Encoding High Dynamic Range and Wide Color Gamut Imagery

Von der Fakultät Informatik, Elektrotechnik und Informationstechnik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von

Jan Fröhlich

aus Freiburg im Breisgau

Hauptberichter: Mitberichter: 2. Mitberichter: Prof. Dr. Daniel Weiskopf Prof. Dr. Bernhard Eberhardt Prof. Dr. Andreas Schilling

Tag der mündlichen Prüfung: 13.12.2017

Visualisierungsinstitut der Universität Stuttgart

2017

Contents

Table of Contents	3
List of Figures	9
List of Tables	13
List of Abbreviations	15
List of Symbols	19
Abstract	21
Zusammenfassung (German Abstract)	23
1 Introduction	27
1.1 Open Challenges in HDR and WCG Imaging	28

	1.2	Reseat	rch Questi	ions	29
	1.3	Outlin	e and Co	ntributions	30
	1.4	Reuse	d and Cop	pyrighted Material	31
2	Bac	kgroun	d and Re	lated Work	33
	2.1	Early	Research	on Perceptual Uniformity and Contrast Sensitivity	34
	2.2	Spatia	l Contrast	Sensitivity	36
	2.3	Lumin	ance Enco	odings	37
		2.3.1	Standard	d Dynamic Range Encodings	38
		2.3.2	HDR Lui	minance Encoding in Video Acquisition	39
		2.3.3	HDR Lui	minance Encoding in Video Distribution	40
			2.3.3.1	DICOM Grayscale Standard Display Function	40
			2.3.3.2	PU Transfer Function	41
			2.3.3.3	PQ Transfer Function	41
			2.3.3.4	HLG Transfer Function	43
			2.3.3.5	Comparison of HDR Transfer Functions	44
	2.4	Color	Difference	e Encodings	45
		2.4.1	Spatial C	Color Contrast Sensitivity	46
		2.4.2	Color Ap	ppearance Models	47
			2.4.2.1	CIE L*a*b*	48

			2.4.2.2	IPT	50
			2.4.2.3	Extended Color Appearance Models	51
		2.4.3	Standard	d Dynamic Range Video Color Encodings	52
		2.4.4	High Dy	namic Range Color Encodings	54
			2.4.4.1	RGBE	54
			2.4.4.2	OpenEXR	55
			2.4.4.3	LogLuv	55
			2.4.4.4	$Y^{\prime\prime}u^{\prime\prime}v^{\prime\prime}~\ldots~\ldots~\ldots~\ldots~\ldots$	56
			2.4.4.5	ITU Rec.2100	57
		2.4.5	Compari	ison of HDR Color Encodings	59
3	HDF	R and V	VCG Vide	o Data Set - 'HdM-HDR-2014'	61
3	HDF 3.1	R and V Relate	V CG Vide d High Dy	o Data Set - 'HdM-HDR-2014' ynamic Range Video Data Sets	61 63
3	HDF 3.1 3.2	R and V Relate Metho	VCG Vide d High Dy ds	o Data Set - 'HdM-HDR-2014' ynamic Range Video Data Sets	61 63 64
3	HDF 3.1 3.2	Relate Metho 3.2.1	VCG Vide d High Dy ds Selection	o Data Set - 'HdM-HDR-2014' ynamic Range Video Data Sets	61 63 64 64
3	HDF 3.1 3.2	Relate Metho 3.2.1 3.2.2	VCG Vide d High Dy ods Selection Set Desi	o Data Set - 'HdM-HDR-2014' ynamic Range Video Data Sets	 61 63 64 64 65
3	HDF 3.1 3.2	Relate Metho 3.2.1 3.2.2 3.2.3	VCG Vide d High Dy ods Selection Set Desig High Dy	o Data Set - 'HdM-HDR-2014' ynamic Range Video Data Sets	 61 63 64 64 65 66
3	HDF 3.1 3.2	Relate Metho 3.2.1 3.2.2 3.2.3 3.2.4	VCG Vide d High Dy ods Selection Set Desig High Dy Postproc	o Data Set - 'HdM-HDR-2014' ynamic Range Video Data Sets	 61 63 64 64 65 66 70
3	HDF 3.1 3.2 3.3	Relate Metho 3.2.1 3.2.2 3.2.3 3.2.4 Limita	VCG Vide d High Dy ods Selection Set Desi High Dy Postproc tions	o Data Set - 'HdM-HDR-2014' ynamic Range Video Data Sets	 61 63 64 64 65 66 70 74

4	Stat	ic WCG	and HDR Color Encodings - ' $IC_A C_B$ ' and ' $IC_T C_P$ '	77
	4.1	Resear	rch Question	79
	4.2	Metho	ds	80
		4.2.1	Model Derivation	80
		4.2.2	Datasets for Optimization	83
			4.2.2.1 Isoluminance Dataset	83
			4.2.2.2 Hue Linearity Dataset	85
			4.2.2.3 JND Dataset	85
		4.2.3	Cost Functions for Optimization	87
		4.2.4	Optimization Process	89
		4.2.5	Parameter Derivation for the IC_AC_B Model \ldots	90
		4.2.6	Parameter Derivation for the $IC_T C_P$ Model	92
			4.2.6.1 Minimal Contribution of the S-Cone to Luminance	92
			4.2.6.2 Crosstalk in LMS-Domain	93
			4.2.6.3 Full $IC_T C_P$ Model	97
	4.3	Result	S	97
		4.3.1	Visual Comparison	98
		4.3.2	Numerical Comparison	105
	4.4	Discus	sion and Outlook	108

5	Con	tent Aware Quantization Scheme - 'CAQ'	111
	5.1	Research Question	112
	5.2	Related Work	112
	5.3	Methods	113
		5.3.1 CAQ Based Quantization	114
		5.3.2 HDR-VDP-2.2 Based Quantization	117
	5.4	Results	117
	5.5	Summary	121
6	Con	clusion and Outlook	123
Bi	Bibliography 1		
A	Арр	endix: HdM-HDR-2014 Overview	149
Α	App A.1	endix: HdM-HDR-2014 Overview Still Life	149 150
A	Арр А.1 А.2	endix: HdM-HDR-2014 Overview Still Life	149 150 151
Α	Арр А.1 А.2 А.3	endix: HdM-HDR-2014 Overview Still Life	149 150 151 156
A	App A.1 A.2 A.3 A.4	endix: HdM-HDR-2014 Overview Still Life Wide Gamut and Moving Lights Low Key Scenes Sunlight Scenes	149150151156158
A	App A.1 A.2 A.3 A.4 A.5	endix: HdM-HDR-2014 Overview Still Life	 149 150 151 156 158 161

8_____

List of Figures

1.1	The HDR and WCG distribution chain from image acquisition and image synthesis to presentation.	27
2.1	Visualization of contrast sensitivity depending on spatial frequency	36
2.2	Derivation of the PQ, PU and DICOM luminance transfer functions from Barten's CSF model.	41
2.3	DICOM, PU, PQ and sRGB luminance transfer functions compared	42
2.4	Needed tonal resolution of the transfer functions PQ, floating-point, logarithmic, gamma and HLG to stay below the visibility threshold	44
2.5	PQ, gamma, logarithmic and HLG transfer functions at 12-bit tonal resolution compared.	45
2.6	Minimum discriminable differences in CIE 1931 xy and CIE 1976 u'v' chromaticity space.	46
2.7	Visualization of color contrast sensitivity depending on spatial frequency.	47
2.8	Example for CIE L*a*b* hue linearity compared to CIECAM02	49

2.9	Bit order of the LabTIFF encoding.	50
2.10	D Example for decorrelation by color difference encoding	52
2.1	1 Bit order of the RGBE encoding	54
2.12	2 Bit order of the OpenEXR half-float encoding	55
2.13	3 Bit order of the LogLuv32 encoding	56
3.1	Dynamic range of current acquisition and distribution devices compared.	62
3.2	Illustration of ghosting artifacts introduced by time-sequential HDR image capture.	67
3.3	Schematic drawing of the mirror rigs used for HDR video acquisition	68
3.4	Simulated and measured signal-to-noise ratio of the image acquisition system.	69
3.5	Overview of the in-camera video processing steps	70
3.6	Limitations to tonal resolution by the camera sensor.	71
3.7	Limitations to tonal resolution by the camera sensor and LogC storage combined	71
3.8	Overview of the video processing steps in postproduction	72
3.9	Content tonal resolution compared to the half-float container used for distributing the reconstructed camera footage.	72
3.10	O Content tonal resolution compared to the PQ 16-bit integer container used for distributing the graded content.	73
3.12	l Spectral power distribution and resulting gamut of the modified DLP cinema projector.	74

3.12	Mirror induced double contours - cause and effect	75
4.1	Overview of a typical HDR video distribution pipeline	78
4.2	Visual pattern used for the isoluminance user study	83
4.3	Study setup for deriving equiluminant color pairs and colors with the same hue	84
4.4	Visual pattern of the hue linearity user study.	85
4.5	Visual pattern of the JND user study.	86
4.6	Study setup for measuring JNDs	86
4.7	Example of JND ellipse representation as half axes and the effect of SVD after transformation.	89
4.8	Trading JND uniformity for hue linearity in $IC_A C_B$ optimization	91
4.9	Influence of crosstalk in linear LMS domain on gamut shape	94
4.10	Example of JND non-uniformity when encoding Rec.2100 RGB primaries with PQ	95
4.11	Spectral sensitivity of of the $IC_A C_B$ and $IC_T C_P$ primaries	96
4.12	Comparison of JND uniformity.	99
4.13	Comparison of hue linearity.	100
4.14	Comparison of isoluminance.	101
4.15	Comparison of desaturation.	103
4.16	Comparison of gamut mapping performance	104

4.17	Illustration of two different tonal resolutions both resulting in visually lossless quantization for the same JND ellipse.	105
4.18	Deriving the largest possible visually lossless quantization step from JND ellipsoids	106
4.19	Encoding color gamuts beyond those spanned by 3 physically realizable primaries.	109
5.1	Block diagram for the application of CAQ	112
5.2	Test pattern for the gradient quantization study.	113
5.3	Results of the gradient quantization study.	115
5.4	CAQ noise and texture estimation block diagram.	116
5.5	Calculation of the CAQ calibration LUT.	117
5.6	Setup for the image quantization study.	118
5.7	Tonal resolution prediction by HDR-VDP-2.2 and CAQ compared to the image quantization study.	120
5.8	Comparison of the predicted minimum quantization of eight example images.	122

List of Tables

2.1	Illustration of the inefficiency of linear lumiance encoding	37
2.2	Comparison of HDR and WCG color encodings for image storage	59
2.3	Comparison of HDR and WCG color encodings for image distribution.	60
3.1	Chromaticity coordinates of the Alexa Wide Gamut color space	70
4.1	Comparison of color space encoding efficiency.	107
4.2	Comparison of color space isoluminance.	108
5.1	Variable parameters of the gradient quantization study	114
5.2	Technical parameters of eight example images from the image based quantization study.	119
A.1	HdM-HDR-2014: 'HDR Test' scene	150
A.2	HdM-HDR-2014: 'Carousel Fireworks' scene	151
A.3	HdM-HDR-2014: 'Beerfest' scene	154

A.4	HdM-HDR-2014: 'Fireplace' scene.	156
A.5	HdM-HDR-2014: 'Smith' scene	157
A.6	HdM-HDR-2014: 'Fishing' scene	158
A.7	HdM-HDR-2014: 'Cars' scene	159
A.8	HdM-HDR-2014: 'Bistro' scene	161
A.9	HdM-HDR-2014: 'Poker' scene	162
A.10	HdM-HDR-2014: 'Showgirl' scene.	163

List of Abbreviations

ACES	Academy Color Encoding System. A framework for image storage, color rendering, tone mapping and gamut mapping.
ARRI	Arnold & Richter Cine Technik. A camera manufacturer.
BBC	British Broadcasting Corporation. A British public service broad- caster.
САМ	Color Appearance Model.
CAQ	C ontent A ware Q antization. A re-quantization method introduced in this thesis.
CIE	C ommission Internationale de l'Éclairage. International standards organization for light, illumination, color and color spaces.
CIE 1931 XYZ	A color space spanned by the CIE 1931 XYZ color matching functions.
СІЕ 1931 ху	CIE 1931 xy chromaticity coordinates.
CIE 1976 u'v'	CIE 1976 $\mathbf{u}'\mathbf{v}'$ chromaticity coordinates.
CIE L*a*b*	CIE 1976 $L^*a^*b^*$. A perceptually uniform color space.
CIE L*u*v*	CIE 1976 $L^*u^*v^*$. A perceptually uniform color space.

CIECAM02	A color appearance model standardized by CIE.
CMOS	Complementary Metal-Oxide-Semiconductor. Image sensor technol- ogy.
CSF	Contrast Sensitivity Function.
DICOM	Digital Imaging and Communications in Medicine standards.
DLP	Digital Light Processing. A light modulation technology.
f-stop	Field Stop . A \log_2 unit used in the domain of film and photography.
fps	Frames p er s econd. Temporal resolution of video.
HDR	High Dynamic Range.
HDR-VDP 2.2	High Dynamic Range Visual Difference Predictor Version 2.2 . A metric for visual differences in HDR images.
HdM-HDR- 2014	Cinematic HDR and WCG video data set introduced in this thesis.
HLG	Hybrid Log Gamma nonlinearity.
НМІ	Hydrargyrum Medium-arc Iodide gas discharge lamp.
IC _A C _B	HDR and WCG color difference encoding introduced in this thesis and based on IPT and PQ.
IC _T C _P	HDR and WCG color difference encoding introduced in this thesis and based on IPT and PQ.
IPT	Perceptually uniform color space named after Intensity and the P rotan and T ritan color deficiencies.
ITU	International Telecommunication Union.

JND	Just Noticeable Difference.
LabTIFF	Standard for storing CIE $L^*a^*b^*$ data in Tagged Image File Format.
LCD	Liquid Crystal Display.
LMS	Color space spanned by the linear cone photoreceptor responses sen- sitive to Long, Medium and Short wavelengths.
LogLuv	Image file format based on ${\bf Log}$ arithmic Luminance encoding and CIE 1976 ${\bf u}'{\bf v}'$ chromaticity.
LUT	Lookup Table.
NHK	Nippon Hōsō Kyōkai. Japan's national public broadcasting organiza- tion.
OLED	Organic Light-Emitting Diode.
OpenEXR	Floating point image file format introduced by Industrial Light and Magic.
PU	Perceptually Uniform transfer function.
PQ	Perceptual Quantizer transfer function.
Rec.2100 HLG	Color space using ITU recommendation 2100 primaries, HLG nonlinearity and $Y'C_BC_R$ color differencing.
Rec.2100 PQ	Color space using ITU recommendation 2100 primaries, PQ nonlinearity and $Y'C_BC_R$ color differencing.
RGB	Color space spanned by additive mixture of Red, Green and Blue light.
RGBE	File format for storing RGB color information logarithmically using a common E xponent.

SDR	Standard Dynamic Range (used as opposite to HDR).
sRGB	Standard Red, Green and Blue color space.
SVD	Singular Value Decomposition.
VDP	Visual Difference Predictor. A metric for visual differences in images.
WCG	Wide Color Gamut.
$\mathbf{Y}'\mathbf{C}_{\mathbf{B}}\mathbf{C}_{\mathbf{R}}$	A color differencing scheme starting from nonlinear encoded RGB.
Y''u''v''	An HDR and WCG color encoding based on PQ nonlinearity and CIE 1976 $u^\prime v^\prime$ chromaticity.

List of Symbols

and chroma.

R, G, B	Quantities of red, green and blue light that span a color space by addi- tive mixture.
f_{\Box}	Nonlinear encoding function. The subscript refers to the corresponding standard for nonlinear encoding. For example f_{PQ} encodes according to SMPTE ST.2084/PQ [184] transfer function.
	A superscript apostrophe added to a color channel denotes that this color is encoded nonlinearly with a gamma function. For example R' refers to a red color channel being gamma coded while R denotes a linear encoding of the red color channel.
\Box^{PQ}, \Box^{HLG}	If the nonlinear encoding is not a pure gamma function or a gamma plus linear slope function, the respective nonlinear encoding is added as superscript. For example R^{PQ} refers to a PQ encoded red color channel.
Y	Luminance in cd/m^2 .
Υ', Ι	Luma, Intensity calculated from nonlinear coded $R'G'B'$ and therefore featuring non-constant luminance, not fully decorrelating luminance

- $\square_{709}, \square_{2020}$ RGB primaries are denoted as subscript, for example R_{2020} refers to an amount of red light having the chromaticity coordinates specified in Rec.2020 [92].
- $C_{\rm B}, C_{\rm R}$, Character subscripts specify whether a chroma channel is the blueyellow color difference channel ($C_{\rm B}$) or the red-green color difference channel ($C_{\rm R}$).
- JCost functions use the letter J. As an example, the cost function on JND
ellipsoid uniformity is named J_{JND} .
- wThe scalar weights for the cost functions use the letter w. For examplethe weighting of JND-uniformity is named w_{JND} .

Abstract

This thesis introduces a cinematic High Dynamic Range (HDR) and Wide Color Gamut (WCG) data set and proposes computational models and methods for encoding HDR and WCG video imagery.

New HDR and WCG image processing algorithms, compression codecs and displays need high quality video sequences for objective and visual evaluation. Hence, a new HDR and WCG video data set containing scenic and documentary scenes with a dynamic range of up to 18 photographic stops is introduced. The individual scenes are designed to pose challenges to tone mapping operators, gamut mapping algorithms, compression codecs and HDR and WCG display devices. The scenes are staged using professional film lighting, make-up and set design. To achieve a cinematic image appearance, digital motion picture cameras with 'Super-35 mm' size sensors are used.

The extended information of HDR and WCG video requires new signal encodings, and improved color spaces compared to standard dynamic range video encodings. Due to the increasing variance in display capabilities, it is desirable to have a color signal encoding that is not only suitable for efficient quantization but also for tone mapping and gamut mapping. While methods for high dynamic range luminance encoding have been introduced, a similar encoding scheme for color difference signals is not yet available. Hence, two novel color space representations are introduced allowing for efficient encoding of HDR and WCG color difference signals as well as tone mapping and gamut mapping applications. These encodings are compared against existing state-of-the-art HDR and WCG color spaces. The introduced encoding schemes allow visually lossless quantization of any HDR and WCG video using three color channels with 12 bits of tonal resolution each.

While quantization of HDR video at 12 bits of tonal resolution is the goal, current mainstream file formats, video interfaces and compression codecs can often only handle lower bit-depths. To leverage this existing infrastructure for the transmission and storage of HDR video, a new content aware baseband quantization scheme is introduced. This quantizer exploits image characteristics like noise and texture to estimate the needed tonal resolution for visually lossless quantization per luminance range and video frame. The proposed method allows for quantization of HDR video with a tonal resolution of 10 bits without introducing visually perceivable differences and requires a lower computing power compared to current HDR visual difference metrics.

Zusammenfassung (German Abstract)

In dieser Dissertation wird ein szenischer Bewegtbilddatensatz mit erweitertem Dynamikumfang (High Dynamic Range, HDR) und großem Farbumfang (Wide Color Gamut, WCG) eingeführt und es werden Modelle zur Kodierung von HDR und WCG Bildern vorgestellt.

Die objektive und visuelle Evaluation neuer HDR und WCG Bildverarbeitungsalgorithmen, Kompressionsverfahren und Bildwiedergabegeräte erfordert einen Referenzdatensatz hoher Qualität. Daher wird ein neuer HDR- und WCG-Video-Datensatz mit einem Dynamikumfang von bis zu 18 fotografischen Blenden eingeführt. Er enthält inszenierte und dokumentarische Szenen. Die einzelnen Szenen sind konzipiert um eine Herausforderung für Tone Mapping Operatoren, Gamut Mapping Algorithmen, Kompressionscodecs und HDR und WCG Bildanzeigegeräte darzustellen. Die Szenen sind mit professionellem Licht, Maske und Filmausstattung aufgenommen. Um einen cinematischen Bildeindruck zu erhalten, werden digitale Filmkameras mit 'Super-35 mm' Sensorgröße verwendet.

Der zusätzliche Informationsgehalt von HDR- und WCG-Videosignalen erfordert im Vergleich zu Signalen mit herkömmlichem Dynamikumfang eine neue und effizientere Signalkodierung. Ein Farbraum für HDR und WCG Video sollte nicht nur effizient quantisieren, sondern wegen der unterschiedlichen Monitoreigenschaften auf der Empfängerseite auch für die Dynamik- und Farbumfangsanpassung geeignet sein. Bisher wurden Methoden für die Quantisierung von HDR Luminanzsignalen vorgeschlagen. Es fehlt jedoch noch ein entsprechendes Modell für Farbdifferenzsignale. Es werden daher zwei neue Farbräume eingeführt, die sich sowohl für die effiziente Kodierung von HDR und WCG Signalen als auch für die Dynamik- und Farbumfangsanpassung eignen. Diese Farbräume werden mit existierenden HDR und WCG Farbsignalkodierungen des aktuellen Stands der Technik verglichen. Die vorgestellten Kodierungsschemata erlauben es, HDR- und WCG-Video mittels drei Farbkanälen mit 12 Bits tonaler Auflösung zu quantisieren, ohne dass Quantisierungsartefakte sichtbar werden.

Während die Speicherung und Übertragung von HDR und WCG Video mit 12-Bit Farbtiefe pro Kanal angestrebt wird, unterstützen aktuell verbreitete Dateiformate, Videoschnittstellen und Kompressionscodecs oft nur niedrigere Bittiefen. Um diese existierende Infrastruktur für die HDR Videoübertragung und -speicherung nutzen zu können, wird ein neues bildinhaltsabhängiges Quantisierungsschema eingeführt. Diese Quantisierungsmethode nutzt Bildeigenschaften wie Rauschen und Textur um die benötigte tonale Auflösung für die visuell verlustlose Quantisierung zu schätzen. Die vorgestellte Methode erlaubt es HDR Video mit einer Bittiefe von 10 Bits ohne sichtbare Unterschiede zum Original zu quantisieren und kommt mit weniger Rechenkraft im Vergleich zu aktuellen HDR Bilddifferenzmetriken aus.

Acknowledgements

First, I would like to thank my PhD advisors Daniel Weiskopf, Andreas Schilling and Bernd Eberhardt for their continuous advice and their trust in my work. I am also grateful to Scott Daly for contributing his exhaustive knowledge in perceptually motivated image encoding literature and for his feedback proofreading our publications.

Second, I would like to thank my coworkers at Dolby Laboratories. Timo Kunkel for proofreading and introducing me to his ET_EX illustration workflow, Robin Atkins for his motivating spirit and for teaching me efficient and robust implementation of image processing algorithms, Greg Ward for the inspiring discussions during lunch, Guan-Ming Su and Peng Yin for introducing me to HDR and WCG image encoding issues in compression, Philip Krätzer for access to the Pulsar Display in Berlin, Scott Miller for sharing his work on PQ with me, Pat Griffis for giving me the strategic big picture and inventing the acronym CAQ and Shane Ruggieri for color grading my contribution to the Hollywood Post Alliance Tech Retreat conference.

I also want to thank the companies ARRI (Harald Brendel) for lending us the two ALEXA M cameras for the HDR-shoot and for pointing me towards the field of gamut mapping in the beginning of the PhD, Christie Digital (Stefan Müller) for lending me a projector for the gamut mapping study, Dolby Laboratories, (Stefan Tiefenbrunner) for lending us an HDR monitor and FilmLight (Wolfgang Lempp) for granting me a full Baselight license. My former CinePostproduction colleagues Michael Sailer and Thomas Ramin provided me with the WCG notch filter, Andreas Minuth graded the first HdM-HDR-2014 version and Daniele Siragusano tought our students NUKE and OCULA used in the HDR reconstruction pipeline. A special thank goes to my PhD peers: David Körner for taking the NumStoch course with me and for discussing color related issues in image synthesis. Matti Grüner and David Körner for the philosophical discussions during lunch and coffee at the Ökumenisches Zentrum. Stefan Reinhardt for conducting the gamut mapping study, Andreas Karge for bringing the Open Film Tools project to great success, Lena Gieseke, Sebastian Herholz and Markus Huber for the most relaxing coffee breaks and board game gatherings.

The Stuttgart Media University (HdM) was an inspiring place: I would like to especially thank Stefan Grandinetti for his outstanding cinematography, Simon Walter for his magic light and OETF-checker and Peter Ruhrmann for our discussions about the best color space for compositing and his tireless help regarding IT problems. Robin Schulte programmed and maintained the HdM-HDR-server and Peter Marquard made it possible to perform research at HdM even if the official procedures got stuck in some cases. Charles Poynton and Barabara Flückiger encouraged me to start the PhD and Katja Schmid and Peter Slansky provided a reference for my PhD application.

Most projects would not have been possible without the help of students: Jascha Vick and Corinna Kübler organized the HdM-HDR-2014 shoot in winter 2012. Ingmar Rieger characterized the Axiom Beta open source cinema camera using the Open Film Tools software, Heike Quosdorf graded the final HDR and WCG version of the HdM-HDR-2014 footage, Philipp Weber helped to prepare the Open Film Tools project with his bachelor thesis and Jonathan Sommer designed the HdM-HDR-2014 web- and SFTP-server. I want to thank all study participants of the gamut mapping study and the quantization studies as well as the students that shot the HdM-HDR-2014 footage in summer term 2013. Additional thank goes to the Silvanesti Team, especially Patrick Heinen for trying out compositing using the ACES Framework as early as in 2012.

I would not have made this way without the trust and values I have inherited from my parents Christine and Jürgen Fröhlich.

This PhD is dedicated to my wife Isabella Fröhlich.

This work was partly supported by 'Kooperatives Promotionskolleg Digital Media' at Stuttgart Media University, University of Tübingen and the University of Stuttgart Chapter 1

Introduction

Audiovisual media plays an important role in communication and culture. Watching movies or sports together, playing computer games, enjoying virtual reality or performing personal video communication, in all these domains, extending image quality is requested to support story-telling, immersion and transport of emotions.

Beyond increasing spatial and temporal resolution, image quality can be increased by capturing, encoding and rendering a higher dynamic range and a wider gamut of colors per pixel. While the fundamental research on High Dynamic Range (HDR) and Wide Color Gamut (WCG) image acquisition, manipulation, display and distribution was carried out in the late 1990s and 2000s, the recent interest in HDR in cinema, television, gaming and virtual reality has revealed new research questions specifically in efficient HDR and WCG image encoding and storage.

Figure 1.1 illustrates the image reproduction pipeline from acquisition and image synthesis to display. This imaging chain is further detailed in Section 1.1.



Figure 1.1: *The HDR and WCG distribution chain from image acquisition and image synthesis to presentation.*

1.1 Open Challenges in HDR and WCG Imaging

While there exists an extensive literature on HDR **image acquisition** techniques [44, 100, 189] and HDR **image synthesis** algorithms [158, 199], as of 2014, there was no cinematic HDR video set available. Reference video content is especially important to evaluate the performance of image manipulation and display technologies.

For **encoding and storage** of HDR and WCG content it is important to use the most efficient quantization scheme to reduce storage space and bandwidth. Optimal quantization is especially crucial when using existing 10-bit infrastructure for HDR and WCG content storage and distribution. Most prior contributions focus on the most efficient HDR grayscale encoding [10, 140, 151]. Existing HDR and WCG color encodings [17, 115, 161] are not yet fully optimized for efficient quantization and for further needs in downstream distribution like tone mapping and gamut mapping.

Video distribution also calls for efficient quantization. In addition, luminance and chroma signals have to be decorrelated for distribution to enable color subsampling and efficient compression. Future HDR and WCG signals should also be distributed in a hue linear color space to simplify gamut mapping in display devices. These needs align with the requirements for video encoding and storage as specified above.

Tone mapping and gamut mapping has to be applied to distributed HDR images, depending on the display capabilities and ambient conditions. While there is a vast literature on tone mapping [7, 51, 52, 63, 117, 133, 163, 164], gamut mapping for video display [68, 134] has come into focus upon the time of writing this thesis. Most scientific tone mapping operators are based on automatic image analysis or a small number of parameters. In the entertainment media context it would be desirable to allow artists to have the full creative control over the tone mapping and gamut mapping process to preserve artistic intent in distribution.

In **display** technology, solutions for HDR and WCG display exist for both television and cinema: Current HDR displays are either based on Liquid Crystal Display (LCD) panels using a dual-modulation backlight [174] or based on Organic Light-Emitting Diode (OLED) technology. Cinema projectors employ laser light sources to achieve a wide color gamut. HDR for cinema was introduced in 2014 using a yet undisclosed technology [47]. The biggest challenge for both television and cinema is to display a high peak luminance, while staying within a reasonable power budget.

1.2 Research Questions

The main research question of this thesis is how to efficiently encode high dynamic range and wide color gamut video.

Because in 2012 there was no **high quality HDR and WCG video data set** available, first, a reference data set has to be acquired. Therefore, the properties needed for making the video data set useful for the evaluation of new display technologies, color spaces, compression codecs and color rendering algorithms have to be found. Subsequently the video data set has to be acquired, edited and all processing steps have to be documented.

Storing such HDR and WCG video imagery with legacy formats like classic $Y'C_BC_R$ color difference encoding yields in low quantization efficiency. Thus, **the most efficient basis for encoding HDR and WCG imagery** has to be found. This includes **finding a suitable color difference scheme**, to be able to use color subsampling. Optimizing such an encoding requires the acquisition of datasets on minimum discriminable differences, hue linearity and isoluminance that span the full visible gamut and feature a high dynamic range.

Given there are efficient encodings for HDR and WCG video data, one of the challenges is using these new encodings on legacy infrastructure. This includes storage formats, transmission interfaces and compression codecs. Legacy infrastructure, originally designed for standard dynamic range video, often features a lower bit-depth than needed for a static HDR and WCG video encoding. Thus, it becomes important to determine how image properties affect the minimum needed amount of code values for quantization, and find how this can be exploited to **reduce tonal resolution needs for HDR and WCG imagery** by means of a content aware encoding.

1.3 Outline and Contributions

The aim of this thesis is to find enhanced quantization schemes for efficient HDR and WCG image storage and distribution. The models and methods described can be applied to all fields that need precise and efficient HDR and WCG color and texture representation like television and cinema, but also in virtual reality, medical imaging and product display.

The following Chapter 2 is not a contribution of this thesis as it summarizes the **his-toric contributions** on perceptually uniform luminance and color representation. It also gives an overview of the **state-of-the-art proposals for encoding HDR and WCG video**.

The original work starts with introducing a novel **HDR and WCG video data set** in Chapter 3. A reference video dataset is needed to be able to verify the quantization schemes introduced in the following chapters. This data set is specifically designed for the evaluation of HDR and WCG video encoding and processing algorithms as well as HDR displays. Chapter 3 is based on the paper describing the HdM-HDR-2014 video data set [71] where Stefan Grandinetti, Simon Walter and Jascha Vick directed the creative cinematography and lighting during the image acquisition. Heike Quosdorf operated the grading system and Stefan Grandinetti contributed to the aesthetic decisions in postproduction. The author of this thesis selected the mirror rig design used to acquire the content, determined the technical and visual requirements for the content and deduced the scenes based on these requirements. The author of this thesis also controlled the technical parameters during acquisition, designed the data-flow and processed the video data set from raw camera output to the reconstructed scene radiance and the color graded version of the video data set.

Chapter 4 presents two new HDR and WCG color encodings, IC_AC_B and IC_TC_P . These color encodings are co-optimized for coding efficiency, decorrelation of luma and chroma and for performing tone mapping and gamut mapping in these spaces. Chapter 4 is mainly based on the paper that describes the tuning of the IC_AC_B color space from psychophysics datasets [73] and also presents the IC_TC_P encoding which is based on modified LMS cone fundamentals, as originally introduced by patent PCT/US2015/051964 [69]. Co-author Robin Atkins first proposed to exchange IPT's original gamma with the Perceptual Quantizer curve, which was incorporated in both models. Co-authors Jaclyn Pytlarz and Timo Kunkel designed the isoluminance study, and co-author Jaclyn Pytlarz conducted the isoluminance study. Co-authors Jaclyn Pytlarz and Robin Atkins introduced the skintone rotation and the scaling of the P channel to IC_TC_p . The author of this thesis designed and conducted the studies on just noticeable differences and hue linearity that are used to tune both color spaces. He also designed and carried out the optimization process and first introduced the concept of crosstalk in linear Long, Medium and Short wavelength (LMS) cone response domain to increase perceptual uniformity of HDR and WCG color representations.

To be able to employ legacy infrastructure, tonal resolution needs in HDR and WCG image encoding have to be reduced beyond the static IC_AC_B and IC_TC_P encodings. Therefore in Chapter 5 a **Content Aware Quantization scheme** (CAQ) is introduced. This dynamic quantization scheme exploits the image-inherent masking properties of noise and texture for quantizing any content at reduced bit-depths, for example 10 bits. Chapter 5 is based on the paper describing the CAQ algorithm [75] and patent PCT/US2016/020230 [74] that generalizes the concept of content aware quantization. The CAQ method is inspired by co-author Scott Daly's Visual Difference Predictor [38]. The author of this thesis designed the CAQ algorithm and performed the studies on masking of quantization artifacts by noise and texture on synthetic gradients and real-world HDR and WCG video imagery.

This thesis closes by looking towards future research needed to close the remaining gaps in the HDR and WCG imaging pipeline in Chapter 6.

1.4 Reused and Copyrighted Material

Material from the following SPIE & IS&T copyrighted paper is partly reused with kind permission of SPIE & IS&T by following the author copyright transfer agreement:

• [71] Copyright 2014 Society of Photo-Optical Engineers and IS&T – The Society for Imaging Science and Technology. Reprinted, with permission, from: Jan Froehlich, Stefan Grandinetti, Bernd Eberhardt, Simon Walter, Andreas Schilling

and Harald Brendel. **Creating Cinematic Wide Gamut HDR-Video for the Evaluation of Tone Mapping Operators and HDR-displays.** *IS&T/SPIE Elec-tronic Imaging*, 90230X-90230X-10. Society for Imaging Science and Technology and International Society for Optics and Photonics, March 2014.

Material from the following IS&T copyrighted paper is partly reused with kind permission of IS&T by following the author copyright agreement:

 [73] Copyright 2015 IS&T – The Society for Imaging Science and Technology. Reprinted, with permission, from: Jan Froehlich, Timo Kunkel, Robin Atkins, Jaclyn Pytlarz, Scott Daly, Andreas Schilling and Bernd Eberhardt. Encoding Color Difference Signals for High Dynamic Range and Wide Gamut Imagery. Color and Imaging Conference, 23: 240–247. Society for Imaging Science and Technology, October 2015.

Material from the following IEEE copyrighted paper is partly reused with kind permission of IEEE by following the agreement for Thesis/Dissertation reuse:

• [75] Copyright 2016 IEEE. Reprinted, with permission, from: Jan Froehlich, Guan-Ming Su, Scott Daly, Andreas Schilling and Bernd Eberhardt. Content Aware Quantization: Requantization of High Dynamic Range Baseband Signals Based on Visual Masking by Noise and Texture. *International Conference on Image Processing (ICIP)*, 884–888 IEEE, September 2016. Chapter 2

Background and Related Work

This chapter provides an overview of models for luminance and color encoding. It starts by describing the early research on achromatic contrast sensitivity by Bouguer, Masson, Weber, Fechner, Plateau and Stevens. These spatial frequency independent models were later extended by Schade and Barten to include a dependency on the spatial frequency of the stimulus resulting in Contrast Sensitivity Functions (CSF). Furthermore, currently used luminance encodings for Standard Dynamic Range (SDR) and HDR video are summarized.

In the mid section of this chapter the focus switches from luminance to color. First Wright's and MacAdam's fundamental studies on color discriminability and the derived chromaticity scales are summarized. Furthermore, current SDR color difference encodings as well as the early color appearance models are described.

The last section of this chapter gives an overview of the current state-of-the-art luminance and color encodings that are specifically tailored for efficient encoding of HDR and WCG video imagery. These encodings build upon the concepts introduced in the beginning of this chapter.

2.1 Early Research on Perceptual Uniformity and Contrast Sensitivity

One of the earliest experiments to find the smallest luminance steps the human visual system can discriminate were done by Bouguer, one of the originators of the domain of photometry. Bouguer discovered a logarithmic relationship between physical stimuli and minimum discriminable steps [21, 22]. His experimental setup was as follows: he illuminated a white paper by candle light, added a second candle and a stick that cast a shadow by partly covering the second candle. By moving the second candle further away he could reduce the light added to the base stimulus from the first candle. Independent of the base luminance, the difference between the area where the second candle adds light and the area where the stick covered the second candle could just be discriminated if the distance of the second candle to the paper was 8 times the distance of the first candle to the paper. Thus, he concluded the minimum fraction that could be detected was $\frac{1}{64}$ of the original stimulus. This fraction is often called Just Noticable Difference (JND).

Roughly a century later, Masson [135] tried to reproduce Bouguer's results with a slightly different apparatus. He darkened a defined part of a disk that rotated with 200 frames per second. Having this speed, the black part of the disk and the white parts melt together temporally. From his observations on contrast visibility using the rotating disks he draws three conclusions:

- The sensitivity of each individual's eye does only deviate minimally over the course of multiple days.
- Every subject could see differences of $\frac{1}{60}$ in luminance whereas some could detect up to $\frac{1}{120}$.
- When written as a fraction of the original stimulus, the eye's contrast sensitivity is independent of the intensity of the original stimulus or the color of the light as long as the intensity is bright enough to read text clearly.

Bouguer's and Massons's photometry paved the way for the field of psychophysics, a term coined by Fechner. Psychophysics aims to define the relationship between a physical stimulus and the perceived response in human perception. Fechner refers to Bouguer and Masson in his fundamental book on human perception [64]. His goal is to derive psychophysical scales for physical stimuli like temperature, tactile sensation and luminance. He starts from 'Weber's law' named after his teacher's, Weber's research on tactile sensations [202]. Weber's law states that the minimum discriminable difference is a constant fraction *c* of the change in physical stimulus ΔR divided by the original physical stimulus *R*:

$$c = \frac{\Delta R}{R} \tag{2.1}$$

Fechner further integrates Weber's law to derive a psychophysical scale γ , where β is the physical stimulus and *b* is the lower threshold for detection:

$$\gamma = k(\log\beta - \log b) = k\log\frac{\beta}{b}$$
(2.2)

For human vision Fechner reports values around k = 100 from his own experiments that follow Bouguer's and Massons's setups. Fechner admits that the constant fraction model does not hold for either very dark stimuli or very bright stimuli like sunspots, which are visible through a smoked glass, but become invisible when looking directly at the sun.

A different model for creating a perceptually uniform scale was proposed by Plateau [159, 160] in 1872 when he published the results of his experiments performed already in 1830. He found the psychophysical scale S not to be a logarithm of the original stimulus but to roughly follow a power law, expressed as a constant A times the physical stimulus E raised by the power of d:

$$S = AE^d \tag{2.3}$$

This function later got known as 'Stevens' Law' [187] named after Stevens, who brought Plateau's findings to a wider audience. The question if lightness perception follows a logarithmic function or power law is still a current research topic [28]. For the discrimination of small differences, there is agreement that a curve of additive JND steps roughly follows a power function for dark stimuli limited through photon shot noise and then turns to a logarithmic relationship in photopic viewing conditions before the curve levels off at very high luminance [14, 140]. The perceptually uniform encoding curves described in Section 2.3.3 will all model this behavior and the masking phenomenon of photon shot noise will be further discussed in Chapter 5.

2.2 Spatial Contrast Sensitivity

All approaches for determining JNDs mentioned in Chapter 2.1 use step edges for discrimination. Schade began to measure visual contrast sensitivity as a function of spatial frequency for modeling human vision by means of a television like system [172]. He found that the minimum discriminable luminance difference is heavily dependent upon spatial frequency. Figure 2.1 illustrates this phenomenon at a glance. Moving from top to bottom, the sine pattern's amplitude is constantly reduced. On the horizontal axis different spatial frequencies from low frequencies at the left to higher frequencies at the right are rendered all with the same amplitude for one vertical position. For a typical reading distance only the mid frequencies have a contrast visibility around 0.1 in CIE L*a*b* L*. Contrast sensitivity of the human visual system is reduced for higher and lower spatial frequencies.



Figure 2.1: Visualization of contrast sensitivity depending on spatial frequency. Preferably watch this illustration on a high spatial and tonal resolution sRGB [87] screen.

In the 1990s Barten developed a formula to predict the contrast sensitivity function of the human visual system based on a physical model of the eye [13, 14, 15]. Barten's model allows one to predict contrast detection for a very wide range of stimuli and viewing environments. He showed that his model aligns well with a large number of studies and data sets [14]. It was used to obtain the **P**erceptual **Q**uantizer function (PQ) [184] referenced in Section 2.3.3.3, which is used as the nonlinearity in the HDR and WCG video encoding models introduced in Chapter 4.
2.3 Luminance Encodings

Following the findings from Chapters 2.1 and 2.2 that luminance perception is highly nonlinear, it becomes clear that a linear digital encoding of luminance would not be very efficient. Table 2.1 shows an example for the inefficiency of linear luminance coding. Assuming a 1% Weber-Fechner fraction as detection limit, a 10-bit encoding can only encode three f-stops of dynamic range without introducing quantization artifacts. Even a 16-bit linear encoding is only visually transparent for the top nine f-stops. For the tenth f-stop below peak white the step-size in quantization becomes more than 1%, thus limiting the visually transparent dynamic range for this encoding to about 900:1 (65353/70). At the same time, the first stop below peak white gets encoded with 32768 code values, whereas 70 code values would be enough assuming a 1% detection limit. This is an over-quantization of more than two orders of magnitude for the brightest f-stop and therefore very inefficient. Hence, different nonlinear transfer functions for luminance have been proposed and are summarized in the following sections.

Table 2.1: Illustration of the inefficiency of linear luminance encoding. Fechnerner refers to the number of code values needed assuming a detection limit of1% linear luminance.10 bits and16 bits show the number of code valuesspent to quantize each f-stop when using a linear integer encoding.

cone values per l'scop										
F-Stop	1	2	3	4	5	6	7	8	9	10
Fechner	70	70	70	70	70	70	70	70	70	70
10-bit	512	256	128	64	32	16	8	4	2	1
16-bit	32768	16384	8192	4096	2048	1023	512	256	128	64

Code values per f-stop

Traditional standard dynamic range Rec.709 gamma can be thought as following Stevens Law. The logarithmic quantization function of LogLuv can be thought of as an implementation of Weber's and Fechner's research. The most recent HDR quantization schemes like Digital Imaging and Communications in Medicine (DICOM), Perceptually Uniform (PU), Perceptual Quantizer and Hybrid Log Gamma (HLG) build upon Barten's research and try to bridge both theories, Stevens' power law and the Weber-Fechner logarithm by roughly following a gamma function near black and then changing into a more logarithmic encoding for brighter stimuli.

2.3.1 Standard Dynamic Range Encodings

Most image encodings for standard dynamic range imagery use a gamma curve to be able to encode luminance efficiently. The gamma curve was originally given by the native cathode ray tube nonlinearity and follows 'Stevens Law' [187]. The sRGB [87] and Rec.709 [94] transfer functions both have a linear toe to reduce the amplitude of high frequency photon shot noise and sensor read noise near black as it typically appears in camera captured images:

$$f_{sRGB}(x) = \begin{cases} 1.055x^{\frac{1}{2.4}} - 0.055, & \text{if } 1 \ge x \ge 0.0031308\\ 12.92x, & \text{if } 0.0031308 > x \ge 0 \end{cases}$$

$$f_{709}(x) = \begin{cases} 1.099x^{0.45} - 0.099, & \text{if } 1 \ge x \ge 0.018\\ 4.5x, & \text{if } 0.018 > x \ge 0 \end{cases}$$

$$(2.4)$$

Instead of limiting quantization towards black by a linear toe, ITU Rec.1886 [90] introduces a small negative offset b to again limit quantization near black. The scalar factor a serves to normalize the signal to its intended range of zero to one again.

$$f_{1886}(x) = \max\left[\left(\frac{x}{a}\right)^{\frac{1}{2.4}} - b, 0\right]$$
(2.6)

Digital cinema [181] uses a pure gamma of 2.6 to increase the tonal resolution near black.

$$f_{\rm DCI}(x) = x^{1/2.6} \tag{2.7}$$

The higher tonal resolution near black is needed in digital cinema because the dark viewing environment leads to darker adaptation and therefore a higher tonal resolution of the human visual system in cinema environments. The higher amplification of noise near black for camera captured content is acceptable because the compression bit-rates of digital cinema [45] are higher compared to television applications. Tonal resolution in digital cinema is 12 bits opposed to 8 and 10 bits in broadcasting. Thus, even smooth and noiseless gradients like they appear in broadcast graphics and computer generated content can be quantized at 12-bit without introducing visible quantization artifacts in typical cinema environments [36].

2.3.2 HDR Luminance Encoding in Video Acquisition

Digital image sensor output from modern sensors is typically delivered in a linear encoding by the analog-to-digital converters. Hence, as illustrated in Table 2.1, the relative tonal resolution varies greatly over luminance.

Following Fechner's logarithmic model of vision, but deriving a minimum contrast sensitivity of $\frac{1}{300}$ from Barten's spatial contrast sensitivity model introduced in Section 2.2 instead of Fechner's 1% findings, 210 code values per f-stop are sufficient for the most critical visual imaging use cases. Thus, it is common to limit raw camera files to a certain number of code values per f-stop via a quantization lookup table. This estimate assumes a limited contrast manipulation in later processing steps. The number of code values per f-stop is typically reduced for the lower part of the quantization curve because of the rapid decrease of tonal resolution from the sensor's linear analog to digital converter. In practice, the reduced tonal resolution from the analog digital conversion for the dark areas is acceptable because of the masking of quantization artifacts by photon shot noise as further discussed in Chapter 5.

A typical example for a logarithmic quantization scheme is the 'LogC' curve [24]. At 12-bit tonal resolution 'LogC' limits the quantization to roughly 256 steps per stop for the midtones and highlights. In the shadows it mimics the toe of analog film comparable to scanned motion picture negative [102]:

$$f_{\text{LogC}}(x) = \begin{cases} c \log_{10}(ax+b) + d, & \text{if } 1 \ge x > 0.010591 \\ ex+f, & \text{if } 0.010591 \ge x \ge 0 \end{cases}$$
(2.8)
$$a = 5.555556 \qquad b = 0.052272 \qquad c = 0.247190 \\ d = 0.385537 \qquad e = 5.367655 \qquad f = 0.092809 \end{cases}$$

Lots of similar pseudo-logarithmic encoding functions are available. Most roughly follow the original Cineon [102] curve. A good place to search for nonlinear encodings of individual camera manufacturers is the 'Colour-Science' project [130].

2.3.3 HDR Luminance Encoding in Video Distribution

In high dynamic range imaging the nonlinear encoding becomes more crucial compared to standard dynamic range because HDR allows for a wide range of adaptation states for the human visual system. Viewers can be dark adapted in a cinema after having watched a couple of dark scenes or be adapted to very bright outdoor viewing environments for mobile devices showing bright content in sunlight. Hence, new transfer functions have to be developed for these new use cases. The first HDR transfer functions were standardized in the domain of medical imaging where analog high density films like those used in x-ray had already offered HDR display before the introduction of digital imagery.

2.3.3.1 DICOM Grayscale Standard Display Function

In 1979 Briggs proposed to derive the best display quantization by adding up JNDs [26]. Johnston et al. advised to apply a similar method to increase interoperability between different medical display devices [98]. Blume et al. [81] extended this approach to be based on Barten's contrast sensitivity function instead of JND visibility experiments [13, 15, 81]. This led to the standardization of the DICOM 'Grayscale Standard Display Function' [150] with the following functional form fitted to the numerical JND data:

$$f_{\text{DICOM}}(x) = A + B \log_{10}(x) + C \log_{10}(x)^2 + D \log_{10}(x)^3 + E \log_{10}(x)^4 +$$
(2.9)

$$F \log_{10}(x)^5 + G \log_{10}(x)^6 + H \log_{10}(x)^7 + I \log_{10}(x)^8$$

$$A = 71.498068 \qquad B = 94.593053 \qquad C = 41.912053$$

$$D = 9.8247004 \qquad E = 0.28175407 \qquad F = -1.1878455$$

$$G = -0.180143493 \qquad H = 0.14710899 \qquad I = -0.017046845$$

The DICOM encoding curve is calculated by adding up the smallest discriminable steps at a frequency of 4 cycles per degree according to the Barten model. Figure 2.2 illustrates this approach. While the contrast sensitivity at 1000 cd/m^2 is between $\frac{1}{200}$ and $\frac{1}{300}$, stimuli at 0.01 cd/m^2 get encoded with a relative step size of about $\frac{1}{7}$. However, at 0.01 cd/m^2 the human visual system features a higher contrast sensitivity of around $\frac{1}{20}$ at lower frequencies below 1 cycle per degree.



Figure 2.2: Derivation of the DICOM, PU and PQ luminance transfer functions from Barten's CSF model. Adapted from [140].

2.3.3.2 PU Transfer Function

In 2006, Mantiuk et al. proposed to encode HDR imagery by using the DICOM curve [132]. Later, Aydın et al. extended the DICOM approach by following the frequencies with the highest respective peak contrast sensitivity [10] in Barten's CSF model. The resulting Perceptually Uniform (PU) curve is designed to prevent visible quantization artifacts at any spatial frequency and any luminance level. The derivation of PU is illustrated in Figure 2.2. While DICOM always samples the smallest detectable contrast step at 4 cycles per degree, PU follows the peak of contrast sensitivity to lower frequencies for darker luminance values.

2.3.3.3 PQ Transfer Function

The PU curve is supplied as a lookup table [10, 11]. For interchange and application in arbitrary precision, it is desirable to have a functional approximation available for HDR encoding. In consequence, Miller et al. fitted a function to the highest contrast sensitivity per luminance range of Barten's spatial contrast sensitivity function [140, 151]. This curve is called **P**erceptual **Q**uantizer (PQ) and is standardized by SMPTE as the standard quantization curve for HDR video distribution in ST.2084 [184].

The functional form of the PQ curve loosely follows the Naka-Rushton equation [146]. x is the linear input in candela per square meter and $f_{PQ}(x)$ provides the nonlinear encoded signal. n, m and c_1, c_2, c_3 are constants derived from fitting the Naka-Rushton equation to the numerical JND data:

$$f_{\rm PQ}(x) = \left(\frac{c_1 + c_2 \times \left(\frac{x}{10000}\right)^n}{1 + c_3 \times \left(\frac{x}{10000}\right)^n}\right)^m$$
(2.10)
$$n = 2610/2^{14} \quad m = 2523/2^5 \quad c_1 = 3424/2^{12} \quad c_2 = 2413/2^7 \quad c_3 = 2392/2^7$$

Figure 2.3 shows the DICOM nonlinearity, PU, PQ and sRGB in comparison. It can be seen that the PU authors have decided to clip all values below about 10^{-2} to stay near the sRGB curve for the 0-100 cd/m² range while PQ is tailored towards current generation displays that often feature darker black values below 10^{-2} . These dark values become especially important in HDR cinema environments. As an example, night scenes are typically color-graded much darker for cinema compared to television because in cinema environments the projected image does not have to compete with potential stray-light and glare from indoor light sources and windows like they are typical in private homes.



Figure 2.3: DICOM, PU, PQ and sRGB luminance transfer functions compared.

2.3.3.4 HLG Transfer Function

When being faced with the challenge to carry out live broadcasting using the PQ transfer function the Japan Broadcasting Corporation (NHK) and the **B**ritish **B**roadcasting **C**orporation (BBC) identified several practical disadvantages:

- When PQ-encoded signals are viewed on legacy SDR monitors, they look low contrast, desaturated and the dark to mid-tones are elevated.
- The high tonal resolution of PQ in the dark areas results in an amplification of camera noise for compression. Hence, non-PQ-aware codecs like H.264 do not assign enough bits to bright areas compared to the dark areas resulting in a reduced compression efficiency.
- The absolute encoding of PQ requires explicit tone mapping on the receiver side for those displays that are not capable of displaying the exact peak luminance of the content.

To solve these problems, the NHK and BBC introduced a new transfer function that follows the findings of both Stevens and Fechner [8]. Stevens' law is applied to dark stimuli where the minimum discriminable steps in human vision roughly follow a power function. For brighter colors, a logarithm is applied as postulated by Fechner. Hence, the HLG nonlinearity is defined piecewise with a gamma of 0.5 for values below diffuse white and a natural logarithm for encoding colors above diffuse white as depicted in Equation (2.11). The linear scene luminance x is normalized to a range of zero to one. Diffuse white is intended to be rendered at $\frac{1}{12}$ of the peak white of the linear input signal resulting in about 3.5 f-stops of dynamic range for specular highlights and self luminous objects that are brighter than diffuse white.

$$f_{\text{HLG}}(x) = \begin{cases} \sqrt{3x}, & \text{if } 0 \le x \le \frac{1}{12} \\ a \ln(12x - b) + c, & \text{if } \frac{1}{12} < x \le 1 \end{cases}$$

$$a = 0.17883277 \quad b = 1 - 4a \quad c = 0.5 - a \ln(4a)$$
(2.11)

Compared to PQ, HLG spends less code values on the darker parts of the tone-scale when encoded with the same bit depth. This results in possible quantization artifacts for the rare case of noiseless and textureless content in dark viewing environments. However, HLG has better backward compatibility to legacy SDR monitors and improved performance with legacy compression codecs. Due to the relative encoding, there is no need for tone mapping HLG signals at the receiver side for typical peak luminances of current HDR television sets between 500 and 2000 cd/m^2 .

2.3.3.5 Comparison of HDR Transfer Functions

Figure 2.4 illustrates the bit-savings of PQ compared to a logarithmic or gamma encoding and the HLG transfer function. PQ closely follows the minimum discriminable



Figure 2.4: Needed tonal resolution of the transfer functions PQ, floatingpoint, logarithmic, gamma and HLG to stay below the visibility threshold according to Barten's CSF model. HLG assumes a black level of 0.01 cd/m², logarithmic and gamma encodings incorporate a black level of 0.0001 cd/m². Adapted from [140] and extend with HLG and minimum black level.

steps from the Barten (ramp) threshold. The Barten (ramp) limit describes the discrimiation limit according to Barten for step-edge stimuli as they occur in the quantization of gradients (ramps). The threshold visibility for step edges is lower compared to the sine pattern described by Barten's spatial CSF model [36]. The Schreiber limit is a model introduced by Schreiber and described in ITU-R Report BT.2246-5 [93]. 12-bit PQ provides a visually lossless encoding for any luminance value between 0 and 10000 cd/m^2 while a logarithmic encoding would need 13 bits and a gamma 3.0 curve would require 14 bits to quantize the same dynamic range without causing visible artifacts. HLG also needs 14 bits to stay below the Barten (ramp) limit. Half-float encoding is included for informational purposes only.

When using a tonal resolution of 12 bits for the PQ, logarithmic, gamma 3.0 and HLG transfer functions as illustrated in Figure 2.4, only PQ can quantize any content without causing artifacts above the visual detection theshold.



Figure 2.5: *PQ, logarithmic, gamma and HLG transfer functions at 12-bit tonal resolution compared to Barten's CSF model. HLG assumes a black level of 0.01 \text{ cd/m}^2, logarithmic and gamma encodings incorporate a black level of 0.0001 \text{ cd/m}^2. Adapted from [140] and extend with HLG and minimum black level.*

2.4 Color Difference Encodings

The current colorimetric reference system started with the definition of the CIE 1931 XYZ color matching functions in 1931. Fairman et al. [27, 62] provide an overview of how the CIE 1931 XYZ color matching functions were derived from the color matching experiments of Guild [77] and Wright [203].

The CIE 1931 XYZ color system provides a common basis to compare minimum discriminable steps in chromaticity. Judd [99] and MacAdam [129] first measured color JND ellipses to derive more constant chromaticity spaces. CIE 1976 u'v' illustrated in Figure 2.6 is a linear transformation of CIE 1931 xy enhancing perceptual uniformity by expanding the blue areas and compressing the green areas.



Figure 2.6: *Minimum discriminable differences in CIE 1931 xy and CIE 1976 u'v' chromaticity space. The MacAdam ellipses [129] for viewer PGN are shown 10 times amplified.*

The formula for converting from CIE 1931 xy to CIE 1976 u'v' is given in Equation (2.12) and (2.13):

$$u' = \frac{4x}{-2x + 12x + 3} \tag{2.12}$$

$$v' = \frac{9y}{-2x + 12y + 3} \tag{2.13}$$

2.4.1 Spatial Color Contrast Sensitivity

In 1984 Mullen extended MacAdam's, Judd's and Wright's measurements that are based on step edges to incorporate spatial frequency. The resulting chromatic con-



Figure 2.7: Visualization of color contrast sensitivity depending on spatial frequency. Preferably watch this illustration on a high spatial and tonal resolution sRGB screen [87].

trast sensitivity functions predict the minimum discriminable step depending on spatial frequency [144]. Figure 2.7 illustrates the dependency of color contrast sensitivity on spatial frequency. As in the visualization of the achromatic CSF in Figure 2.1 the signal amplitude is kept constant for the same position on the amplitude scale. Mullen's chromatic CSF measurements are extended by Kim et al. to higher luminance ranges as they are typical for HDR imaging [103].

2.4.2 Color Appearance Models

Measuring JNDs as described in the preceding section and designing color spaces that are uniform on a visibility threshold level results in color spaces that are optimal for efficient encoding. But adding up JNDs does not automatically result in perceptual uniformity for larger supra threshold differences. A color spaces designed to be perceptually uniform on a supra threshold level is called **C**olor **A**ppearance **M**odel (CAM). 'Color appearance models aim to extend basic colorimetry to specify the perceived color of stimuli in a wide variety of viewing conditions' [61, page 1]. Thus, color appearance models are not primarily optimized for uniformity at the detection level but rather aim to link the appearance dimensions lightness, hue and saturation to the photometric CIE 1931 XYZ color matching functions assuming a certain set of adaptation parameters for the human visual system.

2.4.2.1 CIE L*a*b*

One of the first and most used color appearance models is called CIE L*a*b* [89]. CIE L*a*b* first applies a nonlinearity f_{Lab} :

$$f_{\text{Lab}}(x) = \begin{cases} x^{\frac{1}{3}}, & \text{if } x > \left(\frac{6}{29}\right)^3 \\ x\frac{841}{108} + \frac{4}{29}, & \text{if } x \le \left(\frac{6}{29}\right)^3 \end{cases}$$
(2.14)

to the CIE 1931 XYZ signals scaled by the white point $X_n Y_n Z_n$ ('wrong von Kries' adaptation). The f_{Lab} nonlinearity simulates the nonlinear response function of the human visual system and can be considered as an implementation of Stevens Law as described in Section 2.1. Whereas L^* is a direct result of applying the CIE L*a*b* nonlinearity to Y:

$$L^* = 116 f_{\text{Lab}}(Y/Y_n) - 16 \tag{2.15}$$

 a^* and b^* are calculated by taking the difference between nonlinearly coded Y and X or Z respectively, simulating the color opponent processing in the retina:

$$a^* = 500 \left(f_{\text{Lab}}(X/X_n) - f_{\text{Lab}}(Y/Y_n) \right)$$
(2.16)

$$b^* = 200 \left(f_{\text{Lab}}(Y/Y_n) - f_{\text{Lab}}(Z/Z_n) \right)$$
(2.17)

Chroma C_{ab}^* is calculated as the Euclidean distance from the achromatic axis and hue h_{ab} is defined as the polar angle around the achromatic axis:

$$C_{ab}^* = \sqrt{a^{*2} + b^{*2}} \tag{2.18}$$

$$h_{ab} = \operatorname{atan2}\left(\frac{b^*}{a^*}\right) \tag{2.19}$$

CIE L*a*b* performs respectably well in terms of perceptual uniformity for standard dynamic range imagery and for gamut mapping in printing applications. However, CIE L*a*b* is not as hue linear compared to modern color appearance models. Figure 2.8 shows a linear desaturation of the blue primary color of sRGB in CIE L*a*b* and a newer color appearance model - CIECAM02. When desaturating blue in CIE L*a*b* as illustrated in Figure 2.8 a) the hue changes towards pink when being mapped on a straight line in the direction of the achromatic axis [68]. Figure 2.8 b) shows that CIECAM02 does not suffer from this issue.



Figure 2.8: Example for CIE L*a*b* hue linearity compared to CIECAM02. Best viewed on a color accurate sRGB monitor in a dark environment.

The most common mapping of CIE L*a*b* to digital code values for image encoding is specified in the LabTIFF documentation [2]. For LabTIFF the 0 to 100 range of L^* is scaled to the full 0 to 255 range of 8-bit unsigned integers:

$$L_{\rm LabTIFF8}^* = \lfloor 255 \frac{L^*}{100} \rfloor$$
 (2.20)

The respective scaling is performed for 16-bit encodings:

$$L_{\text{LabTIFF16}}^* = \lfloor 65535 \frac{L^*}{100} \rfloor$$
 (2.21)

For 8-bit encodings the chroma channels a^* and b^* are quantized and clamped to the -127 to 128 range of signed 8-bit integers without any scaling:

$$a_{\text{labTIFF8}}^* = \lfloor a^* \rfloor \tag{2.22}$$

$$b_{\text{LabTIFF8}}^* = \lfloor b^* \rfloor \tag{2.23}$$

For 16-bit encodings a* and b* are multiplied by 256 before clamping and quantization to the -32768 to 32767 range of signed 16-bit integers:

$$a_{\text{labTIFF16}}^* = \lfloor 256a^* \rfloor \tag{2.24}$$

$$b_{\text{LabTIFF16}}^* = \lfloor 256b^* \rfloor \tag{2.25}$$

The LabTIFF encoding is limited to standard dynamic range and does not cover wide gamut color spaces like Rec.2020 [92]. This is by design because LabTIFF is only intended for use in printing applications [2, page 12]: 'Limiting the theoretically unbounded a* and b* ranges to +/- 127 allows encoding in 8 bits without eliminating any but the most saturated self-luminous colors'. The full bit order for LabTIFF is shown in Figure 2.9.

LabTIFF8	Value	LabTIFF16	Value
L* <u>1 2 3 4 5</u>	5 6 7 8	L* <u>123456</u>	7 8 9 10 11 12 13 14 15 16
Sign	Value	Sign	Value
a* <mark>\+/</mark> 1234	5 6 7	a* <u>\+/</u> 12345	6 7 8 9 10 11 12 13 14 15
Sign	Value	Sign	Value
b* [+/-] 1 2 3 4	5 6 7	b* +/- 1 2 3 4 5	6 7 8 9 10 11 12 13 14 15

Figure 2.9: Bit order of the LabTIFF encoding.

2.4.2.2 IPT

IPT was developed by Ebner to address the limited hue linearity of CIE L*a*b* [54]. Instead of applying the nonlinearity on white point adapted CIE 1931 XYZ, IPT first transforms CIE 1931 XYZ values to the LMS cone response domain:

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.4002 & 0.7075 & -0.0807 \\ -0.2280 & 1.1500 & 0.0612 \\ 0.0000 & 0.0000 & 0.9184 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$
(2.26)

and then applies a gamma 0.43 nonlinearity to LMS. This nonlinearity can again be thought as an application of Stevens Law as described in Section 2.1. The 0.43 gamma is the result of an optimization process to generate a most hue linear chroma plane

when using the LMS cone responses as primaries [54]:

$$f_{\rm IPT}(x) = \begin{cases} x^{0.43}, & \text{if } x \ge 0\\ -[(-x)^{0.43}], & \text{if } x < 0 \end{cases}$$
(2.27)

Subsequently, the nonlinear LMS signals are decorrelated into I for Intensity and the color difference channels P and T named after the corresponding color deficiencies in human vision, **P**rotanope and **T**ritanope:

$$\begin{bmatrix} I \\ P \\ T \end{bmatrix} = \begin{bmatrix} 0.4000 & 0.4000 & 0.2000 \\ 4.4550 & -4.8510 & 0.3960 \\ 0.8056 & 0.3572 & -1.1628 \end{bmatrix} \begin{bmatrix} f_{\rm IPT}(L) \\ f_{\rm IPT}(M) \\ f_{\rm IPT}(S) \end{bmatrix}$$
(2.28)

As with CIE L*a*b* chroma C_{PT} can be calculated as the Euclidean distance from the achromatic axis and hue h_{PT} is defined as the polar angle around the achromatic axis:

$$C_{\rm PT} = \sqrt{P^2 + T^2}$$
 (2.29)

$$h_{\rm PT} = {\rm atan} 2 \left(\frac{P}{T}\right) \tag{2.30}$$

IPT is often used for gamut mapping in print applications [143] and was found to be a good candidate for gamut mapping in SDR WCG video applications [68]. To date IPT has not been used as an image storage format.

2.4.2.3 Extended Color Appearance Models

There are newer color appearance models beyond CIE L*a*b* and IPT like CIECAM02 [142]. Fairchild's book on color appearance models [61] contains an overview on recent developments in this area. Unfortunately, the computational complexity of the newer color appearance models is higher compared to CIE L*a*b* and IPT. Also these newer color appearance models typically rely on additional information about the viewing environment and the observer. Therefore, models like CIECAM02 are not suited for general purpose image encoding.

2.4.3 Standard Dynamic Range Video Color Encodings

Typically, video signals are not distributed or stored as nonlinear R'G'B' but they are further decorrelated into one luma and two chroma channels. The term 'luma' is used when the achromatic color channel is calculated from nonlinear R'G'B'. This method for decorrelation of chroma and luma was introduced by Valensi [192] in 1938 for television applications. Figure 2.10 shows an example image as scatterplot. It can be observed that the correlation between R', G' and B' is stronger compared to correlation between Y', C_B and C_R. Having chroma separated from luma facilitates transmitting

Example Image





Figure 2.10: *Example for decorrelation by color difference encoding. Photo: Patrick J. Palmer*

color difference channels at lower resolutions to exploit the lower peak contrast sensitivity of the human visual system for color as measured by Mullen [144]. This concept of subsampling color was introduced by Bedford [16] in the 1950s. Decorrelation also helps in compression by reducing energy in the chroma channels for typical imagery.

High definition television systems in use today still perform $Y'C_BC_R$ color difference coding [94] for subsampling and video compression. Traditional $Y'C_BC_R$ signals start by first transforming to RGB signals relative to the Rec.709 primaries [94]:

$$\begin{bmatrix} R_{709} \\ G_{709} \\ B_{709} \end{bmatrix} = \frac{1}{Y_{\text{white}}} \begin{bmatrix} 3.240969942 & -0.969243636 & 0.055630080 \\ -1.537383178 & 1.875967502 & -0.203976959 \\ -0.498610760 & 0.041555057 & 1.056971514 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$
(2.31)

Then the Rec.709 nonlinearity introduced in Equation (2.5) is applied to RGB:

$$R'_{709} = f_{709}(R_{709}) \qquad G'_{709} = f_{709}(G_{709}) \qquad B'_{709} = f_{709}(B_{709})$$
(2.32)

Further, luma Y' is calculated by a mix of R', G' and B' weighted by the contributions of R, G and B to CIE 1931 luminance:

$$Y'_{709} = 0.2126R'_{709} + 0.7152G'_{709} + 0.0722B'_{709}$$
(2.33)

This is followed by calculating the two color difference channels via subtracting luma from the nonlinearly coded blue (B') and red (R') signals:

$$C'_{\rm B\,_{709}} = \frac{B'_{709} - Y'_{709}}{1.8556} \tag{2.34}$$

$$C_{\rm R\,709}' = \frac{R_{709}' - Y_{709}'}{1.5748} \tag{2.35}$$

These steps can be thought as modeling the color opponent processing of the human visual system. Equations (2.33 to 2.35) can also be written as a linear transformation:

$$\begin{bmatrix} Y'_{709} \\ C'_{B709} \\ C'_{R709} \end{bmatrix} = \begin{bmatrix} 0.2126 & 0.7152 & 0.0722 \\ \frac{-0.2126}{1.8556} & \frac{-0.7152}{1.8556} & \frac{1-0.0722}{1.8556} \\ \frac{1-0.2126}{1.5748} & \frac{-0.7152}{1.5748} & \frac{-0.0722}{1.5748} \end{bmatrix} \begin{bmatrix} R'_{709} \\ G'_{709} \\ B'_{709} \end{bmatrix}$$
(2.36)

Classic Y'C_BC_R is used in almost all bandwidth limited imaging chains, from digital cinema [186] via television [94] to photography [88]. Classic Y'C_BC_R has been extended to support a wider color gamut in the Rec.2020 television standard [92]. Although Rec.2020 supports a wide color gamut, it is still only meant for encoding standard dynamic range imagery.

2.4.4 High Dynamic Range Color Encodings

The very early HDR color encodings originate from the domain of image synthesis, specifically physically based rendering algorithms, where an HDR representation of computer generated images is inherent to the process of image synthesis. Radiance is often quoted to be the first physically based renderer [199]. As most render engines were based on RGB tristimulus colors so are the early HDR color encodings.

2.4.4.1 RGBE

The external picture storage format for the Radiance renderer is named 'pic' from the file extension '.pic' or often called RGBE [198] for Red, Green, Blue and Exponent. For the RGBE format all three R, G, B color channels share one common exponent E that is determined by the maximum value of R, G and B:

$$E = \lceil \log_2(\max(R, G, B)) \rceil + 128$$

$$R = \lfloor 256R/2^{(E-128)} \rfloor \qquad G = \lfloor 256G/2^{(E-128)} \rfloor \qquad B = \lfloor 256B/2^{(E-128)} \rfloor \qquad (2.38)$$

The final storage format is illustrated in Figure 2.11.



Figure 2.11: Bit order of the RGBE encoding.

2.4.4.2 OpenEXR

In current generation, non spectral render engines, colors are typically stored as three floating point numbers representing a mixture of red, green and blue light of a well defined spectral power distribution. There is no need for a nonlinear encoding curve in OpenEXR because floating point numbers are inherently of variable absolute precision. As an example the 10-bit mantissa of IEEE 754-2008 half-float [86] quantizes with a tonal resolution of 1024 steps per f-stop which is below the 0.3% maximum contrast sensitivity of the human visual system according to Barten's spatial CSF model [14]. See Figure 2.4 in Section 2.3.3.5 for an illustration of the precision of floating point encoding compared to other HDR transfer functions.

OpenEXR was developed by Kainz at ILM in 1999 as a storage format for half-float, single-precision and 32-bit integer images to avoid rounding errors and save additional computations when storing images as files [1]. In the further readings the term 'OpenEXR' will refer to the most used flavor, storing red, green and blue color channels in IEEE 754-2008 half-float format [86]. The bit order of half-float OpenEXR is shown in Figure 2.12.

Sign Exponent Red [+/-]1]2]3]4]5]1]2]3]4	Mantissa 5 6 7 8 9 10
Sign Exponent Green +/- 1 2 3 4 5 1 2 3 4	Mantissa 5 6 7 8 9 10
Sign Exponent Blue [+/-]1 2 3 4 5 1 2 3 4	Mantissa 5 6 7 8 9 10

Figure 2.12: Bit order of the OpenEXR half-float encoding.

2.4.4.3 LogLuv

LogLuv was introduced by Ward in 1998 to increase the efficiency of 32 bits per pixel image storage compared to RGBE [115, 116]. LogLuv stores the luminance of an image measured in cd/m^2 via a logarithmic encoding in a 15-bit integer representation

plus one sign bit:

$$Se = sign(Y)$$
 (2.39)

$$Le = \begin{cases} \lfloor 256(\log_2(|Y|) + 64) \rfloor, & \text{if } Y \neq 0\\ 0, & \text{if } Y = 0 \end{cases}$$
(2.40)

LogLuv chromaticity is encoded as CIE 1976 u'v' chromaticity introduced in Section 2.4. u' and v' are scaled and encoded to 8 bits color resolution:

$$ue = \lfloor 410u' \rfloor$$
(2.41)
$$ve = \lfloor 410v' \rfloor$$
(2.42)

The full bit order for LogLuv is shown in Figure 2.13.

Figure 2.13: Bit order of LogLuv32 encoding. Adapted from [116, Figure 3].

There is also a LogLuv variant that uses 24 bits per pixel. This variant will not be described due to its limited dynamic range, lower color resolution and incompatibility with classic compression because of the scan-line encoding scheme for CIE 1976 u'v' chromaticity.

2.4.4. Y''u''v''

Y"u"v" was introduced by Poynton et al. in 2014 [161]. It is based on LogLuv but the logarithmic nonlinearity for encoding luminance is exchanged for the PQ curve [184] as depicted in Equation (2.10). Using PQ reduces the bit depth requirements from 15 bits needed for the Le component of LogLuv to 12 bits with PQ because of the higher coding efficiency of PQ and the lower dynamic range compared to LogLuv.

$$Y'' = f_{PQ}(Y)$$
 (2.43)

The CIE 1976 u' and v' chromaticities are compressed for values below 5 cd/m^2 and subsequently renamed to u" and v":

$$u'' = WP_{u'} + \min\left(\max\left(\frac{Y''}{\text{epsilon}}, \frac{1}{256}\right), 1\right)(u' - WP_{u'})$$
(2.44)

$$v'' = WP_{v'} + \min\left(\max\left(\frac{Y''}{epsilon}, \frac{1}{256}\right), 1\right)(v' - WP_{v'})$$
 (2.45)

$$WP_{u'} = 0.1978$$
 $WP_{v'} = 0.4683$ epsilon = 0.25

Compressing chroma for lower luminance values helps to reduce the color noise without affecting color quantization artifact visibility because the color contrast sensitivity in human vision is reduced for lower luminance values.

2.4.4.5 ITU Rec.2100

Before introducing Rec.2100 the International Telecommunication Union (ITU) first proposed the Rec.2020 system [92] for the next generation of standard dynamic range and wide color gamut Ultra High Definition video. Rec.2020 uses wider color primaries compared to Rec.709 but still applies a gamma curve with a linear toe followed by a color differencing matrix. This is a system similar to regular Rec.709 Y'C_BC_R introduced in Section 2.4.3 though with wider, still physically realizable primaries. Rec.2020 did not find a wide adoption and is superseded by Rec.2100 [96], which adds two HDR transfer curves and a new color differencing scheme to Rec.2020. The primaries for Rec.2020 and Rec.2100 are the same:

$$\begin{bmatrix} R_{2100} \\ G_{2100} \\ B_{2100} \end{bmatrix} = \frac{1}{Y_{white}} \begin{bmatrix} 1.716651188 & -0.666684352 & 0.017639857 \\ -0.355670784 & 1.616481237 & -0.042770613 \\ -0.253366281 & 0.015768546 & 0.942103121 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$
(2.46)

The Rec.2100 standard specifies two alternative curves for HDR encoding. The first curve is the PQ / SMPTE ST.2084 curve presented in Equation (2.10). Using Rec.2100 with PQ curve will be referred to as 'Rec.2100 PQ' within this thesis.

The PQ nonlinearity is applied to R, G, B, resulting in R^{PQ} , G^{PQ} and B^{PQ} :

$$R_{2100}^{PQ} = f_{PQ}(R_{2100}) \qquad G_{2100}^{PQ} = f_{PQ}(G_{2100}) \qquad B_{2100}^{PQ} = f_{PQ}(B_{2100})$$
(2.47)

As in conventional $Y'C_BC_R$ the luma channel Y_{2100}^{PQ} is calculated from the nonlinearly coded R_{2100}^{PQ} , G_{2100}^{PQ} and B_{2100}^{PQ} , channels weighted by the contributions of R, G and B to CIE 1931 luminance:

$$Y_{2100}^{PQ} = 0.2627 R_{2100}^{PQ} + 0.6780 G_{2100}^{PQ} + 0.0593 B_{2100}^{PQ}$$
(2.48)

The color difference channels are derived by subtracting luma Y^{PQ} from B^{PQ} and R^{PQ} :

$$C_{\rm B\,2100}^{\rm PQ} = \frac{B_{2100}^{\rm PQ} - Y_{2100}^{\rm PQ}}{1.8814} \tag{2.49}$$

$$C_{\rm R_{2100}}^{\rm PQ} = \frac{R_{\rm 2100}^{\rm PQ} - Y_{\rm 2100}^{\rm PQ}}{1.4746} \tag{2.50}$$

Equations (2.48) to (2.50) can also be written as a linear transformation:

$$\begin{bmatrix} Y_{2100}^{PQ} \\ C_{B_{2100}}^{PQ} \\ C_{R_{2100}}^{PQ} \end{bmatrix} = \begin{bmatrix} 0.2627 & 0.6780 & 0.0593 \\ \frac{-0.2627}{1.8814} & \frac{-0.6780}{1.8814} & \frac{1-0.0593}{1.8814} \\ \frac{1-0.2627}{1.4746} & \frac{-0.6780}{1.4746} & \frac{-0.0593}{1.4746} \end{bmatrix} \begin{bmatrix} R_{2100}^{PQ} \\ G_{2100}^{PQ} \\ B_{2100}^{PQ} \end{bmatrix}$$
(2.51)

The second curve presented in Rec.2100 is the HLG curve [35] as introduced in Equation (2.11). The HLG encoding scheme applies a piecewise defined function to the linear R, G, B Rec.2100 primaries (2.46) resulting in R^{HLG}, G^{HLG} and B^{HLG}:

$$R_{2100}^{\text{HLG}} = f_{\text{HLG}}(R_{2100}) \qquad G_{2100}^{\text{HLG}} = f_{\text{HLG}}(G_{2100}) \qquad B_{2100}^{\text{HLG}} = f_{\text{HLG}}(B_{2100})$$
(2.52)

The calculation of luma and the two chroma channels follows the Rec.2100 PQ Equation (2.51). In this thesis, the Rec.2100 Hybrid Log Gamma encoding will be referred to as 'Rec.2100 HLG'.

Rec.2100 also specifies a second set of primaries (LMS) and a different color differencing matrix (IC_TC_p). These will only be introduced in Chapter 2.4 because the IC_TC_p encoding scheme is based on contributions of this thesis.

2.4.5 Comparison of HDR Color Encodings

The HDR image encodings presented in this chapter are all tailored for specific applications, but none of them fully satisfies all requirements for entertainment media video distribution. As an example, OpenEXR needs too much bandwidth, RGBE and LogLuv are not suited for compression because the pure logarithmic curve amplifies the energy of the typical photon shot and sensor read noise in the dark areas. Rec.2100 HLG and Rec.2100 PQ are only perceptually uniform near the achromatic axis, quantizing saturated colors inefficiently. Finally, Y"u"v" is not hue linear enough for high quality gamut mapping and needs more computational resources compared to Rec.2100 HLG and Rec.2100 PQ because of the division by a variable component to calculate the u"v" chroma channels. Tables 2.2 and 2.3 give an overview of the parameters for the most common standards and proposals for HDR and WCG file storage and distribution.

	OpenEXR	RGBE	LogLuv	
Bit per pixel	48	32	32	bits
Min. luminance $\neq 0$	5.96 e-8	5.87 e-39	5.44 e-20	cd/m ²
Max. luminance	65504	6.79 e+38	1.84 e+19	cd/m ²
Dynamic range	40	256	128	f-stops
QS at 0.1 cd/m ²	0.061	~ 0.5	0.27	%
QS at 10 cd/m ²	0.078	~ 0.5	0.27	%
QS at $1000 \text{ cd}/\text{m}^2$	0.050	~ 0.5	0.27	%
Negative values	Yes	No	Yes	
Decorrelated	No	No	Yes	
Primary	Rendering and	Storage of con	nputer	
application	postproduction	generated imagery		
Limitations	High bandwidth,	Dynamic range beyond		
	no decorrelation	distribution needs.		
	defined	Quantization in dark		
		areas too high	1.	

Table 2.2: Comparison of HDR and WCG color encodings for image storage. 'QS' refers to one quantization step.

Table 2.3: Comparison of HDR and WCG color encodings for image distribution over 10-bit channels. 'QS' refers to one quantization step when encoding with 10 bits.

	Rec.2100 HLG	Rec.2100 PQ	$\mathbf{Y}''\mathbf{u}''\mathbf{v}''$	
Bit per pixel	30 or 36	30 or 36	30 or 36	bits
Min. luminance $\neq 0$	3.8 e-4	4 e-5	4e-5	cd/m ²
Max. luminance	1200	10000	10000	cd/m ²
Dynamic range	22	28	28	f-stops
QS at 0.1 cd/m ²	12.75	3.76	3.76	%
QS at $10 \text{ cd}/\text{m}^2$	1.24	1.25	1.25	%
QS at $1000 \text{ cd}/\text{m}^2$	0.53	0.9	0.9	%
Negative values	No	No	No	
Decorrelated	Yes	Yes	Yes	
Primary Application	Backwards com-	High quality	High quality	
	patible televison	television	television	
	distribution	distribution	distribution	
Limitations	Quantization	Not hue linear	Not hue line	ar
	artifacts with	enough, inef-	enough, com	npu-
	dark low noise	ficient color	tational com	-
	content, not hue	encoding	plexity too h	igh
	linear enough,			
	inefficient color			
	encoding			



HDR and WCG Video Data Set -'HdM-HDR-2014'

For scientists developing algorithms and hardware for HDR video processing, transmission, storage and display, it is crucial to have high quality HDR content available. These images must be of sufficient spatial, tonal and temporal resolution as well as high dynamic range and wide color gamut, ideally representing the output of future cameras. Following the MPEG definition of HDR [128], high dynamic range starts at a latitude of 16 stops or more. As of 2014, there was no cinematic HDR video data set available. The marginal existence of HDR video sequences stands in great contrast to the domain of still imaging, where HDR-image capture has a long history [9] and is well studied and commonly practiced [44, 165]. The following chapter will introduce a novel HDR and WCG video dataset that can be used to evaluate compression codecs, storage formats, and display technologies. It is based on the paper that describes the creation of the HdM-HDR-2014 dataset [71].

While high resolution and higher frame rate videos can be acquired by using current generation motion picture cameras, there is no single HDR camera with a 'Super 35mm' sized sensor available. Throughout the last decade, professional digital film cameras gained around 4 stops of dynamic range, from about 10 stops in 2001 [188] to 14 stops in 2011 [23]. With high dynamic range displays on the horizon, future image acquisition systems are expected to feature an even higher dynamic range [25]. To

simulate the dynamic range of future cameras today, a combination of two exposures is needed.

Figure 3.1 shows a visual comparison between the dynamic range of a professional motion picture camera, a dual-camera setup and current display devices. While one Alexa camera [6] can capture a higher dynamic range compared to current mainstream display devices, future HDR displays will need even higher dynamic range imagery. Using two Alexa cameras in a mirror rig as explained in the following sections provides a dynamic range satisfying the MPEG requirements for HDR of 16 photographic stops [128] and the needs of state-of-the-art HDR displays.



Figure 3.1: Dynamic range of current acquisition and distribution devices compared. 'Alexa' refers to the ARRI Alexa camera [6], '2*Alexa' to the dual camera rig as used herein, 'TV' to a television set conforming to ITU BT.1886 [90], 'Cinema' to a SMPTE RP 431-2 cinema projection [181] and 'HDR Display' to a Dolby PRM-4220 reference display [46].

Image quality assessments should be performed using high fidelity images. Only if the original images are of significantly higher quality compared to the altered footage, any degradation in image quality caused by a compression algorithm or display device can be clearly assigned to this individual conversion step. The needed image quality is not only determined by the signal quality of the image acquisition system, but also by lighting, depth of field, make-up and staging. As an example, a faithful skin tone reproduction of a non-powdered actor in typical room lighting will not appear lifelike

to most observers. Humans often appear unhealthy, or look tired in reproductions, when filmed without cinematic lighting and makeup. Hence, viewers expect to see staged pictures when evaluating video quality. Thus, staged images are important to avoid misinterpretations, especially when dealing with non-expert observers in user studies. Therefore, a scientific dataset for visual evaluation still has to meet the aesthetic standards of a commercial film production and at the same time deliver a signal quality that is sufficient for research. The required aesthetic quality calls for the use of 'Super 35mm' sized sensors and professional lenses, make-up, set-design and film lighting.

The goal of the HdM-HDR-2014 project is to provide cinematic footage of high visual and technical quality that covers the dynamic range and gamut of future sensors. Using this video content of extended dynamic range and wider gamut, tone mapping algorithms, compression codecs and displays for future HDR and WCG content can be evaluated today.

3.1 Related High Dynamic Range Video Data Sets

Research on HDR video acquisition includes time-sequential approaches [79, 100] and multi sensor camera rigs [189]. One of the earliest HDR video data sets available to the scientific community is the 'Tunnel Sequence' [109]. More recently three additional HDR datasets were released: The Linköping University HDRv Repository [110, 111, 121], captured with a SpheronVR camera, shows everyday scenes on and near the Linköping University campus. The University of British Columbia's Digital Media Lab also released an HDR video dataset named DML-HDR [191]. This footage was acquired using a 'RED SCARLET–X' camera and also features unstaged everyday scenes. Finally, Technicolor contributed a number of outdoor scenes and one animated scene to MPEG [118]. These clips are captured using two Sony F3 or F65 cameras in a mirror rig and also feature unstaged documentary scenes.

As of 2014, no cinematic HDR video has been gathered. All existing HDR data sets feature un-staged scenes with non-powdered actors and without cinematic lighting.

3.2 Methods

In this section the selection and design of scenes will be presented and the dual camera video acquisition system is introduced. The dynamic range of the dual camera rig is evaluated via simulated and measured signal to noise curves. Further, the signal processing steps from sensor via recording and reconstruction to distribution are described and the technical infrastucture for visual verification and color grading is presented.

3.2.1 Selection and Design of Scenes

Five categories of scenes are designed to focus on different challenges in HDR storage, compression, tone mapping, gamut mapping and display:

The **Still Life** scene presented in Appendix A.1 is intended to help develop new camera characterization methods and provide a reference for color rendering algorithms by including a color checker. The dark and bright skin tones and hair provide a challenge to tone mapping operators.

The **Wide Gamut and Moving Lights** scenes presented in Appendix A.2 are designed to include a very large gamut in terms of dynamic range and color saturation at the same time. Bright and saturated colors are a special challenge in tone and gamut mapping because the mapping operator must decide if it preserves saturation or luminance. The large gradients of the colorful lights in the smoke are intended to help verify tonal resolution for wide color gamut baseband encodings and the small saturated lights are designed to pose challenges in color subsampling. Fast moving lights and flashes reduce temporal redundancy and are therefore a challenge to compression algorithms.

The **Low Key Scenes** presented in Appendix A.3 are intended for evaluating monitor technologies. The small bright details unmask the deficiencies of dual modulation [174] monitors. In contrast, RGBW monitors [196] render colored bright objects either too dark or desaturated. Lots of monitor and projection technologies are limited by straylight. The low key scenes are intended to help determine the impact of straylight when removing only the very bright areas of the image and observing the change in the dark areas. The low light scenes are also designed to pose a challenge in tone mapping when rendering the dark details brighter and thus amplifying noise. Especially the fire scenes are designed for evaluating the trade-off tone mapping operators take in rendering the flames bright, but still colorful.

The **Sunlight Scenes** presented in Appendix A.4 are color graded to challenge even the brightest new monitor technologies when displayed at full luminance. They are designed as a difficult subject for tone mapping operators when converting the graded HDR version to lower luminance. They are also intended for fine tuning automatic color rendering algorithms because lowering overall luminance often calls for increased contrast and saturation.

The **High Contrast Skin Tones** scenes presented in Appendix A.5 are designed to challenge the balance of bit-allocation between dark and bright areas in compression. In addition, they are intended to be difficult in tone mapping because humans tend to be most critical when evaluating faces and skin tones.

3.2.2 Set Design, Staging and Lighting

The set design, staging, lighting, lensing and makeup of the HdM-HDR-2014 data set are selected to represent those typically found in the respective production types of the individual scenes, such as documentary, advertising or feature film.

As an example, costumes are chosen as typically used in motion pictures. Only the white shirt in the 'Poker' scene and the glittering dress of the 'Showgirl' are deliberately selected to go beyond average costumes in order to create a challenge for HDRlighting. Studio-sets are built in 180°, with no objects or other lighting equipment bouncing back light from behind the camera. The fill light only comes from visible objects in the scene. This is also true for the 'Fireplace' scene.

The 'Poker' scene is lit by one single 6 kilowatt Hydrargyrum Medium-arc Iodide (HMI) keylight coming straight from above the table. The actor's faces are only lit by bounced

light from table cloth, except if they lean forward and reach the keylight. Like the actors, the background scenery is only illuminated by bounced light from the keylight plus the in-frame candles. This creates a natural behavior of light in a controlled environment. The 'Bistro' scene follows the same concept: a single strong keylight is used to simulate the sunlight shining through a small slot through the window and through curtains, illuminating the whole 'indoor' scenery. For the 'Showgirl' scene, in order to create the dynamic range of 18 stops and a temporal change of different light qualities in one scene, the scene is re-lit throughout the runtime of the shot, changing from a tungsten-lit low-key scenery (girl sitting at mirror) to a high-key scene with daylight simulated by a 6 kilowatt HMI light.

All sceneries are staged to emphasize the perception of light. In the 'Car' scenes the camera travelling is intended to emphasize material appearance by creating moving reflections of bright sunlight on the black car finish. Deep shadows under the car extend the dynamic range of the scene down to the noise floor of the camera system. No additional lighting was used on all outdoor scenes. The 'Fishing' scene represents a typical establishing shot of movies by combining a pan with a camera movement leading to main-scenery. This gives a typical rhythm and feeling of an opening scene and is designed to demonstrate the increased perception of judder in HDR presentation at 24 and 25 frames per second.

Objects are staged to get cinematic images with a shallow depth of field by tightly controlling in-focus areas and rendering the fore- and background out-of-focus. For the scenic production mostly 'Zeiss UltraPrime' lenses are used while those scenes that represent documentary scenes are acquired using 'Zeiss CompactPrime CP.2' lenses and 'Angéniuex Optimo 28-76mm, T.2,6' zoom lenses. Dolly grip is only used in sequences that are representative for movie or advertising shootings. In simulated documentary shots like 'Smith', 'Fireplace' and the 'Beerfest' and 'Carousel' sequences, only a reduced amount of makeup is applied.

3.2.3 High Dynamic Range Video Acquisition

HDR still images and videos are often acquired by capturing multiple images with different exposures, one after the other [44, 100]. When dealing with moving objects,

artifacts can be introduced by not integrating all exposures at the same time. The darkened flames in Figure 3.2 serve as an example of ghosting artifacts introduced by time-sequential HDR image capture using an Apple iPhone 4S mobile phone with 'HDR' feature enabled in the standard camera app. The brightest parts of the flames are replaced from a second shorter exposure. In this case, the flames have already moved at the time the second exposure is captured. Hence, nothing suitable can be inserted, rendering parts of the flames too dark. To avoid these ghosting artifacts with moving subjects, it is essential to capture all exposures simultaneously when aiming for HDR reconstruction of moving objects.



Figure 3.2: Illustration of ghosting artifacts introduced by time-sequential HDR image capture.

To generate HDR video with different exposures captured at the same time, a mirror rig, as shown in Figure 3.3 a) is used. A standard glass pane with anti-reflective coating on the backside is employed as a beam splitter. This results in a ratio of around 1:16 between reflection and transmittance, shifting the camera exposures by about 4 f-stops. To be able to use large sensor motion picture cameras, the mirror is mounted in front of the lens, instead of splitting the light behind the lens, as proposed by Tocci et al. [189]. Thus, aperture, integration time and sensor gain can be kept at identical settings in both cameras. This results in the same depth of field and motion blur, but different signal to noise ratios, which are used to enhance the dynamic range. Both cameras are adjusted mechanically for geometric alignment and the integration times of the camera-sensors are synchronized to record exactly the same fraction of time.



Figure 3.3: Schematic drawing of the mirror rigs used for HDR-video acquisition. Rig a) is used for most shots while some shots are aquired using a mirror with 50% transmission and a 4 f-stop neutral density filter in front of one camera as depicted in b).

When used with lenses of longer focal length, the 1:16 beam splitter generates ghosting artifacts caused by double reflections in the mirror glass. Hence some shots are acquired using an alternative setup shown in Figure 3.3 b). It employs a 1:1 semitransparent mirror instead of the 1:16 mirror. In this case, a neutral density filter has to be mounted in front of one camera to shift the exposure of this camera.

The 'Alexa M' camera, a Complementary Metal-Oxide-Semiconductor (CMOS) sensor based motion picture camera made by Arnold & Richter Cine Technik (ARRI) is chosen for its large dynamic range and accurate color reproduction [6]. The sensor is operated in a dual gain mode resulting in a dynamic range (full-well/read-out noise) of 14.8 f-stops [23]. The exposure is adjusted to capture extended detail in dark and bright areas compared to a typical exposure with a single camera. Frame rates of 24 frames per second (fps) and 25 fps represent both television and cinema applications. The integration time is set to 1/50 second $(172.8^{\circ}/180^{\circ}$ shutter) to achieve a cinematic look. Only the high-speed shots are recorded with a 356° shutter to gain one additional f-stop of exposure.

During in-camera image processing, the signal from the sensor is converted from analog to digital on two paths with different levels of analog amplifications. The shadows are reconstructed from the high gain path and the highlights from the low gain path [4]. The resulting 16-bit 2880 by 1620 resolution RAW Bayer pattern image is then converted to 1920 by 1080 RGB-pixels and coded in 12 bit LogC wide gamut color space [24] to be recorded as video file using near visually lossless 330 Mbit/s ProRes [5] intra-frame compression codec. The theoretical and measured signal-to-noise ratio for the full dual camera rig can be seen in Figure 3.4.



Figure 3.4: *Simulated and measured signal-to-noise ratio of the image acquisition system.*

The simulated signal-to-noise ratio in Figure 3.4 is calculated by simulating noise for both exposures according the document describing the Alexa camera's noise characteristics [23] followed by merging the two exposures as described in the postproduction section 3.2.4. The measured signal-to-noise ratio is determinded by over- and underexposing a reflective gray step wedge chart to extend the 64:1 dynamic range of the chart beyond the dynamic range of the full acquisition system. The measured signalto-noise ratio is calculated from the mean and standard deviation of 10.000 pixels from a 50 by 20 pixel area over 10 frames.

3.2.4 Postproduction

In postproduction, the highlight preserving image is spatially aligned to the lowlightpreserving image by warping through local disparity estimation [80]. Subsequently the colors of the highlight-preserving image are matched to the lowlight-preserving image through multiplication of the individual color channels in linear Alexa Wide Gamut domain by minimizing the sum of squared differences per color channel. All further postproduction steps are also carried out in the Alexa Wide Gamut color space spanned by the primaries listed in Table 3.1.

Table 3.1: *CIE 1931 xy chromaticity coordinates of the primary colors of the Alexa Wide Gamut color space.*

	Red	Green	Blue	White
X	0.6840	0.2210	0.0861	0.3127
У	0.3130	0.8480	-0.1020	0.3290

After spatial exposure alignement, the two images are merged to one HDR frame by blending between both images, depending on the luminance of the individual pixels. Finally the border pixels are set to black, to mask pixels where no highlight pass is available due to the spatial displacement. See Figure 3.5 for an overview of the incamera image-processing pipeline.



Figure 3.5: Overview of the in-camera video processing steps. The quantization is depicted above and below the arrows. If the quantization curve is not linear, the corresponding nonlinearity, for example 'LogC', is noted.

The dual gain read-out depicted in Figure 3.5 provides two times 14 bits of linear tonal resolution merged with an offset of about 4 stops resulting in one 16-bit linear image per camera. Combining this tonal resolution from both cameras is sufficient to cover the full dynamic range of the footage.

As illustrated in Figure 3.6 the highlight preserving camera provides a tonal resolution of 8096 code values for the top photographic stop (-1) below the clipping point (peak white). The lowlight preserving camera still has a tonal resolution of 4 code values at -18 stops below peak white.



Figure 3.6: Limitations to tonal resolution by the camera sensor.

The 12-bit ARRI Alexa LogC color space with its pseudo logarithmic curve summarized in Section 2.3.2 is shown in Figure 3.7. LogC limits the video streams from both cameras to about 256 code values per stop. If RAW recording would have been used, the top eleven stops could have been stored with higher tonal resolution.



Figure 3.7: *Limitations to tonal resolution by the camera sensor and LogC storage combined.*

The postproduction steps are depicted in Figure 3.8. All processing steps are performed in single precision 32-bit floating point per color channel. The tonal resolu-



Figure 3.8: Overview of the video processing steps in postproduction. The quantization is depicted above and below the arrows. For integer representations the corresponding quantization curve, for example 'LogC' or 'G2.6' for a pure gamma of 2.6, is noted.

tion of the final storage containers for the reconstructed but still scene-referred camera footage is shown in Figure 3.9. The final OpenEXR files are still coded relative to the ALEXA Wide Gamut primaries listed in Table 3.1. The half-float OpenEXR file container provides at least 4 times the tonal resolution compared to the 12-bit LogC quantization performed before. A range from 0 up to 500 in OpenEXR is used instead of normalizing white to 1.0 to prevent the shadows from being quantized coarser by the halved tonal resolution for each stop below 2^{-14} in the subnormal numbers of half float [86, Chapters 3.4 and 7.5].



Figure 3.9: Content tonal resolution compared to the half-float container used for distributing the reconstructed camera footage.

The graded content is distributed in 16-bit TIFF container with SMPTE 2084 / PQ nonlinear encoding as described in Section 2.3.3.3. Figure 3.10 illustrates that this format quantizes about 16 times finer compared to what the captured footage con-
tains. Thus, even assuming some contrast enhancements, there are no significant limitations expected from the distribution container of the color graded version of the HdM-HDR-2014 data set.



Figure 3.10: Content tonal resolution compared to the PQ 16-bit integer container used for distributing the graded content.

During postproduction, the rendered image sequences are screened using Filmlight's Baselight color grading system [66] outputting 12-bit pure gamma 2.6 encoded signals on a Dolby PRM-4200 HDR monitor [46]. For quality control the images are displayed three times with exposure offsets of -6, 0 and 6 stops to be able to see the full luminance range near the peak contrast sensitivity of the human visual system and in a range where the pure gamma 2.6 signal and the dual modulation LCD processing of the PRM monitor provides the highest spatial and tonal resolution.

In addition to the PRM monitor a Sony PVM-2541A professional OLED monitor [34] is used for a second pass of quality control to prevent the dual modulation algorithm of the Dolby PRM from masking problems with small bright details.

For wide gamut verification a 'Christie CP4220' **D**igital Light Processing (DLP) digital cinema projector [31] with modified notch-filters is used for quality control and grading. The custom narrow band-pass notch filters that block wavelengths around cyan and yellow allow this projector to feature a gamut very close to the gamut defined in ITU Rec.2100 [96]. The gamut and spectral power distribution of this Xenon DLP cinema projector plus the custom notch filter are illustrated in Figure 3.11.



Figure 3.11: *Spectral power distribution and resulting gamut of the modified DLP cinema projector.*

3.3 Limitations

Recording HDR video by means of a mirror rig makes it possible to capture a dynamic range that single sensor cameras will probably only be capable to capture in the future. However, using a mirror rig comes along with limitations. The handling of the fully rigged recording device is very limited, as it weights about six times more than a single digital film camera. In addition, it has to be kept wired to a recording unit. Therefore, the rig cannot be placed as flexibly as a single camcorder.

Due to the large mirror, the camera rig suffers from stray light and lens flares. Even if these could be reduced using a mirror with a higher-grade coating, stray light and lens flares will always be more severe compared to a single camera due to the much bigger mirror matte box.

The mirror also results in double contours visible in the highlight-protecting pass of the two long shots 'Poker' or 'Showgirl 2'. As illustrated in Figure 3.12 a) these double contours result from a second reflection of the transmitted light rays when leaving the glass of the mirror in the direction of the lower camera. They could be avoided using a thinner beam splitter. The double contours for the 'Showgirl 2' scene are reduced using deconvolution but it is preferred to keep some contours in order to prevent the creation



Figure 3.12: Mirror induced double contours - cause and effect.

of new artifacts originating from deconvolution. The double contours can be observed in Figure 3.12 b) and the result of the deconvolution is shown in Figure 3.12 c).

Besides the double contours, the mirror also introduces polarizing artifacts. This means that polarized light, like from the water surface in the 'Fishing' shots, or the car finish in the 'car' shots, is split with a different ratio compared to the non-polarized light from the surrounding. This results in diverging luminance and color between the different areas in the image depending if they are reconstructed from the lowlight preserving camera or the highlight preserving camera.

In daylight shots, specular highlights and the sun orb are often clipped because lowering the exposure to capture them would have resulted in losing too much detail in the blacks. While this is not a problem for the intended use of the video data set, true radiance maps cannot be recovered using a setup consisting of only two motion picture cameras that are shifted by 4 stops in exposure.

Finally the sensor of the Alexa camera is a rolling shutter sensor. Rolling shutter artifacts can be observed at very fast motion, for example in frame 97031 from the 'Fireworks' scene.

3.4 Results

The HdM-HDR-2014 video dataset is published and freely available at the project website [72]. Thumbnails and image statistics including luminance quantiles and chromaticity plots are presented in Appendix A.

The HdM-HDR-2014 scenes are staged according to common film production techniques. The settings are chosen to present challenges to HDR-presentation such as bright, saturated colors and brightness changes of different speed and magnitude. Human faces including high contrast skin-tones, hair and eyes complete the cinematic settings. As a consequence, the HdM-HDR-2014 test-set can be employed to compare the rendering of material appearance and colors on different HDR-displays or to evaluate temporal video tone mapping operators. Due to the cinematic look and staging, user studies that rely on image quality assessment by non-specialist viewers can be conducted without causing irritations due to non-staged reproductions.

Since its publication, the HdM-HDR-2014 dataset has been used in dozens of research projects. As an example, the MPEG used a subset of scenes for the evaluation of next generation video coding [70]. The HdM-HDR-2014 data set has also been used in research on HDR image acquisition [76], to develop and evaluate tone mapping algorithms [37, 55, 58, 120, 141, 153, 176], compression codecs [56, 57, 82, 107, 119, 123], display technology and metrology [105, 114, 155], image quality assessment [136, 148, 207], metrics for 'HDRness' perception [84], gamut extension algorithms [205], visual attention/saliency prediction [12, 49, 50, 149] and to develop and evaluate new television systems [20, 154].

Within this thesis, the HdM-HDR-2014 data set will be used to visually compare the HDR and WCG encoding schemes introduced in Chapter 4 and to verify the requantization algorithm presented in Chapter 5.



Static WCG and HDR Color Encodings - $(IC_AC_B)^2$ and $(IC_TC_P)^2$

When distributing HDR and WCG video content like the HdM-HDR-2014 video data set an efficient encoding is needed to keep storage space and bandwidth needs as low as possible. Therefore it is crucial to find efficient quantization schemes for HDR and WCG video. As described in Section 2.3.3, luminance encoding for HDR video has been studied extensively. There also have been multiple proposals for HDR and WCG color image encoding, summarized in Section 2.4.4. Common to all of these approaches is a focus on encoding efficiency; they are designed to encode images using a minimum number of code values without introducing visible quantization artifacts or loss of image details.

A modern image encoding scheme should not only be optimized for efficient quantization but should also facilitate downstream image processing steps like color volume mapping to prevent computationally expensive color space conversions for each step. Thus, it is desirable to transmit video signals in an encoding color space that is not only suitable for efficient image encoding, but also for tone mapping and gamut mapping, often referred to as color volume mapping.

Unlike print media, digital cinema and television distribution did not typically require color volume mapping in the past because most display devices had a gamut and dynamic range that matched the encoded signal. Notable exceptions were early LCD displays that did not cover the complete Rec.709 gamut [94].

Future displays will have a larger variance in covered dynamic range and gamut. One reason for this is the trend towards mobile viewing, which extends potential ambient illumination ranging from dark cinema and dim living rooms to bright sunlight for outside viewing. A second reason is that new display technologies have extended capabilities. Laser-illuminated projectors [178] are now becoming available and extend the gamut from DCI-P3 [181] to Rec.2020 [92]. At the same time, major television networks are investigating how to distribute signals in Rec.2020 color space [156] to leverage the wider color gamut of emerging display technologies such as OLED [30], quantum dot displays [127] and already existing multi-primary displays [169, 190]. Thus, there will likely be more variation in device capabilities than ever before. Most of these new displays and cinema projectors will not feature the full Rec.2020 gamut or the peak luminance of high dynamic range mastering displays [152]. This means that mapping between different color volumes will become critically important for consistent best-possible reproduction of TV and cinema imagery. Figure 4.1 depicts the full video distribution pipeline for HDR and WCG video.



Scope of the HDR Color Encoding Presented in this Thesis

Figure 4.1: Overview of a typical HDR video distribution pipeline.

In the following sections, the IC_AC_B color space for both high dynamic range and wide color gamut is introduced. The IC_AC_B color space is co-optimized for encoding efficiency as well as color volume mapping performance. Based on these optimization results, a second model for HDR color encoding named IC_TC_P is introduced. In order to verify the encoding efficiency of the IC_AC_B and IC_TC_P color spaces, they are compared visually and numerically to current state-of-the-art HDR and WCG quantization schemes. Finally, the challenges and limitations associated with the IC_AC_B and IC_TC_P approaches are discussed. This chapter is based on the paper introducing the IC_AC_B color space [73] and the patent disclosing the introduction of crosstalk in LMS domain [69], which is one of the fundamental ideas behind the IC_TC_P color space.

4.1 Research Question

The major goal of image encoding for distribution is to minimize distortions when images are represented with a given number of digital code words, as well as to find the number of code values needed to **prevent visible quantization artifacts**. The best encoding performance is typically achieved when quantization error is distributed perceptually even over the color space. To fully avoid visible quantization errors, the step of one code value should always be below the detection threshold of one JND. Thus, the more uniform in size the JND ellipsoids are throughout the color space, the more efficient the encoding is, as there are less code values wasted to encode sub-JND steps in areas of the color space where JND ellipsoids are larger [113]. This requirement will be denoted as **JND uniformity**.

In addition to JND uniformity, a color space for video encoding should **decorrelate the achromatic axis from the chromatic axes** to enable color sub-sampling that exploits the lower contrast sensitivity of the human visual system for high frequency chroma details as summarized in Section 2.4.1.

Furthermore, a color space used for color volume mapping should be as **hue linear** as possible, as observers perceive changes in hue to be more impactful compared to changes in lightness or chroma. As such, most gamut mapping algorithms either completely avoid or heavily penalize hue changes. Thus, when mapping is performed toward the achromatic axis, or when intensity is changed, the color space should not introduce any hue changes.

As well as maintaining uniformity inside the gamut volume, it is important to consider sufficiently large bounds to encode the necessary gamut and dynamic range. **Rec.2100** gamut [96] is the design goal for the next generation of displays and therefore the minimum requirement for a modern video encoding space.

Short-term adaptive processes can expand the required dynamic range for entertainment imaging beyond the steady state adaptation of the human visual system [112]. A consumer study using an HDR research display in a dark viewing environment identified a **dynamic range of 0.005 to 10000 \text{ cd/m}^2** as required to satisfy 90% of the viewers [40].

Finally, the computational complexity of the transformation from the encoding color space to device RGB should be as low as possible to allow for mass deployment in a wide range of devices. Specifically the transformation should **minimize computa-tions** in linear light and allow separable operations (i.e. functions of a single component rather than multiple components).

4.2 Methods

The goal of this work is to develop color difference encoding models that can be mass deployed. Hence, an existing encoding scheme is chosen. The main work is to find the optimal parameters for this model, rather than starting from scratch without any constraints. The methods introduced here can also be used to optimize other models.

In the following, the color space model will be discussed, the test sets and training sets will be introduced, as well as the cost functions that are used for the optimization of the model parameters. Based on the findings of the optimization process, a second model will be introduced.

4.2.1 Model Derivation

A color space model that follows the processing steps of broadcast video $Y'C_BC_R$ and IPT is chosen because these are a very simple models for the human visual system in daylight vision. $Y'C_BC_R$ and IPT are introduced in Sections 2.4.2.2 and 2.4.3. The $Y'C_BC_R$ and IPT color models first transform from device RGB (R_{Device} , G_{Device} and B_{Device}) to a defined three-dimensional additive color space, like Rec.709 or Rec.2100 for $Y'C_BC_R$ and the LMS cone sensitivities for IPT. For the model introduced herein,

the basis for nonlinear encoding will be called \bar{R} , \bar{G} and \bar{B} . The basis change from any basis like CIE 1931 XYZ or any device RGB to \bar{R} , \bar{G} and \bar{B} can be performed via a linear transformation M_1 shown in Equation (4.1). Subsequently a nonlinear encoding function f that mimics the relation of retinal illumination to cone response is applied in Equation (4.2). The third and last step of the Y'C_BC_R and IPT color models is to simulate the opponent processing in the retina by calculating one achromatic luma channel and two color opponent channels - often implemented by means of a second linear transformation M_2 as depicted in Equation (4.3):

$$\begin{bmatrix} \bar{R} \\ \bar{G} \\ \bar{B} \end{bmatrix} = M_1 \begin{bmatrix} R_{\text{Device}} \\ G_{\text{Device}} \\ B_{\text{Device}} \end{bmatrix}$$
(4.1)

$$\bar{R}' = f(\bar{R})$$
 $\bar{G}' = f(\bar{G})$ $\bar{B}' = f(\bar{B})$ (4.2)

$$\begin{bmatrix} I \\ C_a \\ C_b \end{bmatrix} = M_2 \begin{bmatrix} \bar{R'} \\ \bar{G'} \\ \bar{B'} \end{bmatrix}$$
(4.3)

This model is easily invertible and already implemented in professional broadcast devices. In addition, IPT is known for its excellent hue linearity in standard dynamic range scenarios. The difference between $Y'C_BC_R$ models and IPT is that in IPT the nonlinearity is applied to LMS cone fundamentals, rather than RGB primaries, and IPT uses a different color-differencing matrix compared to $Y'C_BC_R$. Both models have in common that intensity *I* and luma *Y'* are not necessarily exactly weighted to the CIE 1924 photopic $V(\lambda)$ luminosity function [32] for all colors because *I* and *Y'* are calculated from nonlinearly encoded primaries instead of the linear representation.

Having chosen a specific model, next the individual parameters of the model have to be found. First, the nonlinearity function f will be determined by looking only at the achromatic axis. For Y'C_BC_R and IPT colors are achromatic, if all nonlinear components have the same value before applying the decorrelation matrix M_2 . From $f^{-1}(I) = \bar{R} =$ $\bar{G} = \bar{B}$ it can be deduced that if the optimal nonlinear encoding for I is known, this exact function also needs to be applied to \bar{R} , \bar{G} and \bar{B} for the quantization along the achromatic axis to be maximally efficient. The PQ curve introduced in Section 2.3.3 and Equation (2.10) satisfies this requirement. The introduction of PQ as nonlinearity guarantees the quantization to be most efficient along the achromatic axis. However, the coefficients of matrix M_1 and M_2 cannot be directly deduced. Nonetheless, this model suggests some assumptions that lead to the reduction of free parameters. For matrix M_1 the signal range for $\bar{R'}$, $\bar{G'}$ and $\bar{B'}$ should remain equal to the range of the input RGB, because the nonlinear conversion function is only defined for zero to ten thousand. Thus, the sum of the coefficients in each row of matrix M_1 must equal one and colors within the Rec.2100 gamut should not result in negative values. M_1 can therefore be reduced from 9 to 6 parameters p_1 to p_6 as shown in Equation (4.4):

$$M_{1} = \begin{bmatrix} p_{1} & p_{2} & 1 - p_{1} - p_{2} \\ p_{3} & p_{4} & 1 - p_{3} - p_{4} \\ p_{5} & p_{6} & 1 - p_{5} - p_{6} \end{bmatrix}$$
(4.4)

Matrix M_2 can also be reduced to six parameters as follows: Intensity *I* is formed by the first row, I_1 and I_2 are the contributions of nonlinearly encoded $\bar{R'}$ and $\bar{G'}$ to intensity. For peak-white, $I = \bar{G'} = \bar{B'} = \bar{R'} = 1$ is true. Therefore, the sum of all upper row coefficients must again be 1. Thus, the third coefficient of the first row is dependent on I_1 and I_2 . The second and third rows of the color differencing matrix are calculated by subtracting *I* from $\bar{B'}$ and $\bar{R'}$, which mimics the classic color differencing approach from $Y'C_BC_R$ or IPT and guarantees the plane spanned by the chroma axes to be orthogonal to the achromatic axis. The color-differencing matrix is followed by a free transform in the chroma domain via parameters s_1 to s_4 as shown in Equation (4.5). For better visualization, the chroma channels are rotated around the achromatic axis to make different optimization results comparable:

$$M_{2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & s_{1} & s_{2} \\ 0 & s_{3} & s_{4} \end{bmatrix} \begin{bmatrix} I_{1} & I_{2} & 1 - I_{1} - I_{2} \\ -I_{1} & -I_{2} & I_{1} + I_{2} \\ 1 - I_{1} & -I_{2} & -1 + I_{1} + I_{2} \end{bmatrix}$$
(4.5)

The resulting intensity value will be called *I*, and the two resulting color difference channels C_A and C_B . This naming is meant to reference the IPT and Y'C_BC_R models that inspired the IC_AC_B model.

Now that a color encoding model and the nonlinearity function are determined, the matrix parameters p_1 to p_6 , I_1 , I_2 and s_1 to s_4 have to be found using optimization be-

cause, unlike the nonlinearity function, they cannot be directly deduced. The following paragraphs describe the datasets and respective cost functions used to optimize the 12 parameters for hue linearity, isoluminance and uniform minimum discriminability.

4.2.2 Datasets for Optimization

Most existing datasets for hue linearity, isoluminance and discriminability are limited to standard dynamic range and a much smaller gamut compared to Rec.2100. Thus, wide color gamut datasets have to be acquired for optimization. These new data sets are used as training sets while existing standard dynamic range and limited gamut data sets serve as verification sets. The user study set-up for the acquisition of the new data sets is kept as close as possible to existing studies to be able to verify the results by using the overlapping dynamic range and gamut regions between the existing data sets and the newly acquired data sets.

4.2.2.1 Isoluminance Dataset

Equiluminant color pairs are acquired using the negative-face method [106]. The participants of the study are asked to adjust the luminance of the gray part of a duotone image showing a face that is lit with a directional light at 45° from the top. Figure 4.2 a) and 4.2 b) show the study image for two possible user adjustments of the same color patch. In Figure 4.2 a) the achromatic part of the duotone image is adjusted to be brighter compared to the red reference color. This renders the left face in a correct way while the right face is not recognized as a face because the bright lit parts



Figure 4.2: Image used for the isoluminance user study. Photo authored by *Timo Kunkel.*

are rendered darker compared to shadowed areas. When darkening the achromatic area of the study image as shown in Figure 4.2 b) the left face gets rendered uncanny and the right face is rendered correctly. Kindlmann et al. show that the threshold area in which the face 'flips' is very small and thus allows one to determine isoluminant color pairs very fast compared to traditional flicker studies [106].

The projection setup for the isoluminance study is depicted in Figure 4.3. A 2048 by 1080 pixel DLP-cinema projector is equipped with modified notch-filters to allow sampling a color gamut close to Rec.2100. The spectral power distribution and gamut of this projector are depicted in Figure 3.11 in the preceding chapter. The projected image size is adjusted to a diagonal of 50" to reach a peak white of 2000 cd/m² on a 50" 'Da-lite HD Progressive" [139] vinyl screen with a gain of 0.9. The participant is seated 2 m away from the screen to cover a 30° field of view for the projection and 20° for the actual image excluding an achromatic padding. For each color, the achromatic part of the image is matched using a QUEST [200] procedure. After each matching task, both colors, the reference color and the matched achromatic color are displayed fullscreen and measured via an PR-740 spectroradiometer [166] to avoid errors by color drift of the Xenon lamp driven projector. Nine participants with normal color vision performed this study.



Figure 4.3: *Study setup for deriving equiluminant color pairs and colors with the same hue.*

4.2.2.2 Hue Linearity Dataset

Constant hue lines are acquired in a user study that follows the setup of Hung and Berns [85] with the exclusion of the discrimination field but keeping the gray background and bright border for anchoring as shown in Figure 4.4. The mid-row is user



Figure 4.4: Visual pattern of the hue linearity user study.

adjustable in hue where the goal is to adjust the middle patch to have the same perceived hue as the bottom reference. The luminance of all colored patches was kept constant while the difference in hue is reduced via a staircase method. The projection device and study setup is the same as for the isoluminance study as shown in Figures 3.11 and 4.3. Three users that have previously passed the 38 plate Ishihara test [97] adjusted four levels of saturation for 6 hues and two luminance levels. The results for stimuli inside Rec.709 gamut are close to those from Hung and Berns' original results [85] but the new data extends to the full Rec.2020 gamut and to 2000 cd/m² peak white.

4.2.2.3 JND Dataset

Threshold data for the detection criteria (JND) is acquired using a Dolby PRM4200 professional reference monitor with a black level of 0.005 cd/m^2 , a peak white of 600 cd/m^2 and P3 gamut showing a step-edge pattern and using method of adjustment on the edge amplitude. The pattern is shown in Figure 4.5. The adjustment axes



Figure 4.5: Visual pattern of the JND user study.

used are the P3 RGB primaries and the modelfest axes [204]. For verification, the findings of this study are compared with the JND ellipses of MacAdam [129] and Wright [203] as well as to the Barten Model (adapted for step edges) and the contrast sensitivity function measurements for extended luminance ranges by Kim et al. [103]. The results of the study aligned well with these previous results, but newer studies often report smaller chromatic differences which may be due to the high precision display equipment available today compared to the bipartite 2 degree split field used in the 1940s studies by Wright and MacAdam. The viewing distance is 1.6 m and the screen diagonal 42", resulting in a 30° field of view. As the colored background spans 50 % of the image, the field of view for the chromatic area is 15° and one square spans a 2° field of view.



Figure 4.6: Study setup for measuring JNDs.

4.2.3 Cost Functions for Optimization

Cost functions have to be defined for the optimization of the 12 matrix parameters with the respective datasets. The cost functions are designed as simple as possible because it is found that the selection of the dataset and the weighting of the dataset samples have a much stronger influence on the results compared to varying between different reasonable cost functions.

The **isoluminance cost function** J_{IL} (Equation (4.6)) is defined as the mean squared difference between the predicted intensity of *n* color pairs $I_{(i,1)}$ and $I_{(i,2)}$ that have been adjusted by human observers to have the same perceived luminance. The isoluminance cost function is calculated in the nonlinearly encoded domain of the respective color space:

$$J_{\rm IL} = \frac{1}{n} \sum_{i=1}^{n} (I_{i,1} - I_{i,2})^2 \tag{4.6}$$

The **hue linearity cost function** J_{HL} (Equations (4.7 to 4.9)) is defined as the difference between predicted hue $h_{i,j}$ of each color in the hue linearity data set and the mean hue of all samples *j* from the same tuple that have been adjusted by human observers to have the same perceived hue. Index *i* changes for *n* different hues and *j* changes for the perceptually same hue, but *m* differently saturated patches. The hue difference cost is weighted by the saturation $s_{i,j}$ of each color to prevent the over amplification of small hue changes in pastel tones that result in large changes of numerical hue angle. This weighted hue difference is also normalized by the average saturation of all hues in the dataset to normalize between different data sets. The final hue linearity cost function is calculated as the mean squared weighted difference in hue:

$$h_{i,j} = \operatorname{atan2}(C_{a,i,j}, C_{b,i,j})$$
 (4.7)

$$s_{i,j} = \sqrt{C_{a,i,j}^2 + C_{b,i,j}^2}$$
(4.8)

$$J_{\rm HL} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(s_{i,j} - \frac{1}{m} \sum_{w=1}^{m} h_{i,w} - \frac{1}{m} \sum_{w=1}^{m} h_{i,w}}{\frac{1}{nm} \sum_{w=1}^{n} \sum_{v=1}^{m} s_{u,v}} \right)^{2}$$
(4.9)

On its own, this hue linearity cost function would always end up at the trivial solution of projecting to a plane spanned by I and one color vector. The JND uniformity cost function (J_{JND}) introduced in the next paragraph guarantees the color space will be near to uniform in JND size at all stages of optimization (except for the first few iterations). Thus, the hue linearity cost function from Equation (4.9) works without additional constraints.

The JND uniformity cost function J_{JND} (Equation (4.10)) first performs a Singular Value Decomposition (SVD) on the three ellipsoid half axes $q_{i,j}$ to ensure the half axes are orthogonal after transformation to the current optimization stage of the color space. After performing SVD, the sum of the squared differences between 1 and the length of the individual half axes (normalized by the average length of all half axes) is calculated. This is equivalent to calculating the variance of the length of all half-axes with SVD and normalization applied to the half-axes before:

$$J_{JND} = \frac{1}{3n} \sum_{i=1}^{n} \sum_{j=1}^{3} \left(\frac{|q_{i,j}|}{\frac{1}{3n} \sum_{u=1}^{n} \sum_{\nu=1}^{3} |q_{u,\nu}|} - 1 \right)^2$$
(4.10)

Figure 4.7 illustrates the half axes representation of a two-dimensional ellipse and the effect of singular value decomposition. After transformation to the current color space the length of the half axes only represents the maximum eccentricity of the ellipse when SVD is applied. Figure 4.7 a) shows a hypothetical JND ellipse while Figure 4.7 b) shows the result after a linear transformation. To determine the eccentricity of the ellipse, applying SVD to the half axes results in Figure 4.7 c) where the ratio of length between the two half axes is a measure for eccentricity again.

Opposed to Figure 4.7 a typical color space conversion is a nonlinear transformation. However, most JND ellipsoids are very small so that the error introduced by the nonlinearity is acceptable for the optimization. The half axes approach is compared to representing JND ellipsoids as 10000 points that are randomly distributed on the ellipsoid surface. Transferring these points with the corresponding color space transformations and then fitting an ellipsoid again for optimization leads to very close results. As an example, for the ellipsoids used in the numerical verification of JND uniformity and for the final state of optimization in IC_AC_B , the maximum distance of



Figure 4.7: Example of JND ellipse representation as half axes and the effect of SVD after transformation.

the individual points on the ellipsoid surface to the ellipsoid surface defined by the half axes approximation is below 4% of the mean half axes length and the median error is below 0.3% of the mean half axes length. Only converting the half axes compared to converting 10000 points on the ellipsoid surface reduces the time for optimization by more than two orders of magnitude without significantly changing the outcome of the optimization. Hence, for the final optimization, only the elipsoid center and three points on the ellipsoid surface representing the half axes are transformed.

The **total cost function** J_{TOTAL} Equation (4.11) is defined as the weighted sum of all three cost functions J_{IL} , J_{HL} and J_{JND} . Weighting by w_{IL} , w_{HL} and w_{JND} controls the trade-off between partly contradicting requirements like hue linearity and JND uniformity:

$$J_{\rm TOTAL} = J_{\rm IL} w_{\rm IL} + J_{\rm HL} w_{\rm HL} + J_{\rm JND} w_{\rm JND}$$
(4.11)

4.2.4 **Optimization Process**

The optimization process proves to be robust. When all 12 parameters are initialized with uniformly distributed random values from their respective valid ranges, about

80% of the optimization runs converge to the same result (if rotation around the achromatic axis is ignored).

As an example, Figure 4.8 shows the optimization results when changing the weighting of the cost function from hue linearity to JND uniformity. The left column shows the MacAdam JND ellipses for observer PGN. In a color space that is perfectly uniform with respect to JND discriminability for this observer, and given no noise in the acquisition of the data, all ellipses would be rendered as circles and would have the same size. The right column shows one sample from the new hue linearity data set for observer JFR at the respective luminance of near Rec.2100 primaries for 1750 cd/m² peak white. A perfectly hue linear space would result in straight lines.

As shown in Figure 4.8 a) and b) when weighting the hue linearity cost function over JND uniformity, the resulting color space is relatively hue linear, but the lowest JND ellipse in the blue area is squeezed. When putting a stronger weight on JND uniformity like in the bottom row of Figure 4.8 e) and f), the JND ellipses are rendered more uniform relative to Figure 4.8 a) and b), especially in the blue areas, while hue nonlinearities can be observed in the blue areas. These optimization results indicate that having uniform JND ellipsoids in the blue areas and retaining hue linearity at the same time is not possible for the IC_AC_B model. Lissner and Urban [122] suggest this issue is a fundamental problem for any color space model when trying to co-optimize for both JND uniformity and hue linearity.

4.2.5 Parameter Derivation for the IC_AC_B Model

The goal of the optimization process is to get one model as result. Hence, the weights $w_{\rm IL}$, $w_{\rm HL}$ and $w_{\rm JND}$ have to be determined. Changing the weight of the isoluminance cost function $w_{\rm IL}$ does not yield in strong variations. For trading off hue linearity and JND uniformity, it is decided to go with the candidate that offers the best hue linearity with a constraint on the concavity in the blue areas. This concavity is limited so that a projection of each color within Rec.2100 color space to the gray axis always stays inside the Rec.2100 gamut volume. This constraint is introduced to allow the application of hue-preserving gamut mapping algorithms without the risk to map in-gamut colors to out-of-gamut colors. This combination of optimization for hue-linearity and the



Figure 4.8: Trading JND uniformity for hue linearity in $IC_A C_B$ optimization. The JND ellipsoids are the MacAdam ellipses, viewer PGN. The hue linear lines are from the hue-linearity study, viewer JFR.

constraint on concavity in the blue areas yields a compromise between hue linearity and JND uniformity.

The conversion formula from CIE 1931 XYZ 1931 to the IC_AC_B model is given in Equations (4.12) to (4.14). Valid input values include all XYZ values from a color space spanned by the Rec.2100 primaries and D65 10000 cd/m² peak white (X = 9505, Y = 10000 and Z = 10891). f_{PO} is introduced in Chapter 2, Equation (2.10):

$$\begin{bmatrix} \bar{R} \\ \bar{G} \\ \bar{B} \end{bmatrix} = \begin{bmatrix} 0.37613 & 0.70431 & -0.05675 \\ -0.21649 & 1.14744 & 0.05356 \\ 0.02567 & 0.16713 & 0.74235 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$
(4.12)

$$\bar{R'} = f_{PQ}(\bar{R})$$
 $\bar{G'} = f_{PQ}(\bar{G})$ $\bar{B'} = f_{PQ}(\bar{B})$ (4.13)

$$\begin{bmatrix} I \\ C_{\rm A} \\ C_{\rm B} \end{bmatrix} = \begin{bmatrix} 0.4949 & 0.5037 & 0.0015 \\ 4.2854 & -4.5462 & 0.2609 \\ 0.3605 & 1.1499 & -1.5105 \end{bmatrix} \begin{bmatrix} \bar{R'} \\ \bar{G'} \\ \bar{B'} \end{bmatrix}$$
(4.14)

4.2.6 Parameter Derivation for the IC_TC_P Model

When comparing the results from optimizing the IC_AC_B parameters with the original IPT parameters there are three main differences: The basis for nonlinear coding (\bar{R} , \bar{G} and \bar{B}) is different compared to IPT's LMS, the nonlinearity f is PQ opposed to IPT's gamma function and the contributions of the nonlinear coded components \bar{R}' , \bar{G}' to intensity I are roughly 0.5 while \bar{B}' has no significant contribution to intensity. Compared to IC_AC_B , the approach for IC_TC_P is to apply these learnings from the optimization process of IC_AC_B back to the original IPT model instead of directly optimizing the parameters.

4.2.6.1 Minimal Contribution of the S-Cone to Luminance

The original IPT color space uses 0.4: 0.4: 0.2 as contributions of L'M'S' to intensity. Vision science literature indicates that the S-cone's contribution to luminance is only very little [194, 195] or none [59]. The optimization results in Section 4.2.5 are in line

with these findings assigning the S-cone (\overline{B}) virtually no contribution to luminance. While the weighting of L to M are said to be 2:1 [195], the IC_AC_B optimization process yields about equal contributions of \overline{R} and \overline{G} to luminance. This may be a result from the crosstalk term between L and M, or a result from the limitations of the model. When manually constraining the L and M contributions to I to 2:1 during the optimization process, changes to the color space and the metrics are only minor, because of the large overlap between the L and M curves.

As a consequence the parameters I_1 and I_2 from Equation (4.5) are set to 0.5 each for the IC_TC_P model.

4.2.6.2 Crosstalk in LMS-Domain

The second observation is made on the basis for nonlinear encoding \bar{R} , \bar{G} and \bar{B} . To illustrate the different bases for IC_AC_B and IPT, IC_AC_B's matrix M_1 will be split into IPT's XYZ-to-LMS matrix from Equation (2.26):

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.4002 & 0.7075 & -0.0807 \\ -0.2280 & 1.1500 & 0.0612 \\ 0.0 & 0.0 & 0.9184 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$
(2.26)

and the remainder:

$$\begin{bmatrix} \bar{R} \\ \bar{G} \\ \bar{B} \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} -0.0457 & 0.0254 & 0.0204 \\ 0.0204 & -0.0148 & -0.0056 \\ 0.1088 & 0.0784 & -0.1874 \end{bmatrix} \right) \begin{bmatrix} L \\ M \\ S \end{bmatrix}$$
(4.15)

The splitting of the matrix M_1 shown in Equation (4.15) illustrates that the $IC_A C_B$ optimization yields values close to IPT's original LMS values plus a small crosstalk term. Consequently, the primaries for $IC_A C_B$: \bar{R} , \bar{G} and \bar{B} can be seen as LMS cone fundamentals plus crosstalk.

Figure 4.9 illustrates the effect of the crosstalk term in linear LMS domain. It compresses the chroma plane depending on the saturation of the color. Hence, looking at the JND ellipsoids in Figure 4.9 a), c) and e) it can be observed that the tonal resolu-



Figure 4.9: Influence of crosstalk in linear LMS domain on gamut shape, JND elipsoids and hue linear lines. The JND ellipses are 10 times amplified examples from the JND study, participant JFR. The hue linear lines are one sample from the new hue linearity data set for observer JFR at the respective luminance of near Rec.2100 primaries for 1750 cd/m² peak white.

tion of the color difference channels stays about the same for achromatic colors, while saturated colors are compressed with increasing addition of crosstalk in linear LMS domain. This results in a more uniform quantization along the chroma planes. Figures 4.9 b), d) and f) show that the introduction of a limited amount of crosstalk (4%) improves hue linearity in the red areas, while more crosstalk (10%) worsens hue linearity in the blue areas. The IC_TC_P model will inherit the 4% crosstalk as illustrated in Figure 4.9 c) and d).

Figure 4.10 gives a numerical example for the inefficiency of using grayscale encoding functions like PQ for encoding color primaries like RGB or LMS without crosstalk. For RGB triples near the achromatic axis as shown in Figures 4.10 a) and 4.10 b) adding one 10-bit code-value to the green channel results in luminance and color changes around the visual threshold. However, for saturated colors like in Figure 4.10 c) the change of one 10-bit code value in the green channel is multiple orders of magnitude below the detection threshold because the human visual system is adapted to a much higher luminance level resulting in over-quantization of saturated colors. The introduction of crosstalk in linear domain alleviates this issue because the values of the three color channels are evened out for saturated colors.



Figure 4.10: *Example of JND non-uniformity when encoding Rec.2100 RGB primaries with PQ.*





Figure 4.11: Spectral sensitivity of of the IC_AC_B and IC_TC_P primaries \bar{R} , \bar{G} and \bar{B} compared to the LMS cone fundamentals.

it compensates for the fact that the PQ nonlinearity was developed using stimuli near the achromatic axis where L, M and S are approximately the same relative to D65 white. Applying the PQ nonlinearity on each cone separately assumes infinite individual LMS gain (chromatic adaptation). However, chromatic adaptation is limited compared to lightness adaptation [108]. The crosstalk term can also be thought as a different implementation of the noise term in Guth's ATD model [78] or the limitation in chromatic adaptation $p_{L,M,S}$ in Fairchild's chromatic adaptation model [60] from 1991.

From a third perspective, crosstalk can be seen as a step towards calculating color difference channels from linear primaries as done in the CIE 1931 xy and CIE 1976 u'v' models. When theoretically introducing near $\frac{1}{3}$ crosstalk, followed by large coefficients in the second and third row of M_2 the chroma plane converges towards chromaticity calculated from a linear basis.

In conclusion, applying crosstalk to the encoding primaries can be used to optimize the coding efficiency of static color encoding models. Based on these findings a constant 4 % amount of crosstalk between all three linear LMS channels from Equation (4.16) is added to IC_TC_P in comparison to the original LMS primaries of IPT.

4.2.6.3 Full IC_TC_P Model

Besides the addition of 4% crosstalk in linear LMS domain and setting the contributions of L, M and S to intensitiy I to 0.5:0.5:0, the original color differencing matrix of IPT is modified by rotating typical skintones to align with their hue angle in SDR Y'C_BC_R for better backwards compatibility [48]. Also the C_T color difference channel is multiplied by 1.4 to use the full code space of -0.5 to 0.5 for both chroma channels [48] when encoding Rec.2100 gamut. The resulting color differencing matrix is written in Equation (4.18). IC_TC_P is standardized in ITU Rec.2100 [96] as an alternative color differencing scheme to Y'C_BC_R.

$$\begin{bmatrix} \bar{R} \\ \bar{G} \\ \bar{B} \end{bmatrix} = \frac{1}{4096} \begin{bmatrix} 1688 & 2146 & 262 \\ 683 & 2951 & 462 \\ 99 & 309 & 3688 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$
(4.16)

$$\bar{R'} = f_{PQ}(\bar{R})$$
 $\bar{G'} = f_{PQ}(\bar{G})$ $\bar{B'} = f_{PQ}(\bar{B})$ (4.17)

$$\begin{bmatrix} I \\ C_{\rm T} \\ C_{\rm p} \end{bmatrix} = \frac{1}{4096} \begin{bmatrix} 2048 & 2048 & 0 \\ 6610 & -13613 & 7003 \\ 17933 & 17390 & -543 \end{bmatrix} \begin{bmatrix} \bar{R'} \\ \bar{G'} \\ \bar{B'} \end{bmatrix}$$
(4.18)

4.3 Results

In the following section the IC_AC_B and IC_TC_P models are compared visually and numerically to state-of-the-art WCG and HDR encodings Y"u"v", Rec.2100 HLG and Rec.2100 PQ. Rec.2100 HLG and Rec.2100 PQ are standardized by the ITU in Rec.2100 [96] for video coding. Y"u"v" is proposed by Poynton et al. on behalf of Philips [161]. These state of the art encodings are further described in Section 2.4.4. First, a visual comparison of JND uniformity, hue linearity and isoluminance if performed by plotting JND ellipsoids, hue linear lines and by comparing the achromatic channel of the respective color spaces to CIE 1931 Y. This is followed by numerically comparing the encoding performance of the IC_AC_B and IC_TC_P models to Y"u"v", Rec.2100 HLG and Rec.2100 PQ via a new metric for comparing color space quantization efficiency. In addition, isoluminance will be compared by rank order correlation.

4.3.1 Visual Comparison

In this section, JND uniformity, hue linearity and isoluminance are compared by visualization of the reference data sets. Figure 4.12 and 4.13 both show the Rec.2100 1000 cd/m^2 peak white gamut hull in an orthographic view along the achromatic axis. For the visual verification of JND uniformity, all chromaticity ellipses from MacAdam's PGN observer [129] that are located inside the Rec.2100 gamut or very near to the gamut boundary are displayed in Figure 4.12. These JND ellipses are amplified by a factor of 10. The more regular the ellipses are in size the more uniform the space is with regard to JNDs. Y''u''v'', IC_AC_B and IC_TC_P render the minimum discriminability ellipses more uniform than Rec.2100 HLG and Rec.2100 PQ.

In Figure 4.13 the Hung and Berns hue linearity data set is overlaid to the Rec.2100 1000 cd/m^2 peak white gamut hull. Y''u''v'', Rec.2100 HLG, IC_AC_B and IC_TC_P perform better compared to Rec.2100 PQ for the luminance of the Hung and Berns data set. According to the wide color gamut hue linearity data the hue curvature for Rec.2100 PQ and Rec.2100 HLG intensifies for higher luminance, especially in the blue areas while hue linearity for Y''u''v'', IC_AC_B and IC_TC_P stays about the same.

For comparison of isoluminance Figure 4.14 shows CIE 1931 Y luminance plotted versus the luma/intensity value of the respective color space for 10000 random colors inside the Rec.2100 1000 cd/m² peak white gamut. All axes are converted to the PQ domain. Rec.2100 HLG and Rec.2100 PQ represent saturated colors with a large offset in luminance. For saturated blues, the error between luma and luminance in these two color spaces can be multiple f-stops. Y''u''v'' does not introduce isoluminance errors by design because Y'' is calculated from linear RGB. IC_AC_B and IC_TC_P only introduce low isoluminance errors below a quarter of an f-stop for IC_AC_B and below one f-stop for IC_TC_P.

In addition to the visual and numerical analysis, color grading as well as gamut mapping and tone mapping are performed using the wide gamut scenes from the HdM-



Figure 4.12: Comparison of JND uniformity between current state-of-the-art HDR and WCG video encodings and IC_AC_B and IC_TC_P .



Figure 4.13: Comparison of hue linearity between current state-of-the-art HDR and WCG video encodings and IC_AC_B and IC_TC_P .



Figure 4.14: Comparison of isoluminance between current state-of-the-art HDR and WCG video encodings and IC_AC_B and IC_TC_P .

HDR-2014 dataset [71] as well as a custom database of wide color gamut and high dynamic range image sequences. As an example, Figure 4.15 compares the effect of desaturation in the investigated color spaces. Desaturation is a typical operator happening in color grading, gamut mapping and tone mapping. The example image is frame 962 of the 'Carousel Fireworks 02' scene from the reconstructed scene radiance. Desaturation is rendered as orthogonal projection of the chroma values to the achromatic axis in the respective color space. The rendering from high dynamic range and wide color gamut to the SDR sRGB gamut of this document is performed via the 'ARRI Look File 2' rendering transform available on-line from the ARRI website [53]. As expected from the comparison of isoluminance in Figure 4.14, Rec.2100 HLG and Rec.2100 PQ (Figure 4.15 c) and d)) show visible artifacts, rendering saturated blue, magenta and red areas too dark in the desaturated image. The luminance change of IC_AC_B and IC_TC_P compared to CIE 1931 Y as used in Y''u''v'' can only be seen when toggling between the reference and the versions desaturated in IC_AC_B or IC_TC_P .

To illustrate the gamut mapping performance of Y"u"v", Rec.2100 HLG, Rec.2100 PQ, IC_AC_B and IC_TC_P , Figure 4.16 shows frame 96163 of the 'Carousel Fireworks' sequence from the color graded version of the HdM-HDR-2014 data set. The Rec.2100 $0-4000 \text{ cd/m}^2$ gamut of the original color graded image is mapped to a $0-4000 \text{ cd/m}^2$ RGB gamut with Rec.709 primaries. To fit into the luminance range of the SDR sRGB gamut of this thesis, the peak white of the mapped images is linearly scaled to the maximum luminance of this document. 'Weighted minimum delta E' [101] is used as an example gamut mapping algorithm using weights of 2, 1 and 4 on change of luma, chroma and hue-angle. These weights put a higher penalty on hue changes and lower the penalty on chroma differences. Figure 4.16 a) is not gamut mapped but clipped in Rec.709 RGB domain. The clipping algorithm shows hue changes from cyan green to pure green. However it is best in retaining saturation. Gamut mapping in Rec.2100 HLG or Rec.2100 PQ domain using the 'Weighted minimum delta E' algorithm as shown in Figures 4.16 c) and d) introduces hue changes and luminance discontinuities. Y''u''v'', IC_AC_B and IC_TC_P (Figure 4.16b), e) and f)) perform robust without showing typical problems like hue nonlinearities or luminance errors.





c) Rec.2100 HLG



Objects Rendered too Dark

Luminance Discontinuities



Figure 4.15: Comparison of applying the desaturation operator in current state-of-the-art HDR and WCG video encodings and IC_AC_B and IC_TC_P . Best viewed on a color accurate sRGB monitor in a dark environment.



Figure 4.16: Comparison of gamut mapping performance in current state-ofthe-art HDR and WCG video encodings and IC_AC_B and IC_TC_P . Best viewed on a color accurate sRGB monitor in a dark environment.

4.3.2 Numerical Comparison

In this section the IC_AC_B and IC_TC_P models are compared numerically to state-of-theart WCG and HDR encodings Y''u''v'', Rec.2100 HLG and Rec.2100 PQ. The used reference data sets for numerical comparison are existing datasets to keep the training data for IC_AC_B (the data sets acquired herein as described in Section 4.2.2) separate from the data sets for verification.

For the numerical comparison of the coding efficiency, JND ellipsoids are built at different luminance levels by replicating the MacAdam PGN ellipses [129] to 0.02, 0.2, 2, 40 and 200 cd/m² and scaling them by the maximum color contrast sensitivity function findings from Kim et al. [103] in the chromaticity domain relative to the results at 40 cd/m² from the corrected version of this paper. The altered version is available from co-author Mantiuk's website [104].

Encoding efficiency can be calculated by dividing the needed range per axis by the largest allowed quantization step below the visibility threshold. Rec.2100 with 1000 cd/m^2 peak white gamut is chosen to determine the minimum and the maximum expected values per axis. In contrast to the minimum and maximum value per axis, the largest visually lossless quantization step cannot be separately determined per axis because the needed quantization along one axis depends on the quantization along the other axes. Figure 4.17 illustrates this by showing two visually lossless quantization step sizes for the same JND ellipse.



Figure 4.17: Illustration of two different tonal resolutions both resulting in visually lossless quantization for the same JND ellipse.

In Figure 4.17 a) the coarse quantization along axis 2 demands a finer quantization along axis 1 if all combinations of one code value should stay below the visibility threshold while in Figure 4.17 b) the fine quantization along axis 2 relaxes the tonal resolution requirements along axis 1. To guarantee any quantization error stays below the detection threshold, every combination of changes of one code value along all axes must fall inside all JND ellipsoids.

With 3 dimensions used for typical color difference encodings this translates into higher quantization needs per axis compared to only looking at one axis. As an example, Figure 4.18 shows a uniform sphere and the largest axis aligned box that fits inside this sphere. For this example sphere, the maximum quantization along each axis that is below one JND for all 1 code value step combinations is half the size of the axis aligned box inside the ellipsoid. For this example of a sphere it is $\frac{1}{\sqrt{3}}$ the radius of the sphere. It can be concluded that when measuring minimum quantization per axis, the needed tonal resolution for the full three dimensional color space is larger compared to the quantization needed along each separate axis.



Figure 4.18: Deriving the largest possible visually lossless quantization step from JND ellipsoids by scaling the axis aligned bounding box to fit inside the JND ellipsoid.

To determine the amount of quantization steps needed for visually lossless encoding of the color spaces compared herein, the axis-aligned bounding boxes of each ellipsoid are scaled to fit inside the respective ellipsoid. The needed quantization step is then determined by the minimum of half the length of all boxes along each of the three axes. Different metrics can yield different results, for example trading quantization efficiency along the achromatic axis in favor of needed code values for chroma (or vice versa). Table 4.1 compares the encoding efficiency for the investigated color spaces using the scaled bounding box metric as described above. Fewer needed code values indicate a more efficient encoding.

Code Values Needed to Encode:							
Color Space	Lum	a/Intensity	Colo	or Difference 1	Colo	or Difference 2	
Y″u″v″	Y″:	2370	u″:	1829	v":	942	
Rec.2100 HLG	Y':	7009	C_{B} :	9432	C_{R} :	7554	
Rec.2100 PQ	Y':	2816	C _B :	3526	C_{R} :	3526	
IC _A C _B	I:	2147	C _A :	1621	C _B :	832	
IC _T C _P	I:	1857	C_{T} :	1171	C_p :	1843	

Table 4.1: Comparison of color space encoding efficiency. Minimum number of code values needed for encoding Rec.2100 1000 cd/m^2 peak white gamut.

Due to the PQ nonlinearity on the achromatic axis, Y"u"v", IC_AC_B and IC_TC_P provide the most efficient luma encoding. Y"u"v", IC_AC_B and IC_TC_P also need less code values for chroma compared to the other color spaces investigated here. The high number of code values needed to encode the chroma channels of Rec.2100 PQ results from the skewed JND ellipsoids for saturated colors compared to the small JND ellipsoids near the achromatic axis. The high amount of code values for Rec.2100 HLG is due to the comparatively coarse quantization of the square-root transfer curve in the dark areas of the Rec.2100 HLG color space. Y"u"v", IC_AC_B and IC_TC_P need 11 to 12 bits for visually lossless encoding while Rec.2100 PQ requires 12 bits and Rec.2100 HLG needs 13 to 14 bits of tonal resolution per channel for visually lossless encoding of any content.

As a metric for isoluminance, Table 4.2 shows the Spearman rank order correlation for 10000 random colors within Rec.2100 1000 cd/m² peak white gamut. As already observed in Figure 4.14 Rec.2100 PQ perfomes worst, followed by Rec.2100 HLG. Y''u''v'' does not introduce any isoluminance error by design and IC_AC_B and IC_TC_P represent colors with an isoluminance error of about a quarter and half an f-stop. An evaluation of compression efficiency of the IC_TC_P color space has been performed, using the HdM-HDR-2014 data set introduced in Chapter 3. It is shown that IC_AC_B

	Spearman Rank	Maximum Luminance
Color Space	Order Correlation	Error in F-Stops
Y″u″v″	1	0
Rec.2100 HLG	0.9396	5.41
Rec.2100 PQ	0.8721	10.06
IC _A C _B	0.9994	0.17
IC _T C _P	0.9971	0.54

Table 4.2: Comparison of color space isolumance. Spearman rank order correlation between CIE 1931 luminance and luma of the respective color space for 10000 random colors within Rec.2100 1000 cd/m² peak white gamut.

and IC_TC_P also enhance coding efficiency in scenarios beyond baseband encoding of uncompressed image streams [125, 206].

4.4 Discussion and Outlook

The performance of the proposed $IC_A C_B$ and $IC_T C_P$ color spaces for encoding high dynamic range and wide gamut video and performing color volume mapping is promising [162]. Coding efficiency is better compared to approaches with the same computational complexity like Rec.2100 HLG and Rec.2100 PQ and on the same level with Y''u''v'', which has a higher computational complexity. Hue linearity is improved compared to all other approaches and isoluminance is close to constant luminance like employed in Y''u''v''.

Future display systems may feature an extended gamut compared to Rec.2100. Figure 4.19 shows the gamut of visible colors (between 380 and 780nm), below 56.6 $\frac{w}{sr \ cm^2}$ and below 10000 cd/m² in Y''u''v'', Rec.2100 HLG, Rec.2100 PQ, IC_AC_B and IC_TC_P. It can be observed that Y''u''v'', IC_AC_B and IC_TC_P provide an efficient encoding for the full spectral gamut limited to a certain luminance and energy. This makes these color spaces future proof. Still, JND uniformity, isoluminance and hue linearity beyond the Rec.2100 gamut have to be verified. Also the resulting code ranges for the full spectral gamut have to be implemented correctly.


Figure 4.19: Encoding color gamuts beyond color spaces spanned by three physically realizable primaries. Rec.2100 10000 cd/m² gamut is compared to encoding all visible colors (between 380 and 780nm), below 56.6 $\frac{w}{sr \text{ cm}^2}$ and below 10000 cd/m².

While the $IC_A C_B$ color space and its sibling $IC_T C_P$ improve upon the existing proposals for high dynamic range video encoding presented in Chapter 2.4.4, still there is room for improvement. It would be desirable to evaluate and eventually optimize the transformation matrices again with wider data sets for isoluminance, JND uniformity and hue linearity. After the publication of the original paper [73] Safdar et al. optimized the same model using additional datasets [170]. Still no JND data sets that span a wide color gamut near to the spectral locus are available except for the step-edge data from Wright [203]. Thus, the minimum JND study should be repeated for the full spectral gamut using a more robust psychometric method such as two-alternative forced choice instead of the method of adjustment, that was utilized in the minimum JND study herein.

A limitation of the IC_AC_B and IC_TC_P color spaces is their optimization on noiseless signals. Camera captured images are subject to photon shot noise. This noise masks quantization artifacts and therefore reduces bit-depth requirements for dark parts of the image. In the following chapter the implications of noise and texture on the needed tonal resolution of the I channel of IC_AC_B and IC_TC_P are studied. Ideally, the methods presented in the next chapter should be part of the color space definition to eliminate these separate post-processing steps.

Chapter 5

Content Aware Quantization Scheme -'CAQ'

As presented in Chapter 4 visually lossless HDR video encoding for noiseless content needs more than 10 bits of tonal resolution. Still current video distribution pipelines and storage formats are limited to 10 bits in their mainstream flavors. In this chapter, a novel content aware quantization scheme will be introduced. It can be used for reducing quantization needs by exploiting noise and texture inherent to most images. This chapter is based on the paper introducing the **C**ontent **A**ware **Q**uantization scheme (CAQ) [75] and the related patent [74].

While advances in image quality previously focused on increasing spatial resolution [94], the focus nowadays shifts to extending dynamic range [185], tonal resolution (i.e., bit-depth) [128] and color gamut [92]. There has been extensive research on the most efficient quantization scheme for encoding HDR baseband signals [151]. As described in Section 2.3.3.3 the resulting PQ encoding curve [184] needs 12 bits of tonal resolution to quantize any content in its 0-10000 cd/m² range without introducing visible quantization artifacts. The bit-depth and tonal nonlinearity of PQ are determined by threshold visibility criteria and the most demanding imagery (low gradients having zero noise and no texture). Current monitors can reproduce this tonal resolution of 12 bits by using a 10-bit LCD panel with a local backlight array [152].

5.1 Research Question

While 12 bits of tonal resolution is the goal for HDR and WCG image distribution, there are key technologies requiring a lower bit-depth. As examples, interfaces like Display-Port, HDMI and HD-SDI [182] only support 10 bits of tonal resolution for typical applications. File formats like MXF [179] and DPX [183] are also limited to 10 bits in their most used flavors, as are compression codecs like H.264 [95] and H.265 (HEVC) [91]. Therefore it would be desirable to have a quantization scheme available that can quantize any content at 10 bits or lower without introducing visible artifacts. Figure 5.1 illustrates the need for reducing tonal resolution and shows the locations for the bit-depth reduction and expansion in the image processing chain.



Figure 5.1: Block diagram for the application of CAQ.

5.2 Related Work

When aiming to reduce the number of code values needed for HDR image quantization below 12 bits, knowledge about the content, the observer or the viewing environment must be exploited [151]. Prior approaches to this problem either assume static parameters for content and viewing environment [18, 19] or are histogram based [124], prioritizing tonal resolution on those parts of the tone scale with many pixels. Other approaches operate in the frequency-domain [145, 168] and are therefore only applicable to compression, whereas the goal herein is uncompressed baseband transmission and file storage or pre-processing for compression. It is known from the literature on image difference metrics [38, 197] that the visibility of small differences (like they occur in quantization) depends on the local noise and texture. Since the PQ transfer function is intentionally designed to exclude noise, CAQ will be designed to exploit the phenomenon of 'masking' [157, 201] of small differences by noise and texture to reduce the needed tonal resolution per image. For camera-captured images, Poisson distributed photon shot noise [173, 175] is a strong contributor to masking of small differences.

5.3 Methods

In this section the dependency of needed tonal resolution on image properties will be examined by conducting a study on visibility of quantization artifacts for synthetic gradient images. From the findings of this study the CAQ re-quantization method will be derived and compared to determining quantization thresholds by using a current state-of-the-art image difference metric as well as an image based user study.

To explore the relationship between noise and required quantization, first, a study with synthetic shallow gradients as test target is performed. A variable amount of spatially uncorrelated additive white Gaussian noise is added to the gradient, starting with zero noise. As can be seen in Figure 5.2 the same tonal resolution reduction can show visible contouring (5.2 a) to b)) or be visually lossless (5.2 i) to j)) depending on the noise amplitude on the gradient prior to quantization. In this study, the slopes of the gradients are calculated so that the spacing of the quantization steps covers a range of frequencies around the peak contrast sensitivity of the human eye at the respective luminance and the viewing distance of one picture height.



Figure 5.2: Test pattern for the gradient quantization study in order to evaluate the visibility of false contours. The variance of noise σ^2 is denoted relative to the magnitude of one quantization step.

The three participants of this study are asked to identify the coarsest quantized image for which no change could be spotted when self-toggling [83] between the original gradient with added Gaussian noise (Figure 5.2 a), 5.2 e) and 5.2 i)) and the same gradient at reduced tonal resolution (Figure 5.2 b), 5.2 f) and 5.2 j)). The investigated variable parameters are listed in Table 5.1. One participant had 20/20 vision while the other two participants wore glasses. The participants spanned age corridors from 20-30, 30-40 and 40-50. The study setup is depicted in Figure 4.6 in the preceding chapter. The surround was kept dark to allow for up to 3 minutes of cone adaptation for the dark stimuli.

Noise parameters for gradient quantization study		
Mean luminance	$0.01, 0.1, 1, 10, 100, 300 cd/m^2$	
Temporal frequency	0 fps (still image), 24 fps	
Spatial bandwidth	20, 10, 5 cycles per degree	
Amplitude	0, 1, 2, 4, 8, 16, 32, 64 standard	
	deviation σ in 12-bit code-values	
Quantization (tonal resolution)	$q = 5$ to 12 for 2^q code values	
	to encode the full PQ range	

 Table 5.1: Variable parameters of the gradient quantization study.

All combinations of the parameters in Table 5.1 are evaluated. The results of the study are shown in Figure 5.3. The black dots correspond to the minimum bit-depth needed for visually lossless quantization of the gradients. While luminance, temporal and spatial frequency of the noise (Figure 5.3 a), b) and c)) have a low impact on required tonal resolution, it can be seen in Figure 5.3 d) that the required bit-depth is inversely correlated to the amplitude of the noise. As an example, PQ-encoded image areas containing white Gaussian noise with a standard deviation of four 12-bit code values can be quantized using 9 bits of tonal resolution without showing any visual difference for all the other parameter combinations.

5.3.1 CAQ Based Quantization

In this section the CAQ quantization scheme is introduced. CAQ is designed to exploit the masking of quantization artifacts by noise and texture as shown in Figure 5.3 d)



Figure 5.3: *Results of the gradient quantization study. To illustrate the correlation of the needed bit-depth for visually lossless encoding to the parameter on the x-axis, lines connect study results that only differ in this parameter.*

to reduce the number of code values needed to quantize an individual image. As for camera-captured images the photon shot noise is related to luminance, CAQ predicts the required quantization for 8 equally spaced intensity segments from 0 to $\frac{8}{9}$ in the PQ domain (0 to 3524 cd/m^2). Commercial implementations may feature more segments. Since CAQ predicts the required quantization per intensity segment, but spatially global, it can be applied to the image and removed by means of a simple one-dimensional Lookup Table (LUT). For modern codecs like H.265/HEVC the reverse LUT can be embedded as SEI message [91] to undo a variable tone curve at the receiver side. SDI supports ancillary data [180] and DPX and MXF also support user defined meta data for transport of the decoding LUT. The block diagram of the CAQ analysis is shown in Figure 5.4. The input for the algorithm is the PQ-encoded intensity channel. To estimate local noise and texture an isotropic Gaussian high-pass filter (Figure 5.4 b)) with a standard deviation of 2.5 pixels for 1920 by 1080 high definition images is applied to the PQ encoded intensity image (Figure 5.4 a)). After rectification (Figure 5.4 c)), the result is blurred again in Figure 5.4 d) to get a robust estimate of the local masking of quantization artifacts per pixel. These steps are consistent with models [39, 65, 126] of the phase uncertainty properties of the human visual system [29].



Figure 5.4: CAQ noise and texture estimation block diagram.

To calculate the minimum allowed quantization level per pixel, a calibration LUT is applied in Figure 5.4 e). This LUT is derived from the gradient quantization study. The calculation of the calibration LUT is illustrated in Figure 5.5. Each of the data points corresponds to one gradient from the gradient quantization study. The x-axis location shows the CAQ analysis result from Figure 5.4 block d) while the location on the y-axis corresponds to the needed bit-depth for this gradient as found in the gradient quantization study. Applying the LUT in Figure 5.4 e) assigns each pixel a bit-depth that is sufficient to quantize this pixel without a visual difference. To find the minimum allowed quantization per intensity range per image, the needed bit-depth predictions are sorted by the intensity of the corresponding original image pixel into image-dependent histogram bins in Figure 5.4 f). Finally, the minimum allowed quantization for each segment is determined by calculating the maximum needed bit-depths for each bin in Figure 5.4 g).



Figure 5.5: Calculation of the CAQ calibration LUT by determining the minimum required bit-depth for each CAQ analysis results.

5.3.2 HDR-VDP-2.2 Based Quantization

The CAQ design is inspired by the visual difference predictor [38]. A new version of the visual difference predictor for HDR imagery is available as High Dynamic Range Visual Difference Predictor 2.2 (HDR-VDP-2.2) [147]. To evaluate how CAQ performs compared to HDR-VDP 2.2, the needed quantization for each of the 8 intensity segments is also determined by calculating the coarsest quantization that stays below HDR-VDP 2.2's 50% visibility threshold for all pixels. For this application, HDR-VDP 2.2 is preferred over SSIM [197] with PU curve [10] because HDR-VDP can be used in calibrated luminance mode.

5.4 Results

To compare CAQ with re-quantization based on HDR-VDP-2.2, an image based quantization study is performed. The study set-up is shown in Figure 5.6. For this study achromatic still frames are shown on a dual modulation HDR LCD display [152]. Still frames are selected as a worst case scenario because uncorrelated temporal noise and motion also mask quantization artifacts [3]. During image evaluation, the user can adjust the quantization per intensity segment in real-time using physical sliders. To help spot the quantization artifacts, a temporal linearly increasing offset of 1/10 of the current quantization step is added per frame before quantization and subtracted after quantization. This phase-shift of the quantization threshold results in continuously moving contouring artifacts. It helps to find the exact detection threshold because of the 'pop-out' effect [138] of motion. The phase-shift simulates a typical worst case scenario for quantization when an image is followed by a slightly darker version of itself as it occurs in 'fade to black' dissolves.



Figure 5.6: Setup for the image quantization study. The hangar image is © Dolby Laboratories Inc.

The study is performed in a dark room to make sure veiling glare is kept at a minimum level. The viewing distance is one picture height, and the participants are allowed to move their head parallel to the display surface. Eight post-production experts from the TV- and film industry who perform image evaluation tasks every day, adjust the quantization for the 8 intensity segments on 12 images. These images are selected to originate from different technologies (analog film, digital cameras and computer animation) and include Hollywood movies, commercials and TV programs. Table 5.2 shows the technical parameters for a subset of example images used for the image

based quantization study. The results of CAQ, HDR-VDP 2.2 and the image based quantization study for these eight images are shown in Figure 5.8 at the end of this section.

	Scene Name	Acquisition	Image Description
	Camera	Medium	
a)	'Hangar'	ARRIRAW 2.8K	View from a hangar into the
	ARRI Alexa		sun with a pilot's silhouette.
b)	'Fantasy Flight'	ARRIRAW 2.8K	Man standing in front of a
	ARRI Alexa		painting.
c)	'2002 B-Movie'	Kodak Vision	Women dancing on stage in
	Arricam LT and ST	200T, 500T	front of a wall of light bulbs.
d)	'2014 A-Movie'	ARRIRAW 2.8K	Man lying on the ground in a
	ARRI Alexa		dark atmosphere
e)	'2009 Kids Film'	CG Animation,	Dark animated jungle
	Toon Boom Harmony	Rendering, 2K	illuminated by fireflies
f)	'2006 A-Movie'	Kodak Vision2	Sorcerer on stage illuminated
	Panavision Mill. XL2	250D, 500T	by blue searchlights
g)	'Flirting with Fire'	4K RAW	Explosive flame and fireball
	Phantom Flex4K		
h)	'Showgirl'	2*ProRes	Young woman illuminated by
	2*ARRI Alexa	4:4:4 HD	directional stagelight

Table 5.2: Technical parameters of eight example images from the image based quantization study shown in Figure 5.8.

Three participants of the image based quantization study have 20/20 eyesight, four users perform the study with glasses, and one user wears contact lenses. The participants have a mean age of 40. The study setup is the same as depicted in Figure 4.6 in the preceding chapter but with a slightly closer viewing distance of one picture height.

The correlation between the image quantization study and CAQ as well as the HDR-VDP-2.2 based quantization is compared in Figure 5.7. Predictions below the dashed line may result in visible quantization errors, because the predicted tonal resolution is less than required by the observers of the image quantization study. Compared to the image quantization study and except for the darkest segment, CAQ delivers a better prediction of the needed quantization compared to the HDR-VDP-2.2 based requantization, CAQ predictions tend to be closer to the visibility threshold and mostly above, as needed for worst-case design. For the darkest segments CAQ often over-



Figure 5.7: Tonal resolution prediction by HDR-VDP-2.2 and CAQ compared to the image quantization study.

predicts the required bit-depth compared to the verification study because CAQ does not include an observer glare model [33, 167]. This is by design because observer glare depends on the position, orientation and age of the observer [193]. In addition to the uncertainty about the observer, cropping parts of the image, as performed in pan and scan operations or in the processing for small mobile screens, can fully remove sources of glare from the image. Consequently, relying on observer glare for re-quantization would have worked for the controlled set up of the image quantization study, but would fail for the intended usage scenarios of CAQ. Therefore, CAQ only exploits image inherent local texture for re-quantization, to stay independent of the observer and display conditions.

The Spearman rank order correlation between all results from the verification study is 0.78 for CAQ and 0.74 for the HDR-VDP-2.2 based quantization. When omitting the darkest segment, the correlation for CAQ improved to 0.87 while HDR-VDP-2.2 yields 0.70.

The typical results for camera-captured images that are subject to photon shot noise can be observed in Figures 5.8 a) and b). In the 'Fantasy Flight' image in Figure 5.8 b),

the darkest segment requires a tonal resolution of about 6 bits while the bit-depth requirements for the brighter segments constantly rise to about 8 bits of tonal resolution. The 'Showgirl' image 5.8 h) has a bimodal histogram. In this case, CAQ predicts a much higher needed tonal resolution for the darkest segment compared to the user study and HDR-VDP-2.2. This is due to the intended lack of glare prediction in CAQ.

The results from the image based verification study confirm the expectation from the gradient and noise based study. When using variable quantization, most camera captured images need 50-200 code values, while computer generated noiseless content typically needs 200 to 500 code values for visually lossless representation. Using CAQ, all images from the image based study could be quantized to less than 1024 code values. For an extended image set of 200 low-noise video frames, two computer-generated images containing very smooth gradients over the full luminance range needed more than 10 bits of tonal resolution according to CAQ. To avoid visible artifacts in these cases, the local masking prediction map calculated in Figure 5.4 e) can be used to apply local dithering solely to those areas where the needed quantization bit-depth cannot be reached. On an Intel 'Xeon E5 1620 v3' processor the non-optimized MATLAB implementation of CAQ runs about three orders of magnitude faster compared to the HDR-VDP-2.2 MATLAB version available from the HDR-VDP project website [131].

5.5 Summary

High dynamic range imagery with a tonal resolution of 12 bits and more needs to be quantized at bit-depths of 10 bits or less to fit into current image storage and transmission formats. CAQ - a robust and fast method to determine the needed tonal resolution of high dynamic range images by exploiting local noise and texture is presented. By applying the CAQ algorithm, images can be re-quantized by means of a simple one-dimensional lookup table to 10 bits or less. In addition to the lookup table CAQ delivers a map of the necessary bit-depth per pixel. This map can be used to locally apply dithering for those rare cases where more than 10 bits of tonal resolution are needed for visually lossless representation. Compared to using a state-of-the-art image difference predictor, CAQ performs better for estimating needed quantization and needs significantly less computing power.



Figure 5.8: Comparison of the predicted minimum quantization of eight example images from the image based quantization study.

Chapter 6

Conclusion and Outlook

This thesis starts with giving an overview of the open problems in the field of HDR and WCG image acquisition, distribution, storage and display. It then focuses on image encoding by describing the history of luminance and color encoding. This includes an overview of state-of-the-art standard dynamic range and high dynamic range luminance quantization schemes and color spaces.

The original work starts with introducing a new high dynamic range and wide color gamut video data set. This data set has a dynamic range of 18 stops and covers nearly the full visible gamut. Two versions are available to the scientific community, the reconstructed scene radiance as well as a color graded version featuring the full Rec.2100 gamut and a dynamic range of 0.005 to 4000 cd/m². Since its introduction, the HdM-HDR-2014 data set is used to develop new HDR and WCG aware video compression codecs, camera characterization methods, image processing algorithms, video quality metrics, eye tracking data sets, tone mapping algorithms and to evaluate different display technologies.

Subsequently two new HDR and WCG color difference encoding models for quantizing HDR and WCG video like the HdM-HDR-2014 data set are introduced. The IC_AC_B and IC_TC_P color spaces are co-optimized for the decorrelation of luminance and chroma as well as hue linearity and encoding efficiency. The test data and training data is described as well as the cost functions for optimization. It is shown that by using the

proposed optimization methods color encodings can be tailored to satisfy specific requirements like hue linearity and uniform quantization by weighting the cost functions accordingly. Based on the findings from the optimization of the IC_AC_B color space the concept of crosstalk in LMS domain is introduced. Crosstalk in LMS domain serves to emulate limitations in adaptation to static color encoding models. The IC_TC_P color space is derived by applying the learnings about crosstalk in linear LMS domain and new contributions of LMS to intensity I back to the original IPT color space. Both color spaces, IC_AC_B and IC_TC_P are further compared to state-of-the-art HDR and WCG color encodings that are also proposed for the use in HDR and WCG video coding. IC_AC_B and IC_TC_P perform better in both hue linearity and coding efficiency compared to existing models.

These new color spaces still require a tonal resolution of 11 to 12 bits. However, current mainstream video infrastructure is limited to 10 bits. Hence, a re-quantization algorithm is developed to further reduce bit-depth requirements of HDR video. The CAQ texture and noise aware quantization scheme reduces tonal resolution needs for visually lossless storage and transmission of HDR video by only exploiting content characteristics but without restricting presentation size or viewing environment.

In conclusion this thesis provides an HDR and WCG video data set and methods to encode HDR and WCG content for efficient storage, distribution and compression via current and next generation infrastructure.

Remaining Challenges in HDR and WCG Imaging

In **image acquisition** sensors capable of capturing more than 14 f-stops of dynamic range while staying linear for precise color acquisition are not yet available. Reference video content featuring HDR, WCG and spatial resolution beyond high definition and temporal resolution beyond 24 fps will be needed to evaluate the performance of future image manipulation and display technologies. Gamut mapping is especially critical for low-noise gradients. Thus, a computer generated WCG and HDR reference data set is needed.

With increasing computational power in display devices it will be worth searching for more complex **encoding and storage** models. Higher computational complexity may improve the coding efficiency, de-correlation and hue linearity compared to $IC_A C_B$ and $IC_T C_P$ while still staying below the computational complexity of color appearance models. To optimize these new models with more degrees of freedom a large amount of data on human perception is needed. Compared to currently available datasets, this data should extend to the spectral locus and feature an extended dynamic range.

In **distribution** efficient quantization schemes like the ones introduced in this thesis will stay important. Ultra high definition Phase 2 (7680 by 4320 pixels at 120 fps and 12 bits) needs ~ 11 GByte/s even when subsampled to 4:2:2 spatial resolution. This data-rate is beyond the capabilities of current HDMI 2.0 and DisplayPort 1.4b interfaces. Therefore, decorrelation for subsampling, efficient quantization and preprocessing for compression will stay important topics, not only for content distribution over large distances but even for low distance applications like low latency display stream compression for transmission from playback device to display. It is desirable to extend the re-quantization concept introduced in Chapter 5 to include the color difference axes. Higher spatial and temporal resolution video and new acquisition methods like the quanta image sensor [67] will need new color spaces and new concepts for bit-rate reduction.

There is a large amount of literature on **tone mapping and gamut mapping**. As of 2016 Google Scholar lists 1220 results for the term "tone mapping operator" and 990 results for "tone mapping algorithm" with 380 papers being common to both queries. The vast majority of these proposed algorithms work fully or semi-automatic. This contradicts the requirements in applications for TV and cinema where the preservation of artistic intent is the priority and tone mapping and image manipulation algorithms featuring intuitive artist controls verified by user studies are still lacking.

In **display** technology, achieving higher peak brightness in TV, cinema and mobile devices while staying within a tight power budget is one of the major challenges. For TV replacing RGB-LEDs with blue lighting in combination with quantum dot filters [177] promises a higher efficiency and providing a wider gamut at the same time. In cinema prototypes using light redistribution [41, 42, 43] in combination with laser illumi-

nation [178] or self illuminated screens [171] may be used to achieve higher peak brightness.

At the end of the day, it is not technology, but stories that drive us into the cinema or in front of the TV. For all the new technologies mentioned here the most important aspect that will decide if the technology gets adopted or fails will not only be the best technical implementation, but whether the creative talents will find these technologies useful for story-telling or not. The first tests look promising, but whether the technologies presented here will help to tell stories has yet to be seen in the future.

Bibliography

- [1] Academy of Motion Picture Arts and Sciences. The 79th scientific & technical awards 2006 | 2007, February 2007. URL https://www.oscars.org/ sci-tech/ceremonies/2007.
- [2] Adobe Systems Inc. Adobe Photoshop® TIFF technical notes, March 2002. URL http://partners.adobe.com/public/developer/ en/tiff/TIFFphotoshop.pdf.
- [3] A. J. Ahumada, Jr., B. L. Beard, and R. Eriksson. Spatiotemporal discrimination model predicts temporal masking functions. *Proceedings of the SPIE*, 3299:120– 127, 1998. URL https://doi.org/10.1117/12.320103.
- [4] S. Andriani, H. Brendel, T. Seybold, and J. Goldstone. Beyond the Kodak image set: A new reference set of color image sequences. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 2289–2293. IEEE, September 2013. URL https://doi.org/10.1109/ICIP.2013. 6738472.
- [5] Apple Inc. Apple ProRes white paper (019-00323-a). Technical report, Apple Inc., April 2017. URL https://www.apple.com/final-cut-pro/ docs/Apple_ProRes_White_Paper.pdf.
- [6] Arnold & Richter Cine Technik. ALEXA M data sheet. Technical report, Arnold & Richter Cine Technik, November 2011. URL http://www.arri.com/de/camera/alexa/kameras/ kamera_details/alexa-m/subsection/downloads/.
- [7] M. Ashikhmin. A tone mapping algorithm for high contrast images. In Proceedings of the 13th Eurographics Workshop on Rendering, EGRW '02, pages 145–156, Aire-la-Ville, Switzerland, 2002. Eurographics Association. URL http://dl.acm.org/citation.cfm?id=581896.581916.

- [8] Association of Radio Industries and Businesses (ARIB). Arib std-b67, essential parameter values for the extended image dynamic range television (EIDRTV) system for programme production. Technical report, ARIB, Nittochi Bldg. 11F, 1-4-1, Kasumigaseki, Chiyoda-ku, Tokyo 100-0013, Japan, July 2015. URL https://www.arib.or.jp/english/html/overview/doc/2-STD-B67v1_0.pdf.
- [9] S. Aubenas and G. Baldwin. *Gustave Le Gray, 1820-1884*. Getty Publications, 1200 Getty Center Drive Suite 1200, Los Angeles, CA 90049, February 2002.
- [10] T. O. Aydın, R. Mantiuk, and H.-P. Seidel. Extending quality metrics to full luminance range images. *Proceedings of the SPIE*, 6806:68060B-68060B-10, 2008. URL https://doi.org/10.1117/12.765095.
- [11] T. O. Aydın, R. Mantiuk, and H.-P. Seidel. Extending quality metrics to full luminance range images - supplementary materials. Technical report, Max Planck Institute for Informatics, 2008. URL http://resources.mpiinf.mpg.de/hdr/fulldr_extension/.
- [12] A. Banitalebi-Dehkordi, Y. Dong, M. T. Pourazad, and P. Nasiopoulos. A learning-based visual saliency fusion model for high dynamic range video (LBVS-HDR). In Proceedings of the 23rd European Signal Processing Conference (EUSIPCO), pages 1541–1545, August 2015. URL https://doi.org/10.1109/ EUSIPC0.2015.7362642.
- [13] P. G. J. Barten. Physical model for the contrast sensitivity of the human eye. Proceedings of the SPIE, 1666:57–72, 1992. URL https://doi.org/10. 1117/12.135956.
- [14] P. G. J. Barten. Contrast sensitivity of the human eye and its effects on image quality, volume 72. SPIE press, SPIE, PO Box 10, Bellingham WA 98227-0010, USA, December 1999. URL http://spie.org/Publications/ Book/353254.
- [15] P.G. J. Barten. Formula for the contrast sensitivity of the human eye. Proceedings of the SPIE, 5294:231–238, 2003. URL https://doi.org/10.1117/ 12.537476.
- [16] A. V. Bedford. Mixed highs in color television. Proceedings of the IRE, 38 (9):1003-1009, September 1950. URL https://doi.org/10.1109/ JRPROC.1950.229518.
- [17] Blu-ray Disc Association. Coding constraints on HEVC video streams for BD-ROM version 3.0. Technical report, Blu-ray Disc Association, Burbank, October 2011. URL http://www.blu-raydisc.com/ Assets/Downloadablefile/BD-ROM-AV-WhitePaper_HEVC_ 150608a-clean.pdf.

- [18] T. Borer. WHP 283 non-linear opto-electrical transfer functions for high dynamic range television. Technical report, British Broadcasting Corporation, July 2014. URL http://downloads.bbc.co.uk/rd/pubs/whp/whppdf-files/WHP283.pdf.
- [19] T. Borer and A. Cotton. WHP 309 a 'display independent' high dynamic range television system. Technical report, British Broadcasting Corporation, September 2015. URL http://downloads.bbc.co.uk/rd/pubs/ whp/whp-pdf-files/WHP309.pdf.
- [20] T. Borer, A. Cotton, M. Pindoria, and S. Thompson. Approaches to high dynamic range video. In *Proceedings of the Digital Media Industry Academic Forum* (*DMIAF*), pages 71–76, July 2016. URL https://doi.org/10.1109/ DMIAF.2016.7574905.
- [21] P. Bouguer. *Essai d'optique sur la gradation de la lumière*. Claude Jombert, Ruë S. Jacques, au coin de la ruë des Mathurins, à l'Image Notre-Dame, 1729.
- [22] P. Bouguer. *Traité d'optique sur la gradation de la lumière*. H. L. Guerin & L.F. Delatour, Rue Saint Jacques, à Saint Thomas d'Aquin, 1760.
- [23] H. Brendel. Noise characteristic of the ALEXA camera. Technical report, Arnold & Richter Cine Technik GmbH & Co Betriebs KG, October 2011.
- [24] H. Brendel. ALEXA LogC curve Usage in VFX Reference manual. Technical report, Arnold & Richter Cine Technik GmbH & Co Betriebs KG, April 2012. URL https://www.arri.com/?eID=registration& file_uid=8026.
- [25] H. Brendel. 55-3: Invited paper: Delivering content for HDR. SID Symposium Digest of Technical Papers, 48(1):827–830, 2017. URL https://doi.org/ 10.1002/sdtp.11774.
- [26] S. J. Briggs. Photometric technique for deriving a 'best gamma' for displays. Proceedings of the SPIE, 0199:134–145, 1979. URL https://doi.org/ 10.1117/12.958051.
- [27] M. H. Brill. How the CIE 1931 color-matching functions were derived from Wright-Guild data - errata. *Color Research & Application*, 23(4):259–259, 1998. URL https://doi.org/10.1002/(SICI)1520-6378(199808) 23:4<259::AID-COL18>3.0.CO;2-7.
- [28] M. H. Brill and R. C. Carter. Does lightness obey a log or a power law? Or is that the right question? *Color Research & Application*, 39(1):99–101, 2014. URL https://doi.org/10.1002/col.21777.

- [29] T. Caelli, M. Huebner, and I. Rentschler. The detection of phase shifts in twodimensional images. *Perception & Psychophysics*, 37(6):536–542, 1985. URL https://doi.org/10.3758/BF03204919.
- [30] C.-H. Chang, H.-C. Cheng, Y.-J. Lu, K.-C. Tien, H.-W. Lin, C.-L. Lin, C.-J. Yang, and C.-C. Wu. Enhancing color gamut of white OLED displays by using microcavity green pixels. *Organic Electronics*, 11(2):247 – 254, February 2010. URL https://doi.org/10.1016/j.orgel.2009.11.002.
- [31] Christie Digital Systems USA Inc. Christie CP4220/4230 datasheet. Technical report, Christie Digital Systems USA Inc., March 2016. URL https: //www.christiedigital.com/en-us/cinema/cinemaproducts/digital-cinema-projectors/christiecp4220-dlp-digital-cinema-projector.
- [32] Commission Internationale de l'Éclairage (CIE). Principales decisions (CIE sixième session, 1924). In Commission Internationale de l'Éclairage Proceedings, pages 67—69, Geneva, Switzerland, July 1926. Cambridge University Press. URL http://www.cie.co.at/index.php?i_ca_id=563& pubid=95.
- [33] Commission Internationale de l'Éclairage (CIE). CIE 146:2002 CIE equations for disability glare. Color Research & Application, 27(6):457–458, 2002. URL https://doi.org/10.1002/col.10102.
- [34] E. A. Cooper, H. Jiang, V. Vildavski, J. E. Farrell, and A. M. Norcia. Assessment of OLED displays for vision research. *Journal of Vision*, 13:1534–7362, October 2013. URL https://doi.org/10.1167/13.12.16.
- [35] A. Cotton, T. Borer, M. Pindoria, S. Thompson, S. S. Matteo Naccari, and M. Mrak. M36249 - BBC's response to CfE for HDR video coding (category 3a). Technical Report ISO/IEC JTC1/SC29/WG11, British Broadcasting Corporation (BBC), Warsaw, Poland, June 2015. URL http://mpeg. chiariglione.org/meetings/112.
- [36] M. Cowan, G. Kennel, T. Maier, and B. Walker. Contrast sensitivity experiment to determine the bit depth for digital cinema. *SMPTE Motion Imaging Journal*, 113(9):281–292, September 2004. URL https://doi.org/10.5594/ J11549.
- [37] P. Cyriac, D. Kane, and M. Bertalmío. Automatic, viewing-condition dependent contrast grading based on perceptual models. In *Proceedings of the SMPTE Annual Technical Conference and Exhibition*, pages 1–11, October 2016. URL https://doi.org/10.5594/M001687.

- [38] S. Daly. The visible differences predictor: An algorithm for the assessment of image fidelity. In A. B. Watson, editor, *Digital Images and Human Vision*, pages 179–206. MIT Press, Cambridge, MA, USA, 1993. URL http://dl.acm. org/citation.cfm?id=197765.197783.
- [39] S. Daly. A visual model for optimizing the design of image processing algorithms. In Proceedings of the IEEE International Conference on Image Processing (ICIP), volume 2, pages 16–20, November 1994. URL https://doi.org/ 10.1109/ICIP.1994.413522.
- [40] S. Daly, T. Kunkel, X. Sun, S. Farrell, and P. Crum. Preference limits of the visual dynamic range for ultra high quality and aesthetic conveyance. *Proceedings of the SPIE*, 8651:86510J–86510J–11, 2013. URL https://doi.org/10. 1117/12.2013161.
- [41] G. Damberg, A. Ballestad, E. Kozak, R. Kumaran, and J. Minor. Efficient, high brightness, high dynamic range projection. In *Proceedings of the ACM SIGGRAPH* 2014 Emerging Technologies, SIGGRAPH '14, pages 18:1–18:1, New York, NY, USA, 2014. ACM. URL http://doi.acm.org/10.1145/2614066. 2614078.
- [42] G. Damberg, A. Ballestad, E. Kozak, J. Minor, R. Kumaran, and J. Gregson. High brightness HDR projection using dynamic phase modulation. In *Proceedings of the ACM SIGGRAPH 2015 Emerging Technologies*, SIGGRAPH '15, pages 13:1– 13:1, New York, NY, USA, 2015. ACM. URL http://doi.acm.org/10. 1145/2782782.2792487.
- [43] G. Damberg, J. Gregson, and W. Heidrich. High brightness HDR projection using dynamic freeform lensing. ACM Transactions on Graphics (TOG), 35(3):24:1– 24:11, May 2016. URL http://doi.acm.org/10.1145/2857051.
- [44] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the ACM SIGGRAPH 1997 on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 369–378, New York, NY, USA, 1997. ACM. URL http://dx.doi.org/10.1145/258734. 258884.
- [45] Digital Cinema Initiatives LLC. Digital cinema system specification v.1.2. Technical report, DCI, Hollywood, California, USA, August 2012. URL http://dcimovies.com/specification/DCI_ DCSS_v12_with_errata_2012-1010.pdf.
- [46] Dolby Laboratories Inc. Dolby PRM-4200 professional reference monitor whitepaper. Technical report, Dolby Laboratories Inc., San Francisco, 2012. URL https://mwa-nova.com/Support/PDF/Dolby_PRM-4200_White_Paper.pdf.

- [47] Dolby Laboratories Inc. Dolby cinema the total cinema experience., March 2016. URL http://www.dolby.com/us/en/platforms/dolbycinema.html.
- [48] Dolby Laboratories Inc. ICtCp whitepaper. Technical report, Dolby Laboratories Inc., 1275 Market Street, San Francisco, CA 94103-1410, USA, April 2016. URL https://www.dolby.com/us/en/technologies/dolbyvision/ictcp-white-paper.pdf.
- [49] Y. Dong, E. Nasiopoulos, M. T. Pourazad, and P. Nasiopoulos. High dynamic range video eye tracking dataset. In *Proceedings of the second International Conference on Electronics, Signal processing and Communications (ESPC)*, pages 56–59, Athens, Greece, November 2014. URL http://www.inase.org/ library/2014/athens/bypaper/ELECT/ELECT-08.pdf.
- [50] Y. Dong, M. T. Pourazad, and P. Nasiopoulos. Human visual system-based saliency detection for high dynamic range content. *IEEE Transactions on Multimedia*, 18(4):549–562, April 2016. URL https://doi.org/10.1109/ TMM.2016.2522639.
- [51] F. Drago, K. Myszkowski, T. Annen, and N. Chiba. Adaptive logarithmic mapping for displaying high contrast scenes. *Computer Graphics Forum*, 22(3):419–426, 2003. URL https://doi.org/10.1111/1467-8659.00689.
- [52] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamicrange images. In ACM Transactions on Graphics (TOG), SIGGRAPH '02, pages 257–266, New York, NY, USA, 2002. ACM. URL http://doi.acm.org/ 10.1145/566570.566574.
- [53] S. Duschl and J. Heugel. ALEXA and AMIRA Using the ARRI Look File 2. Technical report, Arnold & Richter Cine Technik, Türkenstraße 89, 80799 München, Germany, March 2017. URL http://www.arri.com/ camera/amira/tools/arri_color_tool/.
- [54] F. Ebner. Derivation and modelling hue uniformity and development of the IPT color space. PhD thesis, Rochester Institute of Technology, Rochester, NY, USA, July 1998. URL http://scholarworks.rit.edu/theses/2858.
- [55] G. Eilertsen, R. Mantiuk, and J. Unger. Real-time noise-aware tone mapping. ACM Trans. Graph., 34(6):198:1–198:15, October 2015. URL http://doi. acm.org/10.1145/2816795.2818092.
- [56] G. Eilertsen, R. Mantiuk, and J. Unger. Luma HDRv: An open source high dynamic range video codec optimized by large-scale testing. In *Proceedings of the ACM SIGGRAPH 2016 Talks*, SIGGRAPH '16, pages 17:1–17:2, New York, NY, USA, 2016. ACM. URL http://doi.acm.org/10.1145/2897839. 2927447.

- [57] G. Eilertsen, R. Mantiuk, and J. Unger. A high dynamic range video codec optimized by large-scale testing. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 1379–1383, September 2016. URL https://doi.org/10.1109/ICIP.2016.7532584.
- [58] G. Eilertsen, R. Mantiuk, and J. Unger. A comparative review of tone-mapping algorithms for high dynamic range video. *Computer Graphics Forum*, 36(2): 565–592, 2017. URL https://doi.org/10.1111/cgf.13148.
- [59] A. Eisner and D. I. A. MacLeod. Blue-sensitive cones do not contribute to luminance. Journal of the Optical Society of America (JOSA), 70(1):121–123, January 1980. URL http://www.osapublishing.org/abstract. cfm?URI=josa-70-1-121.
- [60] M. D. Fairchild. Formulation and testing of an incomplete-chromatic-adaptation model. Color Research & Application, 16(4):243–250, 1991. URL https: //doi.org/10.1002/col.5080160406.
- [61] M. D. Fairchild. Color Appearance Models. The Wiley-IS&T Series in Imaging Science and Technology. John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA, 3rd edition edition, August 2013. URL http://eu.wiley. com/WileyCDA/WileyTitle/productCd-0470030321.html.
- [62] H. S. Fairman, M. H. Brill, and H. Hemmendinger. How the CIE 1931 color-matching functions were derived from Wright-Guild data. Color Research & Application, 22(1):11–23, 1997. URL https: //doi.org/10.1002/(SICI)1520-6378(199702)22:1<11:: AID-COL4>3.0.CO;2-7.
- [63] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. In ACM Transactions on Graphics (TOG), SIGGRAPH '02, pages 249–256, New York, NY, USA, 2002. ACM. URL http://doi.acm.org/ 10.1145/566570.566573.
- [64] G. T. Fechner. *Elemente der Psychophysik*. Breitkopf & Härtel, Leipzig, 1860.
- [65] X. Feng and S. Daly. Automatic JPEG compression using a color visual model. In Proceedings of the Society for Imaging Science and Technology (IS&T)'s PICS CON-FERENCE, pages 29–32. Society for Imaging Science and Technology (IS&T), 2003. URL http://www.imaging.org/IST/store/epub.cfm? abstrid=8494.
- [66] FilmLight Ltd. Baselight Software Technical Datasheet. Technical Report FL-BL-DS-0615, FilmLight Ltd., April 2014. URL http://www.filmlight.ltd.uk/support/documents/baselight/datasheets_bl.php.

- [67] E. Fossum. The quanta image sensor (QIS): Concepts and challenges. In Imaging and Applied Optics, page JTuE1. Optical Society of America, 2011. URL http://www.osapublishing.org/abstract.cfm? URI=COSI-2011-JTuE1.
- [68] J. Froehlich, A. Schilling, and B. Eberhardt. Gamut mapping for digital cinema. In Proceedings of the SMPTE Annual Technical Conference and Exhibition, pages 1–11, October 2013. URL https://doi.org/10.5594/M001489.
- [69] J. Froehlich, R. Atkins, Q. Wang, G.-M. Su, and P. Yin. Encoding and decoding perceptually-quantized video content. United States Patent: PCT/US2015/051964, September 2014. URL https://patentscope. wipo.int/search/en/detail.jsf?docId=W02016049327.
- [70] J. Froehlich, S. Grandinetti, and B. Eberhardt. M35044 description of HDR and WCG test sequences proposed by stuttgart media university. Technical Report ISO/IEC JTC1/SC29/WG11, Stuttgart Media University (HdM), Strasbourg, France, October 2014. URL https://www.hdm-stuttgart. de/~froehlichj/m35044_HdM-Stuttgart_content.pdf.
- [71] 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. *Proceedings of the SPIE*, 9023:90230X–90230X–10, February 2014. URL https://doi.org/10.1117/12.2040003.
- [72] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel. Website: Download the HdM-HDR-2014 HDR-Video data set, February 2014. URL https://hdr-2014.hdm-stuttgart.de/.
- [73] J. Froehlich, T. Kunkel, R. Atkins, J. Pytlarz, S. Daly, A. Schilling, and B. Eberhardt. Encoding color difference signals for high dynamic range and wide gamut imagery. *Color and Imaging Conference (CIC)*, 2015(1):240–247, October 2015. URL http://www.ingentaconnect.com/content/ ist/cic/2015/00002015/0000001/art00044.
- [74] J. Froehlich, G.-M. Su, R. Atkins, S. Daly, and S. Miller. Content-adaptive perceptual quantizer for high dynamic range images. United States Patent: PCT/US2016/020230, March 2015. URL https://patentscope. wipo.int/search/en/detail.jsf?docId=WO2016140954.
- [75] J. Froehlich, G.-M. Su, S. Daly, A. Schilling, and B. Eberhardt. Content aware quantization: Requantization of high dynamic range baseband signals based on visual masking by noise and texture. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 884–888, September 2016. URL https://doi.org/ICIP.2016.7532484.

- [76] Y. Gryaditskaya, T. Pouli, E. Reinhard, K. Myszkowski, and H.-P. Seidel. Motion aware exposure bracketing for HDR video. *Comput. Graph. Forum*, 34(4):119– 130, July 2015. URL https://doi.org/10.1111/cgf.12684.
- [77] J. Guild. The colorimetric properties of the spectrum. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 230:149–187, 1932. URL http://www.jstor.org/ stable/91229.
- [78] S. L. Guth. Model for color vision and light adaptation. Journal of the Optical Society of America (JOSA), 8(6):976–993, June 1991. URL http://josaa. osa.org/abstract.cfm?URI=josaa-8-6-976.
- [79] B. Guthier, S. Kopf, and W. Effelsberg. Optimal shutter speed sequences for real-time HDR video. In 2012 IEEE International Conference on Imaging Systems and Techniques Proceedings, pages 303–308, July 2012. URL https://doi. org/10.1109/IST.2012.6295505.
- [80] L. Hallpike. The role of ocula in stereo post production. Technical report, The Foundry, August 2008. URL https://www.fxguide.com/ wp-content/uploads/2008/09/stereowhitepaper.pdf? c20ef6.
- [81] E. M. Hartwig R. Blume, Scott J. Daly. Presentation of medical images on CRT displays: a renewed proposal for a display function standard, 1993. URL https://doi.org/10.1117/12.146970.
- [82] M. Hernández-Cabronero, V. Sanchez, F. Aulí-Llinàs, and J. Serra-Sagristà. Lossy compression of natural HDR content based on multi-component transform optimization. In *Proceedings of the Digital Media Industry Academic Forum* (*DMIAF*), pages 23–28, July 2016. URL https://doi.org/10.1109/ DMIAF.2016.7574894.
- [83] D. M. Hoffman and D. Stolitzka. A new standard method of subjective assessment of barely visible image artifacts and a new public database. *Journal* of the Society for Information Display, 22(12):631–643, 2014. URL https: //doi.org/10.1002/jsid.297.
- [84] V. Hulusic, G. Valenzise, E. Provenzi, K. Debattista, and F. Dufaux. Perceived dynamic range of HDR images. In Proceedings of the Eighth International Conference on Quality of Multimedia Experience (QoMEX), pages 1–6, June 2016. URL https://doi.org/10.1109/QoMEX.2016.7498953.
- [85] P.-C. Hung and R. S. Berns. Determination of constant hue loci for a CRT gamut and their predictions using color appearance spaces. *Color Research & Application*, 20(5):285–295, 1995. URL https://doi.org/10.1002/col. 5080200506.

- [86] Institute of Electrical and Electronics Engineers (IEEE). IEEE standard for floating-point arithmetic. Technical report, IEEE Std 754-2008, August 2008. URL https://doi.org/10.1109/{IEEESTD}.2008.4610935.
- [87] International Electrotechnical Commission (IEC). Multimedia systems and equipment-colour measurement and management-part 2-1: Colour management-default RGB colour space-sRGB. Technical report, IEC 61966-2-1, 1999. URL https://webstore.iec.ch/publication/6169.
- [88] International Organization for Standardization (ISO). Information technology - Digital compression and coding of continuous-tone still images: Requirements and guidelines. ISO 10918-1:1994 (ITU T.81), Geneva, Switzerland, February 1994. URL http://www.iso.org/iso/catalogue_detail. htm?csnumber=18902.
- [89] International Organization for Standardization (ISO). Colorimetry - Part 4: CIE 1976 L*a*b* Colour Space. ISO 11664-4:2008 (CIE S 014-4/E:2007), Geneva, Switzerland, November 2008. URL http://www.iso.org/iso/home/store/catalogue_tc/ catalogue_detail.htm?csnumber=52497.
- [90] International Telecommunication Union (ITU). Recommendation ITU-R BT.1886, reference electro-optical transfer function for flat panel displays used in HDTV studio production. Technical report, International Telecommunication Union (ITU), Geneva, March 2011. URL http://www.itu.int/rec/ R-REC-BT.1886/en.
- [91] International Telecommunication Union (ITU). Recommendation h.265: High efficiency video coding. Technical report, International Telecommunication Union (ITU), Geneva, April 2015. URL http://www.itu.int/rec/T-REC-H.265/.
- [92] International Telecommunication Union (ITU). Recommendation ITU-R BT.2020-2, parameter values for ultra-high definition television systems for production and international programme exchange. Technical report, International Telecommunication Union (ITU), Geneva, October 2015. URL http: //www.itu.int/rec/R-REC-BT.2020/en.
- [93] International Telecommunication Union (ITU). Report ITU-R BT.2246-5, the present state of ultra-high definition television. Technical report, International Telecommunication Union (ITU), Geneva, July 2015. URL http://www. itu.int/pub/R-REP-BT.2246.
- [94] International Telecommunication Union (ITU). Recommendation ITU-R BT.709-6, parameter values for the HDTV standards for production and international programme exchange. Technical report, International Telecommuni-

cation Union (ITU), Geneva, June 2015. URL http://www.itu.int/ rec/R-REC-BT.709/en.

- [95] International Telecommunication Union (ITU). Recommendation H.264: Advanced video coding for generic audiovisual services. Technical report, International Telecommunication Union (ITU), Geneva, February 2016. URL http://www.itu.int/rec/T-REC-H.264.
- [96] International Telecommunication Union (ITU). Recommendation ITU-R BT.2100-0, image parameter values for high dynamic range television for use in production and international programme exchange. Technical report, International Telecommunication Union (ITU), Geneva, July 2016. URL http: //www.itu.int/rec/R-REC-BT.2100.
- [97] S. Ishihara. Tests for color-blindness. Handaya, Tokyo, Hongo Harukicho, 1917.
- [98] R. E. Johnston, J. B. Zimmerman, D. C. Rogers, and S. M. Pizer. Perceptual standardization. *Proceedings of the SPIE*, 0536:44–49, 1985. URL https: //doi.org/10.1117/12.947328.
- [99] D. B. Judd. Estimation of chromaticity differences and nearest color temperature on the standard 1931 ici colorimetric coordinate system. Journal of the Optical Society of America (JOSA), 26(11):421-426, November 1936. URL http://www.osapublishing.org/abstract.cfm? URI=josa-26-11-421.
- [100] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. In ACM Transactions on Graphics (TOG), SIGGRAPH '03, pages 319– 325, New York, NY, USA, 2003. ACM. URL http://doi.acm.org/10. 1145/1201775.882270.
- [101] N. Katoh and M. Ito. Gamut mapping for computer generated images (II). Color and Imaging Conference, 1996(1):126–128, 1996. URL http://www.ingentaconnect.com/content/ist/cic/ 1996/00001996/00000001/art00034.
- [102] G. Kennel and D. Snider. SMPTE tutorial: Gray-scale transformations of digital film data for display, conversion, and film recording. SMPTE Journal, 102 (12):1109–1119, December 1993. URL https://doi.org/10.5594/ J01652.
- [103] K. J. Kim, R. Mantiuk, and K. H. Lee. Measurements of achromatic and chromatic contrast sensitivity functions for an extended range of adaptation luminance. *Proceedings of the SPIE*, 8651:86511A–86511A–14, 2013. URL https://doi.org/10.1117/12.2002178.

- [104] K. J. Kim, R. Mantiuk, and K. H. Lee. Measurements of achromatic and chromatic contrast sensitivity functions for an extended range of adaptation luminance - corrected version. *Proceedings of the SPIE*, 8651:86511A-86511A-14, 2013. URL https://www.cl.cam.ac.uk/~rkm38/ pdfs/kiml3mcsf.pdf.
- [105] S. Kim, D. Kang, J. Lee, J. Kim, Y. Park, T. Han, S. Jung, J. J. Yoo, M. Lim, and J. Baek. Image quality evaluation of LCDs based on novel RGBW sub-pixel structure. *Proceedings of the SPIE*, 9396:939613 1–7, 2015. URL https: //doi.org/10.1117/12.2077071.
- [106] G. Kindlmann, E. Reinhard, and S. Creem. Face-based luminance matching for perceptual colormap generation. In *Proceedings of the Conference on Visualization '02*, VIS '02, pages 299–306, Washington, DC, USA, 2002. IEEE Computer Society. URL http://dl.acm.org/citation.cfm?id=602099. 602145.
- [107] P. Korshunov, P. Hanhart, T. Richter, A. Artusi, R. Mantiuk, and T. Ebrahimi. Subjective quality assessment database of HDR images compressed with jpeg xt. In Proceedings of the Seventh International Workshop on Quality of Multimedia Experience (QoMEX), pages 1–6, May 2015. URL https://doi.org/10. 1109/QoMEX.2015.7148119.
- [108] J. Krauskopf and G. Karl. Color discrimination and adaptation. Vision Research, 32(11):2165–2175, 1992. URL https://doi.org/10.1016/0042-6989(92)90077-V.
- [109] G. Krawczyk. Tunnel video sequence, 2005. URL http://www.mpiinf. mpg.de/resources/hdr/video/.
- [110] J. Kronander, S. Gustavson, G. Bonnet, and J. Unger. Unified HDR reconstruction from raw CFA data. In Proceedings of the IEEE International Conference on Computational Photography (ICCP), pages 1–9, April 2013. URL https://doi.org/10.1109/ICCPhot.2013.6528315.
- [111] J. Kronander, S. Gustavson, G. Bonnet, A. Ynnerman, and J. Unger. A unified framework for multi-sensor HDR video reconstruction. Signal Processing: Image Communication, 29(2):203–215, 2014. URL http://www.sciencedirect.com/science/article/pii/ S0923596513001355.
- [112] T. Kunkel and E. Reinhard. A reassessment of the simultaneous dynamic range of the human visual system. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, APGV '10, pages 17–24, New York, NY, USA, 2010. ACM. URL http://doi.acm.org/10.1145/1836248. 1836251.

- [113] T. Kunkel, G. Ward, B. Lee, and S. Daly. HDR and wide gamut appearancebased color encoding and its quantification. In *Proceedings of the Picture Coding Symposium (PCS), 2013*, pages 357–360, December 2013. URL https:// doi.org/10.1109/PCS.2013.6737757.
- [114] K. J. Kwon, M. B. Kim, C. Heo, S. G. Kim, J. S. Baek, and Y. H. Kim. Wide color gamut and high dynamic range displays using RGBW LCDs. *Displays*, 40: 9 - 16, 2015. URL http://www.sciencedirect.com/science/ article/pii/S0141938215000967.
- [115] G. W. Larson. LogLuv encoding for full-gamut, high-dynamic range images. *Journal of Graphics Tools*, 3(1):15–31, 1998. URL https://doi.org/10. 1080/10867651.1998.10487485.
- [116] G. W. Larson. Overcoming gamut and dynamic range limitations in digital images. Color and Imaging Conference (CIC), 1998(1):214–219, 1998. URL http://www.ingentaconnect.com/content/ist/ cic/1998/00001998/0000001/art00046.
- [117] G. W. 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, October 1997. URL https: //doi.org/10.1109/2945.646233.
- [118] S. Lasserre, F. LeLéannec, and E. Francois. Description of HDR sequences proposed by technicolor. ISO/IEC JTC1/SC29/WG11 JCTVC-P0228, IEEE, San Jose, USA, 2013. URL http://wg11.sc29.org/.
- [119] M. Le Pendu, C. Guillemot, and D. Thoreau. Inter-layer prediction of color in high dynamic range image scalable compression. *IEEE Transactions on Image Processing*, 25(8):3585–3596, August 2016. URL https://doi.org/10. 1109/TIP.2016.2571559.
- [120] L. Lenzen and M. Christmann. Subjective viewer preference model for automatic HDR down conversion. *Electronic Imaging*, 2017(12):191–197, January 2017. URL http://www.ingentaconnect.com/content/ist/ ei/2017/00002017/00000012/art00028.
- [121] Linköping University Computer Graphics and Image Processing, Visual Computing Laboratory. LiU HDRv Repository, 2013. URL http://www.hdrv. org/Resources.php#HDRv.
- [122] I. Lissner and P. Urban. How perceptually uniform can a hue linear color space be? Color and Imaging Conference (CIC), 2010(1):97–102, 2010. URL http://www.ingentaconnect.com/content/ist/ cic/2010/00002010/00000001/art00018.

- [123] L. Litwic, O. Baumann, P. White, and M. S. Goldman. Bit rate requirements for high dynamic range video. *SMPTE Motion Imaging Journal*, 125(5):52–60, July 2016. URL https://doi.org/10.5594/JMI.2016.2569920.
- [124] S. P. Lloyd. Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2):129–137, March 1982. URL https://doi.org/10. 1109/TIT.1982.1056489.
- [125] T. Lu, F. Pu, P. Yin, T. Chen, W. Husak, J. Pytlarz, R. Atkins, J. Froehlich, and G.-M. Su. ITP colour space and its compression performance for high dynamic range and wide colour gamut video distribution. *ZTE Communications*, February 2016. URL http://wwwen.zte.com.cn/endata/ magazine/ztecommunications/2016/1/articles/201603/ t20160311_448972.html.
- [126] A. Lukin. Improved visible differences predictor using a complex cortex transform. In Proceedings of the International Conference on Computer Graphics and Vision (GraphiCon), pages 145–150, 2009. URL http://www.graphicon.ru/html/2009/conference/sel1/ 125/125_Paper.pdf.
- [127] Z. Luo, Y. Chen, and S.-T. Wu. Wide color gamut LCD with a quantum dot backlight. Optics Express, 21(22):26269-26284, November 2013. URL http://www.opticsexpress.org/abstract.cfm? URI=oe-21-22-26269.
- [128] A. Luthra, E. François, and W. Husak. N15084 requirements and use cases for HDR and WCG content distribution. Technical report, ISO/IEC JTC1/SC29/WG11, February 2015. URL http: //mpeg.chiariglione.org/standards/exploration/ high-dynamic-range-and-wide-colour-gamut-contentdistribution/requirements-and.
- [129] D. L. MacAdam. Visual sensitivities to color differences in daylight. Journal of the Optical Society of America (JOSA), 32(5):247-274, May 1942. URL http://www.osapublishing.org/abstract.cfm? URI=josa-32-5-247.
- [130] T. Mansencal, M. Mauderer, M. Parsons, L. Canavan, S. Cooper, N. Shaw, K. Wheatley, and K. Crowson. Colour science for Python., August 2017. URL http://colour-science.org/.
- [131] R. Mantiuk. HDR visual difference predictor 2.2.1 download page, 2016. URL https://sourceforge.net/projects/hdrvdp/ files/hdrvdp/2.2.1/.

- [132] R. Mantiuk, K. Myszkowski, and H.-P. Seidel. Lossy compression of high dynamic range images and video. *Proceedings of the SPIE*, 6057:60570V–60570V– 10, 2006. URL https://doi.org/10.1117/12.639140.
- [133] R. Mantiuk, S. Daly, and L. Kerofsky. Display adaptive tone mapping. In ACM Transactions on Graphics (TOG), SIGGRAPH '08, pages 68:1–68:10, New York, NY, USA, 2008. ACM. URL http://doi.acm.org/10.1145/ 1399504.1360667.
- [134] K. Masaoka, Y. Kusakabe, T. Yamashita, Y. Nishida, T. Ikeda, and M. Sugawara. Algorithm design for gamut mapping from UHDTV to HDTV. *Journal of Display Technology*, PP(99):1–1, February 2016. URL https://doi.org/ 10.1109/JDT.2016.2527039.
- [135] A. Masson. Elektro-photometrische Studien. Annalen der Physik, 139(9):158-161, 1844. URL https://doi.org/10.1002/andp. 18441390919.
- [136] S. T. McCarthy. A biologically-inspired approach to making HDR video quality assessment easier. SMPTE Motion Imaging Journal, 126(4):47–58, May 2017. URL https://doi.org/10.5594/JMI.2017.2673718.
- [137] J. S. McElvain and W. Gish. Camera color correction using twodimensional transforms. Color and Imaging Conference (CIC), 2013(1):250– 256, 2013. URL http://www.ingentaconnect.com/content/ ist/cic/2013/00002013/00000001/art00046.
- [138] P. McLeod, J. Driver, Z. Dienes, and J. Crisp. Filtering by movement in visual search. Journal of Experimental Psychology: Human Perception and Performance, 17(1):55–64, 1991. URL https://doi.org/10.1037/0096-1523. 17.1.55.
- [139] Milestone AV Technologies LLC. Milestone Da-Lite HD Progressive 0.9 Specifications, August 2017. URL https://www.milestone.com/ en/dalite/screen_surfaces/front_projection/high_ resolution/hd_progressive_09.
- [140] S. Miller, M. Nezamabadi, and S. Daly. Perceptual signal coding for more efficient usage of bit codes. In *Proceedings of the Annual Technical Conference Exhibition, SMPTE 2012*, pages 1–9, October 2012. URL https://doi.org/10.5594/M001446.
- [141] S.-E. Moon and J.-S. Lee. Perceptual experience analysis for tone-mapped HDR videos based on EEG and peripheral physiological signals. *IEEE Transactions* on Autonomous Mental Development, 7(3):236–247, September 2015. URL https://doi.org/10.1109/TAMD.2015.2449553.

- [142] N. Moroney, M. D. Fairchild, R. W. Hunt, C. Li, M. R. Luo, and T. Newman. The CIECAM02 color appearance model. *Color and Imaging Conference (CIC)*, 2002(1):23-27, 2002. URL http://www.ingentaconnect.com/ content/ist/cic/2002/00002002/00000001/art00006.
- [143] J. Morovič. Color gamut mapping, volume 10 of The Wiley-IS&T Series in Imaging Science and Technology. John Wiley & Sons Inc., The Atrium, Southern Gate, Chichester, West Sussexx PO19 8SQ, England, June 2008. URL http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470030321.html.
- [144] K. T. Mullen. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology*, 359(1):381– 400, 1985. URL https://doi.org/10.1113/jphysiol.1985. sp015591.
- [145] M. J. Nadenau, J. Reichel, and M. Kunt. Wavelet-based color image compression: exploiting the contrast sensitivity function. *IEEE Transactions on Image Processing*, 12(1):58–70, January 2003. URL https://doi.org/10.1109/TIP.2002.807358.
- [146] K.-I. Naka and W. A. H. Rushton. S-potentials from luminosity units in the retina of fish (cyprinidae). *The Journal of physiology*, 185(3):587, 1966. URL http: //www.ncbi.nlm.nih.gov/pmc/articles/PMC1395832/.
- [147] M. Narwaria, R. Mantiuk, M. P. Da Silva, and P. Le Callet. HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images. *Journal of Electronic Imaging*, 24(1):010501, 2015. URL https://doi.org/10.1117/1.JEI.24.1.010501.
- [148] M. Narwaria, M. Perreira Da Silva, and P. Le Callet. Study of high dynamic range video quality assessment. *Proceedings of the SPIE*, 9599:95990V–95990V–13, 2015. URL https://doi.org/10.1117/12.2189178.
- [149] E. Nasiopoulos, Y. Dong, and A. Kingstone. Evaluation of high dynamic range content viewing experience using eye-tracking data. In Proceedings of the 10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, pages 13–17, August 2014. URL https://doi. org/10.1109/QSHINE.2014.6928653.
- [150] National Electrical Manufacturers Association (NEMA). Digital imaging and communications in medicine (DICOM), part 3.14, grayscale standard display function. Technical report, Technical Report, ACR/NEMA, 1998. URL http: //dicom.nema.org/Dicom/2011/11_14pu.pdf.

- [151] M. Nezamabadi, S. Miller, S. Daly, and R. Atkins. Color signal encoding for high dynamic range and wide color gamut based on human perception. *Proceedings* of the SPIE, 9015:90150C-90150C-12, 2014. URL https://doi.org/ 10.1117/12.2042893.
- [152] A. Ninan and S. Daly. Evolution of displays through advances in next generation video. In *Proceedings of the International Display Workshops (IDW)*, page 1312, Toki Messe Niigata Convention Center Niigata Japan, 2014. Society for Information Display (SID). URL https://www.idw.or.jp/ IDW14FP.pdf.
- [153] M. Oskarsson. Temporally consistent tone mapping of images and video using optimal k-means clustering. Journal of Mathematical Imaging and Vision, 57(2):225–238, 2017. URL https://doi.org/10.1007/s10851-016-0677-1.
- [154] D. Pajuelo, P. Cardoso, R. Barbieri, and S. C. e Yuzo Iano. Proposal for broadcast high dynamic range content transmission. In *Proceedings of the IEEE International Symposium on Consumer Electronics (ISCE)*, pages 69–70, September 2016. URL https://doi.org/10.1109/ISCE.2016.7797376.
- [155] Y. Park and J.-u. Kwon. 62-4: Invited paper: Consideration of display metrology for HDR and WCG standards based on real content. SID Symposium Digest of Technical Papers, 48(1):923–926, 2017. URL https://doi.org/10. 1002/sdtp.11787.
- [156] M. Pedzisz. Beyond BT.709. In Proceedings of the Annual Technical Conference Exhibition, SMPTE 2013, pages 1–13, October 2013. URL https://doi. org/10.5594/M001507.
- [157] D. G. Pelli. Effects of visual noise. PhD thesis, University of Cambridge, 1981. URL http://ethos.bl.uk/OrderDetails.do?uin=uk. bl.ethos.279618.
- [158] M. Pharr, W. Jakob, and G. Humphreys. *Physically Based Rendering, From Theory To Implementation*. Morgan Kaufmann, 30 Corporate Drive, Suite 400, Burlington, MA 01803, USA, November 2016. URL http://www.pbrt.org.
- [159] J. Plateau. Sur la mesure des sensations physiques, et sur la loi qui lie l'intensité de la cause excitante, volume 33. Bulletin de L'Académie Royale des Sciences, des Lettres et des Beaux Arts de Belgique, Belgique, 1872.
- [160] J. Plateau. Über die Messung physischer Empfindungen und das Gesetz, welches die Stärke dieser Empfindungen mit der Stärke der erregenden Ursache verknüpft. Annalen der Physik, 226(11):465–476, 1874. URL https: //doi.org/10.1002/andp.18742261114.

- [161] C. Poynton, J. Stessen, and R. Nijland. Deploying wide color gamut and high dynamic range in HD and UHD. SMPTE Motion Imaging Journal, 124(3):37–49, April 2015. URL https://doi.org/10.5594/j18530.
- [162] J. Pytlarz, K. Thurston, D. Brooks, P. Boon, and R. Atkins. Realtime crossmapping of high dynamic range images. SMPTE Motion Imaging Journal, 126(3):37–44, April 2017. URL https://doi.org/10.5594/JMI. 2017.2665298.
- [163] E. Reinhard and K. Devlin. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics*, 11 (1):13–24, January 2005. URL https://doi.org/10.1109/TVCG. 2005.9.
- [164] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. In ACM Transactions on Graphics (TOG), SIG-GRAPH '02, pages 267–276, New York, NY, USA, 2002. ACM. URL http: //doi.acm.org/10.1145/566570.566575.
- [165] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting. Morgan Kaufmann Publishers Inc., 30 Corporate Drive, Suite 400, Burlington, MA 01803, USA, 2005. URL http://store.elsevier.com/product.jsp?isbn= 9780080478319.
- [166] P. Research. Pr-7 series brochure. Technical report, Photo Research Inc., 9731 Topanga Canyon Place, Chatsworth, CA 91311-4135, USA, 2012. URL http://www.photoresearch.com/sites/default/ files/PR-7%20Series%20%20Brochure_0.pdf.
- [167] A. Rizzi and J. J. McCann. Glare-limited appearances in HDR images. *Journal of the Society for Information Display*, 17(1):3–12, 2009. URL https://doi.org/10.1889/JSID17.1.3.
- [168] R. Rosenholtz and A. B. Watson. Perceptual adaptive JPEG coding. In Proceedings of the IEEE International Conference on Image Processing (ICIP), volume 1, pages 901–904, September 1996. URL https://doi.org/10.1109/ ICIP.1996.559645.
- [169] S. Roth, N. Weiss, M. B. Chorin, I. B. David, and C. H. Chen. Multi-primary LCD for TV applications. SID Symposium Digest of Technical Papers, 38(1):34–37, 2007. URL https://doi.org/10.1889/1.2785219.
- [170] M. Safdar, G. Cui, Y. J. Kim, and M. R. Luo. Perceptually uniform color space for image signals including high dynamic range and wide gamut. *Optics Express*, 25(13):15131–15151, June 2017. URL https://doi.org/10.1364/ OE.25.015131.
- [171] Samsung Electronics. Samsung Electronics Premieres the Theater of the Future with New Cinema Screen Technology. Technical re-Samsung Electronics, Samsung Digital City, Samsungno 129, port, Yeongtong District, Suwon[1], South Korea, Maetan-dong, March URL https://news.samsung.com/global/samsung-2017. electronics-premieres-the-theater-of-the-futurewith-new-cinema-screen-technology.
- [172] O. H. Schade. Optical and photoelectric analog of the eye. Journal of the Optical Society of America (JOSA), 46(9):721-739, September 1956. URL http://www.osapublishing.org/abstract.cfm? URI=josa-46-9-721.
- [173] M. Schöberl, A. Brückner, S. Foessel, and A. Kaup. Photometric limits for digital camera systems. *Journal of Electronic Imaging*, 21(2):020501-1-020501-3, 2012. URL https://doi.org/10.1117/1.JEI.21.2.020501.
- [174] 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. URL https://doi.org/10.1889/1. 1832558.
- [175] T. Seybold, C. Keimel, M. Knopp, and W. Stechele. Towards an evaluation of denoising algorithms with respect to realistic camera noise. In *Proceedings of the 2013 IEEE International Symposium on Multimedia (ISM)*, pages 203–210, December 2013. URL https://doi.org/10.1109/ISM.2013.39.
- [176] H. Shahid, D. Li, A. Fanaswala, M. T. Pourazad, and P. Nasiopoulos. A new hybrid tone mapping scheme for high dynamic range (HDR) videos. In *Proceedings of the IEEE International Conference on Consumer Electronics (ICCE)*, pages 351–352, January 2015. URL https://doi.org/10.1109/ICCE.2015.7066441.
- [177] Y. Shirasaki, G. J. Supran, M. G. Bawendi, and V. Bulović. Emergence of colloidal quantum-dot light-emitting technologies. *Nature Photonics*, 7(1):13–23, 2013. URL https://doi.org/10.1038/nphoton.2012.328.
- [178] B. D. Silverstein, A. F. Kurtz, J. R. Bietry, and G. E. Nothhard. A laser-based digital cinema projector. SID Symposium Digest of Technical Papers, 42(1):326– 329, 2011. URL https://doi.org/10.1889/1.3621311.
- [179] Society of Motion Pictures and Television Engineers (SMPTE). SMPTE ST 377-1:2011 - Material Exchange Format (MXF) – File Format Specification. Technical report, Society of Motion Pictures and Television Engineers (SMPTE), White Plains, NY, June 2011. URL https://doi.org/10.5594/SMPTE. ST377-1.2011.

- [180] Society of Motion Pictures and Television Engineers (SMPTE). SMPTE ST 291-1:2011 – Ancillary Data Packet and Space Formatting. Technical report, Society of Motion Pictures and Television Engineers (SMPTE), White Plains, NY, September 2011. URL https://doi.org/10.5594/SMPTE. ST291-1.2011.
- [181] Society of Motion Pictures and Television Engineers (SMPTE). SMPTE RP 431-2:2011 – D-Cinema Quality – Reference Projector and Environment. Technical report, Society of Motion Pictures and Television Engineers (SMPTE), White Plains, NY, April 2011. URL https://doi.org/10.5594/SMPTE. RP431-2.2011.
- [182] Society of Motion Pictures and Television Engineers (SMPTE). SMPTE ST 292-1:2012 – 1.5 Gb/s Signal / Data Serial Interface. Technical report, Society of Motion Pictures and Television Engineers (SMPTE), White Plains, NY, January 2012. URL https://doi.org/10.5594/SMPTE.ST292-1.2012.
- [183] Society of Motion Pictures and Television Engineers (SMPTE). SMPTE ST 268:2014 – File Format for Digital Moving-Picture Exchange (DPX). Technical report, Society of Motion Pictures and Television Engineers (SMPTE), White Plains, NY, December 2014. URL https://doi.org/10.5594/ SMPTE.ST268.2014.
- [184] Society of Motion Pictures and Television Engineers (SMPTE). SMPTE ST 2084:2014 – High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays. Technical report, Society of Motion Pictures and Television Engineers (SMPTE), White Plains, NY, August 2014. URL https: //doi.org/10.5594/SMPTE.ST2084.2014.
- [185] Society of Motion Pictures and Television Engineers (SMPTE). Study group report: High-dynamic-range (HDR) imaging ecosystem. Technical report, Society of Motion Pictures and Television Engineers (SMPTE), September 2015. URL https://www.smpte.org/sites/default/files/Study% 20Group%20On%20High-Dynamic-Range-HDR-Ecosystem.pdf.
- [186] Society of Motion Pictures and Television Engineers (SMPTE). SMPTE ST 2085:2015 $f Y^{D}_{Z}D_{X}$ Color-Difference Computations for High Dynamic Range X'Y'Z' Signals. Technical report, Society of Motion Pictures and Television Engineers (SMPTE), White Plains, NY, June 2015. URL https://doi.org/10.5594/SMPTE.ST2085.2015.
- [187] S. S. Stevens. On the psychophysical law. Psychological Review, 64(3):153, 1957. URL https://doi.org/10.1037/h0046162.

- [188] L. J. Thorpe. Sony HDCAM: Exposure latitude issues of dynamic range 24p technical seminar #3. Technical report, Sony Corporation, 2001. URL https://pro.sony.com/bbsccms/ext/cinealta/docs/ 24PTechnicalSeminar3.pdf.
- [189] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen. A versatile HDR video production system. In ACM Transactions on Graphics (TOG), SIGGRAPH '11, pages 41:1– 41:10, New York, NY, USA, 2011. ACM. URL http://doi.acm.org/10. 1145/1964921.1964936.
- [190] S. Ueki, K. Nakamura, Y. Yoshida, T. Mori, K. Tomizawa, Y. Narutaki, Y. Itoh, and K. Okamoto. Five-primary-color 60-inch LCD with novel wide color gamut and wide viewing angle. SID Symposium Digest of Technical Papers, 40(1):927–930, 2009. URL https://doi.org/10.1889/1.3256949.
- [191] University of British Columbia. Digital media lab high dynamic range (DML-HDR) video dataset, 2014. URL http://dml.ece.ubc.ca/data/ DML-HDR/.
- [192] G. Valensi. Procédé de télévision en couleurs. Patent FR 841335, May 1939. URL http://worldwide.espacenet.com/ publicationDetails/biblio?CC=FR&NR=841335&KC=&FT= E&locale=en_EP.
- [193] J. J. Vos. On the cause of disability glare and its dependence on glare angle, age and ocular pigmentation. *Clinical and Experimental Optometry*, 86(6): 363-370, 2003. URL https://doi.org/10.1111/j.1444-0938. 2003.tb03080.x.
- [194] J. J. Vos and P. L. Walraven. On the derivation of the foveal receptor primaries. Vision Research, 11(8):799-818, 1971. URL http://www.sciencedirect.com/science/article/pii/ 0042698971900034.
- [195] J. J. Vos, O. Estévez, and P. L. Walraven. Improved color fundamentals offer a new view on photometric additivity. *Vision Research*, 30(6):937–943, 1990. URL https://doi.org/10.1016/0042-6989(90)90059-T.
- [196] L. Wang, Y. Tu, L. Chen, K. Teunissen, and I. Heynderickx. 27.2: Trade-off between luminance and color in RGBW displays for mobile-phone usage. *SID Symposium Digest of Technical Papers*, 38(1):1142–1145, May 2007. ISSN 2168-0159. URL https://doi.org/10.1889/1.2785510.
- [197] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004. URL https://doi.org/ 10.1109/TIP.2003.819861.

- [198] G. Ward. Real pixels. In J. Arvo, editor, *Graphics Gems II*, pages 80–83. Academic Press Professional Inc., San Diego, CA, USA, 1991. URL http: //dl.acm.org/citation.cfm?id=532136.
- [199] G. J. Ward. The RADIANCE lighting simulation and rendering system. In ACM Transactions on Graphics (TOG), SIGGRAPH '94, pages 459–472, New York, NY, USA, 1994. ACM. URL http://doi.acm.org/10.1145/192161. 192286.
- [200] A. B. Watson and D. G. Pelli. QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33(2):113–120, Mar 1983. URL https: //doi.org/10.3758/BF03202828.
- [201] A. B. Watson and J. A. Solomon. Model of visual contrast gain control and pattern masking. Journal of the Optical Society of America (JOSA), 14(9):2379– 2391, September 1997. URL http://josaa.osa.org/abstract. cfm?URI=josaa-14-9-2379.
- [202] E. H. Weber. *Die Lehre vom Tastsinne und Gemeingefühle auf Versuche gegründet*. Friedrich Vieweg und Sohn, Braunschweig, 1851.
- [203] W. D. Wright. The sensitivity of the eye to small colour differences. Proceedings of the Physical Society, 53(2):93–112, 1941. URL https://doi.org/10. 1088/0959-5309/53/2/301.
- [204] S. M. Wuerger, A. B. Watson, and A. J. Ahumada, Jr. Towards a spatio-chromatic standard observer for detection. *Proceedings of the SPIE*, 4662:159–172, 2002. URL https://doi.org/10.1117/12.469512.
- [205] S. W. Zamir, J. Vazquez-Corral, and M. Bertalmío. Perceptually-based gamut extension algorithm for emerging wide color gamut display and projection technologies. In *Proceedings of the SMPTE Annual Technical Conference and Exhibition*, pages 1–11, October 2016. URL https://doi.org/10.5594/ M001713.
- [206] E. Zerman, V. Hulusic, G. Valenzise, R. Mantiuk, and F. Dufaux. Effect of color space on high dynamic range video compression performance. In 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), pages 1–6, May 2017. URL https://doi.org/10.1109/QoMEX.2017. 7965632.
- [207] E. Zerman, G. Valenzise, and F. Dufaux. An extensive performance evaluation of full-reference HDR image quality metrics. *Quality and User Experience*, 2(1):5, 2017. URL https://doi.org/10.1007/s41233-017-0007-4.



Appendix: HdM-HDR-2014 Overview

In the following appendix, all scenes from the HdM-HDR-2014 data set introduced in Chapter 3 are summarized. First the scene name, length and frame rate are listed. Then, a thumbnail color graded for standard dynamic range sRGB color space is shown alongside a brief description of the scene. Finally, luminance quantiles and chromaticity plots of the reconstructed scene radiance visualize the dynamic range and gamut of the individual scenes.

The HdM-HDR-2014 data set consist of the reconstructed scene radiance and a color graded version. The reconstructed scene radiance is delivered as Open EXR files in Alexa Wide Gamut color space and is not meant for direct display as it is the equivalent to RAW camera files. The reconstructed scene radiance can be used for working on new camera characterization algorithms and to verify image processing chains. It can also be used for research on automatic color rendering, tone mapping and gamut mapping. To be displayed properly, the reconstructed scene radiance has to be color graded.

The color graded version of the HdM-HDR-2014 content is most useful for evaluating displays or compression algorithms. As an example, it can be used to evaluate tone mapping and gamut mapping algorithms that map down from the 0-4000 cd/m² Rec.2100 base format to lower luminance levels like standard dynamic range and smaller gamuts like Rec.709. The following plots illustrate the dynamic range and gamut of the reconstructed scene radiance. For the chromaticity plots, values outside the spectral locus can occur because of the matrix based camera characterization used in the HdM-HDR-2014 project. New algorithms for camera characterization will help to avoid out of locus values in the future [137].

A.1 Still Life

The still life scene listed in Table A.1 contains high contrast and standard skin tones, all in one reference image. This can be useful when comparing monitors or checking image processing pipelines.

Table A.1: HdM-HDR-2014: 'HDR Test' scene.

HDR Testimage, Night / Interior481 frames, 25 fpsLongshot: A couple with dark and pale skin tones is stand-
ing behind a color-checker and a transmittance gray scale.A couple with dark and pale skin tones and clothes is
lighted by means of a high contrast backlight. This pro-
vides bright highlights on skin, hair and clothes as well
as dark areas. The square above the gray scale on top
of the Ulbricht sphere represents a black/stray-light refer-
ence with virtually no radiance emitted.HDR Testimage



A.2 Wide Gamut and Moving Lights

An annual fair is filmed on location to provide saturated highlights and fast moving colorful objects. The saturated lights are dominant light sources that illuminate the actors with different colorful shades. The overall brightness and color of the scenes changes very fast, both outside on the fair shown in Table A.2, as well as inside the beer hall depicted in Table A.3. To provide even more light changes, the scenes are edited together as a sequence with multiple cuts.

Table A.2: HdM-HDR-2014: 'Carousel Fireworks' scene.

Carousel Fireworks, Ni	ght / Exterior	2536 frames, 25 fps
	Establishing Longshot: Crowded str	eet on an annual fair
	with illuminated fun rides in the bac	ckground
A A A	Fullshot: Moving carousel with color	red lights
	Mediumshot: Girl watching carousel	l
	High Angle Longshot: Fireworks	
× Zerro	'Carousel Fireworks' is a sequence	of shots acquired un-
	der available light at an annual fair	during the night. The
S S S S S S S S S S S S S S S S S S S	distinction of this scenery is to pre-	sent colorful self illu-
	minated objects and dark surroundi	ngs at the same time.
	Changing colored light sources illur	ninate the scenery in-
	cluding cloth and skin tones of the ac	ctors. The moving car-
MIL	riages of the carousel create blurred	light sources that are
	both filmed in standard speed and s	low-motion. The fire-
	work provides bright colored highlig	ghts glittering against
TANK C	a uniform black sky with the moon	in frame.







Appendix

The quantization issues shown in the chromaticity diagram of 'Carousel Fireworks 08' are due to the camera being limited to 8 bits of tonal resolution when shooting high frame rates.









Log Pixel Density

Table A.3: HdM-HDR-2014: 'Beerfest' scene.

Beerfest Lightshow, Night / Interior

2035 frames, 25 fps



Multiple Longshots: Crowded hall with fast moving lights 'Beerfest Lightshow' is filmed on location in a smoky beer hall while a light show is performed. This light show includes various kinds of fast switched and moving lights that send out bright and colorful light beams. Additionally, a mirror bowl reflects neutral colored light beams. Laser beams flash up and strobe lights as well as blinder lights brighten up the scenery temporarily.

Beerfest Lightshow 01











A.3 Low Key Scenes

The following low-key scenes are also filmed in with flickering light sources on location. But in this case, the light sources are mostly black body radiators covering a wide range of color temperatures. The mood of the sceneries is dominated by the natural single light sources in combination with dim low-key sceneries.

 Table A.4: HdM-HDR-2014: 'Fireplace' scene.

Fireplace, Dawn to Nigl	nt / Exterior	952 frames, 24 fps
	Fullshot: Tilt down from defocused	branches to group of
	persons at campfire.	
	Medium Fullshot: Persons standing	beside and behind
	flames at a fire site stoking up the fir	e.
	The warm light of the campfire illu	minates the persons
	that are surrounded by a snowy scen	ery. The fire provides
	a strong color contrast to the blue ar	nbient light at dawn.
	Torchlights and flying sparks against	the dark background
	provide high contrasts and with fast	movements.
Fireplace 01		



Fireplace 02



Table A.5: HdM-HDR-2014: 'Smith' scene.

Smith Welding, Mixed	Light / Interior	1102 frames, 25 fps
	Fullshot: A smith creates a lig	ht arc and flying sparks by
	welding iron.	
	In the dark mixed light of a b	lacksmith's shop, the low-
	key scenery is brightened up l	by an intense, fast moving
	point light source. The blue we	elding arc is reflected on the
	scenery and forms a color cont	rast to the yellow spraying

sparks and the warm fire in the background.

Smith Welding



Smith Hammering, Mixed Light / Interior 467 frames, 25 fps



Fullshot: A smith carries a forging blank from a fire to his anvil. Sparks fly when the iron is pounded.

Hammering incandescent iron at red heat creates bright spraying sparks that dominate the scenery in front of a dark background. In this setup, the fire and the forging blank are heated up to 800°C (1470°F).

Smith Hammering



A.4 Sunlight Scenes

The 'Sunlight Scenes' represent typical conditions of documentary filming. They are captured in the field under natural light conditions and regular sunlight. Their dynamic range exceeds the latitude of the dual camera image acquisition system. Hence, the framed sun orb and specular highlights are partly left in the clipping range, in order to save details in the shades. A travelling camera accentuates the appearance of textures. This can be seen on natural objects in the landscape scenery like in the 'Fishing' scen and on synthetic materials in an environment of architecture like in the 'Cars' scene. In both longshots of the following table, the brightness of the scene changes substantially over time, to provide a challenge for temporal tone mapping operators.

Table A.6: HdM-HDR-2014: 'Fishing' scene.

Fishing Longshot, S	Sunrise /	Exterior
---------------------	-----------	----------

834 frames, 25 fps



Longshot: A fisherman stands in front of a lake. The camera is travelling towards him.

'Fishing Longshot' represents a typical sunrise in nature. It shows scenery with a lake, trees and a fisherman in front of a sunny sky. The rising sun is just coming up and shines on surfaces like wood and moving water. Shafts of sunlight illuminate the morning haze in soft gradients and are mirrored on the water while specular highlights contrast to dark foreground objects.





Fishing Closeshot, Sun	rise / Exterior	371 frames, 25 fps
	Closeshot: A fishhook is thrown	into a lake and pulled out.
	In 'Fishing Closeshot', sunlight	reflections are glistening
	on the water surface stimulated	d by a fishhook. The expo-

pulled out. e glistening . The exposure is set to highlight-protection excluding the center of the reflection of the sun.

Fishing Closeshot



Table A.7: HdM-HDR-2014: 'Cars' scene.

Cars Longshot, Day / Exterior



820 frames, 25 fps

Longshot: Cars and flags on a stone paced plaza in front of a building made of steel and glass.

'Cars Longshot' is a back-lit scene showing the sun in frame. The sunlight emphasizes textures of stone, steel, glass and the flags by generating specular reflections. The camera pans from the sun to the entrance of a building. Contrast increases throughout the shot, as stray light is reduced at the end of the camera pan.





Cars Fullshot, Day / Exterior

442 frames, 25 fps

414 frames, 25 fps



Fullshot: Black car stands on plaza and is captured by a moving camera.

'Cars Fullshot' shows directional sunlight on a black car. The material appearance of the car finish and glass windows is emphasized by a camera travelling. The exposure is set to highlight protection excluding the specular reflections to preserve the dark shades under the car.

Cars Fullshot



Cars Closeshot, Day / Exterior



Closeshot: Details of a standing black car are captured by a moving camera.

'Cars Closeshot' shows specular reflections of directional sunlight moving over the front of a car. The surface feel of the car finish and the glass windows are emphasized through a close framing. The exposure is set to highlight protection excluding the specular reflections of the sun.

Cars Closeshot



A.5 High Contrast Skin Tones

Frame Number

The following sequences focus on the reproduction of skin tones in different lighting situations, ranging from documentary settings in the 'Bistro' scene to movie settings like the 'Poker' scene. They are set up under controlled lighting conditions in a studio. Faces, hands, hair and textiles, are partially lit up and exposed to the maximum luminance available in the latitude of the recording device, but they are as well presented in marginal illumination. The mean luminance of the 'Poker Travelling Slowmotion' scene changes over time by temporarily covering the brightest areas in the image while the brightness and type of the illumination is changed in the 'Showgirl 2' scene.

Table A.8: HdM-HDR-2014: 'Bistro	'scene.
----------------------------------	---------

Bistro, Day / Interior	969 frames, 24 fps		
	Medium Fullshot: A man sits at a table with the sun shining		
	on him through a window.		
	<i>Fullshot:</i> A waiter steps from shade into sunlight followed		
	by a woman coming from the dark part of the room. She		
	points at him with a gun.		
The second	The 'Bistro' sequence simulates an available-light situation,		
	where the sun shines through a window as a single source		
	light. This scenery combines local bright sunlight at the		
	window with a dark guest room resulting in a high con-		
	trast scenery that represents a difficult lighting situation		
	typically encountered in documentary filming. The set-ups		
	are staged to show skin tones, hair and textures like glass,		
	water, wood and textiles in sunlight and in the shade.		
Bistro 01			
10^{3} 10^{2} 10^{3} 10^{2} 10^{1} 10^{2} 10^{1} 1	$\begin{array}{c} \textbf{Percentiles} \\ 100.0\% \\ 99.9\% \\ 99.0\% \\ 99.0\% \\ 99.0\% \\ 10.0\% \\ 10.0\% \\ 10.0\% \\ 1.0\% \\ 0.1\% \\ 0 \end{array} \begin{array}{c} 0.6 \\ 0.5 \\ 0.4 \\ 0.6 \\ 0.7 \\ 0.4 \\ 0.7 \\ 0.$		

CIE 1976 u'



Table A.9: HdM-HDR-2014: 'Poker' scene.





10⁻⁴

200

400

600

Frame Number

800



Table A.10: HdM-HDR-2014:	'Showgirl' scene.
---------------------------	-------------------

Showgirl 1, Night / Inte	erior	776 frames, 25 fps	
	Closeshot: Actress sitting in front of	a makeup mirror. Il-	
	luminated by warm tungsten lightin	.g.	
	'Showgirl 1' shows an actress sitting	g in front of a makeup	
	mirror. Tungsten light bulbs illuminate her skin tones and		
	create specular highlights on her costume, makeup, jew-		
	elry and other reflecting props. This glamorous mood is		
	often intended in film productions to emphasize the beauty		
	of an actress.		
Showgirl 1			
$(10^{3})^{2}$	Percentiles 0.6 100.0% 99.9% 99.0% 90.0% 90.0% 0.4 90.0% 0.3 10.0% 0.2 10.0% 0.1 0.1% 0.1%	Camera Gamut	

0

0

0.1 0.2 0.3 0.4 0.5 0.6

CIE 1976 u

Showgirl 2, Night / Interior

341 frames, 25 fps



light changes from tungsten to daylight. The 'Showgirl 2' scene executes a light change from tungsten light to bright stagelight from an HMI-lamp. Thus the skin tone of the actress is shown in two extreme lighting situations throughout one take. The dull feather boa serves as a diffuse white reference, whereas the glistening of the costume and jewelry is brighter than diffuse white.

Closeshot: Actress standing up from makeup-table while



A.6 Additional Ressources

All HdM-HDR-2014 scenes can be downloaded from the project website [72], both the reconstructed scene radiance in scene referred state and the color graded version in display referred state. The color graded version is graded for Rec.2100 gamut with 4000 cd/m^2 peak white. The web site provides additional technical details like exposure and the choice of lenses.