image regions. The rest of the image is taken from the input LDR image, by applying an inverse CRF. As discussed in Section 5.1, pixel saturation is the most prominent problem in inferring HDR from LDR and also tractable to attempt to solve. By focusing on this, we are able to provide better dynamic range recovery as compared to the other methods. For example, the presented results of the concurrent methods have not demonstrated successful reconstruction of intense highlights and light sources, which are important in e.g. IBL and a number of post-processing applications. Moreover, we argue that, by focusing on saturated pixels, the quality level of reconstructed colors and details demonstrated in Paper **E** is not possible to achieve with other currently existing methods.

5.3 Deep learning reconstruction

5.3.1 CNN design

Paper **E** uses a CNN in an auto-encoder design, as illustrated in Figure 5.4. The encoder takes a display-referred LDR image as input and transforms it to a latent representation of 512 feature maps. The encoder uses the convolutional layers of the VGG16 network [227], where pooling operations down-sample the $W \times H$ pixels image to an encoded resolution of $W/32 \times H/32$ pixels. The decoder reconstructs the image from the encoded representation, using a number of consecutive convolutional and up-sampling layers. Since the architecture is a *fully convolutional network* (FCN), it is able to process any image that has horizontal and vertical resolutions that are multiples of 32.

While the encoder processes display-referred pixels, the decoder reconstructs scene-referred values in the log domain. However, at some point in the deeper layers of the auto-encoder the concepts of display-referred and scene-referred are lost, and the latent feature representation cannot be said to have a particular calibration. This allows for a connection directly between the different domains of the encoder and decoder.

In order to provide better details for the decoder's reconstruction, there are skip-connections that by-pass information from the encoder to decoder. These include a transformation, $D = \log\left(f^{-1}(E)\right)$, as demonstrated in Figure 5.4, which accounts for the different domains of the encoder and decoder. The

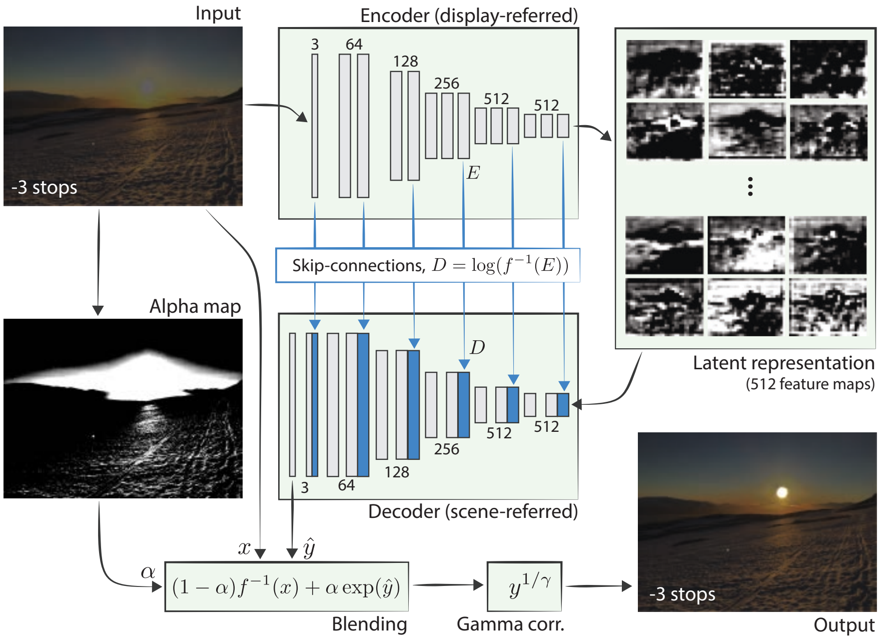


Figure 5.4: Overview of the single-exposure deep learning pipeline from Paper **E**. The encoder transforms the display-referred input to a latent feature representation. The decoder reconstructs a scene-referred HDR image in the log domain. The final image is computed from interpolating between input and CNN output using the blend map α. The inverse CRF, f^{-1}, transforms from display-referred to linear pixel values. The numbers for the CNN layers specify the depth of each layer used in the different levels.

transformation linearizes the encoder layer E using a static inverse CRF, f^{-1}, followed by transforming to the decoder layer D in the log domain. However, as mentioned above, the concept of calibration is lost somewhere in the deeper layers of the auto-encoder. Thus, while the domain transformation makes for a better starting point in the top-most layers, so that only the residual has to be reconstructed, it is probably not needed for the deepest layers. However, it is not clear how deep into the architecture the transformations are beneficial, so they are applied to all skip-connections.

In order to focus only on the saturated regions, as discussed in Section 5.1, the output from the decoder is blended with the input. To this end, the LDR image x is linearized using a static inverse CRF, $f^{-1}(x)$, while the reconstructed image $\hat{y}$ is given in linear values by transforming from the log domain, $\exp(\hat{y})$. The blend map, α, draws pixels from the reconstructed image around saturated areas, while retaining the input image for non-saturated pixels.

5.3.2 Training

The objective function that the CNN is trained to minimize, is split into two separate terms. One compares the log illumination of the CNN output, $\log(\hat{I})$, and the ground truth, $\log(I)$. The second term compares the log reflectance of reconstruction, $\log(\hat{R})$, and ground truth, $\log(R)$. The separation of the images into illumination and reflectance is performed using a Gaussian filtering of the log luminance, similar to how detail/base layer decomposition is done in tone-mapping. As described in Section 2.4.2, separate consideration of these attributes is motivated from a perceptual standpoint. For the purpose of learning to reconstruct HDR images, the illumination-reflectance loss can produce results that are visually more robust, with less visible artifacts. Moreover, it provides an option for prioritizing the different terms depending on the application, e.g. to provide better illumination approximation in IBL applications.

Given the image decomposition, the layers are combined to a scalar loss, $\mathcal{L}$, according to

$$\mathcal{L} = \lambda \sum \left| \alpha \left(\log(\hat{I}) - log(I) \right) \right|^2 + (1 - \lambda) \sum \left| \alpha \left(\log(\hat{R}) - \log(R) \right) \right|^2, \qquad (5.1)$$

where the scalar λ controls the relative importance of illumination and reflectance. The blend map, α, is used to limit the loss to only the areas around saturated pixels, as illustrated in Figure 5.4. Specific pixel indices are dropped for readability, but the summations average across all pixels of the respective layers. The weights of the CNN are optimized for minimal loss, from back-propagation using gradient descent with momentum, employing the ADAM (adaptive moment estimation) optimizer [130].

The CNN is trained on a gathered set of ≈1.1K HDR images and ≈2.6K HDR video frames. The videos are sampled by selecting every 10th frame from a total of 67 HDR video clips. In order to simulate LDR images for training input, and to augment the database with more samples, the concept of a virtual camera is employed. This simulates a number of LDR images from each input HDR image scene in a stochastic procedure, where each of the following parameters is randomly sampled:

1. Position and size of a cropped area.
2. Horizontal image flipping.
3. Exposure setting, selected so that 5-15% of the total number of pixels are saturated and clipped.

4. Two settings of a parametric camera response function.
5. Standard deviation of added image noise.
6. Color hue and saturation.

In total, the HDR images are augmented to create a training set of ≈125K, 320 × 320 pixels, LDR-HDR image pairs. Nevertheless, the amount of training data may still be a limiting factor. In order to provide a better starting point for the optimization, the network is pre-trained on a larger set of simulated HDR images. The images are selected from the Places database [279], by only choosing images that are not saturated. An image is considered not saturated if less than a very small fraction of the image uses the highest pixel value. In total, a subset of around 600K images satisfy the criteria. These are subsequently linearized from assuming a static CRF, followed by processing with the same virtual camera procedure as above. However, only one LDR-HDR image pair is created from each image and the exposure is selected to saturate 10-30% of the image pixels.

By performing a two-stage training procedure, it is possible to achieve a significant increase in reconstruction quality as compared to only training on the native HDR data. The first stage uses the simulated HDR dataset to optimize over a very wide variety of images and pixel saturation situations. However, the simulated images are limited in dynamic range, which means that it is not possible to reconstruct intense highlights and light sources after this training phase. In the second stage, the optimization is fine-tuned on the native HDR data, which allows for training the architecture to recreate a significantly higher dynamic range.

5.3.3 Weight initialization

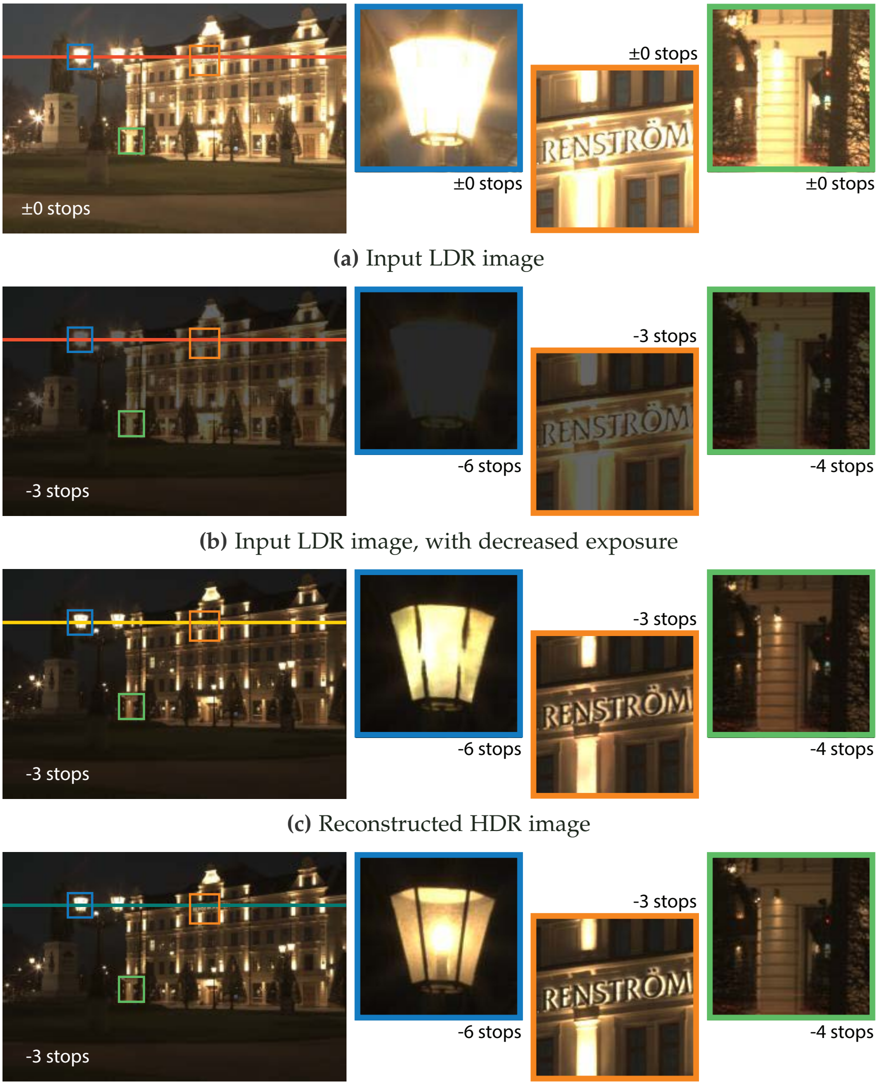
A deep neural network is a complex model, specified from a vast amount of trainable parameters. Since the objective function, Equation 5.1, is non-convex over the parameter space, finding a global minimum is in practice impossible. However, this is in general not a problem for optimization of neural networks, since most of the local minima tend to have costs close to the global minimum [101]. Still, there might be a significant difference between minima at different locations in the parameter space, so that a difference in starting point for the optimization can affect the final result. This is especially true if there is a limitation in the amount of training data available for optimization. A common strategy for selecting starting point is to use pre-trained weights, which may have been optimized for a completely different task. Since the basic feature extraction that is performed by a CNN often is similar for different tasks, and application-specific processing mainly happens in deeper layers, this can facilitate finding a good local minimum.

For the training of the CNN in Paper E, we make a number of design and training choices that are intended to improve on the starting point for the optimization. As described in Section 5.3.2, weights are first optimized over a simulated HDR dataset. In order to provide a good starting point for this pre-training, the encoder convolutional layers are initialized from VGG16 weights pre-trained for classification on the Places database [279]. The decoder up-sampling filters are initialized to perform bilinear interpolation. Moreover, the skip-connected layers provide a better starting point by including the domain transformation shown in Figure 5.4. These layers are concatenated with the decoder layers, and then combined by learning how to fuse the information. The fusion is initialized to perform an addition of the layers, similar to how residual networks do. The remaining weights, for convolution within the latent representation and for the final layers of the decoder, are specified using the Xavier initialization method [99].

5.3.4 Results

The single-exposure HDR reconstruction CNN can provide convincing predictions of over-exposed pixels in a standard LDR image, as shown in the example in Figure 5.5. In the input LDR image, Figure 5.5a, the exposure is set to reproduce details of the darker foreground. Figure 5.5b shows that when decreasing the exposure by 3 stops in post-processing, of the already captured image, it becomes clear that there is no information available in the brighter parts of the image. A single exposure is incapable of registering both details in darker regions of the scene and the high-intensity lighting. In fact, many different exposures are required to capture the dynamic range of the scene, as illustrated in Figure 1.1. The reconstruction using the trained CNN is displayed in Figure 5.5c, where the enlarged regions show how colors, details, and high intensities can be inferred with high quality. The limitations are also evident. While smaller spatial neighborhoods can be reproduced to be visually indistinguishable from HDR images without ground truth reference, the larger area of the lamp lack in reconstructed details. However, the reconstruction still offers a huge improvement as compared to the input image, and would allow for high-quality results in many HDR applications, including IBL.

While large areas with saturation in all color channels are difficult to reconstruct, it is a much easier problem if there is some information left in one of the channels. This is demonstrated in Figure 5.6, where the input image shows well-exposed buildings, but where a large portion of the sky is over-exposed. In the input image with decreased exposure, it is evident that there are many pixels with lost information due to sensor saturation. In Figure 5.6c, the over-exposed pixels have been color coded to show which color channels are saturated. The saturated channels have a value of 0 and the others have maximum value.



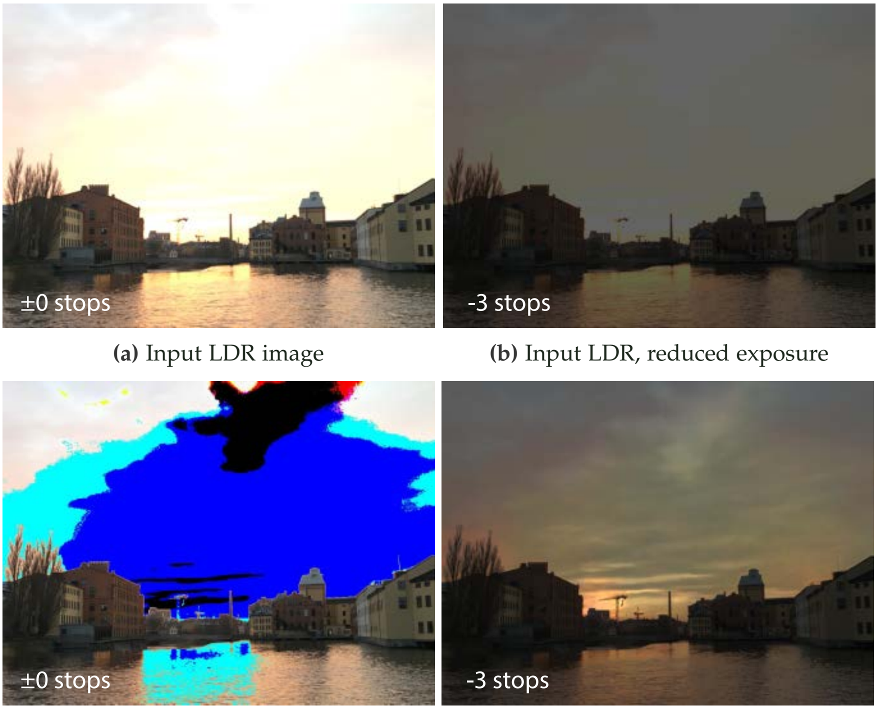
(a) Input LDR image

(b) Input LDR image, with decreased exposure

(c) Reconstructed HDR image

(d) Ground truth HDR image

Figure 5.5: Reconstruction of a high resolution LDR image (1920×1280 pixels). The exposures of the images and enlarged regions have been reduced according to the specified numbers of stops. The HDR images are displayed after applying gamma correction. The colored lines correspond to the scanline plots in Figure 5.3.



(a) Input LDR image

(b) Input LDR, reduced exposure

(c) Saturated pixels

(d) Reconstructed HDR image

Figure 5.6: Reconstruction from an iPhone 6S camera. The input image experiences a total of 34.5% saturated pixels, as seen by the large over-exposed area in **(a-b)**. However, many pixels still have information in the blue color channel, as visualized in **(c)**, where only the black areas show saturation in all color channels. The information allows for successful reconstruction, **(d)**.

This means that black corresponds to saturation in all channels. In cyan areas, only the red channel has saturated, and in blue areas both red and green channels are saturated. Although a total of 34.5% of the pixels are saturated, the saturation is distributed according to 53.5%, 42.7%, and 7.1% saturation in the red, green, and blue channels, respectively. With the information that is left in the blue channel, reconstruction is much simplified, as can be seen in the reconstruction in Figure 5.6d. The example also demonstrates successful generalization, where the image is captured by a smartphone camera with unknown CRF and post-processing applied.

5.3.5 Compression artifacts

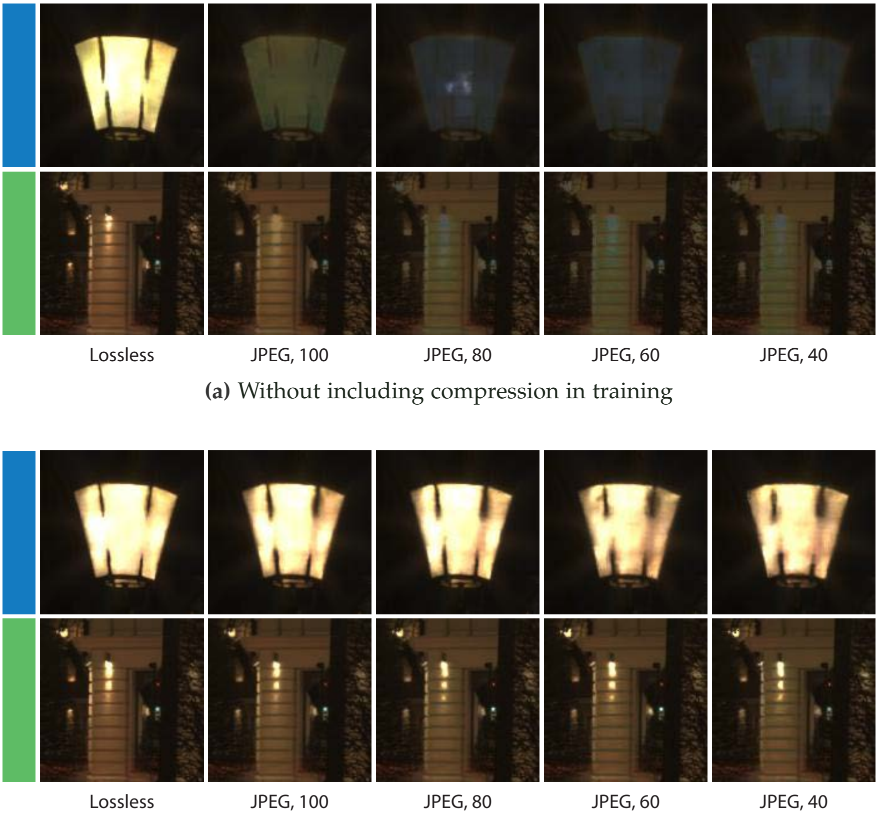
The trained model in Paper E is limited in how much lossy image compression, and the artifacts thereof, can be tolerated. The reconstruction is highly sensitive to differences in the information around saturated image areas, and even a visually imperceptible degradation caused by compression artifacts can completely break the reconstruction.

In order to account for compression artifacts in the reconstruction, we have complemented the training from Paper E with a new dimension of data augmentation. In addition to the list of parameters of the virtual camera in Section 5.3.2, the final image is stored with JPEG compression, choosing a random quality level in the range 30-100. Training with the updated augmentation is done only in the second phase, on the native HDR dataset, initializing the CNN with the same pre-training parameters as in Paper E.

In Figure 5.7, a comparison is made between reconstruction with and without including lossy compression in the training, for a number of different JPEG quality levels. In this example, when compression has not been considered in training, the reconstruction is heavily affected already at the maximum JPEG quality level. With lowered quality level, the reconstruction soon shows very little improvement over the input LDR image. Including compression artifacts in the training leads to a significant improvement in reconstruction quality on JPEG images. However, inspecting the reconstructions on lossless LDR images, the original weights can better reproduce image details. One possible explanation for this is that the CNN trained with compression learns to perform a selective low-pass filtering of the images. Since the ground truth HDR images in the training do not contain compression artifacts, the network attempts to suppress the blocking artifacts produced by the JPEG encoder.

5.3.6 Adversarial training

As mentioned in relation to the example in Figure 5.5, there is a limit to how large areas with all channels saturated that can be convincingly reconstructed. This limit is highly dependent on image content, and therefore hard to quantify. The difficulty in reconstructing content in large areas is an inherent problem with image-to-image training using a pixel-wise loss, as the L2 loss in Equation 5.1. Even if there was an infinite number of different HDR images for training, and a successful optimization could be made across those, the result would lack in details. This is due to the fact that the reconstruction will be optimal in an L2 sense, comparing to all the provided possible solutions. Consequently, the best reconstruction is the average across all these solutions. If it was possible to select, for each image, only a single solution out of all the



(a) Without including compression in training

(b) With JPEG compressed training images

Figure 5.7: Reconstruction of JPEG compressed LDR images at different quality levels, using the image from Figure 5.5. The color codes refer to the marked areas in Figure 5.5. The lamp and facade are displayed at -6 and -3 stops, respectively, followed by gamma correction.

possibilities, there would be a higher loss over the database, despite providing more convincing representations of true scenes.

The concept of adversarial training, using *generative adversarial networks* (GANs) [100], can be thought of as forcing the solution of a neural network towards one particular mode, thus alleviating the averaging problems with a direct loss. This is achieved by having one generative model that attempts at capturing samples from a certain data distribution, and one discriminative model that estimates the probability of the sample coming from the training data as opposed to being generated. Both models are trained simultaneously. The generator attempts to fool the discriminator, and the latter tries to separate generated samples from true training data. This training strategy has been applied to CNN generators, using *deep convolutional GANs* (DCGANs) [206], which map a vector of uniformly distributed noise, through a set of convolutional up-sampling layers, to a natural image output.

In order to apply DCGANs in a supervised training setting, one possible solution is to formulate a combined loss, containing one pixel-wise term and one adversarial term. While the pixel-wise loss assures that the image output complies with a ground truth, the adversarial term promotes solutions that faithfully capture the image statistics. In practice, this means that more sharp features and details, which better convey a convincing solution, can be reproduced. The strategy has shown promising results e.g. for the purpose of inpainting [196], which is similar to our problem.

To confirm that adversarial training also can aid in reconstructing larger saturated image regions, we modify the *context encoder* (CE) used by Pathak et al. for inpainting [196]. This uses an auto-encoder generator network, where the latent representation is stored in a 1D fully connected layer with 4000 numbers. The generator takes a display-referred image as input and predicts an inpainted image in the same domain. We complement the network with skip-connections between encoder and decoder, so that fine details can bypass the deeper layers. Also, the loss is evaluated over the complete image, instead of only in a rectangular region with missing information.

We follow the recommendations on how to construct and train a DCGAN that is reasonably stable to train [206], with input/output layers specifying pixel values in the range $[-1,1]$, batch normalizations, and leaky ReLu. Since the range of the output is limited, we scale the intensity of the input image by a factor 1/3. This enables learning of highlights that are at most 3 times brighter than the input, in display-referred values. For training, we use the same subset of Places images that were used for pre-training in Section 5.3.2. The images are captured by the virtual camera in Section 5.3.2, where the exposure is set so that 20-40% of the pixels are saturated. 2 captures are made in each of the images, for a total of 1.2M training images.

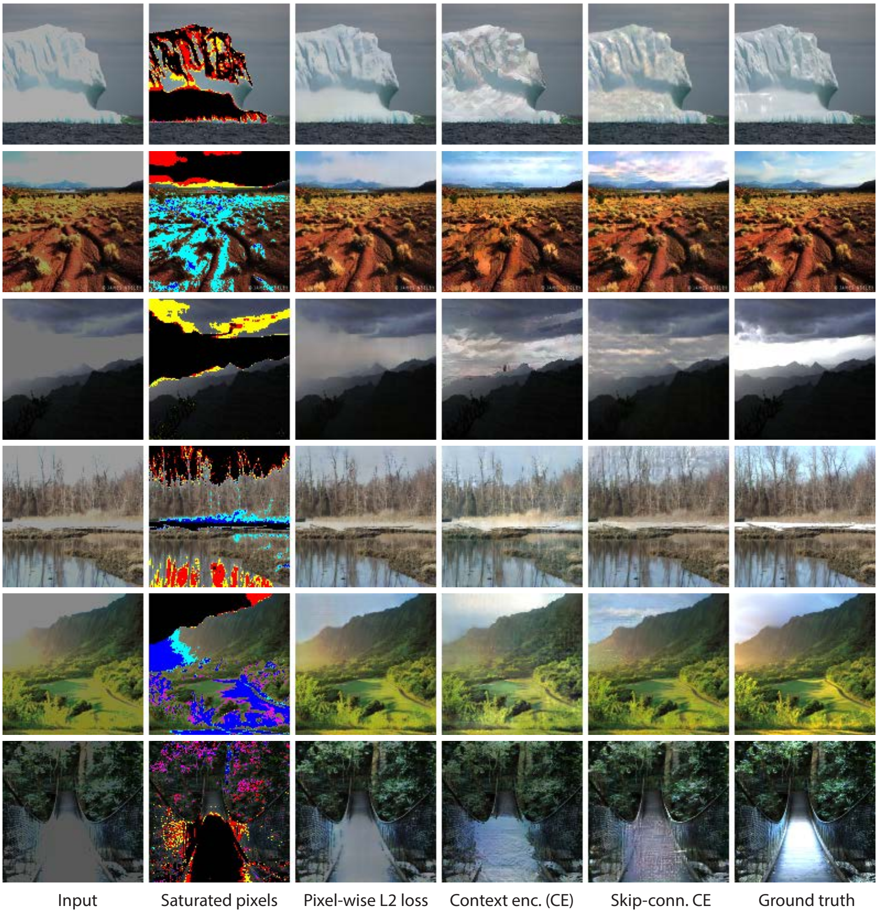


Figure 5.8: Adversarial highlight hallucination on test images from the Places database. The test images have been clipped such that 30% of the pixel information is lost in the highlights, as can be seen in the second column, which is visualized in the same way as in Figure 5.6. Compared to using only a pixel-wise L2 loss, the context encoders (CE) can hallucinate visually plausible image features (e.g. specular reflections on iceberg and clouds on sky). Skip-connections make the CE better preserve details in areas around highlights (see for example mountains and plants in the second row). In some cases the hallucinated structures from adversarial training can be perceived as artifacts (bottom row). Images are shown at ≈3 stops exposure reduction.

Some results of the adversarial highlight inpainting are shown in Figure 5.8. The figure also includes examples of training without the adversarial term and with the original inpainting context encoder, without skip-connections. The examples clearly show that adversarial training can help in hallucinating new and plausible information, such as the specular highlights of the iceberg and the clouds on the skies. This is not possible by using only the pixel-loss, which generates blurry features and lack in detail. Finally, comparing the skip-connection network to the original context encoder, it can much better preserve details when only one or two color channels have saturated, since the available information can be passed to the decoder without having to be compressed within the latent feature vector.

Although visually convincing results can be generated with the adversarial training methodology, it is also highly unpredictable. Results can be widely different from one training iteration to the next and in many cases predicted structures can be perceived as artifacts, as shown in the bottom example in Figure 5.8. Furthermore, the fully connected latent representation of the auto-encoder makes it limited to the image resolution used in the training (128x128 pixels in our case). Finally, the predictions are restricted to a maximum output intensity, which is a significant limitation for HDR images.

When it comes to image resolution, there are examples of fully convolutional networks that incorporate adversarial training, and which can predict high resolution images in high quality. One example is the deep ResNet for single image super-resolution by Ledig et al. [147]. However, the problem of super-resolution is significantly different, and more local in nature, as compared to inpainting large areas in the spatial domain. There are also some promising recent methods for inpainting, which utilize FCNs together with adversarial training [120, 153, 275]. These attempt to include a more high-level understanding of the images in order to create features that are semantically meaningful. In general, there is a lot of research interest around GANs, with new training strategies appearing frequently. Hence, it is expected that it is only a matter of time before robust single-exposure HDR image reconstruction can be performed at high resolution with aid of adversarial training. For the work in Paper E, we opted not to include GANs, to allow for a more robust and less limited reconstruction.

5.4 Summary

The work in Paper E makes it possible to reconstruct HDR images from single-exposure LDR images in a wide range of situations. Compared to previous methods, the results show unprecedented quality in terms of details, colors and intensities of the reconstructed saturated image regions. From the subjective experiment in Paper E, performed on an HDR display, it is also confirmed that the CNN reconstruction provides convincing HDR images. In most cases, the predictions are comparable to the ground truth HDR images in terms of perceived naturalness.

The presented method excels at recovering intensities of small highlights, such as specular reflections and street lights, which would require many exposures with classical capturing methods. For the purpose of IBL, this enables renderings that are very close to what the ground truth HDR images would give. This has not been possible previously, where methods for inverse tone-mapping only can attempt to boost saturated image regions in order to provide renderings that are visually more appealing, but not necessarily true to nature. Also, in situations where dynamic HDR panoramas are required for IBL, the CNN HDR reconstruction can potentially be useful in order to increase the dynamic range of already captured HDR videos. Here, the reconstruction can help in recovering high intensity highlights that are outside the range of what can be captured in a HDR video camera.

The HDR reconstruction CNN is made available online[1], together with trained weights, so that inference can be made using any LDR images. Additional weights, which have been trained with compression artifacts included (Section 5.3.5), are also provided. Finally, the code has also been complemented with training script and virtual camera code, so that the model can be trained with different data, and possibly tweaked for improved results.

5.4.1 Limitations and future work

While the method in Paper E can predict very high intensities in smaller saturated regions and highlights, there is still some under-estimation of the very brightest pixels. This can, for example, be seen in Figure 5.3, where the brightest reconstructed pixels are around 100 times more intense than in the input LDR image. However, in the ground truth HDR image there are pixels with more than 1,000 times larger luminance. Also, as demonstrated in Figure 1.1, the most intense pixels should be even larger, since there are some saturated pixels in the shortest exposure. This HDR image has been captured with 7 different exposures, which is more than for many of the HDR images in the training dataset. For example, many of the images are taken from HDR videos, which often can be more limited in dynamic range as compared to static images. Thus, there is an inherent difficulty in learning how to reconstruct the extreme pixels, as the training data also experience saturation.

One of the most central aspects of successful learning is the training data. While HDR images are starting to become available in larger quantities, the number

[1] https://github.com/gabrieleilertsen/hdrcnn

of images is still not comparable to the very large image databases, e.g. Places and Imagenet, which are used for learning other imaging tasks. There are other datasets that potentially can be used in order to improve performance, such as from Google's HDR+ project [112], SJTU HDR Video Sequences [228], the RAISE [59] and FiveK [45] datasets of RAW images, etc. All these sources provide images at an increased bit-depth. However, the images show relatively limited extensions in dynamic range and/or are saturated in high-intensity image regions. The latter issue can potentially make the under-estimation of extreme intensities more pronounced. One possible option for incorporating the data in the training would be to use it for pre-training of the network, or by attempting to only select images that do not include saturated regions.

Another natural extension of Paper E is to consider the reconstruction of video sequences. This would require investigating how to ensure temporal coherence in the reconstruction. Moreover, the added dimension makes it interesting to explore how to perform the reconstruction of a frame given predictions of the previous frames, placing a conditioning in the temporal domain.

In the current reconstruction pipeline, there is no mechanism for quality control of the output. In certain situations, e.g. when large areas are saturated, the reconstructed pixels can experience artifacts. In order to improve robustness, it would be of interest to ensure that the reconstruction result makes sense, so that it cannot be of visually lower quality than the input image.

Finally, as discussed in Section 5.3.6, a very interesting avenue for future work is to employ adversarial training, in order to be able to hallucinate image content in larger saturated regions.

Chapter 6

Conclusions

In this thesis and the included papers, we have presented a set of contributions at different steps of the HDR imaging pipeline. Starting from a high-level introduction to the concept of high dynamic range imaging and video, the background follows up with an overview of research and production related to HDR imaging. The field of HDR imaging has grown rapidly over the last two decades (see Figure 6.1), and the background attempts to give a broad description in relation to the different components of the HDR imaging pipeline. Following, in the paper specific chapters the thesis work is then discussed. This is done in an attempt to not repeat the details that can be found in the individual papers, but rather to provide a higher level discussion around the motivation, contributions, implications, limitations, and possible directions for future work. To this end, a set of new examples and results help in mediating this information and complement the thesis papers with some new insights.

This chapter concludes the thesis with a final summary of the contributions of the thesis papers in Section 6.1. We also summarize some of the new insights and results that were provided throughout the first part of the thesis. Finally, Section 6.2 reflects over the current and future situation in HDR imaging

6.1 Contributions

Contributions have been presented within three of the software components of the HDR imaging pipeline, as illustrated in Figure 1.4. For each of the components, the thesis also contributes with a number of complementing discussions, details, and results related to the papers.

6.1.1 Tone-mapping

The first contribution to tone-mapping is the qualitative and subjective evaluation of video TMOs presented in Paper B. This demonstrates that, at the time of the work, there were a number of challenges that needed to be addressed in order to allow for robust tone-mapping of content captured with HDR camera systems. We believe that this evaluation has had a distinct impact on the subsequent research on video tone-mapping, where it often is used in order to motivate the need for developing new HDR video processing algorithms.

The second contribution is the novel tone-mapping operator in Paper C, which is specifically tailored considering the challenges in tone-mapping of natural HDR video sequences. The method produces high levels of detail and local contrast, without revealing spatial and temporal artifacts. The dynamic range is compressed by minimizing the distortions of contrasts in the mapping, where special considerations are made in order not to reveal visible noise. All the computations run in real-time on high-resolution videos, by implementing the method for hardware acceleration.

The third contribution is the literature review and quantitative evaluation in Paper A. The work serves both as an up-to-date comprehensive reference and categorization of video TMOs, and as a comparative assessment of the latest development in tone-mapping for HDR video.

In addition to the contributions of the individual tone-mapping papers, the thesis provides a number of additional insights:

1. First, by discussing the papers in combination, it is evident how they follow a natural chain of development. The work starts with an evaluation of existing techniques for tone-mapping of HDR video, then uses the findings in order to develop a new and improved video TMO. Finally, the improvements of the new method are confirmed in the literature study and quantitative evaluation.
2. We show additional details of the technique used for calibration of video TMOs. This is an important topic, as it can to a large extent affect the outcome of an evaluation. The technique enables using interpolation between a sparse set of videos, by sampling in a linearized parameter space. Then, optimization is performed by means of a conjugate gradient search in the parameter space, using perceptual judgments as the objective function.
3. In Paper B it is noted that there seems to be a correlation between the qualitative ratings and the subjective preferences measured from the pairwise comparison experiment. To confirm this, we compile the results from the two experiments and illustrate them side-by-side for each evaluated TMO. This demonstrates an evident correlation between the experiments, where a higher artifact and attribute rating predicts a lower end subjective preference.

4. Finally, we also show a clear correlation between the quantitatively measured temporal incoherence in Paper **A** and the qualitative incoherence ratings from Paper **B**. This observation confirms that the quantitative approach indeed provides a good measure of temporal artifacts.

6.1.2 Distribution

The work in Paper **D** shows two contributions to the area of HDR video distribution. First, a set of methods involved in preparing HDR data for encoding are compared. The performance is evaluated in terms of two objective metrics, computed over a set of 33 different HDR video sequences. The results demonstrate that the perceptually based luminance and color encodings allow for a significant increase in quality for a given bit-rate. Second, based on the results of the comparisons, the Luma HDRv codec and API is built by using the best performing techniques as the default settings. The software is made available on open source terms, and to our knowledge it was the first freely available HDR video codec.

In the thesis, we complement the paper with additional results and insights:

1. We perform a new evaluation, which compares the Lu'v' and YC_bC_r color spaces at two different combinations of luma/chroma bit-depth. The results indicate that YC_bC_r benefits from 10/10 bits luma/chroma instead of the combination 11/8 that was used in Paper **D**. The results also show that, as expected, the 11/8 combination is the better choice for Lu'v'. Although the performance increased for YC_bC_r, the Lu'v' encoding still provides a better rate-distortion trade-off.
2. Also, due to the activity around HDR TV, we provide a discussion on the recent developments in HDR video codecs. While open source alternatives for HDR video encoding are starting to become available, we discuss how Luma HDRv still can provide a useful tool. It can provide a versatile HDR encoding abstraction layer, which can be used with different encoders under the hood.

6.1.3 Reconstruction

With the work in Paper **E**, we contribute by providing a novel solution to the difficult problem of inferring HDR pixels in a single-exposed LDR image. By utilizing deep learning strategies, we are able to demonstrate results that far exceed what was previously possible. While we are restricted to reconstruction in saturated image regions, the results produced by the trained CNN allow for using LDR images in a larger number of HDR applications than was previously possible.

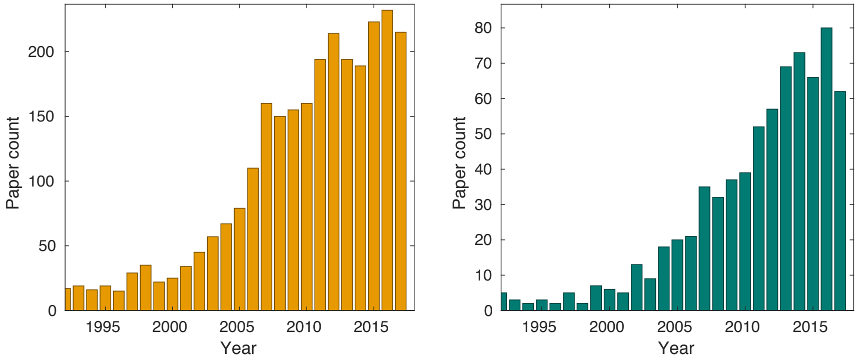
In order to complement the paper, the thesis provides some new results and discussions:

1. A thorough discussion and analysis of the single-exposure HDR reconstruction problem is provided, which is facilitated by extracting statistics from a large HDR image dataset. The analysis motivates focusing on the saturated image regions, as there is a high gain in dynamic range for few successfully reconstructed pixels.
2. An additional optimization is performed, which includes JPEG compressed training images. It shows substantial improvements in reconstruction quality on images that are degraded by compression artifacts. The trained parameters are made available online, together with code for running inference and for performing optimization of the CNN.
3. Finally, GANs are discussed and shown to generate promising results. As opposed to using only a pixel-wise loss, complementing with an adversarial loss allows for reconstructing sharp hallucinated image features in large saturated image areas. However, there are many limitations to be overcome with adversarial training, but we believe that it is only a matter of time before this can be done.

6.2 Outlook

The high activity around HDR imaging within the research community can be seen from the increasing number of publications each year. Figure 6.1 shows the yearly count of publications over the last 25 years, according to Google Scholar, which contain the specified search phrases in the title. There might be publications that use “high dynamic range” or “tone mapping” in the title, but do not consider images or video. It is also most certainly the other way around, where publications treat HDR imaging but do not specify this in the title. However, the plots give an indication of the increasing interest in HDR imaging. With this past and ongoing research, we will most likely see a rapid increase in the use of HDR images and video in the near future.

For display of HDR images, the recent introduction and popularity of HDR TVs mean that the format has truly been established on the consumer market. This trend is likely to continue, and not only for display on HDR capable TVs. For a conventional display device, there are also great benefits in supporting HDR material. The format allows for flexibility when preparing an image or video for the display, taking into account the certain display parameters and environmental factors (Equation 1.1). For example, higher contrast and brightness is required in a bright environment in order to match the viewing experience in a dark room as close as possible. Thus, performing tone-mapping



(a) "high dynamic range"

(b) "tone mapping" or "tone reproduction"

Figure 6.1: The number of publications per year that contain the different terms in the title, according to Google Scholar. Citations and patents are not included in the searches.

on location can make for minimal differences in the viewed content in different situations. It also allows for personal adjustments, maximizing the subjective quality on a per-unit level. While LDR images also can be tweaked to some extent in order to satisfy the aforementioned goals, the HDR format allows for significantly more extensive processing.

On the capturing side, in the future there will probably be more options that utilize multiple sensors, also for consumer level products. One possible direction of development is to combine different types of sensors. For example, a larger conventional sensor can capture the majority of details, while a log sensor registers a low-resolution HDR image. The fusion of the different types of sensory data will likely use a machine learning approach, in order to better handle areas with missing information. In general, there will most likely also be a continuing increase in learning based post-processing methods for improving image quality. This is especially expected for mobile devices, where there are physical constraints on optics and sensors due to the limited size. In order to enable extensive image reconstruction algorithms, custom chips for image processing could be put on-board the device. This has already been realized with Google's Pixel Visual Core chip in the Pixel 2 smartphone, but will likely be common in the future. Given such development, it could, for example, be possible to reconstruct HDR images directly in the device using neural networks, such as the method presented in Paper E.

Bibliography

[1] A. Adams, N. Gelfand, J. Dolson, and M. Levoy. Gaussian KD-trees for fast high-dimensional filtering. *ACM Transactions on Graphics*, 28(3): 21:1–21:12, 2009. [page 37]

[2] M. Aggarwal and N. Ahuja. Split aperture imaging for high dynamic range. In *Proceedings of IEEE International Conference on Computer Vision (ICCV 2001)*, volume 2, pages 10–17, 2001. [page 23]

[3] M. Aggarwal and N. Ahuja. Split aperture imaging for high dynamic range. *International Journal of Computer Vision*, 58(1):7–17, 2004. [page 23]

[4] C. Aguerrebere, A. Almansa, Y. Gousseau, J. Delon, and P. Musé. Single shot high dynamic range imaging using piecewise linear estimators. In *Proceedings of IEEE International Conference on Computational Photography (ICCP 2014)*, pages 1–10, 2014. [page 24]

[5] W. Ahn and J.-S. Kim. Flat-region detection and false contour removal in the digital TV display. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME 2005)*, pages 1338–1341, 2005. [page 25]

[6] A. O. Akyüz and E. Reinhard. Color appearance in high-dynamic-range imaging. *Journal of Electronic Imaging*, 15(3):033001, 2006. [page 36]

[7] A. O. Akyüz and E. Reinhard. Noise reduction in high dynamic range imaging. *Journal of Visual Communication and Image Representation*, 18(5): 366–376, 2007, Special issue on High Dynamic Range Imaging. [page 40]

[8] A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bulthoff. Do HDR displays support LDR content? A psychophysical evaluation. *ACM Transactions on Graphics*, 26(3), 2007. [pages 26 and 41]

[9] V. G. An and C. Lee. Single-shot high dynamic range imaging via deep convolutional neural network. In *Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2017)*, pages 1768–1772, 2017. [page 86]

[10] A. Artusi, R. K. Mantiuk, T. Richter, P. Hanhart, P. Korshunov, M. Agostinelli, A. Ten, and T. Ebrahimi. Overview and evaluation of the JPEG XT HDR image compression standard. *Journal of Real-Time Image Processing*, pages 1–16, 2015. [page 32]

[11] A. Artusi, T. Pouli, F. Banterle, and A. O. Akyüz. Automatic saturation correction for dynamic range management algorithms. *Signal Processing: Image Communication*, 63:100–112, 2018. [page 39]

[12] M. Ashikhmin. A tone mapping algorithm for high contrast images. In *Proceedings of Eurographics Workshop on Rendering (EGWR 2002)*, pages 145–156, 2002. [page 37]

[13] M. Ashikhmin and J. Goyal. A reality check for tone mapping operators. *ACM Transactions on Applied Perception*, 3(4):399–411, 2006. [page 41]

[14] M. Aubry, S. Paris, S. W. Hasinoff, J. Kautz, and F. Durand. Fast local Laplacian filters: Theory and applications. *ACM Transactions on Graphics*, 33(5):167:1–167:14, 2014. [pages 35 and 37]

[15] V. Aurich and J. Weule. Non-linear gaussian filters performing edge preserving diffusion. In *Mustererkennung 1995*, pages 538–545, 1995. [pages 37 and 58]

[16] Axis Communications white paper. WDR solutions for forensic value. `https://www.axis.com/files/whitepaper/wp_wide_dynamic_range_70788_en_1710_lo.pdf`, 2017, Accessed: 2018-04-15. [page 21]

[17] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel. Dynamic range independent image quality assessment. *ACM Transactions on Graphics*, 27 (3):69:1–69:10, 2008. [page 41]

[18] T. O. Aydin, R. Mantiuk, and H.-P. Seidel. Extending quality metrics to full luminance range images. In *Proceedings of SPIE, Human Vision and Electronic Imaging XIII*, volume 6806, 2008. [page 73]

[19] T. O. Aydin, M. Čadík, K. Myszkowski, and H.-P. Seidel. Video quality assessment for computer graphics applications. *ACM Transactions on Graphics*, 29(6):161:1–161:12, 2010. [page 41]

[20] T. O. Aydin, N. Stefanoski, S. Croci, M. Gross, and A. Smolic. Temporally coherent local tone mapping of HDR video. *ACM Transactions on Graphics*, 33(6):1–13, 2014. [pages 37, 40, and 64]

[21] M. Azimi, R. Boitard, B. Oztas, S. Ploumis, H. R. Tohidypour, M. T. Pourazad, and P. Nasiopoulos. Compression efficiency of HDR/LDR content. In *Proceedings of Seventh International Workshop on Quality of Multimedia Experience (QoMEX 2015)*, pages 1–6, 2015. [page 32]

[22] S. Bae, S. Paris, and F. Durand. Two-scale tone management for photographic look. *ACM Transactions on Graphics*, 25(3):637–645, 2006. [page 37]

[23] A. Banitalebi-Dehkordi, M. Azimi, M. T. Pourazad, and P. Nasiopoulos. Compression of high dynamic range video using the HEVC and H.264/AVC standards. In *Proceedings of International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine 2014)*, pages 8–12, 2014. [page 86]

[24] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers. Inverse tone mapping. In *Proceedings of International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia (GRAPHITE 2006)*, pages 349–356, 2006. [page 26]

[25] F. Banterle, P. Ledda, K. Debattista, A. Chalmers, and M. Bloj. A framework for inverse tone mapping. *The Visual Computer*, 23(7):467–478, 2007. [page 26]

[26] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers. Expanding low dynamic range videos for high dynamic range applications. In *Proceedings of Spring Conference on Computer Graphics (SCCG 2008)*, pages 33–41, 2008. [page 26]

[27] F. Banterle, P. Ledda, K. Debattista, M. Bloj, A. Artusi, and A. Chalmers. A psychophysical evaluation of inverse tone mapping techniques. *Computer Graphics Forum*, 28(1):13–25, 2009. [page 26]

[28] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers. *Advanced high dynamic range imaging*. CRC press, 2017. [page 15]

[29] H. Barrow and J. Tenenbaum. Recovering intrinsic scene characteristics from images. *Computer vision systems*, pages 3–26, 1978. [page 37]

[30] P. G. J. Barten. Formula for the contrast sensitivity of the human eye. In *Proceedings of SPIE, Image Quality and System Performance*, volume 5294, 2003. [page 31]

[31] M. Bätz, T. Richter, J.-U. Garbas, A. Papst, J. Seiler, and A. Kaup. High dynamic range video reconstruction from a stereo camera setup. *Signal Processing: Image Communication*, 29(2):191–202, 2014, Special Issue on Advances in High Dynamic Range Video Research. [page 23]

[32] E. P. Bennett and L. McMillan. Video enhancement using per-pixel virtual exposures. *ACM Transactions on Graphics*, 24(3):845–852, 2005. [pages 40, 48, and 50]

[33] A. Benoit, D. Alleysson, J. Herault, and P. L. Callet. Spatio-temporal tone mapping operator based on a retina model. In *Proceedings of Computational*

Color Imaging Workshop (CCIW 2009), pages 12–22, 2009. [pages 34, 40, 48, and 50]

[34] R. S. Berns. Methods for characterizing CRT displays. *Displays*, 16(4): 173–182, 1996. [page 7]

[35] S. Bhagavathy, J. Llach, and J. f. Zhai. Multi-scale probabilistic dithering for suppressing banding artifacts in digital images. In *Proceedings of IEEE International Conference on Image Processing (ICIP 2007)*, volume 4, pages 397– 400, 2007. [page 25]

[36] C. Bist, R. Cozot, G. Madec, and X. Ducloux. Tone expansion using lighting style aesthetics. *Computers & Graphics*, 62:77–86, 2017. [page 26]

[37] R. Bogart, F. Kainz, and D. Hess. OpenEXR image file format. *ACM SIGGRAPH 2003, Sketches & Applications*, 2003. [pages 28 and 77]

[38] R. Boitard, K. Bouatouch, R. Cozot, D. Thoreau, and A. Gruson. Temporal coherency for video tone mapping. In *Proceedings of SPIE, Applications of Digital Image Processing XXXV*, volume 8499, 2012. [pages 35, 40, and 50]

[39] R. Boitard, R. Cozot, D. Thoreau, and K. Bouatouch. Zonal brightness coherency for video tone mapping. *Signal Processing: Image Communication*, 29(2):229–246, 2014. [pages 40 and 64]

[40] R. Boitard, R. Cozot, D. Thoreau, and K. Bouatouch. Survey of temporal brightness artifacts in video tone mapping. In *Proceedings of Second International Conference and SME Workshop on HDR imaging (HDRi 2014)*, 2014. [page 86]

[41] R. Boitard, R. K. Mantiuk, and T. Pouli. Evaluation of color encodings for high dynamic range pixels. In *Proceedings of SPIE, Human Vision and Electronic Imaging XX*, volume 9394, 2015. [pages 31, 72, 74, and 75]

[42] N. Bonneel, J. Tompkin, K. Sunkavalli, D. Sun, S. Paris, and H. Pfister. Blind video temporal consistency. *ACM Transactions on Graphics*, 34(6): 196:1–196:9, 2015. [page 40]

[43] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, volume 2, pages 2:60–65, 2005. [page 25]

[44] A. Buades, B. Coll, and J.-M. Morel. Nonlocal image and movie denoising. *International Journal of Computer Vision*, 76(2):123–139, 2008. [pages 25 and 40]

[45] V. Bychkovsky, S. Paris, E. Chan, and F. Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011)*, pages 97–104, 2011. [page 100]

[46] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi. Image attributes and quality for evaluation of tone mapping operators. *Proceedings of Pacific Graphics (PG 2006)*, pages 35–44, 2006. [pages 41 and 57]

[47] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi. Evaluation of HDR tone mapping methods using essential perceptual attributes. *Computers & Graphics*, 32(3), 2008. [pages 41, 42, and 57]

[48] A. Chalmers, G. Bonnet, F. Banterle, P. Dubla, K. Debattista, A. Artusi, and C. Moir. High-dynamic-range video solution. In *ACM SIGGRAPH ASIA 2009, Art Gallery & Emerging Technologies: Adaptation*, pages 71–71, 2009. [pages 20, 23, and 48]

[49] A. Chalmers, P. Campisi, P. Shirley, and I. Olaizola, editors. *High Dynamic Range Video: Concepts, Technologies and Applications*. Academic Press, 2016. [page 15]

[50] J. Chen, S. Paris, and F. Durand. Real-time edge-aware image processing with the bilateral grid. *ACM Transactions on Graphics*, 26(3):103:1–103:9, 2007. [page 37]

[51] K. Chiu, M. Herf, P. Shirley, S. Swamy, C. Wang, K. Zimmerman, et al. Spatially nonuniform scaling functions for high contrast images. In *Graphics Interface*, pages 245–245, 1993. [pages 35 and 37]

[52] P. Choudhury and J. Tumblin. The trilateral filter for high contrast images and meshes. In *Proceedings of Eurographics workshop on Rendering (EGWR 2003)*, pages 186–196, 2003. [pages 35 and 37]

[53] Contrast Optical. Fathom 4K HDR product specification. `https://www.contrastoptical.com/fathom-4k`, Accessed: 2018-04-15. [page 20]

[54] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. [pages 25 and 40]

[55] S. J. Daly and X. Feng. Bit-depth extension using spatiotemporal microdither based on models of the equivalent input noise of the visual system. In *Proceedings of SPIE, Color Imaging VIII: Processing, Hardcopy, and Applications*, volume 5008, pages 455–466, 2003. [page 25]

[56] S. J. Daly and X. Feng. Decontouring: prevention and removal of false contour artifacts. In *Proceedings of SPIE, Human Vision and Electronic Imaging IX*, volume 5292, pages 130–149, 2004. [page 25]

[57] G. Damberg, H. Seetzen, G. Ward, W. Heidrich, and L. Whitehead. 3.2: High dynamic range projection systems. *SID Symposium Digest of Technical Papers*, 38(1):4–7, 2007. [page 44]

[58] G. Damberg, J. Gregson, and W. Heidrich. High brightness HDR projection using dynamic freeform lensing. *ACM Transactions on Graphics*, 35(3): 24:1–24:11, 2016. [page 44]

[59] D.-T. Dang-Nguyen, C. Pasquini, V. Conotter, and G. Boato. RAISE: A raw images dataset for digital image forensics. In *Proceedings of ACM Multimedia Systems Conference (MMSys 2015)*, pages 219–224, 2015. [page 100]

[60] P. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of SIGGRAPH 1998, Annual Conference Series*, pages 189–198, 1998. [page 9]

[61] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of SIGGRAPH 1997, Annual Conference Series*, pages 369–378, 1997. [pages 21 and 22]

[62] P. B. Delahunt, X. Zhang, and D. H. Brainard. Perceptual image quality: Effects of tone characteristics. *Journal of Electronic Imaging*, 14(2), 2005. [page 41]

[63] J. M. DiCarlo and B. A. Wandell. Rendering high dynamic range images. In *Proceedings of SPIE, Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications*, volume 3965, pages 392–401, 2000. [page 37]

[64] P. Didyk, R. Mantiuk, M. Hein, and H. Seidel. Enhancement of bright video features for HDR displays. *Computer Graphics Forum*, 27(4):1265–1274, 2008. [page 26]

[65] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *Proceedings of European Conference on Computer Vision (ECCV 2014)*, pages 184–199, 2014. [page 86]

[66] X. Dong, B. Bonev, Y. Zhu, and A. L. Yuille. Region-based temporally consistent video post-processing. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015)*, 2015. [page 40]

[67] F. Drago, W. Martens, K. Myszkowski, and H.-P. Seidel. Perceptual evaluation of tone mapping operators with regard to similarity and preference. Research Report MPI-I-2002-4-002, Max-Planck-Institut für Informatik, 2002. [page 41]

[68] F. Drago, K. Myszkowski, T. Annen, and N. Chiba. Adaptive logarithmic mapping for displaying high contrast scenes. *Computer Graphics Forum*, 22:419–426, 2003. [pages 35 and 38]

[69] J. Duan, M. Bressan, C. Dance, and G. Qiu. Tone-mapping high dynamic range images by novel histogram adjustment. *Pattern Recognition*, 43(5): 1847–1862, 2010. [page 38]

[70] F. Dufaux, G. J. Sullivan, and T. Ebrahimi. The JPEG XR image coding standard [standards in a nutshell]. *IEEE Signal Processing Magazine*, 26(6): 195–204, 2009. [page 30]

[71] F. Dufaux, P. L. Callet, R. K. Mantiuk, and M. Mrak, editors. *High Dynamic Range Video: From Acquisition, to Display and Applications*, volume 1. Academic Press, 2016. [pages ix and 15]

[72] F. Durand and J. Dorsey. Interactive tone mapping. In *Proceedings of Eurographics Workshop on Rendering Techniques*, pages 219–230. Springer-Verlag, 2000. [page 40]

[73] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics*, 21(3):257–266, 2002. [pages 35, 37, and 58]

[74] G. Eilertsen, J. Unger, R. Wanat, and R. K. Mantiuk. Survey and evaluation of tone mapping operators for HDR video. In *ACM SIGGRAPH 2013 Talks*, pages 11:1–11:1, 2013. [pages ix and 53]

[75] G. Eilertsen, R. Wanat, R. K. Mantiuk, and J. Unger. Evaluation of tone mapping operators for HDR-video. *Computer Graphics Forum (Proceedings of Pacific Graphics 2013)*, 32(7):275–284, 2013. [pages ix, 41, 43, and 48]

[76] G. Eilertsen, J. Unger, R. Wanat, and R. K. Mantiuk. Perceptually based parameter adjustments for video processing operations. In *ACM SIGGRAPH 2014 Talks*, pages 74:1–74:1, 2014. [pages ix, 42, and 50]

[77] G. Eilertsen, R. K. Mantiuk, and J. Unger. Real-time noise-aware tone mapping. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2015)*, 34(6):198:1–198:15, 2015. [pages x, 35, 37, 38, 40, 48, and 64]

[78] G. Eilertsen, R. K. Mantiuk, and J. Unger. Real-time noise-aware tone-mapping and its use in luminance retargeting. In *Proceedings of IEEE International Conference on Image Processing (ICIP 2016)*, pages 894–898, 2016.

[79] G. Eilertsen, R. K. Mantiuk, and J. Unger. A high dynamic range video codec optimized by large-scale testing. In *Proceedings of IEEE International Conference on Image Processing (ICIP 2016)*, pages 1379–1383, 2016. [pages x and 31]

[80] G. Eilertsen, R. K. Mantiuk, and J. Unger. Luma HDRv: an open source high dynamic range video codec optimized by large-scale testing. In *ACM SIGGRAPH 2016 Talks*, pages 17:1–17:2, 2016. [page x]

[81] G. Eilertsen, J. Unger, and R. K. Mantiuk. Evaluation of tone mapping operators for HDR video. In F. Dufaux, P. L. Callet, R. K. Mantiuk, and M. Mrak, editors, *High Dynamic Range Video: From Acquisition, to Display and Applications*, chapter 7, pages 185–207. Academic Press, 2016. [pages ix and 49]

[82] G. Eilertsen, P.-E. Forssén, and J. Unger. BriefMatch: Dense binary feature matching for real-time optical flow estimation. In *Proceedings of Scandinavian Conference on Image Analysis (SCIA 2017)*, pages 221–233, 2017.

[83] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger. HDR image reconstruction from a single exposure using deep CNNs. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2017)*, 36(6):178:1–178:15, 2017. [pages x and 27]

[84] G. Eilertsen, R. K. Mantiuk, and J. Unger. A comparative review of tone-mapping algorithms for high dynamic range video. *Computer Graphics Forum (Proceedings of Eurographics 2017)*, 36(2):565–592, 2017. [pages ix, 40, 43, and 48]

[85] Y. Endo, Y. Kanamori, and J. Mitani. Deep reverse tone mapping. *ACM Transactions on Graphics*, 36(6):177:1–177:10, 2017. [pages 27 and 86]

[86] M. D. Fairchild and G. M. Johnson. iCAM framework for image appearance, differences, and quality. *Journal of Electronic Imaging*, 13(1):126–138, 2004. [page 36]

[87] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics*, 27(3):67:1–67:10, 2008. [pages 35 and 37]

[88] R. Fattal. Edge-avoiding wavelets and their applications. *ACM Transactions on Graphics*, 28(3):1–10, 2009. [page 37]

[89] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, 21(3):249–256, 2002. [pages 35 and 36]

[90] G. Fechner. *Elements of psychophysics*. Holt, Rinehart & Winston, 1860/1965. [pages 31 and 36]

[91] P. Ferschin, I. Tastl, and W. Purgathofer. A comparison of techniques for the transformation of radiosity values to monitor colors. In *Proceedings of IEEE International Conference on Image Processing (ICIP 1994)*, volume 3, pages 992–996, 1994. [pages 35 and 38]

[92] J. Ferwerda and S. Luka. A high resolution, high dynamic range display for vision research. *Journal of Vision*, 9(8):346–346, 2009. [page 44]

[93] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg. A model of visual adaptation for realistic image synthesis. In *Proceedings of SIGGRAPH 1996, Annual Conference Series*, pages 249–258, 1996. [pages 3, 30, 34, 48, and 50]

[94] E. François, C. Fogg, Y. He, X. Li, A. Luthra, and A. Segall. 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):63–75, 2016. [page 9]

[95] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel. Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays. In *Proceedings of SPIE, Digital Photography X*, volume 9023, 2014. [pages 20, 23, 28, 76, and 86]

[96] M.-A. Gardner, K. Sunkavalli, E. Yumer, X. Shen, E. Gambaretto, C. Gagné, and J.-F. Lalonde. Learning to predict indoor illumination from a single image. *ACM Transactions on Graphics*, 36(6):176:1–176:14, 2017. [page 86]

[97] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, 2016. [page 86]

[98] A. Gilchrist and A. Jacobsen. Perception of lightness and illumination in a world of one reflectance. *Perception*, 13(1):5–19, 1984. [page 37]

[99] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of International Conference*

on Artificial Intelligence and Statistics (AISTATS 2010), volume 9, pages 249–256, 2010. [page 91]

[100] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Proceedings of International Conference on Neural Information Processing Systems (NIPS 2014)*, pages 2672–2680, 2014. [page 96]

[101] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016, `http://www.deeplearningbook.org`. [page 90]

[102] N. Goodnight, R. Wang, C. Woolley, and G. Humphreys. Interactive time-dependent tone mapping using programmable graphics hardware. In *Proceedings of Eurographics Workshop on Rendering (EGWR 2003)*, pages 26–37, 2003. [pages 40 and 48]

[103] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H. P. Seidel, and H. P. A. Lensch. Optimal HDR reconstruction with linear digital cameras. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, pages 215–222, 2010. [pages 22 and 40]

[104] J. Gu, Y. Hitomi, T. Mitsunaga, and S. Nayar. Coded rolling shutter photography: Flexible space-time sampling. In *Proceedings of IEEE International Conference on Computational Photography (ICCP 2010)*, pages 1–8, 2010. [page 24]

[105] D. Guo, Y. Cheng, S. Zhuo, and T. Sim. Correcting over-exposure in photographs. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, pages 515–521, 2010. [pages 26 and 82]

[106] Y. Guo, Z. Xie, W. Zhang, and L. Ma. Efficient high dynamic range video using multi-exposure CNN flow. In Y. Zhao, X. Kong, and D. Taubman, editors, *International Conference on Image and Graphics (ICIG 2017)*, pages 70–81, 2017. [page 86]

[107] B. Guthier, S. Kopf, M. Eble, and W. Effelsberg. Flicker reduction in tone mapped high dynamic range video. In *Proceedings of SPIE, Color Imaging XVI: Displaying, Processing, Hardcopy, and Applications*, volume 7866, 2011. [pages 35 and 40]

[108] D. Hafner, O. Demetz, and J. Weickert. Simultaneous HDR and optic flow computation. In *Proceedings of 22nd International Conference on Pattern Recognition (ICPR 2014)*, pages 2065–2070, 2014. [page 22]

[109] S. Hajisharif, J. Kronander, and J. Unger. HDR reconstruction for alternating gain (ISO) sensor readout. In *Eurographics 2014 Short Papers*, 2014. [pages 22 and 24]

[110] S. Hajisharif, J. Kronander, and J. Unger. Adaptive dualISO HDR reconstruction. *EURASIP Journal on Image and Video Processing*, 2015(41), 2015. [page 24]

[111] R. Harvey. Optical beam splitter and electronic high speed camera incorporating such a beam splitter, 1998. [Online]. Available: `https://www.google.com/patents/US5734507`, US Patent 5,734,507. [page 23]

[112] S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen, and M. Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics*, 35 (6):192:1–192:12, 2016. [pages 20 and 100]

[113] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397–1409, 2013. [pages 35 and 37]

[114] F. Heide, M. Steinberger, Y.-T. Tsai, M. Rouf, D. Pająk, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian, J. Kautz, and K. Pulli. FlexISP: A flexible camera image processing framework. *ACM Transactions on Graphics*, 33(6): 231:1–231:13, 2014. [page 22]

[115] Y. Hold-Geoffroy, K. Sunkavalli, S. Hadap, E. Gambaretto, and J.-F. Lalonde. Deep outdoor illumination estimation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, 2017. [page 86]

[116] B. K. Horn. Determining lightness from an image. *Computer Graphics and Image Processing*, 3(4):277–299, 1974. [page 37]

[117] X. Hou, J. Duan, and G. Qiu. Deep feature consistent deep image transformations: Downscaling, decolorization and HDR tone mapping. *arXiv preprint arXiv:1707.09482*, 2017. [page 86]

[118] J. Hu, O. Gallo, K. Pulli, and X. Sun. HDR deghosting: How to deal with saturation? In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2013)*, pages 1163–1170, 2013. [page 22]

[119] S. Iizuka, E. Simo-Serra, and H. Ishikawa. Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Transactions on Graphics*, 35(4):110:1–110:11, 2016. [page 86]

[120] S. Iizuka, E. Simo-Serra, and H. Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics*, 36(4):107:1–107:14, 2017. [page 98]

[121] P. Irawan, J. A. Ferwerda, and S. R. Marschner. Perceptually based tone mapping of high dynamic range image streams. In *Proceedings of Eurographics Conference on Rendering Techniques 16 (EGSR 2005)*, pages 231–242, 2005. [pages 34, 48, and 50]

[122] S. Jia, Y. Zhang, D. Agrafiotis, and D. Bull. Blind high dynamic range image quality assessment using deep learning. In *Proceedings of IEEE International Conference on Image Processing (ICIP 2017)*, pages 765–769, 2017. [page 86]

[123] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7):965–976, 1997. [page 37]

[124] F. Kainz, R. Bogart, and P. Stanczyk. Technical introduction to OpenEXR. *Industrial light and magic*, 2009. [page 28]

[125] N. K. Kalantari and R. Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Transactions on Graphics*, 36(4):144:1–144:12, 2017. [pages 22 and 86]

[126] N. K. Kalantari, E. Shechtman, C. Barnes, S. Darabi, D. B. Goldman, and P. Sen. Patch-based high dynamic range video. *ACM Transactions on Graphics*, 32(6):202:1–202:8, 2013. [pages 20 and 23]

[127] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. *ACM Transactions on Graphics*, 22(3):319–325, 2003. [pages 20, 22, 40, and 48]

[128] S. Kavadias, B. Dierickx, D. Scheffer, A. Alaerts, D. Uwaerts, and J. Bogaerts. A logarithmic response CMOS image sensor with on-chip calibration. *IEEE Journal of Solid-State Circuits*, 35(8):1146–1152, 2000. [page 19]

[129] M. H. Kim, T. Weyrich, and J. Kautz. Modeling human color perception under extended luminance levels. *ACM Transactions on Graphics*, 28(3): 27:1–27:9, 2009. [page 36]

[130] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. [page 89]

[131] R. P. Kovaleski and M. M. Oliveira. High-quality brightness enhancement functions for real-time reverse tone mapping. *The Visual Computer*, 25(5): 539–547, 2009. [page 26]

[132] R. P. Kovaleski and M. M. Oliveira. High-quality reverse tone mapping for a wide range of exposures. In *Proceedings of 27th Conference on Graphics, Patterns and Images (SIBGRAPI 2014)*, pages 49–56, 2014. [page 26]

[133] G. Krawczyk, K. Myszkowski, and H.-P. Seidel. Lightness perception in tone reproduction for high dynamic range images. *Computer Graphics Forum*, 24(3):635–645, 2005. [page 35]

[134] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Proceedings of International Conference on Neural Information Processing Systems (NIPS 2012)*, pages 1097–1105, 2012. [page 86]

[135] J. Kronander, S. Gustavson, G. Bonnet, and J. Unger. Unified HDR reconstruction from raw CFA data. In *Proceedings of IEEE International Conference on Computational Photography (ICCP 2013)*, pages 1–9, 2013. [pages 20, 22, 40, and 49]

[136] J. Kronander, S. Gustavson, G. Bonnet, A. Ynnerman, and J. Unger. A unified framework for multi-sensor HDR video reconstruction. *Signal Processing : Image Communications*, 29(2):203–215, 2014. [pages 20, 21, 22, 23, 48, 49, and 86]

[137] J. Kuang, H. Yamaguchi, G. M. Johnson, and M. D. Fairchild. Testing HDR image rendering algorithms. In *Proceedings of IS&T/SID 12th Color Imaging Conference*, pages 315–320, 2004. [page 41]

[138] J. Kuang, G. M. Johnson, and M. D. Fairchild. iCAM06: A refined image appearance model for HDR image rendering. *Journal of Visual Communication and Image Representation*, 18(5):406–414, 2007, Special issue on High Dynamic Range Imaging. [pages 35 and 36]

[139] J. Kuang, H. Yamaguchi, C. Liu, G. M. Johnson, and M. D. Fairchild. Evaluating HDR rendering algorithms. *ACM Transactions on Applied Perception*, 4(2), 2007. [pages 41, 42, and 57]

[140] J. Kuang, R. Heckaman, and M. D. Fairchild. 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. [page 41]

[141] T. Kunkel and E. Reinhard. A reassessment of the simultaneous dynamic range of the human visual system. In *Proceedings of Symposium on Applied Perception in Graphics and Visualization (APGV 2010)*, pages 17–24, 2010. [page 4]

[142] P. H. Kuo, C. S. Tang, and S. Y. Chien. Content-adaptive inverse tone mapping. In *Proceedings of Visual Communications and Image Processing (VCIP 2012)*, pages 1–6, 2012. [page 26]

[143] H. Landis. Production-ready global illumination. *ACM SIGGRAPH 20012 Course Notes*, 16, 2002. [page 26]

[144] M. Lang, O. Wang, T. Aydin, A. Smolic, and M. Gross. Practical temporal consistency for image-based graphics applications. *ACM Transactions on Graphics*, 31(4):34:1–34:8, 2012. [page 40]

[145] P. Ledda, L. P. Santos, and A. Chalmers. A local model of eye adaptation for high dynamic range images. In *Proceedings of International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa 3 (AFRIGRAPH 2004)*, pages 151–160, 2004. [pages 34, 40, and 50]

[146] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen. Evaluation of tone mapping operators using a high dynamic range display. *ACM Transactions on Graphics*, 24(3), 2005. [page 41]

[147] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802*, 2016. [page 98]

[148] C. Lee and C.-S. Kim. Gradient domain tone mapping of high dynamic range videos. In *Proceedings of International Conference on Image Processing (ICIP 2007)*, 2007. [pages 36 and 40]

[149] C. Lee and C. S. Kim. Rate-distortion optimized compression of high dynamic range videos. In *Proceedings of 16th European Signal Processing Conference (EUSIPCO 2008)*, pages 1–5, 2008. [page 32]

[150] J. W. Lee, B. R. Lim, R.-H. Park, J.-S. Kim, and W. Ahn. Two-stage false contour detection using directional contrast and its application to adaptive false contour reduction. *IEEE Transactions on Consumer Electronics*, 52(1): 179–188, 2006. [page 25]

[151] S. Lee, G. H. An, and S.-J. Kang. Deep chain HDRI: Reconstructing a high dynamic range image from a single low dynamic range image. *arXiv preprint arXiv:1801.06277*, 2018. [pages 27 and 86]

[152] H. Li and P. Peers. CRF-net: Single image radiometric calibration using CNNs. In *Proceedings of European Conference on Visual Media Production (CVMP 2017)*, pages 5:1–5:9, 2017. [page 86]

[153] H. Li, G. Li, L. Lin, and Y. Yu. Context-aware semantic inpainting. *arXiv preprint arXiv:1712.07778*, 2017. [page 98]

[154] J. Li, O. Skorka, K. Ranaweera, and D. Joseph. Novel real-time tone-mapping operator for noisy logarithmic CMOS image sensors. *Journal of Imaging Science and Technology*, 60(2):1–13, 2016. [page 40]

[155] P. Lichtsteiner, C. Posch, and T. Delbruck. A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits*, 43(2):566–576, 2008. [page 19]

[156] H.-Y. Lin and W.-Z. Chang. High dynamic range imaging for stereoscopic scene representation. In *Proceedings of IEEE International Conference on Image Processing (ICIP 2009)*, pages 4305–4308, 2009. [page 23]

[157] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015)*, 2015. [page 86]

[158] T. Lule, M. Wagner, M. Verhoeven, H. Keller, and M. Bohm. 100000-pixel, 120-dB imager in TFA technology. *IEEE Journal of Solid-State Circuits*, 35 (5):732–739, 2000. [page 19]

[159] G. Luzardo, J. Aelterman, H. Luong, W. Philips, and D. Ochoa. Real-time false-contours removal for inverse tone mapped HDR content. In *Proceedings of ACM on Multimedia Conference (ACMMM 2017)*, pages 1472–1479, 2017. [page 25]

[160] B. C. Madden. Extended intensity range imaging. Technical Report MS-CIS-93-96, University of Pennsylvania, Department of Computer and Information Science, 1993. [pages 21 and 22]

[161] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward, and W. Heidrich. Optimizing a tone curve for backward-compatible high dynamic range image and video compression. *IEEE Transactions on Image Processing*, 20(6):1558–1571, 2011. [page 32]

[162] A. Manakov, J. F. Restrepo, O. Klehm, R. Hegedüs, E. Eisemann, H.-P. Seidel, and I. Ihrke. A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging. *ACM Transactions on Graphics*, 32(4):47:1–47:14, 2013. [page 24]

[163] S. Mangiat and J. Gibson. High dynamic range video with ghost removal. In *Proceedings of SPIE, Applications of Digital Image Processing XXXIII*, volume 7798, 2010. [pages 20 and 22]

[164] S. Mann and R. Picard. 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*, pages 442–448, 1995. [pages 21 and 22]

[165] S. Mann, C. Manders, and J. Fung. Painting with looks: Photographic images from video using quantimetric processing. In *Proceedings of ACM International Conference on Multimedia (ACMMM 2002)*, pages 117–126, 2002. [page 22]

[166] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.-P. Seidel. Perception-motivated high dynamic range video encoding. *ACM Transactions on Graphics*, 23(3):733–741, 2004. [pages 30, 31, 72, 74, and 75]

[167] R. Mantiuk, K. Myszkowski, and H.-P. Seidel. A perceptual framework for contrast processing of high dynamic range images. In *Proceedings of Applied perception in graphics and visualization (APGV 2005)*, pages 87–94, 2005. [page 36]

[168] R. Mantiuk, A. Efremov, K. Myszkowski, and H.-P. Seidel. Backward compatible high dynamic range MPEG video compression. *ACM Transactions on Graphics*, 25(3):713–723, 2006. [page 32]

[169] R. Mantiuk, G. Krawczyk, R. Mantiuk, and H.-P. Seidel. High dynamic range imaging pipeline: Perception-motivated representation of visual content. In *Proceedings of SPIE, Human Vision and Electronic Imaging XII*, volume 6492, 2007. [page 77]

[170] R. Mantiuk, S. Daly, and L. Kerofsky. Display adaptive tone mapping. *ACM Transactions on Graphics*, 27(3):68:1–68:10, 2008. [pages 35, 38, 48, and 50]

[171] R. Mantiuk, R. Mantiuk, A. Tomaszewska, and W. Heidrich. Color correction for tone mapping. *Computer Graphics Forum*, 28(2):193–202, 2009. [page 39]

[172] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on Graphics*, 30(4):40:1–40:14, 2011. [pages 41, 65, 73, and 74]

[173] R. K. Mantiuk. Practicalities of predicting quality of high dynamic range images and video. In *Proceedings of IEEE International Conference on Image Processing (ICIP 2016)*, pages 904–908, 2016. [page 73]

[174] R. K. Mantiuk, A. Tomaszewska, and R. Mantiuk. Comparison of four subjective methods for image quality assessment. *Computer Graphics Forum*, 31(8), 2012. [page 42]

[175] R. K. Mantiuk, K. Myszkowski, and H.-P. Seidel. *High Dynamic Range Imaging*. John Wiley & Sons, Inc., 2015. [pages 7 and 15]

[176] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista. ExpandNet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. *arXiv preprint arXiv:1803.02266*, 2018. [pages 27 and 87]

[177] B. Masia, S. Agustin, R. W. Fleming, O. Sorkine, and D. Gutierrez. Evaluation of reverse tone mapping through varying exposure conditions. *ACM Transactions on Graphics*, 28(5):160:1–160:8, 2009. [pages 26 and 82]

[178] B. Masia, A. Serrano, and D. Gutierrez. Dynamic range expansion based on image statistics. *Multimedia Tools and Applications*, 76(1):631–648, 2017. [page 26]

[179] S. Z. Masood, J. Zhu, and M. F. Tappen. Automatic correction of saturated regions in photographs using cross-channel correlation. *Computer Graphics Forum*, 28(7):1861–1869, 2009. [pages 26 and 82]

[180] M. McGuire, W. Matusik, H. Pfister, B. Chen, J. F. Hughes, and S. K. Nayar. Optical splitting trees for high-precision monocular imaging. *IEEE Computer Graphics and Applications*, 27(2):32–42, 2007. [page 23]

[181] M. Melo, M. Bessa, K. Debattista, and A. Chalmers. Evaluation of HDR video tone mapping for mobile devices. *Signal Processing: Image Communication*, 29(2):247–256, 2014, Special Issue on Advances in High Dynamic Range Video Research. [pages 41 and 43]

[182] L. Meylan, S. Daly, and S. Süsstrunk. The reproduction of specular highlights on high dynamic range displays. *Color and Imaging Conference*, 2006(1):333–338, 2006. [page 26]

[183] L. Meylan, D. Alleysson, and S. Süsstrunk. Model of retinal local adaptation for the tone mapping of color filter array images. *Journal of the Optical Society of America A*, 24(9):2807–2816, 2007. [pages 26 and 34]

[184] N. J. Miller, P. Y. Ngai, and D. D. Miller. The application of computer graphics in lighting design. *Journal of the Illuminating Engineering Society*, 14(1):6–26, 1984. [page 33]

[185] S. Miller, M. Nezamabadi, and S. Daly. Perceptual signal coding for more efficient usage of bit codes. *SMPTE Motion Imaging Journal*, 122(4), 2013. [pages 30, 31, 74, and 75]

[186] T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 1999)*, volume 1, pages 374–380. IEEE, 1999. [page 22]

[187] R. Mukherjee, K. Debattista, T. Bashford-Rogers, P. Vangorp, R. Mantiuk, M. Bessa, B. Waterfield, and A. Chalmers. Objective and subjective evaluation of high dynamic range video compression. *Signal Processing: Image Communication*, 47:426–437, 2016. [page 32]

[188] K. Naka and W. Rushton. S-potentials from colour units in the retina of fish (cyprinidae). *The Journal of physiology*, 185(3):536–555, 1966. [page 38]

[189] S. G. Narasimhan and S. K. Nayar. Enhancing resolution along multiple imaging dimensions using assorted pixels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):518–530, 2005. [page 24]

[190] S. K. Nayar and T. Mitsunaga. High dynamic range imaging: spatially varying pixel exposures. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000)*, volume 1, pages 472–479, 2000. [page 24]

[191] M. Okuda and N. Adami. Two-layer coding algorithm for high dynamic range images based on luminance compensation. *Journal of Visual Communication and Image Representation*, 18(5):377–386, 2007, Special issue on High Dynamic Range Imaging. [page 32]

[192] A. Oppenheim, R. Schafer, and T. Stockham. Nonlinear filtering of multiplied and convolved signals. *IEEE Transactions on Audio and Electroacoustics*, 16(3):437–466, 1968. [page 33]

[193] D. Pajak, M. Čadík, T. O. Aydin, K. Myszkowski, and H.-P. Seidel. Visual maladaptation in contrast domain. In *Proceedings of SPIE, Human Vision and Electronic Imaging XV*, volume 7527, 2010. [page 34]

[194] S. Paris and F. Durand. A fast approximation of the bilateral filter using a signal processing approach. In *Proceedings of European Conference on Computer Vision (ECCV 2006)*, pages 568–580, 2006. [page 37]

[195] S. Paris, S. W. Hasinoff, and J. Kautz. Local Laplacian filters: edge-aware image processing with a Laplacian pyramid. *ACM Transactions on Graphics*, 30(4):68:1–68:12, 2011. [pages 35 and 37]

[196] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. Efros. Context encoders: Feature learning by inpainting. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, pages 2536–2544, 2016. [page 96]

[197] S. Pattanaik and H. Yee. Adaptive gain control for high dynamic range image display. In *Proceedings of Spring Conference on Computer Graphics (SCCG 2002)*, pages 83–87, 2002. [page 37]

[198] S. N. Pattanaik, J. A. Ferwerda, M. D. Fairchild, and D. P. Greenberg. A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of SIGGRAPH 1998, Annual Conference Series*, pages 287–298. ACM, 1998. [pages 34, 36, 37, and 38]

[199] S. N. Pattanaik, J. Tumblin, H. Yee, and D. P. Greenberg. Time-dependent visual adaptation for fast realistic image display. In *Proceedings of SIGGRAPH 2000, Annual Conference Series*, pages 47–54, 2000. [pages 34, 48, and 50]

[200] M. Perez-Ortiz and R. K. Mantiuk. A practical guide and software for analysing pairwise comparison experiments. *arXiv preprint arXiv:1712.03686*, 2017. [page 55]

[201] J. Petit and R. K. Mantiuk. Assessment of video tone-mapping : Are cameras' S-shaped tone-curves good enough? *Journal of Visual Communication and Image Representation*, 24, 2013. [pages 41, 43, and 57]

[202] Photons to Photos. Photographic dynamic range versus ISO setting. `http://www.photonstophotos.net/Charts/PDR.htm`, Accessed: 2018-04-15. [page 18]

[203] T. Pouli and E. Reinhard. Progressive histogram reshaping for creative color transfer and tone reproduction. In *Proceedings of International Symposium on Non-Photorealistic Animation and Rendering (NPAR 2010)*, pages 81–90, 2010. [page 38]

[204] T. Pouli, A. Artusi, F. Banterle, A. O. Akyüz, H.-P. Seidel, and E. Reinhard. Color correction for tone reproduction. In *Proceedings of Color and Imaging Conference (CIC 2013)*, volume 2013, pages 215–220, 2013. [page 39]

[205] M. J. D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, 7(2), 1964. [page 52]

[206] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015. [page 96]

[207] A. Radonjić, S. R. Allred, A. L. Gilchrist, and D. H. Brainard. The dynamic range of human lightness perception. *Current Biology*, 21(22):1931–1936, 2011. [page 4]

[208] Z.-u. Rahman, D. J. Jobson, and G. A. Woodell. A multiscale retinex for color rendition and dynamic range compression. In *Proceedings of SPIE, International Symposium on Optical Science, Engineering and Instrumentation, Applications of Digital Image Processing XIX*, volume 2847, pages 183–191, 1996. [pages 35 and 37]

[209] S. D. Ramsey, J. T. Johnson III, and C. Hansen. Adaptive temporal tone mapping. In *Proceedings of IASTED International Conference on Computer Graphics and Imaging (CGIM 2004)*, pages 124–128. Citeseer, 2004. [pages 40 and 48]

[210] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. *ACM Transactions on Graphics*, 21(3): 267–276, 2002. [pages 35 and 37]

[211] E. Reinhard, G. Ward, S. N. Pattanaik, P. E. Debevec, W. Heidrich, and K. Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting (2nd ed.)*. Morgan Kaufmann, 2010. [page 15]

[212] E. Reinhard, T. Pouli, T. Kunkel, B. Long, A. Ballestad, and G. Damberg. Calibrated image appearance reproduction. *ACM Transactions on Graphics*, 31(6), 2012. [pages 35, 36, and 50]

[213] K. Rematas, T. Ritschel, M. Fritz, E. Gavves, and T. Tuytelaars. Deep reflectance maps. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, pages 4508–4516, 2016. [page 86]

[214] A. G. Rempel, M. Trentacoste, H. Seetzen, H. D. Young, W. Heidrich, L. Whitehead, and G. Ward. Ldr2Hdr: On-the-fly reverse tone mapping of legacy video and photographs. *ACM Transactions on Graphics*, 26(3), 2007. [page 26]

[215] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Proceedings of International Conference on Neural Information Processing Systems (NIPS 2015)*, pages 91–99, 2015. [page 86]

[216] M. Řeřábek and T. Ebrahimi. Comparison of compression efficiency between HEVC/H.265 and VP9 based on subjective assessments. In *Proceedings of SPIE, Optical Engineering + Applications*, volume 9217, 2014. [page 75]

[217] M. A. Robertson, S. Borman, and R. L. Stevenson. Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging*, 12(2):219–229, 2003. [page 22]

[218] M. Rouf, R. Mantiuk, W. Heidrich, M. Trentacoste, and C. Lau. Glare encoding of high dynamic range images. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011)*, pages 289–296, 2011. [page 24]

[219] Rtings.com. Peak brightness of TVs. `https://www.rtings.com/tv/tests/picture-quality/peak-brightness`, Accessed: 2018-04-15. [page 45]

[220] C. Schlick. Quantization techniques for visualization of high dynamic range pictures. In *Photorealistic Rendering Techniques*, pages 7–20, 1995. [pages 35, 36, and 38]

[221] M. Schöberl, A. Belz, J. Seiler, S. Foessel, and A. Kaup. High dynamic range video by spatially non-regular optical filtering. In *Proceedings of IEEE International Conference on Image Processing (ICIP 2012)*, pages 2757–2760, 2012. [page 24]

[222] H. Seetzen, L. A. Whitehead, and G. Ward. 54.2: A high dynamic range display using low and high resolution modulators. *SID Symposium Digest of Technical Papers*, 34(1):1450–1453, 2003. [page 43]

[223] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs. High dynamic range display systems. *ACM Transactions on Graphics*, 23(3):760–768, 2004. [pages 9 and 43]

[224] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman. Robust patch-based HDR reconstruction of dynamic scenes. *ACM Transactions on Graphics*, 31(6):203:1–203:11, 2012. [page 22]

[225] A. Serrano, F. Heide, D. Gutierrez, G. Wetzstein, and B. Masia. Convolutional sparse coding for high dynamic range imaging. *Computer Graphics Forum*, 35(2):153–163, 2016. [page 24]

[226] F. D. Simone, G. Valenzise, P. Lauga, F. Dufaux, and F. Banterle. Dynamic range expansion of video sequences: A subjective quality assessment study. In *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP 2014)*, pages 1063–1067, 2014. [pages 26 and 82]

[227] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. [pages 86 and 87]

[228] L. Song, Y. Liu, X. Yang, G. Zhai, R. Xie, and W. Zhang. The SJTU HDR video sequence dataset. In *Proceedings of International Conference on Quality of Multimedia Experience (QoMEX 2016)*, 2016. [page 100]

[229] Q. Song, G. M. Su, and P. C. Cosman. Hardware-efficient debanding and visual enhancement filter for inverse tone mapped high dynamic range images and videos. In *Proceedings of IEEE International Conference on Image Processing (ICIP 2016)*, pages 3299–3303, 2016. [page 25]

[230] Sony. Sony news releases January 09, 2018. `https://www.sony.net/SonyInfo/News/Press/201801/18-002E/index.html`, 2018, Accessed: 2018-04-15. [page 44]

[231] K. E. Spaulding. Using a residual image to extend the color gamut and dynamic range of an sRGB image. *Proceedings of IS&T PICS Conference, 2003*, pages 307–314, 2003. [page 32]

[232] L. Spillmann and J. S. Werner. *Visual perception: The neurophysiological foundations*. Elsevier, 2012. [page 4]

[233] S. Stevens. *Psychophysics: Introduction to Its Perceptual, Neural, and Social Prospects*. John Wiley & Sons, 1975. [page 41]

[234] K. Subr, C. Soler, and F. Durand. Edge-preserving multiscale image decomposition based on local extrema. *ACM Transactions on Graphics*, 28 (5):147:1–147:9, 2009. [page 37]

[235] N. Sun, H. Mansour, and R. Ward. HDR image construction from multi-exposed stereo LDR images. In *Proceedings of IEEE International Conference on Image Processing (ICIP 2010)*, pages 2973–2976, 2010. [page 23]

[236] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen. A versatile HDR video production system. *ACM Transactions on Graphics*, 30(4):41:1–41:10, 2011. [pages 20, 23, and 48]

[237] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of International Conference on Computer Vision (ICCV 1998)*, pages 839–846, 1998. [pages 37 and 58]

[238] A. Tomaszewska and R. Mantiuk. Image registration for multi-exposure high dynamic range image acquisition. In *Proceedings of International Conference on Computer Graphics, Visualization and Computer Vision (WSCG 2007)*, 2007. [page 22]

[239] A. Troccoli, S. B. Kang, and S. Seitz. Multi-view multi-exposure stereo. In *Third International Symposium on 3D Data Processing, Visualization, and Transmission*, pages 861–868, 2006. [page 23]

[240] Y. Tsin, V. Ramesh, and T. Kanade. Statistical calibration of CCD imaging process. In *Proceedings IEEE International Conference on Computer Vision (ICCV 2001)*, volume 1, pages 480–487, 2001. [page 22]

[241] J. Tumblin and H. Rushmeier. Tone reproduction for realistic images. *IEEE Computer Graphics and Applications*, 13(6):42–48, 1993. [pages 33, 35, and 38]

[242] J. Tumblin and G. Turk. LCIS: a boundary hierarchy for detail-preserving contrast reduction. In *Proceedings of SIGGRAPH 1999, Annual Conference Series*, pages 83–90, 1999. [page 37]

[243] J. Tumblin, J. K. Hodgins, and B. K. Guenter. Two methods for display of high contrast images. *ACM Transactions on Graphics*, 18(1):56–94, 1999. [page 38]

[244] O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem. The state of the art in HDR deghosting: A survey and evaluation. *Computer Graphics Forum*, 34(2):683–707, 2015. [page 22]

[245] J. Unger and S. Gustavson. High-dynamic-range video for photometric measurement of illumination. In *Proceedings of SPIE, Sensors, Cameras, and Systems for Scientific/Industrial Applications VIII*, volume 6501, 2007. [pages 20 and 24]

[246] J. Unger, S. Gustavson, M. Ollila, and M. Johannesson. A real time light probe. In *Proceedings of Eurographics Annual Conference, Short Papers and Interactive Demos*, pages 17–21, 2004. [page 20]

[247] J. Unger, J. Kronander, P. Larsson, S. Gustavson, J. Löw, and A. Ynnerman. Spatially varying image based lighting using HDR-video. *Computers and Graphics*, 37(7), 2013. [page 9]

[248] J. Unger, F. Banterle, G. Eilertsen, and R. K. Mantiuk. The HDR-video pipeline - from capture and image reconstruction to compression and tone mapping. In *Eurographics 2016 Tutorials*, 2016.

[249] S. D. Upstill. *The Realistic Presentation of Synthetic Images: Image Processing in Computer Graphics*. PhD thesis, University of California, Berkeley, 1985. [page 33]

[250] J. H. van Hateren. Encoding of high dynamic range video with a model of human cones. *ACM Transactions on Graphics*, 25:1380–1399, 2006. [pages 34, 40, 48, and 50]

[251] P. Vangorp, K. Myszkowski, E. W. Graf, and R. K. Mantiuk. A model of local adaptation. *ACM Transactions on Graphics*, 34(6):166:1–166:13, 2015. [page 4]

[252] C. Villa and R. Labayrade. Psychovisual assessment of tone-mapping operators for global appearance and colour reproduction. In *Proceedings of Colour in Graphics Imaging and Vision (CIC 2010)*, 2010. [page 41]

[253] R. Wanat, J. Petit, and R. Mantiuk. Physical and perceptual limitations of a projector-based high dynamic range display. In *Theory and Practice of Computer Graphics*. The Eurographics Association, 2012. [page 44]

[254] H. Wang, R. Raskar, and N. Ahuja. High dynamic range video using split aperture camera. In *Proceedings of IEEE 6th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras (OMNIVIS 2005)*, 2005. [pages 20, 23, 36, and 48]

[255] L. Wang, L.-Y. Wei, K. Zhou, B. Guo, and H.-Y. Shum. High dynamic range image hallucination. In *Proceedings of Eurographics Conference on Rendering Techniques (EGSR 2007)*, pages 321–326, 2007. [pages 27 and 82]

[256] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *Proceedings of Asilomar Conference on Signals, Systems Computers, (ACSSC 2003)*, volume 2, pages 1398–1402, 2003. [pages 41 and 73]

[257] G. Ward. Real pixels. *Graphics Gems II*, pages 80–83, 1991. [page 28]

[258] G. Ward. A contrast-based scalefactor for luminance display. *Graphics gems IV*, pages 415–421, 1994. [pages 35 and 38]

[259] G. Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of Graphics Tools*, 8 (2):17–30, 2003. [page 22]

[260] G. Ward and M. Simmons. Subband encoding of high dynamic range imagery. In *Proceedings of Symposium on Applied Perception in Graphics and Visualization (APGV 2004)*, pages 83–90, 2004. [page 32]

[261] G. Ward and M. Simmons. JPEG-HDR: A backwards-compatible, high dynamic range extension to JPEG. In *Proceedings of Color and Imaging Conference (CIC 2005)*, volume 2005, pages 283–290, 2005. [page 32]

[262] G. J. Ward. The RADIANCE lighting simulation and rendering system. In *Proceedings of SIGGRAPH 1994, Annual Conference Series*, pages 459–472, 1994. [page 28]

[263] G. Ward Larson. LogLuv encoding for full-gamut, high-dynamic range images. *Journal of Graphics Tools*, 3(1):15–31, 1998. [pages 29 and 74]

[264] G. Ward Larson, H. Rushmeier, and C. Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics*, 3(4):291–306, 1997. [pages 35 and 38]

[265] B. Weiss. Fast median and bilateral filtering. *ACM Transactions on Graphics*, 25(3):519–526, 2006. [page 37]

[266] S. Wu, J. Xu, Y. Tai, and C. Tang. End-to-end deep HDR imaging with large foreground motions. *arXiv preprint arXiv:1711.08937*, 2017. [page 86]

[267] D. Xu, C. Doutre, and P. Nasiopoulos. Correction of clipped pixels in color images. *IEEE Transactions on Visualization and Computer Graphics*, 17 (3):333–344, 2011. [pages 26 and 82]

[268] R. Xu, S. N. Pattanaik, and C. E. Hughes. High-dynamic-range still-image encoding in JPEG 2000. *IEEE Computer Graphics and Applications*, 25(6): 57–64, 2005. [page 29]

[269] Q. Yang. Recursive bilateral filtering. In *Proceedings of European Conference on Computer Vision (ECCV 2012)*, pages 399–413, 2012. [page 37]

[270] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar. Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum. *IEEE Transactions on Image Processing*, 19(9):2241–2253, 2010. [page 24]

[271] H. Yeganeh and Z. Wang. Objective quality assessment of tone-mapped images. *IEEE Transactions on Image Processing*, 22(2):657–667, 2013. [page 41]

[272] H. Yeganeh, S. Wang, K. Zeng, M. Eisapour, and Z. Wang. Objective quality assessment of tone-mapped videos. In *Proceedings of IEEE International Conference on Image Processing (ICIP 2016)*, pages 899–903, 2016. [page 41]

[273] A. Yoshida, V. Blanz, K. Myszkowski, and H.-P. Seidel. Perceptual evaluation of tone mapping operators with real world scenes. In *Proceedings of SPIE, Human Vision and Electronic Imaging X*, volume 5666, 2005. [pages 41, 42, and 50]

[274] A. Yoshida, R. Mantiuk, K. Myszkowski, and H.-P. Seidel. Analysis of reproducing real-world appearance on displays of varying dynamic range. *Computer Graphics Forum*, 25(3), 2006. [page 41]

[275] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. Generative image inpainting with contextual attention. *arXiv preprint arXiv:1801.07892*, 2018. [page 98]

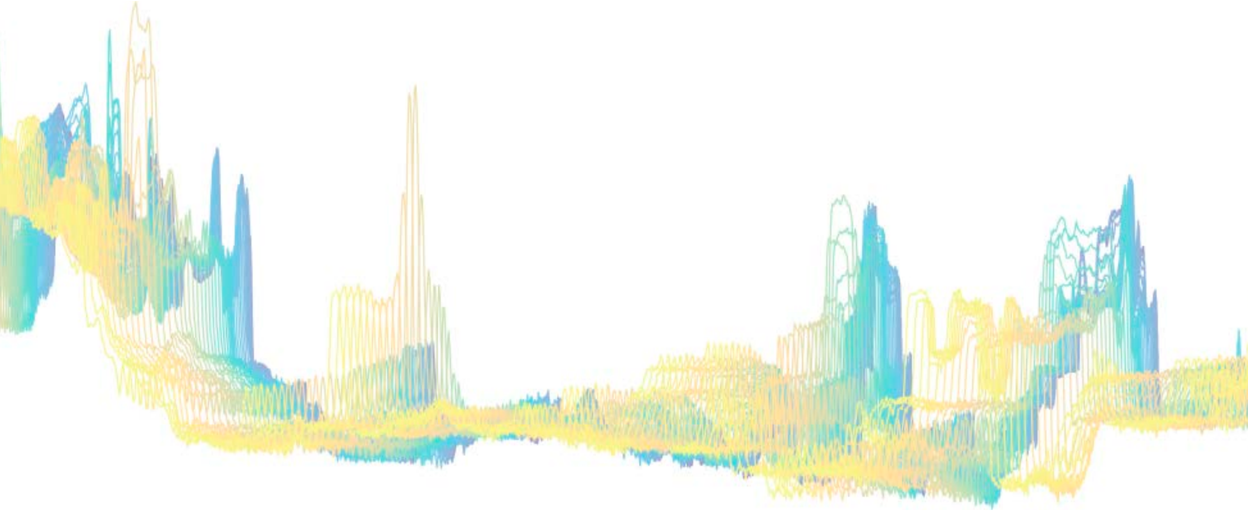
[276] J. Zhang and J.-F. Lalonde. Learning high dynamic range from outdoor panoramas. In *Proceedings of IEEE International Conference on Computer Vision (ICCV 2017)*, 2017. [pages 27 and 86]

[277] R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. In *Proceedings of European Conference on Computer Vision (ECCV 2016)*, pages 649–666, 2016. [page 86]

[278] X. Zhang and D. H. Brainard. Estimation of saturated pixel values in digital color imaging. *Journal of the Optical Society of America A*, 21(12): 2301–2310, 2004. [pages 26 and 82]

[279] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning deep features for scene recognition using places database. In *Proceedings of International Conference on Neural Information Processing Systems (NIPS 2014)*, pages 487–495, 2014. [pages 90 and 91]

[280] H. Zimmer, A. Bruhn, and J. Weickert. Freehand HDR imaging of moving scenes with simultaneous resolution enhancement. *Computer Graphics Forum*, 30(2):405–414, 2011. [page 22]



Publications

Publications