

TECHNISCHE UNIVERSITÄT ILMENAU

DISSERTATION

**Konzept zur Einführung von HDR im
Broadcast mithilfe präferenzbasierter
Kontrastkompression**

Autor:
M.Eng. Lucien Lenzen

Erstgutachter:
Prof. Dr.-Ing. Dr. rer. nat. h.c.
mult. Karlheinz Brandenburg

zur Erlangung des akademischen Grades Doktor-Ingenieur (Dr.-Ing.)

vorgelegt der

Fakultät für Elektrotechnik und Informationstechnik

23. November 2020

urn:nbn:de:gbv:ilm1-2021000124

Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen:

1. In den durch mich betreuten Abschlussarbeiten von Mark Benyamin [9], Yannick Hoff [46], Pascal Kutschbach [66], Daniel Schmitt [110] und Nico Schultz [111] wurden subjektive Untersuchungen vorbereitet, durchgeführt und / oder dazugehörige Daten erhoben. Dies ist an den entsprechenden Stellen in der Arbeit kenntlich gemacht.

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt.

Ich bin darauf hingewiesen worden, dass die Unrichtigkeit der vorstehenden Erklärung als Täuschungsversuch bewertet wird und gemäß § 7 Abs. 10 der Promotionsordnung den Abbruch des Promotionsverfahrens zur Folge hat.

Fulda, November 2020

”Wenn wir fernsehen, schauen wir auf elektronische Pixel: die schönste Sinnestäuschung seit es Augen gibt.”

Manfred Poisel

Abstract

Konzept zur Einführung von HDR im Broadcast mithilfe präferenzbasierter Kontrastkompression

von M.Eng. Lucien Lenzen

HDR (High Dynamic Range) provides the opportunity to capture a much wider scene contrast range than it would be the case in today's HD broadcast. Details in the shadows and highlights can be preserved at the same time. But at the display side the possibilities are heterogenous and more limited. An adaption (contrast compression) is needed to let all viewers benefit from the enhanced recording quality.

Manual techniques for contrast compression are known from cinematic post productions, whereas automatic methods are used at computer graphics. But due to the specific requirements of broadcast they cannot be taken over easily. One fundamental challenge is the viewers preference.

Therefore, the work aims at quantifying the viewer preference in terms of brightness and colorfulness. Based on the results, an algorithmic solution for modifying contrast compression for broadcast applications is proposed. It is outlined that the enhancement in image quality was be verified by objective and subjective studies. Finally, a sustainable path for a comprehensive implementation of HDR is developed. Exemplary workflow implementations and field tests results for the demonstration and verification of the findings are presented.

Abstract

Konzept zur Einführung von HDR im Broadcast mithilfe präferenzbasierter Kontrastkompression

von M.Eng. Lucien Lenzen

HDR (High Dynamic Range) ermöglicht es, einen weitaus größeren Kontrastumfang einer Szene einzufangen als es im HD-Broadcast der Fall wäre. In der Folge können Details sowohl in den Lichtern als auch in den Schatten erhalten werden. Allerdings sind die Möglichkeiten zur Wiedergabe sehr heterogen und meist deutlich limitierter. Um trotzdem alle Zuschauer von der gesteigerten Aufnahmequalität profitieren zu lassen, wird eine Anpassung - auch Kontrastkompression genannt - nötig.

Manuelle Techniken zur Kontrastkompression sind aus der filmischen Postproduktion bekannt, während automatische Verfahren in der Computergrafik Anwendung finden. Aufgrund der speziellen Anforderungen des Broadcast lassen sich diese jedoch nicht einfach übertragen. Eine grundlegende Herausforderung besteht dabei in der Präferenz des Zuschauers.

Das Ziel der Arbeit ist es deshalb, die Zuschauerpräferenz bezüglich der Helligkeits- und Farbwahrnehmung zu quantifizieren und anschließend auf diesen Ergebnissen eine algorithmische Lösung zur Anpassung der Kontrastkompression für die Anwendung beim Broadcast anzubieten. Mithilfe von objektiven und subjektiven Untersuchungen soll gezeigt werden, wie sich hierdurch die Bildqualität signifikant steigern lässt. Abschließend gilt es anhand von beispielhaften Workflows und Feldversuchen einen Weg für die flächendeckende Einführung von HDR aufzuzeigen.

Danksagungen

Mein Dank gilt an erster Stelle Herrn Prof. Brandenburg, der das kooperative Verfahren und somit diese Dissertation erst möglich gemacht hat.

Besonderer Dank gilt auch Herrn Prof. Christmann und Herrn Prof. Hedtke, durch welche ich eine enge Betreuung genoss. Der intensive Austausch und Diskurs haben die Arbeit maßgeblich geprägt. Auch ihr außerordentlicher Einsatz für das dahinterstehende Forschungsprojekt war eine nötige Grundlage.

Besonders erwähnen möchte ich auch Herrn Prof. Ruppel, welcher die wissenschaftliche Betreuung an der Hochschule RheinMain komplettierte.

Bedanken möchte ich mich ebenso bei meiner Familie und hier zuerst bei meiner Frau, welche mich stets motiviert und ermutigt hat. Auch war es hilfreich, in ihr eine Mitstreiterin gefunden zu haben. Der weiteren Familie danke ich vor allem für das Korrekturlesen einer nicht unbedingt trivialen Lektüre, welche für den Laien herausfordernd sein muss.

Abschließend gilt mein Dank den studentischen Mitarbeitern sowie meinen Projekt-, Bachelor- und Masterabreitern für ihr Engagement im Forschungsprojekt.

Inhaltsverzeichnis

Erklärung	i
Abstract	iii
Danksagungen	v
Inhalt	vi
Abbildungsverzeichnis	ix
Tabellenverzeichnis	xi
Abkürzungsverzeichnis	xii
Veröffentlichungen	xv
1 Einleitung, Grundlagen und Definitionen	1
1.1 Einleitung und Überblick	1
1.2 Definitionen und Grundlagen	4
1.2.1 Lichtstrom, Lichtstärke, Leuchtdichte, Helligkeit, Luma und Beleuchtungsstärke	4
1.2.2 Kontrastumfang, Dynamikumfang, Kontrastkompression, Kontrast und <i>Kontrastparadoxon</i>	6
1.2.3 Verhältniszahl, Blendenstufen und Dekaden	7
1.2.4 LDR, SDR und HDR	8
1.2.5 Buntheit, Chroma, Sättigung und Farbton	8
1.2.6 Farbeindruck	9
2 Menschliche Wahrnehmung und Stand der Fernsehtechnik	10
2.1 Motivation	10
2.2 Menschliche Wahrnehmung	11
2.2.1 Spatiale Auflösung	11
2.2.2 Kontrastumfang	12
2.2.3 Simultaner Kontrastumfang	13
2.2.4 Fotorezeptorreaktionskurve und Adaption	13
2.2.5 t.v.i., CSF und JND	14
2.3 Technischer Fortschritt bei Bildaufnahme- und -wiedergabesystemen	16

2.3.1	Entwicklung im Bereich der Kameras	16
2.3.2	Entwicklung im Bereich der Displays	17
2.3.3	Entwicklung im Bereich der Fernsehnormen	19
2.4	Zielsetzung der Arbeit	20
2.5	Potenzielle Workflows und ihre Verwendbarkeit	22
2.5.1	Autarker HDR- und SDR-Workflow	22
2.5.2	Synergetischer HDR-SDR-Workflow	24
2.5.3	Beispiel PQ und Dolby Vision	26
2.5.4	Beispiel HLG	30
2.5.5	Zwischenfazit bezüglich der Broadcast-Workflows	31
3	Tonemapping und Broadcast-Anforderungen	33
3.1	Funktionsweise von Tonemapping-Operatoren	33
3.1.1	Einleitung	33
3.1.2	Globale Operatoren	35
3.1.3	Lokale Operatoren	36
3.1.4	Temporaler Aspekt	37
3.2	Tonemapping und Broadcast-Anforderungen	39
3.2.1	Echtzeit	39
3.2.2	Szenenwechselerkennung und temporale Interpolation	40
3.2.3	Universale Verwendbarkeit und Robustheit gegen Artefakte	41
3.2.4	Akzeptanz durch den Zuschauer	44
3.2.5	Zusammenfassung der Broadcast-Anforderungen	47
4	Visuelle Präferenz bezüglich der Helligkeitswahrnehmung	49
4.1	Genauere Untersuchung zur simultanen Kontrastwahrnehmung	49
4.1.1	Kritische Betrachtungen zum Stand der Forschung	49
4.1.2	Testdesign und -durchführung	51
4.1.3	Resultate	53
4.1.4	Zwischenfazit bezüglich des simultanen Kontrastumfangs	57
4.2	Stand der Forschung bezüglich der Präferenz	58
4.3	Qualitätsgewinn von HDR im Vergleich zum heutigen SDR	62
4.4	Coloristen-Test	64
4.4.1	Aufbau und Durchführung Coloristen-Test	64
4.4.2	Ergebnisse Coloristen-Test	65
4.5	Probanden-Grading	69
4.5.1	Aufbau und Durchführung Probanden-Grading	69
4.5.2	Ergebnisse Probanden-Grading	73
4.5.3	Zusammenfassung Probanden-Grading	80
4.6	Zwischenfazit bezüglich der Helligkeitspräferenz	81
5	Visuelle Präferenz bezüglich der Farbwahrnehmung	83
5.1	RGB-Farbgamut	83
5.2	Klassische Sättigungskompensation auf RGB	85
5.3	XYZ-Farbraum	87
5.4	IPT-Farbraum	88
5.5	Sättigungskompensation nach Pouli	88

5.6	hdrIPT, IC_tC_p und YC_bC_r	89
5.7	Farberscheinungsmodelle	91
5.8	Erste Messungen der subjektiven Präferenz bezüglich der Sättigungswahrnehmung	92
5.8.1	Sättigung beim Coloristentest	92
5.8.2	Sättigung oder Chroma	93
5.8.3	Sättigung beim Probanden-Grading	95
5.9	Probandentest Farbpatches	98
5.9.1	Aufbau und Durchführung	98
5.9.2	Ergebnisse und Fehlerbetrachtung	100
5.10	Konstante Buntheit	102
5.10.1	Theorie zur konstanten Buntheit	102
5.10.2	Testaufbau zur konstanten Buntheit	108
5.10.3	Ergebnisse zur konstanten Buntheit	109
5.11	Zwischenfazit bezüglich der Präferenz bei der Farbwahrnehmung	111
6	EVI: Theoretisches Modell zum TMO-Tuning	113
6.1	Virtuelle Blende	113
6.2	Sektionales Tonemapping und bilaterale Filterung	119
6.3	Temporale Glättung und Szenenwechsel	124
6.4	Chroma-Korrektur	126
6.5	Zusammenfassung EVI	127
7	Nachweis der Qualitätssteigerung und praxistauglicher Workflow	128
7.1	Auswirkungen auf den PSNR	128
7.1.1	Stand der Forschung bezüglich des PSNR	129
7.1.2	Versuchsaufbau und -durchführung (PSNR)	130
7.1.3	Auswertung (PSNR)	132
7.2	Auswirkungen auf den MOS	135
7.2.1	Versuchsaufbau und -durchführung (MOS)	135
7.2.2	Auswertung (MOS)	138
7.3	Workflow mit Gamma- und / oder HLG-Distribution	139
7.4	GPU-basierter Echtzeitprototyp mit GUI-Steuerung	143
7.5	Praxistest und Anwendungsfälle	145
7.6	Zwischenfazit Praxis	148
8	Fazit und Ausblick	149
A	Transformationsmatrizen	154
A.1	Transformation nach XYZ	154
A.2	Transformation aus XYZ	154
A.3	Transformationen Dolby	155
	Literaturverzeichnis	156

Abbildungsverzeichnis

2.1	Auswirkung unterschiedlicher Blendeneinstellungen	11
2.2	Naka-Rushton-Gleichung	14
2.3	Unterschiedliche CSE- bzw. t.v.i.-Funktionen im Vergleich	15
2.4	Parallele SDR- und HDR-Produktion	22
2.5	Szene bei einem Fußballspiel bei starker Sonneneinstrahlung	23
2.6	HDR nach SDR Workflow	24
2.7	Dolby PQ	26
2.8	Ursprünglicher Workflow bei Dolby Vision	27
2.9	Codierung bei Dolby Vision	28
3.1	Die Skyline von Frankfurt mit starken Halo-Artefakten	41
3.2	4 Beispiele des iCAM06-Operators	43
3.3	Szene aus einem Fußballspiel, die mit dem Drago TMO komprimiert wurde	44
3.4	Hinterhof als natürlichste und als geschmacklich präferierte Variante	46
4.1	Testbilder zur Evaluierung der menschlichen simultanen Kontrastwahrnehmung	53
4.2	Ergebnisse des Probandentests bezüglich des simultanen Kontrastumfangs	54
4.3	Näherungsformel für wahrnehmbaren Kontrastumfang	56
4.4	Die Ergebnisse des Coloristentests	66
4.5	Die Ergebnisse des Coloristentests im direkten Vergleich	66
4.6	Korrelation zwischen Probandenbewertung und Kontrast, Helligkeit oder Clipping	68
4.7	Oberfläche für das Probandengrading	70
4.8	3D-Plot mit Helligkeit, Display- und Umgebungsleuchtdichte	73
4.9	Gesamtauswertung des Probadens-Gradings	74
4.10	Vertrauensintervalle im Probadens-Gradings	77
4.11	Korrelation von eingestellten Werten mit Szenenparametern	78
5.1	Verschiedene Farbgamuts auf der CIE-Normfarbtafel	84
5.2	Korrelation zwischen I und C_p	90
5.3	Korrelation zwischen Probandenbewertung und Sättigung	93
5.4	Auswertung des Probadens-Gradings bezüglich der Sättigung	96
5.5	Vertrauensintervalle des Sättigungsparameteres beim Probadens-Grading	96
5.6	Korrelation der Sättigungseinstellung mit den Szenenparametern	97
5.7	Vergleich der Chroma-Korrektur von Pouli mit der Koeffizienten-Methode	98
5.8	Farbpatches zur Messung der Chroma-Kompensation pro Farbton	99
5.9	Farbpatches zur Messung der Chroma-Kompensation im CIE-xy-Diagramm	100

5.10 Ergebnisse der subjektiven Untersuchung mit Farbpatches	101
5.11 Konstante Buntheit beim lokalen Tonemapping	106
5.12 Vergleichsbilder mit verschiedenen Chroma-Kompensationen	107
5.13 Ergebnisse der Probandenuntersuchung für die fünf verschiedenen Verfahren zur Chroma-Kompensation	109
6.1 Schritte der virtuellen Blende	116
6.2 Beispiele der virtuellen Blende	118
6.3 Auswirkung der virtuellen Blende	119
6.4 Sektorenweise Berechnung der virtuellen Blende	120
6.5 Auswirkung der sektionalen Blende	120
6.6 Einzelschritte der bilateralen Filterung	123
6.7 Vergleich zwischen der bikubischen und der bilateralen Filterung	123
7.1 Vergleich der TV-Simulationen mit den EVI-Versionen	133
7.2 PSNR bei konstanter Datenrate	133
7.3 Verhältnis der I-Frame- zu B-Frame-Datenrate	133
7.4 Korrelationen mit den die Szene beschreibenden Parametern	134
7.5 Ablauf eines Durchlaufs beim DSCQS und Bewertungsbogen	136
7.6 Vergleich der klassischen SDR-Variante mit EVI	137
7.7 Vergleich von EVI zu klassischem SDR anhand des MOS	138
7.8 EVI-Workflow bei reiner SDR-Ausspielung	140
7.9 EVI-Workflow bei Konvertierung eines HDR-Signals am Ende der Kette	141
7.10 EVI-Workflow mit Homogenisierung von HDR und SDR	142
7.11 EVI-GUI	144
7.12 Vergleich von EVI und AJA FS-HDR bei der Produktion von DSDS	146
7.13 Vergleich von EVI und Sony bei der Produktion eines Openair-Konzertes	146

Tabellenverzeichnis

2.1 Übersicht der HDR-Formate	26
3.1 Übersicht über die globalen TMOs	35
3.2 Übersicht über die lokalen TMOs	37
4.1 Übersicht über die Leuchtdichtewerte der Grautreppe	52
4.2 Übersicht über die Variationen der drei Testparameter	52
7.1 Übersicht der Testsequenzen für die Datenratenanalyse	130

Abkürzungsverzeichnis

ACES	Academy Color Encoding Specification
APS	aktiver Pixelsensor
BBC	British Broadcasting Company
BR	Bayrischer Rundfunk
CCD	Charge Coupled Device
CIE	Commission Internationale de l'Éclairage
CMOS	Complementary Metal Oxide Semiconductor
CMU	Content Mapping Unit
CRT	Kathodenstrahlröhre
CPU	zentrale Verarbeitungseinheit
CSF	Contrast Sensitivity Function
DLP	Digital Light Projector
DMA	Direct Memory Access
DSCQS	Double-Stimulus Continuous Quality-Scale
EBU	European Broadcasting Union
EDR	Extended Dynamic Range
EOTF	elektro-optische Transferfunktion
EU	Europäische Union
EV	Exposure Value
EVI	Enhanced Video Imaging
GPU	Grafikprozessor

GUI	Graphical User Interface
HEVC	High Efficiency Video Coding
HD	High Definition
HDR	High Dynamic Range
HDTV	High Definition Television
HFR	High Frame Rate
HLG	Hybrid-Log-Gamma
HVS	Human Visual System
IBC	International Broadcasting Conference
IET	Institution of Engineering and Technology
JND	Just Noticeable Difference
JPEG	Joint Photographic Experts Group
KI	künstliche Intelligenz
LCD	Liquid Crystal Display
LDR	Low Dynamic Range
LED	Leuchtdiode
LKW	Lastkraftwagen
LMS	long-medium-short cone response data
LUT	Lookup-Table
MOS	Mean Opinion Score
NAB	National Association of Broadcasters
NHK	Japan Broadcasting Corporation
OETF	opto-elektrische Transferfunktion
OLED	organische Leuchtdiode
OOTF	opto-optische Transferfunktion
SDR	Standard Dynamic Range
SDI	Serial Digital Interface
SMPTE	Society of Motion Picture and Television Engineers

TMO	Tonemapping-Operator
PAL	Phase Alternating Line
PSNR	Peak Signal-to-Noise Ratio
PQ	Perceptual Quantizer
PCIe	Peripheral Component Interconnect Express
RTL	Radio Television Luxembourg
t.v.i.	Threshold Versus Intensity Function
UHDTV	Ultra High Definition Television
WDR	Westdeutscher Rundfunk

Veröffentlichungen

Teilergebnisse der vorliegenden Arbeit wurden bereits vorab wie folgt publiziert:

Fachzeitschriften

LENZEN, L. ; HEDTKE, R. ; CHRISTMANN, M. : How Tone Mapping Influences the Bit Rate and the Bit Depth of Coded Sequences. In: *SMPTE Motion Imaging Journal 127 (2018)*, June, Nr. 5, S. 38-43

LENZEN, L. ; HEDTKE, R. ; CHRISTMANN, M. : HDR in Consideration of the Abilities of the Human Visual System. In: *SMPTE Motion Imaging Journal 128 (2019)*, June, Nr. 5, S. 40-45

Konferenzen

LENZEN, L. : HDR for legacy displays using Sectional Tone Mapping. In: *Proceedings of the IBC conference (2016)* und *The Best of IET and IBC (2016)*

LENZEN, L. ; CHRISTMANN, M. : Subjective viewer preference model for automatic HDR down conversion. In: *Proceedings of the IS&T International Symposium on Electronic Imaging (2017)*

LENZEN, L. ; HEDTKE, R. ; CHRISTMANN, M. : How Tone Mapping influences the bit rate and the bit depth of coded sequences. In: *Proceedings of the IBC conference (2017)*

LENZEN, L. ; HEDTKE, R. ; CHRISTMANN, M. : HDR in consideration of the abilities of the human visual system. In: *Proceedings of the IBC conference (2018)*

Patentschriften

LENZEN, L. ; CHRISTMANN, M. : Patent DE102015119137 B3 - Erfindung betreffend Verfahren zur Blendeneinstellung. 2017

Meiner Frau

Kapitel 1

Einleitung, Grundlagen und Definitionen

1.1 Einleitung und Überblick

Die moderne Medienwelt entwickelt laufend neue Begrifflichkeiten. Eine lautet Immersion. Der Zuschauer soll sich als Teil der Handlung verstehen und im Idealfall ganz und gar darin versinken. Die reale Sehsituation, zuhause im Wohnzimmer auf der Couch sitzend, ist in der Folge vollkommen ausgeblendet. Damit das Eintauchen in eine künstliche Welt funktioniert, muss natürlich die inhaltliche und erzählerische Umsetzung ansprechend und fesselnd sein. Aber auch die technischen Gegebenheiten spielen eine entscheidende Rolle. Denn je besser das Abbild der realen Welt gelingt, desto eher ist der Zuschauer geneigt zu glauben, sich eben dort zu befinden. So definiert Helge Seetzen das Ziel, dass die Betrachtung einer Szene durch den Fernseher mit der durch ein Fenster identisch sein soll. [112]

Vier technologische Neuerungen sind Bestandteil des Standards für Ultra High Definition Television (**UHDTV**) und somit auch Parameter für die Immersion.^[1] Der wohl bekannteste Bestandteil ist die Erhöhung der örtlichen Auflösung. Das Schlagwort heißt hier 4k. Wurde die Anzahl der Pixel von **PAL** zu High Definition (**HD**) verfünffacht, folgt nun eine weitere Vervierfachung auf 2160 x 3840 Pixel. Die Japan Broadcasting Corporation (**NHK**) wird sogar die Olympischen Spiele 2021 in Tokyo in 8k (4320 x 7680) produzieren und übertragen. Dies erinnert ein wenig an das Megapixelwettrennen, das sich die Kamerahersteller zu Beginn des neuen Jahrtausends lieferten, bis die

¹Die Immersion kann auch rein wiedergabeseitig gesteigert werden. So wird sie z. B. durch die physikalische Größe des Wiedergabedisplay beeinflusst. Dieser Parameter ist aber für die Senderseite unerheblich und wird deshalb nicht weiter verfolgt.

Einsicht siegte, dass die Anzahl der Pixel nicht den einzigen wichtigen Parameter für die Bildqualität darstellt.

Beim Bewegtbild kann neben der örtlichen auch die zeitliche Auflösung erhöht werden. Diese so genannte High Frame Rate (**HFR**) sorgt für eine flüssigere und schärfere Bewegungsdarstellung. Die Figuren wirken in der Folge plastischer. Dass hiervon ein deutlicher subjektiver Qualitätsgewinn ausgeht, konnte, neben anderen, auch der Autor selbst in seiner Masterarbeit nachweisen. [71]

Untersuchungsergebnisse z. B. der European Broadcasting Union (**EBU**) [47] zeigen allerdings, dass die Erhöhung des Kontrastumfangs - High Dynamic Range (**HDR**) - von allen bisher angesprochenen Neuerungen das größte Potenzial bietet; vor allem dann, wenn jener mit dem vierten Bestandteil der im Zusammenhang mit UHDTV steht, einem erweiterten Farbraum, einhergeht. Denn aktuelle Videostandards decken nur einen Bruchteil dessen ab, was der Mensch an Helligkeitsunterschieden und Farben in der Natur wahrnehmen kann. Die Verknüpfung der beiden Aspekte ist deshalb sinnvoll, da sie eine starke Korrelation aufweisen. So wird der Farbeindruck auch über die Helligkeit mitgeprägt. [131]

Bis vor wenigen Jahren war die Aufzeichnung von solchem Material nur mit aufwendigen Tricks möglich (z. B. durch halbtransparente Spiegel [41]). Doch durch den enormen Leistungssprung der Kamerasensoren dürfte es in naher Zukunft möglich sein, fast jede wahrnehmbare Farbe in einem sehr weiten Helligkeitsspektrum aufzuzeichnen. Die Reproduktion auf einem Display hingegen ist technisch deutlich eingeschränkter. Zwar sind inzwischen die ersten so genannten **HDR**-Displays erhältlich, diese stehen den Kameras aber sowohl im Helligkeits- als auch im Farbumfang deutlich nach.

Damit sind die Fernsehanstalten erstmals mit dem Problem konfrontiert, dass das Format bei der Akquisition sich grundsätzlich vom dem der Distribution unterscheiden kann.² Konnte man zu Zeiten der Kathodenstrahlröhre (**CRT**) noch davon ausgehen, dass die Bedingungen bei der Abnahme im Studio und bei der Wiedergabe an den Heimgeräten bezüglich der Leuchtdichte und der Farben ähnlich limitiert waren, kann dies in Zukunft nicht mehr garantiert werden. Trotz zentraler Ausstrahlung eines Kanals muss allerdings weiterhin gewährleistet sein, dass jeder Empfänger ein bestmögliches Bild erhält. Dies gilt in besonderem Maße für die öffentlich-rechtlichen Fernsehanstalten, die laut Rundfunkstaatsvertrag [1] zur Teilhabe und somit auch zur Kompatibilität mit älteren Geräten verpflichtet sind.

Als Folge wird unter anderem eine Downkonvertierung benötigt. Im Vergleich zur Erhöhung der räumlichen oder zeitlichen Auflösung ist diese weitaus weniger trivial. Übersteigt

²Zwar wurden in der Übergangsphase zu **HDTV** auch Konvertierungen notwendig, diese sind aber mit den hier erörterten Anforderungen nicht vergleichbar.

die räumliche Auflösung des Videomaterials die des Displays, können z. B. durch Mittelwertbildung mehrerer Pixel die Anzahl eben jener unkompliziert verkleinert werden. Übersteigt die Anzahl der Bilder pro Sekunde die Wiedergabefrequenz des Displays, können im einfachsten Fall Zwischenbilder ausgelassen werden. Übersteigen aber Kontrastumfang und Farbraum die Möglichkeiten der Wiedergabe, muss eine Stauchung des Kontrastumfangs und eine Transformation des Farbraumes vorgenommen werden. Eine massive Beeinflussung oder im schlimmsten Falle Beeinträchtigung des Bildeindrucks können die Folge sein, wie an späterer Stelle auch gezeigt werden wird. Auf der anderen Seite besteht allerdings auch die Chance, das heutige Fernsehbild deutlich aufzuwerten, indem es aus einem qualitativ hochwertigeren Quellmaterial erzeugt wird.

Zielsetzung der Arbeit ist es, die flächendeckende Einführung von **HDR** im Broadcast zu ermöglichen, indem eine Lösung erarbeitet wird, die einen hohen Grad an Kompatibilität bietet, gleichzeitig aber alle potenziellen Empfänger von der gesteigerten Qualität des Quellmaterials profitieren lässt. Das Wort Broadcast ist dabei als Überbegriff für all jene zu verstehen, welche Programme nach dem Prinzip der Eins-zu-Viele-Kommunikation betreiben und verbreiten oder daran beteiligt sind. Charakteristisch für dieses Feld sind fest definierte Videosignalwege und damit verbundene Geräte und Arbeitsabläufe. Diese Ketten werden als Workflows bezeichnet. Die in der Arbeit gewonnenen Erkenntnisse sollen am Ende entsprechend in anwendbare Workflows münden.

Zu Beginn der Arbeit müssen zunächst die nötigen Grundlagen erarbeitet werden. Dies beinhaltet Begriffsdefinitionen, einen Einblick in die menschliche Wahrnehmung sowie die Betrachtung des Stands der Fernsehtechnik. Letzterer beinhaltet bereits Vorschläge für **HDR**-Workflows. Es wird sich zeigen, dass dabei die automatische Kontrastkompression den Kernaspekt darstellt. Allerdings werden die derzeit verfügbaren Lösungen als unzureichend erachtet. Auch wenn die Herausforderung der Kontrastkompression für den Fernsbereich neu ist, so hat es sie bereits auf anderen Gebieten gegeben. Dem vorliegenden Fall am ähnlichsten ist die Problematik des analogen Films. Auch hier war es möglich bei der Akquisition mehr Kontrastumfang einzufangen, als auf die meisten Wiedergabemedien gebannt hätte werden können. Ähnlich verhält es sich auch bei der Computergrafik, bei der jede beliebige Leuchtdichte künstlich errechnet werden kann.

In Kapitel **3** sollen deshalb diverse, aus der Computergrafik stammende Tonemapping-Operatoren (TMOs) hinsichtlich ihrer Verwendbarkeit im vorliegenden Kontext analysiert werden. Dazu werden die Anforderungen des Broadcast-Marktes erhoben und mit dem Stand der Technik abgeglichen. Es wird sich zeigen, dass neben anderen Aspekten (z. B. Filterung in der zeitlichen Domäne oder Robustheit gegen Artefakte) gerade die Präferenz des Zuschauers eine ungelöste Problematik darstellt.

Im Zentrum dieser Arbeit soll deshalb die Quantifizierung der Präferenz bezüglich der Helligkeits- und Farbwahrnehmung bei unterschiedlichen äußeren Bedingungen und Bildvorlagen stehen. Hierfür werden in den Kapiteln 4 und 5 mehrere Probandenuntersuchungen durchgeführt. So sind die Teilnehmer unter anderem dazu aufgerufen, eine vereinfachte Kontrastkompression selbst durchzuführen. Weil Zielkonflikte bestehen, müssen dabei Abwägungsentscheidungen getroffen werden.

Aus den Ergebnissen soll in Kapitel 6 ein Algorithmus abgeleitet werden, welcher das Tonemapping hinsichtlich des Bildeindrucks und der im Broadcast-Bereich benötigten Robustheit optimiert. Dass das Verfahren zu einem nachweislichen Qualitätsgewinn führt, gilt es in Kapitel 7 zu belegen. Darüber hinaus werden dort weitere Aspekte betrachtet, die für den Übergang zur Praxis von Relevanz sind. Es wird aufgezeigt, wie der Algorithmus in reale Workflows eingebettet werden kann und wie ein Einführungsszenario für die flächendeckende Verwendung von HDR aussehen könnte, das auch noch Jahre später Gültigkeit besäße. Auf diese Weise soll letztendlich jeder Zuschauer von HDR profitieren, egal welches Display er besitzt.

1.2 Definitionen und Grundlagen

Zunächst gilt es einige Begriffe zu definieren und sie klar voneinander abzugrenzen. Vor allem in der Umgangssprache wird nicht immer auf die korrekte Verwendung der einzelnen Termini geachtet. Aber auch in der Fachliteratur finden sich unterschiedliche Angaben. Deshalb gilt es die für diese Arbeit gebräuchliche Verwendung zu klären, um im Folgenden eindeutige Aussagen zu gewährleisten. Das englische Äquivalent wird zusätzlich in Klammern genannt. Die Termini bezeichnen dabei sowohl messtechnische als auch wahrnehmungspsychologische Größen und grenzen diese zunächst bei achromatischen und später bei chromatischen Aspekten voneinander ab. Viele Definitionen gehen dabei auf Fairchild [33] zurück, der sich selbst wiederum am *International Lighting Vocabulary* [19] orientiert hat.

1.2.1 Lichtstrom, Lichtstärke, Leuchtdichte, Helligkeit, Luma und Beleuchtungsstärke

Lichtstrom (luminous flux): Der Lichtstrom gibt an, wie viel für das menschliche Auge wahrnehmbares Licht eine Lichtquelle pro Zeiteinheit abstrahlt. Dazu wird die Strahlungsleistung mit der Empfindlichkeit des Auges hinsichtlich unterschiedlicher Wellenlängen gewichtet. Die Maßeinheit ist Lumen (lm).

Lichtstärke (luminous intensity): Die Lichtstärke gibt den Lichtstrom pro Raumwinkel an und zeigt somit, wie intensiv eine Lichtquelle in eine bestimmte Richtung leuchtet, beispielsweise ins Auge des Betrachters. Sie wird in Candela (*cd*) gemessen.

Leuchtdichte (luminance): Die für diese Arbeit wichtigste Kenngröße für die Menge an Licht ist die Leuchtdichte. Sie ist definiert über die Lichtstärke im Bezug zur Größe der leuchtenden Fläche. Die Einheit ist Candela/Quadratmeter (cd/m^2), wobei im englischsprachigen Raum hierfür auch oft die Einheit *nits* verwendet wird ($1 \text{ nit} = 1 \text{ cd}/\text{m}^2$). [45] Auf diese Weise werden die ursprüngliche Szene und die darin befindlichen Photonen genau beschrieben. Deshalb wird auch von *scene referred* gesprochen. Das Symbol ist *L* oder *Y*.

Absolute Helligkeit (brightness): Die absolute Helligkeit ist die visuell wahrgenommene und bewertete Leuchtdichte. Es wird somit berücksichtigt, dass das menschliche Auge kein lineares System ist und der doppelte Einfall an Photonen auf der Netzhaut nicht als doppelt so hell empfunden wird. Es handelt sich um eine absolute Größe.

Normierte Helligkeit (lightness): Die normierte Helligkeit ist die absolute Helligkeit, betrachtet in Relation zur absoluten Helligkeit einer weißen Fläche unter den gleichen Beleuchtungsbedingungen. [33] Es handelt sich somit um eine relative Größe. In der Mehrzahl der Literatur wird nicht zwischen absoluter Helligkeit und normierter Helligkeit unterschieden und daher beides in der Folge als Helligkeit bezeichnet. Ist eine genaue Unterscheidung relevant, wird diese in dieser Arbeit auch entsprechend kenntlich gemacht. Das Symbol ist L^* .

Luma (luma): Durch Codierung liegen Helligkeitsinformationen in der Praxis oft als gewichtete Summe der nichtlinearen Farbkanäle (Rot, Grün und Blau) vor. Das bedeutet, dass die Leuchtdichte für die Speicherung oder Übertragung aus Gründen der effektiveren Ausnutzung der Bandbreite über eine Funktion, beispielsweise eine Potenz oder ein Logarithmus, vorentzerrt ist. Hierfür findet sich der Begriff Luma. In diesem Fall wird das Symbol Y' verwendet. Durch die Vorentzerrung spiegeln die so gespeicherten Werte gegebenenfalls sehr nahe die subjektive Gleichabständigkeit wider, dabei haben sie keinerlei Entsprechung mit den Leuchtdichten der realen Szenen, sondern sind so gewählt, dass für eine bestimmte Sehsituation ein möglichst ansprechendes Bild auf dem Display reproduziert werden kann. Deshalb wird auch von *display referred* gesprochen. [99] [98]

Beleuchtungsstärke (lux): Die Beleuchtungsstärke (in lm/m^2 oder auch in lx) gibt an, wie viel Lichtstrom (lm) auf eine Flächeneinheit (m^2) fällt. Es handelt sich somit um eine reine Empfängergröße. In dieser Arbeit ist das Umgebungslicht von Bedeutung. Umgebungslicht, das für eine Aufhellung auf einem Display sorgt, kann sowohl als Empfängergröße (Raumlicht, welches auf das Display fällt) als auch als Sendergröße

(Raumlicht, welches vom Display zum Auge reflektiert wird) gesehen werden. Entsprechend finden sich hierfür einerseits Angaben als Beleuchtungsstärke und andererseits als Leuchtdichte.

1.2.2 Kontrastumfang, Dynamikumfang, Kontrastkompression, Kontrast und *Kontrastparadoxon*

Kontrastumfang (dynamic range): Der Kontrastumfang ist das Verhältnis von maximaler zu minimaler Leuchtdichte in der wahrgenommenen Szene. Dabei kann es sich auch um eine Reproduktion auf dem Display oder auf einem analogen Film handeln.

[45]

Dynamikumfang (dynamic range): Sobald eine Wandlung eines optischen (oder akustischen) Signals in ein elektrisches Signal durchgeführt wurde, spricht man von Dynamik bzw. Dynamikumfang. Aufnahmeseitig, also z. B. bei einem Kamerasensor, ist der Dynamikumfang das Verhältnis von der maximalen zur minimalen (vom Rauschen zu differenzierbaren) Anzahl von Elektronen, die aus den einfallenden Photonen generiert werden können. Während der gesamten elektrischen Verarbeitung spricht man dann nur noch von Dynamikumfang und nicht von Kontrastumfang. Dies gilt sowohl für rein analoge-elektrische als auch für digitale Signale. [45]

Im Englischen wird in diesem Zusammenhang einzig von *dynamic range* gesprochen, was fälschlicherweise zu der verbreiteten Annahme führt, dass 8 Bit einem Kontrastumfang der Szene von 256:1 entspricht. [99] Dabei kann es vorkommen, dass die Szene einen größer Kontrastumfang aufweist als abgebildet werden kann, aber auch, dass nicht alle Bits mit Werten gefüllt sind. Ausschlaggebend ist allerdings die Tatsache, dass Helligkeitsinformationen in den meisten Fällen als Luma vorliegen und somit nichtlinear codiert sind. Je nach Art der Kamera und der Nachbearbeitung ist es auch möglich, mit 8 Bit einen Kontrastumfang der Szene von z. B. mehr als 1.000:1 einzufangen.

Kontrastkompression (dynamic range compression, contrast compression): Stimmt der Kontrastumfang, welcher aus einer Szene eingefangen wurde, nicht mit dem Kontrastumfang den ein Display wiedergeben kann überein, muss eine Kontrastkompression durchgeführt werden, um die Anzeige zu ermöglichen.

Kontrast (contrast): Der Kontrast beschreibt, wie die Reproduktion der Szene auf den Zuschauer wirkt. Besitzt das Bild überwiegend schwarze und weiße Anteile, wird es als kontrastreich wahrgenommen. Überwiegen allerdings die Mitten, wird es kontrastarm – man spricht auch von *flau* – wahrgenommen.

Kontrastparadoxon: Um die Problematik bei der Abgrenzung der Begrifflichkeiten zu illustrieren, sei folgendes *Kontrastparadoxon* angenommen: Steigt der Kontrastumfang der Aufzeichnung, während der Kontrastumfang des Wiedergabedisplay konstant bleibt, wirkt das Bild kontrastärmer, obwohl ein höherer Kontrastumfang der ursprünglichen Szene dargestellt wird. Bei der Darstellung von **HDR**-Inhalten auf **SDR**-Displays stellen somit die Steigerung des Kontrastumfangs und die Erhöhung des Kontrasts zwei gegensätzliche Ziele dar, zwischen denen eine Abwägungsentscheidung getroffen werden muss.

1.2.3 Verhältniszahl, Blendenstufen und Dekaden

Verhältniszahl (ratio): Die einfachste Darstellung des Kontrastumfangs ist, das Verhältnis vom dunkelsten zum hellsten Punkt über eine Verhältniszahl auszudrücken. Dabei wird der dunkelste Wert auf 1 normiert. CRTs konnten üblicherweise Werte von $0,1 \text{ cd/m}^2$ bis 100 cd/m^2 darstellen. Es ergibt sich ein Kontrastumfang von 1000:1. Diese Darstellungsweise findet sich z. B. bei Spezifikationen von Fernsehern.³

Blendenstufen (f-stops) oder Exposure Values: Wird der Kontrastumfang in Blendenstufen (engl. *f-stops*) angegeben, handelt es sich um den Logarithmus Dualis der oben genannten Verhältniszahl. In der Folge besteht ein quadratischer Zusammenhang zwischen Leuchtdichte und Blendenstufen. Für das obige Beispiel ergibt sich $\text{ld}(1000) = 9,97$ Blendenstufen. Diese Darstellung impliziert darüber hinaus, wie viele Bits für die Codierung der Szene mindestens benötigt werden. Allerdings besitzt dieser (technische) lineare Zusammenhang nur selten Gültigkeit, da es sich im Falle einer Vorentzerrung um kein lineares System handelt. In dieser Arbeit wird das Licht oft als Leuchtdichte beschrieben, wodurch die Interpretation wieder eine Legitimität besitzt. Im Bereich der Fotografie wird dagegen eher von Exposure Values (EVs) gesprochen. Dies impliziert, dass eine Verdopplung oder Halbierung der Belichtung nicht nur aufgrund der Blendeneinstellung, sondern auch aufgrund der Belichtungszeit möglich ist. Beide Begriffe werden im Rahmen dieser Arbeit synonym verwendet.

Dekaden (orders of magnitude): Bei sehr hohen Kontrastumfängen ist eine ebenfalls häufig gebräuchliche Form die Darstellung als Zehnerpotenz. Der dekadischen Logarithmus liefert die am einfachsten zu handhabenden Zahlenwerte. So ist $\log(1000) = 3$.⁴ Im Englischen wird dem Wert die Bezeichnung *orders of magnitude* beigefügt.

³Auch liefert die Verhältniszahl das größte und somit werbewirksamste Erscheinungsbild.

⁴1000:1 = 9,97 EVs = 3 orders of magnitude

1.2.4 LDR, SDR und HDR

LDR und **SDR**: Das klassische, im Kontrastumfang recht eingeschränkte Fernsehbild wird als Standard Dynamic Range (**SDR**) oder wahlweise auch Low Dynamic Range (**LDR**) bezeichnet. Es basiert auf den Grenzwerten der **CRT**. Als Richtwerte werden dabei eine maximale Displayleuchtdichte von 100 cd/m^2 , eine gedimmte Umgebungsleuchtdichte (ca. 10 cd/m^2) und oft eine Quantisierung von 8 Bit, auf der eine so genannte Gamma-Vorentzerrung liegt, angenommen.

HDR: Laut der Definition der UHD Alliance⁵ dürfen Displays, die eine maximale Leuchtdichte von 1000 cd/m^2 bei einer minimalen Leuchtdichte von $0,05 \text{ cd/m}^2$ (alternativ 540 cd/m^2 und $0,0005 \text{ cd/m}^2$) erreichen, als **HDR** bezeichnet werden. [126] Für Kameras gibt es einen solchen Grenzwert nicht. Inoffiziell wird aber von 14 EVs gesprochen. In Teilen der Literatur wird der Begriff **HDR** allerdings anders verstanden. **HDR** bezeichnet dabei ein Rohbild, das die realen Leuchtdichtewerte der Szene gespeichert hat, ohne dass diese durch eine Blendeneinstellung auf ein darstellbares Maß begrenzt worden wären. Diese Arbeit folgt der ersten Definition.

1.2.5 Buntheit, Chroma, Sättigung und Farbton

Die bis zu diesem Zeitpunkt definierten Begriffe beschränken sich einzig auf eine Darstellung der Helligkeitskomponente. Für die Beschreibung eines visuellen Reizes muss allerdings die Farberscheinung mitberücksichtigt werden. Dass Lichtmenge und farbiges Erscheinen nur schwer isoliert voneinander betrachtet werden können, wird an einigen späteren Stellen noch zu Problemen führen.

Buntheit (colorfulness): Die Buntheit gibt an, wie farbig eine Oberfläche wahrgenommen wird. [33]

Chroma (chroma): Das Chroma ist die Buntheit einer Oberfläche betrachtet in Relation zur absoluten Helligkeit einer weißen Fläche unter den gleichen Beleuchtungsbedingungen. [33] Das Symbol ist C .

Sättigung (saturation): Die Sättigung ist die Buntheit einer Oberfläche betrachtet in Relation zu ihrer absoluten Helligkeit. [33] Das Symbol ist S .

Farbton (hue): Der Farbton gibt an, inwieweit eine Oberfläche einer der Farben – rot, gelb, grün und blau – oder einer Kombination aus benachbarten Paaren dieser Farben entspricht. Dazu werden die Farben in einem Kreis angeordnet. [33] Das Symbol ist h .

⁵Die UHD Alliance ist eine Vereinigung von Displayherstellern - darunter Samsung, LG, Philips, Toshiba, Hisense oder Vestel.

1.2.6 Farbeindruck

Nayatani et al. [89] konnten zeigen, dass fünf Erscheinungsattribute für eine komplette Beschreibung des Farbeindrucks benötigt werden, die da wären: Absolute Helligkeit, normierte Helligkeit, Buntheit, Farbton und Sättigung bzw. Chroma. Sättigung und Chroma verhalten sich dabei, wie oben gezeigt, redundant zueinander. Chroma ergibt sich aus der Multiplikation von normierter Helligkeit und Sättigung. Wird nun eine natürliche Szene auf einem Display reproduziert, ist es selten möglich, absolute Helligkeit und Buntheit zu erhalten, da die maximal zur Verfügung stehende Leuchtdichte geringer ist. In diesem Fall kann es nur das Ziel sein, normierte Helligkeit, Farbton und Sättigung bzw. Chroma zu bewahren. Kommt es nun darüber hinaus noch zu einer Kontrastkompression, um die Abbildung eines größeren Kontrastumfangs zu ermöglichen, wird die Helligkeitskomponente und somit die normierte Helligkeit bewusst verändert. Das bedeutet auch, dass es nicht möglich ist, sowohl Sättigung als auch Chroma in diesem Fall konstant zu halten. Eines der Attribute oder auch beide müssen ihren Wert ändern, damit die Gleichung aus normierter Helligkeit, Sättigung und Chroma wieder aufgeht.

Kapitel 2

Menschliche Wahrnehmung und Stand der Fernsehtechnik

2.1 Motivation

Die Betrachtung einer natürlichen Szene oder der Blick durch ein Fenster lösen andere Sinneseindrücke aus, als es bei der Reproduktion auf einem Fernsehgerät der Fall ist. Letztere bleibt hinter den Limitationen der menschlichen Wahrnehmung zurück. Das Ziel neuer technologischer Entwicklungen auf diesem Gebiet ist es, die Lücke fortwährend zu schließen und die Grenzen zwischen Realität und Reproduktion immer weiter verschwimmen zu lassen.

Die vielleicht größte Diskrepanz besteht bei der Kontrastwahrnehmung. Bis vor kurzem war es nur möglich einen Bruchteil aller vom Auge unterscheidbaren Leuchtdichtewerte aufzuzeichnen und wiederzugeben. Dieser beschränkte Kontrastumfang wird auch als Standard Dynamic Range (**SDR**) bezeichnet. Er wird immer dann offensichtlich, wenn eine Szene sowohl aus hellen, als auch aus dunklen Bildanteilen besteht. Klassische Beispiele sind hierbei ein Fußballspiel, bei dem die Tribüne einen Schatten auf das Spielfeld wirft, oder eine Szene, die zwar in einem Innenraum spielt, durch Fenster aber der Blick nach draußen gegeben ist. In diesen Fällen liegen die Leuchtdichtewerte der unterschiedlich hellen Bildanteile so weit auseinander, dass sie nicht mehr im **SDR**-Bereich abgebildet werden können. Optisch äußert sich dies durch eine fehlerhafte Belichtung von mindestens einem der beiden Teile der Szenerie. Sprich, wird die Belichtung so gewählt, dass der Rasen im Schatten gut zu sehen ist, brennen die Sonnenbereiche aus und werden viel zu hell bis komplett weiß dargestellt. Liegt der Fokus hingegen auf den hellen Bereichen, kommt es zu einer Unterbelichtung der Schatten. In der Folge sind diese fast oder ganz schwarz. Die nachstehende Abbildung illustriert diesen Zusammenhang.



ABBILDUNG 2.1: Die gleiche Szene mit zwei verschiedenen Blendeneinstellungen aufgenommen. Links kommt es zu einer Überbelichtung und rechts zu einer Unterbelichtung.

Um Abhilfe zu schaffen wird deshalb seit einigen Jahren daran gearbeitet, den verarbeitbaren Kontrastumfang deutlich zu steigern, welches als High Dynamic Range (**HDR**) bezeichnet wird. Durch enorme Fortschritte auf dem Gebiet der Kamera- und Displaytechnologie ist es heute möglich, die Limitationen der menschlichen Wahrnehmung auszureizen, wie im nun Folgenden genauer beleuchtet wird. Die sich hieraus ergebenden Herausforderungen bilden die Motivation für diese Arbeit.

2.2 Menschliche Wahrnehmung

Um die Fortschritte auf dem Gebiet der Bildaufnahme- und Wiedergabesysteme einordnen zu können, ist es nötig den Prozess des Sehens genauer zu betrachten, der durch das Zusammenspiel verschiedener Komponenten und nicht einzig vom Auge allein, zustande kommt. Der englische Begriff Human Visual System (**HVS**) trägt dem Rechnung.

2.2.1 Spatiale Auflösung

Auch wenn im hier vorliegenden Fall der wahrnehmbare Kontrastumfang im Fokus steht, soll zunächst kurz auf das spatiale Auflösungsvermögen eingegangen werden. Mit dessen Hilfe kann errechnet werden, wie viele Bildzeilen, und somit auch Pixel, bei welchem Betrachtungsabstand gerade noch unterschieden werden können. Es ist allgemein anerkannt, dass die Grenze des spatialen Auflösungsvermögens bei etwa einer Winkelminute liegt. [109] Auf dieser Grundlage lässt sich der maximale Betrachtungsabstand a in Vielfachen der Displayhöhe H ermitteln, indem der Betrachtungswinkel α mithilfe der Winkelminute δ sowie der Anzahl der Bildzeilen n_z ausgedrückt wird.

$$n_z \tan\left(\frac{\delta}{2}\right) = \tan\left(\frac{\alpha}{2}\right) \quad (2.1)$$

$$a = \frac{H}{2 \tan\left(\frac{\alpha}{2}\right)} = \frac{H}{0,000291 n_z} \quad (2.2)$$

Hieraus ergibt sich die Faustformel, dass bei 1080 Zeilen der Betrachter nicht weiter als dreimal die Displayhöhe vom Fernsehgerät entfernt sitzen sollte. Bei 2160 Zeilen, also 4k, ist es folglich nur noch das 1,5-fache. Für ein 50"-Display liegt der ideale Betrachtungsabstand dann bei etwa einem Meter. Wird hingegen der heute übliche Betrachtungsabstand von etwa 2,7 m als gegeben angenommen [90], ergibt sich eine Displaygröße von 140 Zoll.

Nicht ganz außer Acht sollte in diesem Zusammenhang der Effekt der *Hyperacuity* gelassen werden. Durch die Kombination von Optik und Neurobiologie ist der Mensch in der Lage, kleine örtliche Versetzungen von gut sichtbaren Bildebenen zu erkennen, auch wenn diese kleiner sind, als der Durchmesser der Fotorezeptoren es eigentlich zulassen würde. [85] Hierdurch lässt sich erklären, wieso es auch bei kleineren Displays zu einem unterschiedlichen Schärfeeindruck zwischen HD und 4k kommen kann.

Trotzdem zeigen die Zahlen warum die Erhöhung der spatialen Auflösung bei UHDTV nur bedingt zur Steigerung der Bildqualität beitragen kann und andere Parameter ein größeres Potenzial versprechen.

2.2.2 Kontrastumfang

Im Vergleich zur spatialen Auflösung ist die menschliche Kontrastwahrnehmung der Wiedergabe des HD-Fernsehbildes weit überlegen. In der Literatur finden sich hierzu leicht unterschiedliche Angaben [62] [104] [38] [48], die aber folgendermaßen zusammengefasst werden können:

Die Netzhaut des menschlichen Auges verfügt über zwei Typen von Fotorezeptoren, die Stäbchen und die Zapfen. Beim so genannten photopischen Sehen, dem Tagsehen, sind alleine die Zapfen aktiv. Da es von ihnen drei verschiedene Ausprägungen gibt, die jeweils unterschiedlich auf bestimmte Wellenlängen des Lichts reagieren, ermöglichen sie uns die Wahrnehmung von Farbeindrücken. Sie sind etwa in einem Bereich von 0,01 bis 10^8 cd/m^2 aktiv. Die Stäbchen hingegen sind besonders lichtempfindlich und ermöglichen das Sehen in dunkler Umgebung. Ihr Aktivitätsbereich erstreckt sich in etwa von 10^{-6} bis 10 cd/m^2 . Beim so genannten skotopischen Sehen, bei dem nur die Stäbchen aktiv sind, findet lediglich eine Differenzierung bezüglich der Leuchtdichte statt. Farben können nicht mehr wahrgenommen werden.

Der Bereich, in dem sich die Aktivität der beiden Fotorezeptortypen überschneidet, wird mesopisches Sehen genannt. Anhand von Experimenten hat Hunt [52] versucht, diesen

genauer zu bestimmen. Nach seinen Angaben beginnt bei $3,4 \text{ cd/m}^2$ die Sensitivität der Zapfen nachzulassen, bis sie schließlich bei $0,03 \text{ cd/m}^2$ nicht mehr relevant ist.

2.2.3 Simultaner Kontrastumfang

Der gesamte Kontrastumfang der menschlichen Wahrnehmung erstreckt sich, wie gerade gezeigt, etwa über 14 Dekaden oder 43 EVs. Doch dies ist nur durch mechanische, fotochemische und neuronale Adaption möglich. [37] Bis sich das Auge vollkommen auf eine Umgebungsleuchtdichte eingestellt hat, können bis zu 30 Minuten vergehen. Die Teilmenge des gesamten Kontrastumfangs, die zu einem Zeitpunkt verfügbar ist, wird simultaner oder stationärer Kontrastumfang genannt. Die Literaturlage hierzu ist allerdings dünn. Die aussagekräftigste und meistzitierte Studie stammt von Kunkel et al. aus dem Jahr 2010 [64]. Sie kommt zu dem Ergebnis, dass simultan mindestens 3,7 Dekaden oder 13 EVs wahrgenommen werden können. Es besteht dabei eine Abhängigkeit vom Kontrast der Stimuli, von der Dauer, die sie gezeigt werden und vom Leuchtdichtewert des Hintergrunds, auf dem sie gezeigt werden. Die Studie unterlag der Einschränkung, dass das zur Verfügung stehende Display, ein Dolby-Prototyp, einen Kontrastumfang von 4,7 Dekaden aufwies.

Eine neuere Untersuchung [17] deutet allerdings darauf hin, dass der simultane Kontrastumfang bei realen Vorlagen aufgrund von Maskierungseffekten deutlich geringer ausfallen kann. So überstrahlen sehr helle Bildelemente geringe Abstufungen in den angrenzenden dunklen Bereichen, die in der Folge nicht mehr wahrgenommen werden können. Da das Sichtfeld beim Blick auf ein Display in der Regel eingeschränkter ist als das bei der Betrachtung einer realen Szene, werden die einzelnen Objekte in geringerer räumlicher Nähe wiedergegeben, wodurch der Effekt eine hohe Relevanz im Bezug auf HDR hat. Allerdings liegen derzeit keine verlässlichen Zahlen vor. Es wäre denkbar, dass Abstufungen in der Helligkeit zwar auf dem Display reproduziert werden können, diese für die Zuschauer aber nicht sichtbar sind oder nur unter Anstrengung differenziert werden können. Unter Umständen empfinden die Zuschauer sich bei solch inhomogen ausgeleuchteten Szenen auch unangenehm geblendet. Entsprechend soll der Effekt zu Beginn von Kapitel 4 genauer untersucht werden.

2.2.4 Fotorezeptorreaktionskurve und Adaption

Durch den Adaptionsprozess wird die Lage der Übertragungsfunktion der Fotorezeptoren beeinflusst. Dabei bleibt die Form der Kurve erhalten. Es findet lediglich eine Verschiebung entlang der Abszissenachse statt, auf der die Leuchtdichtewerte der Szene aufgetragen sind. Die Übertragungsfunktion kann durch die Naka-Rushton-Gleichung

[88] beschrieben werden. Allgemeinhin wird sie in der folgenden Schreibweise angegeben: [104]

$$\frac{R}{R_{max}} = \frac{L^n}{L^n + K^n} \quad (2.3)$$

Dabei ist R die Reaktion auf einen Leuchtdichtewert L und R_{max} die asymptotisch maximale Reaktionsamplitude. K gibt an, an welchem Punkt $R_{max}/2$ erreicht sind. In der Folge wird über K die Verschiebung auf der Abszissenachse und somit der Adaptionstand der Fotorezeptoren wiedergegeben. Werden die Leuchtdichtewerte logarithmisch aufgetragen, ergibt sich eine S-Kurve, wie in Abbildung 2.2 gezeigt. Über n , auch Sensitivitäts-Kontroll-Parameter genannt, kann die Steilheit der Kurve beeinflusst werden. Laut Boynton und Whitten [15] liegt der Wert zwischen 0,7 und 0,73. Dieser Wert wurde anhand einer Affenfovea¹ ermittelt.

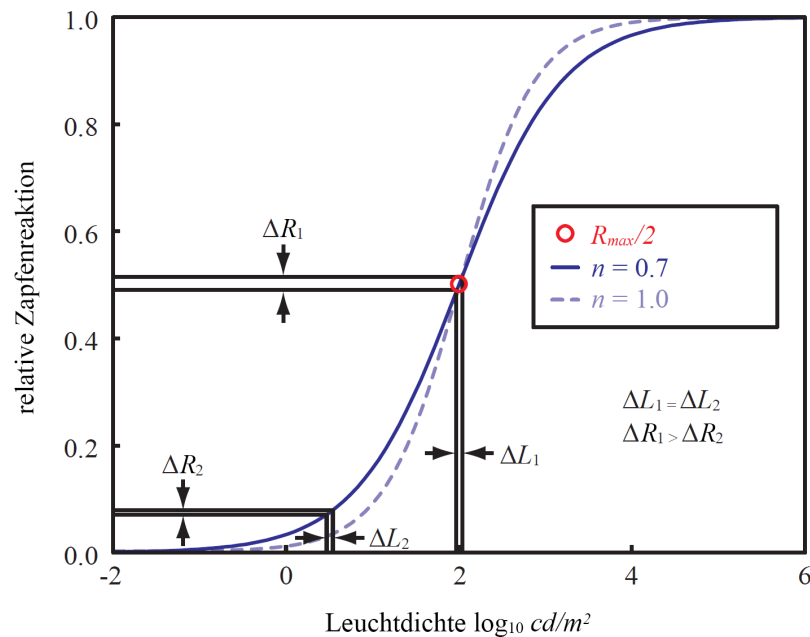


ABBILDUNG 2.2: Naka-Rushton-Gleichung mit Variation des Sensitivitäts-Kontroll-Parameters nach [64]

2.2.5 t.v.i., CSF und JND

Betrachtet man die Naka-Rushton-Gleichung unter verschiedenen Adaptionstufen wird deutlich, dass die benötigte absolute Leuchtdichtedifferenz für die Erzeugung eines gerade noch wahrnehmbaren Unterschieds bei steigenden Leuchtdichten zunimmt. Dies

¹Als Fovea centralis wird der etwa 1,5 mm große Bereich der Netzhaut bezeichnet, der das Zentrum des schärfsten Sehens bildet.

kann über eine so genannte **t.v.i.**-Funktion (threshold versus intensity function) abgebildet werden. Die Stufen werden als Just Noticeable Differences (**JND**s) bezeichnet. Um sie zu messen, wird die Leuchtdichte eines Testspots (Stimulus) sowie die des Hintergrunds variiert. Das Verhältnis von Hintergrund zu Stimulus kann als Quotient angegeben werden. Dieser Quotient wird als der Weber-Fechner-Kontrast bezeichnet. **[104]** Neuere Modelle nutzen anstelle von einfarbigen Stimuli Sinus-Muster oder Gabor-Raster. Hieraus ergeben sich Contrast Sensitivity Functions (**CSF**s). **[104]**

Das Wissen über diese Aspekte wird sowohl bei der Kontrastkompression von so genannten Tonemapping-Operatoren (TMOs) genutzt, als auch von Encodern für eine effiziente Codierung. Denn eine feinere Quantisierung als ein **JND** bringt keinen subjektiven Qualitätsgewinn mit sich. Entsprechend ist es sinnvoll, bei einer in den meisten Fällen auf 10 Bit begrenzten Dynamik die Leuchtdichtewerte visuell bestmöglich auf die 1024 Codewörter zu verteilen. Sowohl der Vorschlag von Dolby **[114]**, als auch der Vorschlag von Philips **[116]** bezüglich solch einer Vorentzerrung, auch opto-elektrische Transferfunktion (**OETF**) genannt, zur Übertragung von **HDR**-Signalen, basiert auf der Barten-Kurve **[5]**. Diese **CSF** ist allgemein anerkannt und scheint gegen einen Diskurs immun. Dabei gibt es einige Alternativen, z. B. **[38]** **[10]** oder **[86]**, wie in Abbildung **2.3** ersichtlich wird. Mantiuk ist sogar der Auffassung, dass die Barten-Kurve für Leuchtdichtewerte unter 1000 cd/m^2 eine zu grobe und für Werte über diesem Punkt eine zu konservative Quantisierung durchführt. **[84]**

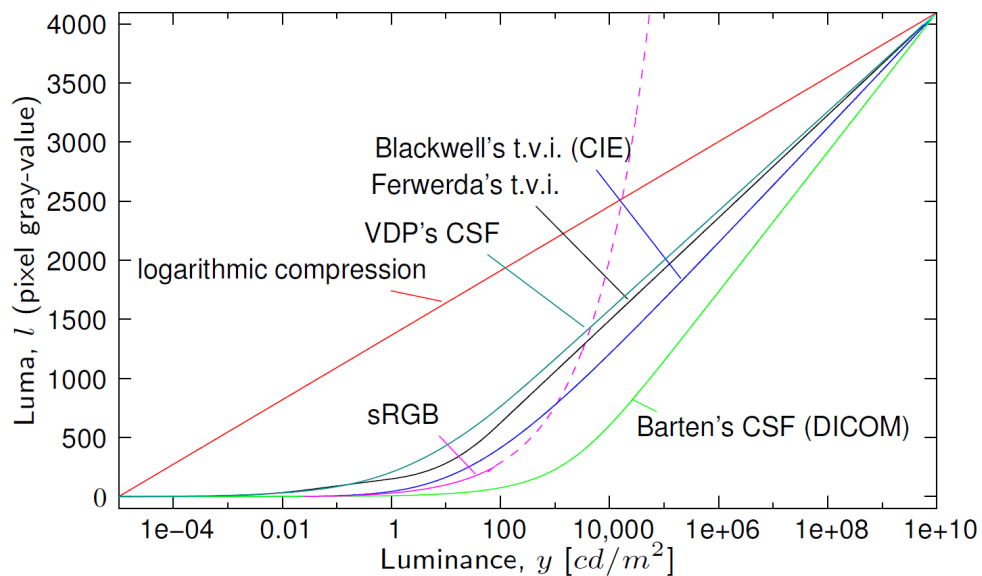


ABBILDUNG 2.3: Unterschiedliche **CSF**- bzw. Threshold Versus Intensity Function (**t.v.i.**)-Funktionen im Vergleich **[84]**

Vom Gesamtverlauf unterscheiden sich die Kurven allerdings nicht. Alle beginnen sehr

flach und sorgen so für prozentual große Sprünge im dunklen Bereich. Die absolute Differenz zwischen zwei Stufen ist an dieser Stelle trotzdem sehr klein. In einem Bereich zwischen 10 und 1.000 cd/m^2 werden die Kurven deutlich steiler und münden in eine Gerade. Sie spiegelt den logarithmischen Zusammenhang zwischen der Anzahl der Photonen und dem absoluten Helligkeitsempfinden wider.

Allgemein ist der Wert 1%, welcher vom Weber-Fechner-Gesetz [128] vorgegeben wird, als Grenzwert anerkannt. Anhand von Abbildung 2.3 ist aber zu erkennen, dass dieser streng genommen nur für den Bereich des Tagsehens eine ausreichende Genauigkeit besitzt. Zur Ausbildung von sichtbaren Treppenstrukturen, auch *Banding* genannt, kommt es laut einer BBC-Untersuchung [14] allerdings erst ab etwa 3%.

2.3 Technischer Fortschritt bei Bildaufnahme- und -wiedergabesystemen

2.3.1 Entwicklung im Bereich der Kameras

Um hochkontrastreiche Foto- und Videoinhalte zu erzeugen, wird traditionell der begrenzte Kontrastumfang der Kameras dadurch erweitert, dass mehrere Aufnahmen mit unterschiedlicher Belichtung gemacht wurden. Die Bilder dieser Belichtungsreihen wurden nachfolgend in der Kamera selber oder in der Postproduktion zusammengeführt. Eine verfeinerte und echtzeitfähige Variante ist z. B. der Algorithmus von Guthier [43]. Einen etwas anderen Weg geht Fröhlich [41]. Im Vergleich zur klassischen Belichtungsreihe, bei der die Aufnahmen zeitlich versetzt und mit einer anderen Blende bzw. Belichtungszeit aufgezeichnet werden, sorgt dort ein halbdurchlässiger Spiegel am Kamera-Rig für unterschiedlich belichtete Bilder trotz identischer Kameraeinstellungen. Ghosting aufgrund der zeitlichen Differenz der Bilder sowie unterschiedlicher Tiefenschärfe bzw. unterschiedlicher Bewegungsunschärfe sind somit nicht mehr zu fürchten. Lediglich die Größe der Apparatur und das nachträgliche Zusammenführen in der Postproduktion beschränken die Einsatzgebiete.

Dabei sind heutige Kameras, auch ohne die genannten Techniken, bereits in der Lage, einen deutlich größeren Kontrastumfang einzufangen, als aktuell genutzt wird, und werden solche Herangehensweisen über kurz oder lang obsolet machen. Mit dem Einzug der Complementary Metal Oxide Semiconductor (CMOS)-Technik in den Videobereich ergaben sich neue Möglichkeiten. Bereits 1963 von Frank Wanlass erfunden [129], sollte es bis ins neue Jahrtausend dauern, bis der Siegeszug des aktiven Pixelsensors (APS), welcher auf der CMOS-Technologie basiert, beginnen konnte. Ein bekannter Vertreter war beispielsweise ARRI's Videokamera D-20. Bis dato war die Lichtausbeute des CMOS

gegenüber dem Charge Coupled Device (**CCD**) durch die zusätzlichen Bauteile nicht konkurrenzfähig gewesen. Immer rauschärmere Elemente und der Einsatz von speziellen Mikrolinsen auf den lichtempfindlichen Flächen, gepaart mit der nichtlinearen Bewertung der Signalwerte (quasi logarithmisch), kompensierten die Nachteile zusehends. Der größte Zugewinn bei den **CMOS**-Sensoren kommt aber von den immer kleiner werdenden lithographischen Prozessen. Mit jeder neuen Fertigungstechnologie und kleiner werdenden Strukturen steigt die Dynamik, weil die nutzbare Fläche immer größer wird. So müssen bei der rückwärtigen Belichtung keine Zwischenräume mehr für den Ladungstransport auf der Vorderseite verwendet werden.

Das aktuelle Modell aus dem Hause ARRI, die Alexa, besitzt inzwischen einen Kontrastumfang von fast 15 EVs. Konkurrenzprodukte wie die RED Epic oder die Sony Venice liegen auf einem ähnlichen Niveau. Somit ist der simultane Kontrastumfang des menschlichen Auges erreicht, wenn nicht sogar schon überschritten.

Die Firma GrassValley nutzt bei dem LDX-Series-Camera-Systems einen weiteren Vorteil des **CMOS** gegenüber dem **CCD** besonders intensiv aus - das *Non-Destructive Readout*. [130] Die aktuelle Ladungsmenge der einzelnen Pixel kann während der Belichtung ausgelesen werden, ohne dass es zu einer Entladung des Pixels kommt. So können immer wieder Zwischenwerte abgefragt werden. Erreicht das Pixel vor dem Ende der Belichtungszeit seine maximale Ladungsmenge, kann aus den Zwischenwerten ein theoretischer Endwert interpoliert werden. Mit diesem Verfahren kann die Empfindlichkeit zukünftiger Kameragenerationen um ein Vielfaches gesteigert werden. Aktuell ist so ein Kontrastumfang von 22 EVs möglich. [12] Dieser überschreitet bei Weitem die Möglichkeiten des Auges und der Consumer-Displays, wie im nächsten Abschnitt gezeigt werden wird.

2.3.2 Entwicklung im Bereich der Displays

Für die Entwicklung der Displays waren die von Seetzen et al. 2004 vorgestellten Ergebnisse wegweisend. [112] Es wurde versucht, die maximale Leuchtdichte zu steigern, ohne dabei den Schwarzwert anzuheben. Nur so kann der Kontrastumfang deutlich erhöht werden. Die Idee war es, zwei Displaytechnologien miteinander zu kombinieren - ein hochauflösendes Liquid Crystal Display (**LCD**) vorne und ein schwachauflösendes Display dahinter. In der Theorie sollten sich so beide Kontrastumfänge multiplizieren. In der Praxis verhindert aber die hohe Absorptionsrate des **LCD** mit gerade einmal einer Transparenz von 3% bis 8% solche Ergebnisse.

Als schwachauflösendes Display, sprich als Hintergrundbeleuchtung, wurden zwei verschiedene Varianten getestet. Im ersten Fall kam ein Digital Light Projector (**DLP**) zum

Einsatz, der aufgrund der kleinen Fläche, die er bespielen musste, und des ausgebauten Farbrades eine hohe Lichtausbeute erzielen konnte. Auf diesem Weg war es möglich, eine minimale Leuchtdichte von $0,05 \text{ cd/m}^2$ und eine maximale von 2.700 cd/m^2 zu erreichen. Das entspricht einem Kontrastumfang von etwa 16 EVs. Den Prototypen, den Dolby für seine Studie [87] bezüglich der vom Zuschauer präferierten Leuchtdichte von Displays eingesetzt hat, basiert auf dem gleichen Vorgehen. Hier liegen die Werte bei $0,004 \text{ cd/m}^2$ bis 20.000 cd/m^2 .

Allerdings bringt diese Konstruktion zahlreiche Nachteile mit sich, die das Einsatzfeld auf die Forschung begrenzen. Hier wären die optische Länge aufgrund des Projektors, die Leistungsaufnahme, die Kosten, die thermischen Probleme sowie die geringe Displaygröße von rund 20 Zoll zu nennen.

Aus diesen Gründen hat sich der zweite Fall durchgesetzt, bei dem Leuchtdioden (LEDs) zum Einsatz kommen, die einzeln angesprochen werden können. Durch diese partielle Ansteuerung konnte im Versuch ein Schwarzwert von weniger als $0,03 \text{ cd/m}^2$ und eine maximale Leuchtdichte von 8.500 cd/m^2 erreicht werden. Dabei waren die im vorherigen Abschnitt erwähnten Probleme nicht in dieser Form vorhanden. Zwar hatte das erste weit verbreitete professionelle **HDR**-Display, der Dolby PRM-4220, das auf dieser Technologie beruht, die Bezeichnung „Flachbildschirm“ nicht mehr wirklich verdient und die Wärmeentwicklung sowie die Stromaufnahme sind ziemlich hoch, dennoch war und ist es praxistauglich. Es erreicht eine maximale Leuchtdichte von 600 cd/m^2 . Sein großer Bruder, der Dolby Pulsar, kommt auf 4.000 cd/m^2 . Dieser benötigt allerdings eine Wasserkühlung.

In der neuesten Generation von Referenzmonitoren wird das partielle Abdunkeln nicht mehr über eine vierstellige Anzahl an LEDs erreicht, sondern über ein weiteres **LCD**-Panel (z. B. Sony BVM-HX310 oder EIZO CG3145) realisiert. Hierdurch kann die angegebene Spitzenleuchtdichte auf der gesamten Fläche erreicht und Artefakte aufgrund der Unterabtastung verhindert werden. Werte oberhalb von 1.000 cd/m^2 sind bei einem Schwarzwert unterhalb von $0,05 \text{ cd/m}^2$ möglich.

Darüber hinaus werden zwei weitere Technologien genutzt, um hochkontrastreiche Bilder zu reproduzieren. Bereits von kleineren Displays und professionellen Referenzmonitoren sind organische Leuchtdioden (OLEDs) bekannt. Da es sich um ein selbstleuchtendes Display handelt, ergibt sich bauartbedingt ein nahezu perfekter Schwarzwert und, daraus resultierend, ein quasi unendlicher Kontrastumfang. Allerdings war dieser lange Zeit in der maximalen Leuchtdichte mit Werten zwischen 200 und 300 cd/m^2 eingeschränkt. 2015 präsentierte Sony mit dem BVM-X300 erstmals ein Modell, das bis zu 1.000 cd/m^2 (auf einer begrenzten Fläche) erreicht. Neben dem Schwarzwert haben OLEDs den Vorteil, dass der verwendete organische Film Licht in einem sehr schmalen Wellenlängenbereich

emittiert. In der Folge können sehr reine Grundfarben erzeugt werden, was die Darstellung eines sehr großen Farbraumes ermöglicht.

Für den Consumer-Bereich geht ein Trend zu Quantum Dots. Dabei handelt es sich um Nano-Kristalle aus Halbleiter-Materialien, die je nach Größe unterschiedliche Farben emittieren, wenn sie durch blaues Licht dazu angeregt werden. Das Emittieren geschieht sogar in einem noch schmalbandigeren Bereich als es bei OLEDs der Fall ist. In Kombination mit einer **LCD**-Hintergrundbeleuchtung lassen sich so hoher Kontrastumfang, großer Farbraum und geringe Energieaufnahme in Einklang bringen.

Obwohl in diesem Abschnitt gezeigt wurde, dass Displays in der Lage sind, weit mehr als 1.000 cd/m^2 zu erreichen, sprechen zwei Gründe dafür, dass sich im Consumer-Bereich Werte zwischen 600 und 1000 cd/m^2 durchsetzen werden. Einerseits haben subjektive Untersuchungen gezeigt, dass einige Zuschauer Werte über 1.000 cd/m^2 als unangenehm empfinden, wie an späterer Stelle noch genauer erläutert wird.² Andererseits kommen ganz praktische Gründe hinzu. Die Europäische Union (**EU**) hat sich auf die Fahne geschrieben, den Stromverbrauch in den privaten Haushalten zu begrenzen; angefangen von dem Verbot von Glühbirnen bis zu Staubsaugern über einer gewissen kW-Zahl. Auch Displays werden hiervon nicht unbetroffen bleiben. Die beschränkte Stromaufnahme schlägt sich auch in der ausgegebenen Leuchtdichte nieder.

2.3.3 Entwicklung im Bereich der Fernschnormen

Aus den vorangegangenen Abschnitten wird deutlich, dass sowohl die Akquisition als auch die Wiedergabe von Videomaterial deutlich heterogener geworden ist. Hinzu kommen die bis hierhin noch unerwähnten Möglichkeiten von Projektion und mobilen Endgeräten. Jedoch spiegelt sich diese Vielfalt nicht in den derzeit fast ausschließlich gebräuchlichen **SDR**-Fernschnormen BT.709 und BT.1886 [54] [55] wider. Sie basieren auf eine an die Limitationen der **CRT** angepasste und abgestimmte Verarbeitungskette. So wird beispielsweise davon ausgegangen, dass bei der Wiedergabe Leuchtdichtewerte in einem Bereich von $0,1 \text{ cd/m}^2$ bis 100 cd/m^2 dargestellt werden können.

Deshalb führten die technologischen Entwicklungen zu zwei neueren Fernschnormen ITU-R BT.2020 [57] und ITU-R BT.2100 [58], die es ermöglichen, einen weitaus größeren Bereich an Leuchtdichtewerten und Farbinformationen aufzuzeichnen und zu übertragen. Bestandteil dieser Normen sind unter anderem zwei neue Arten der Signalvorentzerrung. Diese zwei **OETFs** dienen dazu, im weiterhin auf 10 Bit limitierten Produktionsumfeld

²Die im Webinar [87] erwähnte Studie von Dolby kommt zwar zu einem anderen Ergebnis, sollte aber vor dem Hintergrund betrachtet werden, dass beim Fernsehen die Betrachtung über einen langen Zeitraum erfolgt, welches zu Ermüdungserscheinungen führen kann.

derart große Kontrastumfänge in einem digitalen Signal abzubilden. Um die korrekten linearen Leuchtdichten zurückzugewinnen, ist es nötig die inverse Funktion zu kennen. Diese wird elektro-optische Transferfunktion (EOTF) genannt. Handelt es sich bei OETF und EOTF innerhalb eines Systems nicht um jeweils Inverse, kommt es zu einer Abweichung in der wiedergebenden Leuchtdichte. Diese kann durch eine opto-optische Transferfunktion (OOTF) ungleich 1 beschrieben werden und ist unter Umständen aus visuellen Gründen gewollt. Beim SDR-Fernsehen wird beispielsweise eine OOTF anvisiert, die als Potenzfunktion mit dem Wert 1,2 dargestellt werden kann. In diesem Zusammenhang wird auch von *Over-All-Gamma* gesprochen.

Damit die für HDR gedachten OETFs in der Lage sind, eben jenen großen Kontrastumfang abzubilden, muss sich entsprechend auch deren Kurvenverlauf deutlich von dem der Gamma-Vorentzerrung nach BT.709 [54] unterscheiden. Somit besteht keine Abwärtskompatibilität zwischen dem alten und dem neuen System oder sie ist zumindest stark eingeschränkt. Hintergrund ist, dass ältere Fernsehgeräte nur die EOTF nach BT.1886 [55] kennen und in der Folge auch die HDR-Signale damit entzerren. Das resultierende Bild ist, genauso wie der Kurvenverlauf der OETFs, sehr flach. Erschwerend kommt hinzu, dass auch die Definition der drei Grundfarben, aus welchen der Farbeindruck gemischt wird, nicht identisch ist. Auch gibt es weitere OETFs, wie z. B. Slog3 von Sony [119], die rein für die Akquisition gedacht sind.

2.4 Zielsetzung der Arbeit

Damit ist der Broadcast erstmals mit einem Problem konfrontiert, dass das Format bei der Akquise nicht zwangsläufig dem der Distribution entsprechen muss. Konnte man zu Zeiten der CRT noch davon ausgehen, dass die Bedingungen bei der Abnahme im Studio und bei der Wiedergabe an den Heimgeräten bezüglich der Leuchtdichte und der Farben ähnlich limitiert waren, kann dies in Zukunft nicht mehr garantiert werden. Es ist sogar erstmals denkbar, dass Signale erzeugt werden können, welche die Wahrnehmbarkeitsschwelle des Menschen übersteigen.

Auch wenn diese Herausforderungen für den Broadcast neu sind, so hat es sie bereits auf anderen Gebieten gegeben. So ist es beispielsweise im Audiobereich schon längere Zeit möglich mit einer Dynamik zu arbeiten, die dem menschlichen Hörvermögen überlegen ist. Dem vorliegenden Fall aber ähnlicher ist die Problematik des analogen Films. Auch hier war es möglich bei der Akquise mehr Kontrastumfang einzufangen, als auf den meisten Wiedergabemedien gebannt hätte werden können. Ähnlich verhält es sich auch bei der Computergrafik, bei der jede beliebige Leuchtdichte künstlich errechnet werden kann.

In der nun folgenden Arbeit gilt es, dieses Wissen vor dem Hintergrund der besonderen Anforderungen eines Broadcast-Systems näher zu beleuchten, neu zu kombinieren und zu erweitern. So muss beispielsweise immer auf eine mögliche Echtzeitfähigkeit geprüft werden. Auch ist es beim Broadcast nur möglich, ein einziges Signal pro Fernsehkanal zu übertragen. In Anbetracht der verschiedenen Zielempfänger stellt sich die Frage, wie es gelingen kann, auf allen ein ansprechendes Bild zu produzieren. Des Weiteren sollte berücksichtigt werden, welche Randbedingungen die Einführung von **HDR** mit sich bringt. Bis auf wenige Kanäle wird derzeit kein **HDR**-Programm ausgestrahlt. Hierfür gibt es zwei Hauptgründe: Zum einen ist die Reichweite sehr eingeschränkt, da nur ein sehr geringer Teil der Zuschauer über ein entsprechendes Empfangsgerät verfügt. Zum anderen wären massive Investitionen in die gesamte Distributionskette nötig. Selbst wenn eine entsprechende Kamera bereits vorhanden wäre, bedürfte es noch entsprechender Displays, Kreuzschienen, Encoder und weiterer Geräte.

Der vorliegenden Arbeit liegt deshalb die Frage zu Grunde, wie es unter diesen Voraussetzungen gelingen kann, **HDR** im Broadcast flächendeckend einzuführen und somit von den neuen technologischen Entwicklungen Gebrauch zu machen, ohne die aktuellen Empfänger zu vernachlässigen - sprich auch diese an einer **HDR**-Aufzeichnungsqualität partizipieren zu lassen. Es wird ein Einführungsszenario gesucht, das aber auch noch nach der Übergangsphase Gültigkeit besitzt.

Dazu muss, wie bereits angedeutet, Wissen aus mindestens drei Fachgebieten neuartig kombiniert bzw. erweitert werden. Zunächst sei ein Blick in die Fernsehtechnik geworfen, um mögliche Workflows [22] [14] und deren Schwierigkeiten zu analysieren. Es wird sich zeigen, dass die Kontrastkompression dabei den essentiellen Bestandteil bildet. Eine manuelle Generierung wie beim Film ist aber aus zeitlichen Gründen ausgeschlossen. Auch wäre der kreative Einfluss des jeweiligen Coloristen unter Umständen zu groß.

Als Mittel zur Lösung bieten sich die aus der Computergrafik stammenden Tonemapping-Operatoren (TMOs) an. In der Folge sollen die Möglichkeiten dieses Feldes näher beleuchtet werden. Allerdings wird deutlich, dass die derzeit verfügbaren Operatoren aus unterschiedlichen Gründen nicht für den Einsatz beim Broadcast geeignet sind [31] [96]. Ein Kernproblem neben physikalischen Wahrnehmungsaspekten ist dabei die abweichende durchschnittliche Präferenz des Zuschauers bezüglich Kontrast und Farbsättigung der Darstellung.

Im Mittelpunkt der Arbeit steht in der Folge der Mensch. Es gilt die Limitationen seiner Kontrastwahrnehmung, aber vor allem seine subjektiven Präferenz mithilfe von Probandenstudien zu quantifizieren. Aufbauend auf den Ergebnissen wird es möglich, die vorhandenen TMOs hinsichtlich dieser Aspekte zu optimieren. Dies geschieht in Form einer aus den Untersuchungen abgeleiteten algorithmischen Lösung. Gleichzeitig ist es

das Ziel, durch den Algorithmus die weiteren notwendigen Broadcast-Anforderungen, wie z. B. die Robustheit, zu erfüllen. Anhand einer praktischen Umsetzung inklusive Live-Tests sollen die Erkenntnisse in ein einsatzfähiges Verfahren münden.

2.5 Potenzielle Workflows und ihre Verwendbarkeit

2.5.1 Autarker HDR- und SDR-Workflow

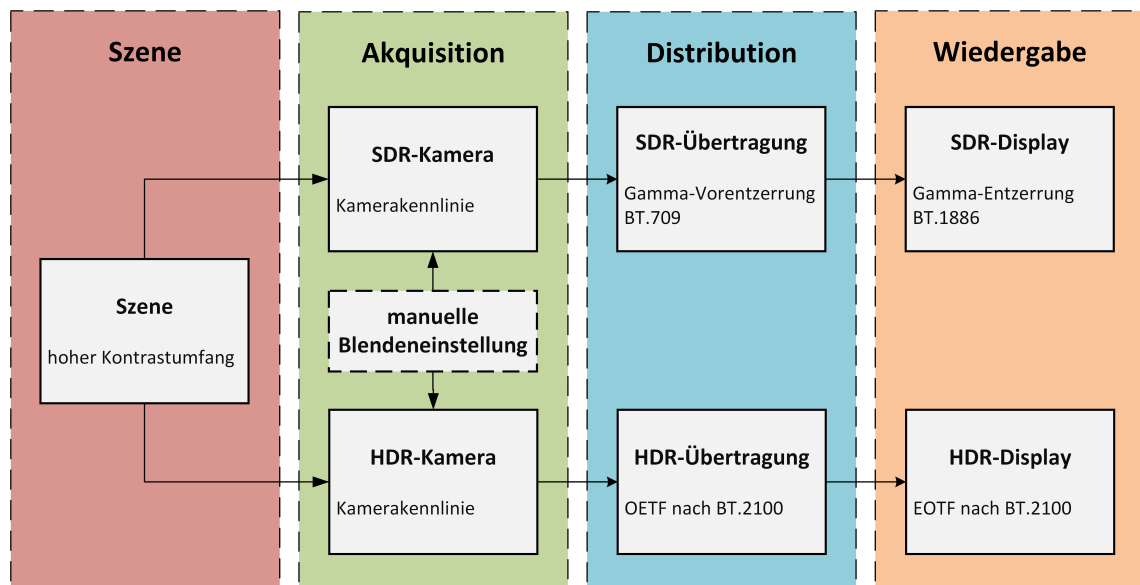


ABBILDUNG 2.4: Workflow für eine parallele SDR- und HDR-Produktion

Zunächst soll der Workflow zur Kontrastkompression einer klassischen Fernsehproduktion betrachtet werden, wie er im oberen Teil von Abbildung 2.4 gezeigt wird. Die Begrenzung des Kontrastumfanges findet hier durch die in der Kamera implementierte Übertragungskurve, auch als Kamerakennlinie bezeichnet, statt. Je nach Modell variiert der eingefangene Kontrastumfang zwischen 7 und 9 EVs. Zur Distribution wird das Signal mit einer OETF nach BT.709 belegt. Der Empfänger nutzt anschließend zur Darstellung die EOTF nach BT.1886.

Um die für den Betrachter wichtigen Bildelemente jederzeit korrekt belichtet zu wissen, kontrolliert und regelt der Bildingenieur, z. B. im Übertragungswagen, die Blendeneinstellung und definiert so den Wirkungsbereich der Kamerakennlinie. Außerhalb des Bereichs kommt es in der Folge zu Clipping. Die Zeichnung in hellen und dunklen Stellen geht verloren.³ In der Praxis ist dieser Effekt noch dadurch verstärkt, da es aufgrund von mangelndem Personal oft nicht zu einem ausreichenden Nachführen der Blende kommt.

³Moderne Kameras versuchen über das so genannte *Knee* (Knie) große Leuchtdichten langsamer ausbrennen zu lassen. Auf diese Weise vergrößert sich der eingefangene Kontrastumfang etwas.

Somit wäre eine (Teil-)Automatisierung an dieser Stelle wünschenswert. Ferner ist das Regulieren der Blende ein diskreter Vorgang. Es stehen nur eine begrenzte Anzahl an Stufen zur Verfügung. Das Umschalten muss sehr gefühlvoll erfolgen, um unschöne, vom Zuschauer wahrnehmbare Helligkeitssprünge von einem Bild zum nächsten zu vermeiden.



ABBILDUNG 2.5: Szene bei einem Fußballspiel bei starker Sonneneinstrahlung. Der Schlagschatten auf der Tribüne und dem Spielfeld lässt die Spieler fast darin verschwinden.

Vor allem beim Live-Broadcast - und hier explizit im Sportbereich - werden die Schwächen des beschriebenen Workflows deutlich. Ein typisches Beispiel ist ein Fußballspiel bei sonnigem Wetter. Das Stadionsdach wirft einen Schlagschatten auf Tribüne und Spielfeld. Ein enormer Kontrastumfang zwischen den direkt von der Sonne beleuchteten und den im Schatten befindlichen Abschnitten ist die Folge. Das menschliche Auge wäre in der Lage, diese Unterschiede deutlich abzumildern und so die Spielhandlung über das gesamte Feld ohne Probleme zu verfolgen. Im klassischen **SDR**-Bild hingegen überstrahlt die eine Bildpartie, während in der anderen einige Spieler regelrecht im Schatten versinken und kaum auszumachen sind. Der beschriebene Effekt ist in Abbildung **2.5** illustriert.

Um mit dem gleichen Ansatz auch **HDR** betreiben zu können, musste der Workflow um einen zweiten, völlig autarken Workflow erweitert werden (s. Abbildung **2.4**), so wie es GrassValley erstmals 2014 bei ihrer Testproduktion bei der MotoGP in Valencia gezeigt haben. **[18]** Neben den **SDR**-Kameras kamen **HDR**-Kameras zum Einsatz - mitsamt den zugehörigen Kameramännern sowie einem zweiten Arbeitsplatz für einen zusätzlichen Bildingenieur. Als Übertragungskanal wurde ein 3D-Kanal gewählt. Die eigentlich für das linke und rechte Auge gedachten Streams wurden somit zweckentfremdet. Heute

käme für die Distribution des Signals als **OETF** die ITU-R BT.2100 [58] zum Einsatz; die entsprechende **EOTE** würde anschließend beim Empfänger zur Darstellung genutzt.

Dass ein solches Vorgehen auf Dauer nicht praktikabel ist, liegt auf der Hand. Aus ökonomischen Gesichtspunkten ist eine doppelte Produktion ohne jegliche Synergie nicht sinnvoll, zumal in der Anfangszeit nur ein Bruchteil der Zuschauer das **HDR**-Signal empfangen und davon profitieren könnte sowie als potenzielle Kunden in Frage käme. Neben diesem offensichtlichen Punkt gibt es aber noch einen zweiten gravierenden Nachteil: Die Bildqualität bei **SDR** bleibt unverändert und in der Folge muss sich der allergrößte Teil der Zuschauer weiterhin mit den in Abbildung 2.5 illustrierten Problemen arrangieren. Der Anspruch sollte sein, auch die **SDR**-Empfänger von der besseren Qualität einer **HDR**-Produktion profitieren zu lassen.

2.5.2 Synergetischer HDR-SDR-Workflow

Sinnvoller wäre es, das **SDR**-Signal automatisch aus dem **HDR**-Signal zu erzeugen, ohne dabei zusätzlich eingreifen zu müssen. In diesem Fall ergibt sich im Vergleich zur oberen Darstellung die Abweichung, dass die Blendeneinstellung nur beim **HDR**-Signal erfolgt. Die noch nötige Kontrastkompression für **SDR** wird mithilfe einer Funktion bzw. eines Algorithmus realisiert. Dieses Vorgehen ist in Abbildung 2.6 gezeigt.

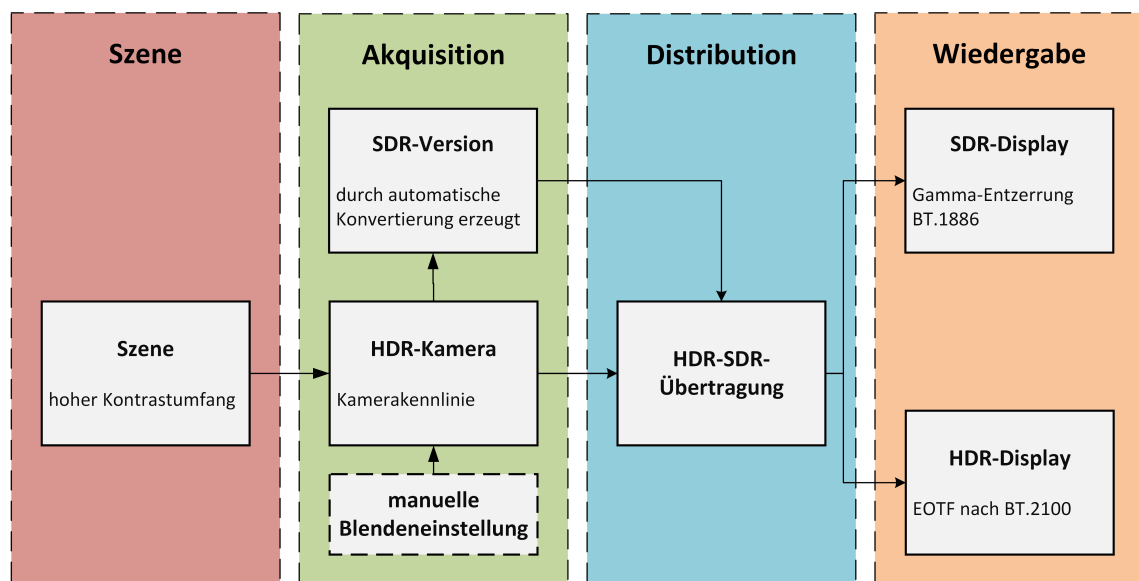


ABBILDUNG 2.6: **HDR** nach **SDR** Workflow

Es ist sogar möglich, noch einen Schritt weiterzugehen und sich einem Film-Workflow zu nähern. Aufgrund der verschiedenen Distributionskanäle mit ihren jeweiligen technischen Möglichkeiten wird dort zunächst ein Masterformat erzeugt, aus welchem sich alle anderen Versionen ergeben. Übertragen auf den Broadcast bedeutet dies, die maximale

zur Verfügung stehende Qualität der Kamera abzugreifen und erst für die Distribution entsprechend limitierte Versionen zu generieren. Dies kann erreicht werden, indem anstelle einer **OETF** nach ITU-R BT.2100 [58], die native logarithmische Kennlinie der jeweiligen Kamera gewählt wird. Der bekannteste Vertreter ist die Slog3-Kurve [119].

Auch bei der Distribution sind für den synergetischen HDR-SDR-Workflow verschiedene Varianten denkbar. Natürlich können weiterhin **HDR** und **SDR** in zwei einzelnen Übertragungskanälen zum Zuschauer gesendet werden. Denkbar wäre aber auch eine Konstruktion mit zwei Layern, welche in ein einziges Signal gepackt sind, oder die Verwendung einer hybriden **OETF**, welche eine Entzerrung mit zwei verschiedenen **EOTFs** erlaubt. Eben jene zwei alternativen Vorschläge sollen in den beiden folgenden Abschnitten genauer vorgestellt und hinsichtlich ihrer Einschränkungen analysiert werden.

Es wird sich zeigen, dass die automatische Kompression des Kontrastumfangs das Herzstück eines solchen Vorgehens darstellt. Je nach Qualität kann es bei dieser Konvertierung gelingen, eine Vielzahl an Informationen, die mit einer **SDR**-Kamera nicht eingefangen werden könnten, zu erhalten und so das **SDR**-Bild zu verbessern. Allerdings wird hiervon derzeit kein Gebrauch gemacht. Die verwendeten statischen Kurven bieten kaum bis keine Verbesserung im Vergleich zu einer klassischen **SDR**-Kamerakennlinie, woraus sich entsprechender Handlungsbedarf ergibt.

Da **HDR** eines der großen Zukunftsthemen und Forschungsgebiete im Bereich des Broadcasts ist, arbeiten mehrere Anbieter und Unternehmen der Branche daran, Vorschläge für die Produktion, Speicherung und Distribution anzubieten. Hier wären Dolby Laboratories, die British Broadcasting Company (**BBC**), Samsung, Philips und Technicolor zu nennen. In teils abgewandelter Weise wurden diese später bei der Society of Motion Picture and Television Engineers (**SMPTE**) als Subgruppen des SMPTE ST 2094 (*Dynamic Metadata for Color Volume Transform* [117]) standardisiert.⁴ Tabelle 2.1 zeigt eine Übersicht der wichtigsten Vorschläge.

Im Folgenden soll zunächst der ursprüngliche Vorschlag von Dolby Vision [94], der Ansatz der Dolby Laboratories, näher betrachtet werden, da die historische Entwicklung die Problematiken einer flächendeckenden Einführung von **HDR** verdeutlicht. Die Ansätze von Philips, Technicolor und Samsung besitzen Parallelen und arbeiten ebenfalls auf Grundlage von Metadaten. Da die **BBC** einen gänzlich anderen Weg beschritt und weiterhin beschreitet, wird dieser Vorschlag im Anschluss vorgestellt.

⁴Dolby Vision ST 2094-10, SL-HDR1 Philips ST 2094-20, Advanced HDR Technicolor ST 2094-30 und HDR10+ Samsung ST 2094-40

Herausgeber	Name	Jahr	Referenz
Dolby Laboratories	Dolby Vision	2014	[22] [94] [117]
Philips / Technicolor	SL-HDR1	2016	[39] [30]
Technicolor	Advanced HDR (heute Bestandteil von „SL-HDR1“)	2015	[117]
Samsung	HDR10+	2017	[117]
BBC	A Display Independent High Dynamic Range Television System	2015	[13]
BBC	Non-linear Opto-Electrical Transfer Functions for High Dynamic Range Television (Vorläufer des „Display Independent High Dynamic Range Television System“)	2014	[14]

TABELLE 2.1: Übersicht der HDR-Formate

2.5.3 Beispiel PQ und Dolby Vision

Die Grundlage für Dolby Vision bildet der Perceptual Quantizer (PQ) [114], der an die in Abschnitt 2.2.5 erläuterten Barten-Kurve [5] angelehnt ist. Er ermöglicht eine effektive Codierung von Leuchtdichtewerten im Bereich zwischen $0,002 \text{ cd/m}^2$ und 10.000 cd/m^2 . Dies entspricht 22 EVs. Die Gamma-Vorentzerrung würde für einen identisch großen Kontrastumfang 15 Bit benötigen. Dem PQ reichen wahlweise 10 oder 12 Bit. Bei 12 Bit werden die JNDs nach dem Barten-Modell an keiner Stelle überschritten, wie in Abbild 2.7 (rote Zone) zu sehen ist. Bei realen Aufnahmen, die ein minimales Rauschen aufweisen, werden auch bei 10 Bit keine Quantisierungsfehler sichtbar. [3] [40]

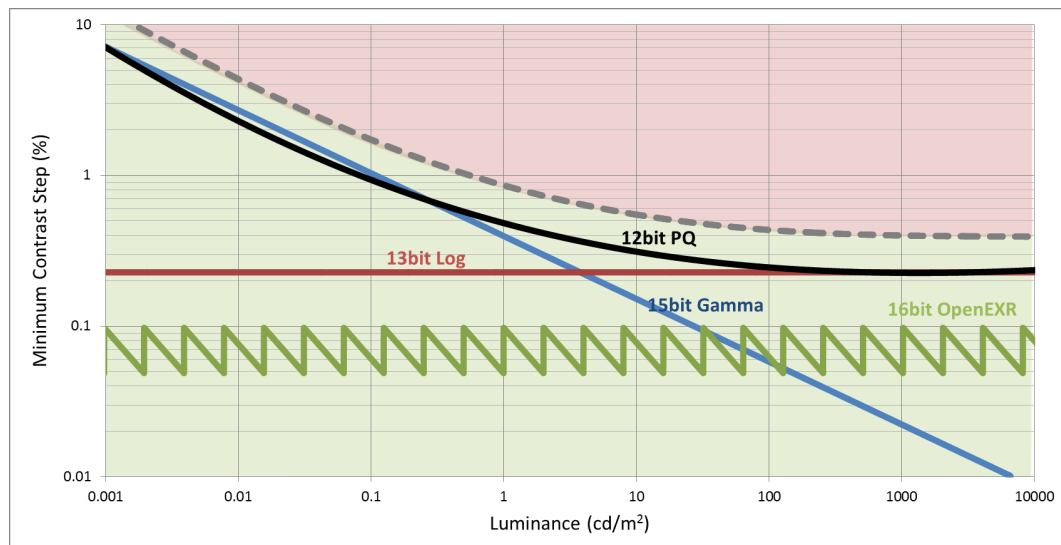


ABBILDUNG 2.7: Der Perceptual Quantizer (PQ) sorgt für eine effektive Codierung von HDR-Material. Er ist an die Barten-Kurve angelehnt, welche Auskunft über die JNDs der menschlichen Kontrastwahrnehmung gibt. [3]

Der **PQ** ist über die **EOTF** definiert. Sie, sowie die Inverse der **EOTF**, sind nachfolgend angeben. Es gilt zu beachten, dass die Inverse der **EOTF** nicht mit der **OETF** identisch ist, da aus gestalterischen Gründen noch eine **OOTEF** hinzugerechnet werden muss. Diese soll das Over-All-Gamma des **SDR**-Fernsehens nachbilden.

$$V = \left(\frac{0,8359375 + 18,8515625 * L^{0,1593017578125}}{1 + 18,6875 * L^{0,1593017578125}} \right)^{78,84375} \quad (2.4)$$

$$L = \left(\frac{V^{\frac{1}{78,84375}} - 0,8359375}{18,8515625 - 18,6875 * V^{\frac{1}{78,84375}}} \right)^{\frac{1}{0,1593017578125}} \quad (2.5)$$

L steht in den Formeln für die Leuchtdichte und V für die Spannung, welche dem Luma entspricht. L ist dabei zunächst durch den maximal möglichen Wert (10.000 cd/m^2) zu dividieren und somit auf den Bereich zwischen 0 und 1 zu normieren. Auch V ist auf diesen Bereich normiert. Im Vergleich zum heutigen System, welches mit einem Referenzweiß arbeitet, und somit relative Werte speichert, handelt es sich hier um absolute Werte. Die in der Formel enthaltenden Konstanten kommen durch ein mathematisches Näherungsverfahren an die Barten-Kurve zu Stande.

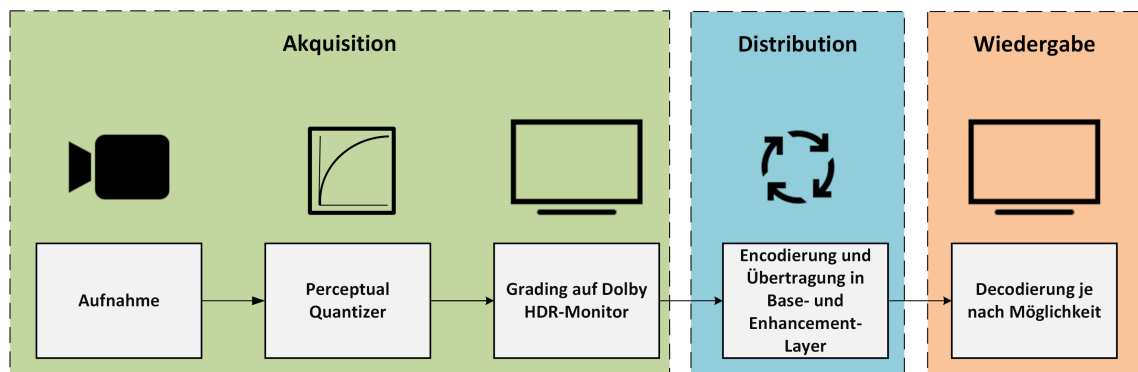


ABBILDUNG 2.8: Der ursprüngliche Workflow bei Dolby Vision von der Aufnahme bis zur Wiedergabe

Dolby Vision sieht für Nicht-Echtzeitanwendungen den in Abbildung 2.8 dargestellten Workflow vor. Die von der Kamera registrierten Spannungswerte werden mithilfe der **OETF** abgespeichert. Ein von Dolby entwickeltes Plugin ermöglicht das Color-Grading in der Postproduktion, an deren Ende ein Extended Dynamic Range (**EDR**)-Master⁵ steht. Das Grading muss an von Dolby lizenzierten Displays durchgeführt werden. Typische Beispiele sind die Dolby PRM-4220 oder der Dolby Pulsar. Die Darstellung erfolgt entsprechend durch Einsatz der **EOTF**.

⁵Mit **EDR** ist bei Dolby ein Signal gemeint, das sowohl einen höheren Kontrastumfang als auch einen erweiterten Farbraum sowie meist eine höhere spatiale Auflösung besitzt.

Nach dem Color-Grading für **HDR** schließt sich die manuelle Erstellung einer **SDR**-Version nach ITU-R BT.709 [54] an. Prinzipiell ist alternativ auch die Analyse des Videomaterials mit Hilfe eines weiteren Plugins und über die sogenannte Content Mapping Unit (**CMU**) denkbar. Hierdurch wird die Abwärtskompatibilität des Systems erreicht. Gleichzeitig wird die **CMU** auch dafür eingesetzt, spezielle Metadaten zu errechnen, um aus dem Standardsignal das Ursprungssignal zurückgewinnen zu können. [94]

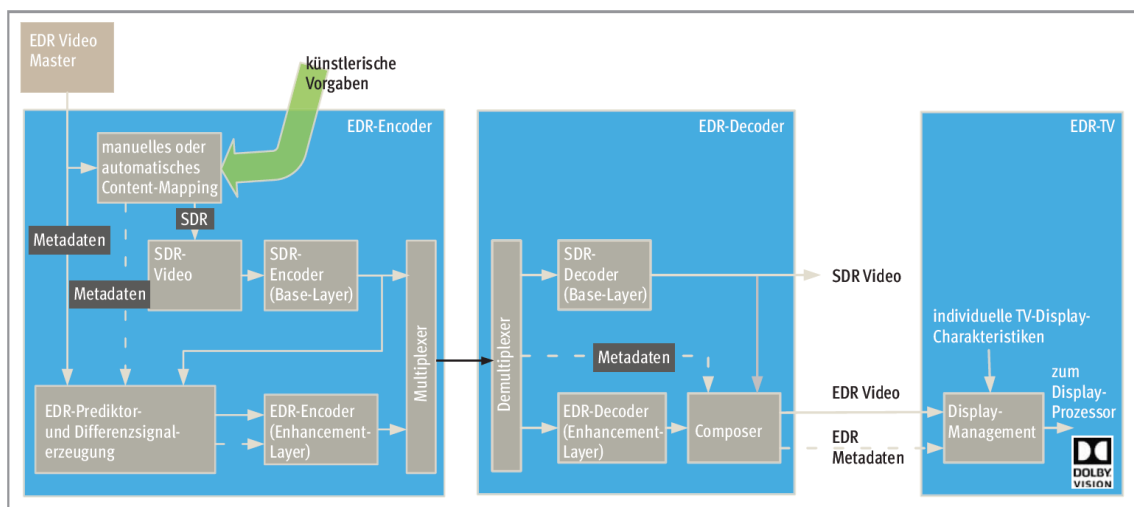


ABBILDUNG 2.9: Bei Dolby-Vision wird das Signal in zwei Layer zerlegt und ist somit abwärtskompatibel. Durch die Metadaten wird eine Anpassung auf lizenzierten Displays erreicht. [94]

Für die Übertragung zum Zuschauer können sowohl der HEVC- als auch der AVC-Standard verwendet werden. Dazu wird das Signal in einen Basis- und einen Enhancement-Layer zerlegt. In Abbildung 2.9 ist die Codierung dargestellt. Das in der Produktion durch die **CMU** entstandene **SDR**-Video wird im Basis-Layer codiert und kann von allen heutigen Empfängern interpretiert werden. Für den Enhancement-Layer vergleicht der **EDR**-Prediktor den Basis-Layer mit dem **EDR**-Master, um ein Differenzsignal zu erzeugen. Dazu werden auch die errechneten Metadaten miteinbezogen. Das Differenzsignal wird anschließend inklusive der Metadaten codiert. [94]

Diese statischen Metadaten sind von der SMPTE [115] standardisiert. Sie ermöglichen eine korrekte Anpassung des Signals an die Displaycharakteristiken der von Dolby lizenzierten Produkte. Dazu werden beide Layer decodiert und mithilfe der Metadaten kombiniert. Alle anderen Displays beschränken sich auf die Interpretation und Darstellung des Basis-Layers.

Wie aus dem Workflow ersichtlich wird, ist dieses System in erster Linie für die Postproduktion und somit für szenische Produktionen gedacht. Durch die Metadaten und die **OETF**, welche absolute Leuchtdichtewerte codiert, kann sichergestellt werden, dass das

Bild bei jedem von Dolby lizenzierten Empfänger quasi identisch aussieht. Auf diese Weise kann der *Creative Intent* des Coloristen, auf dessen manuelles Eingreifen und kreatives Entscheiden das System ausgelegt ist, sichergestellt werden. Der Leuchtdichtewert des Pixels in der Abnahme ist auch der Leuchtdichtewert des Pixels in der Wiedergabe. Sollte das Display bei Spitzlichtern nicht in der Lage sein, einen entsprechend hohen Wert zu erzeugen, wird dieser auf den größtmöglichen geclippt. Dies kann zu dem Paradoxon führen, dass das **HDR**-Signal dunkler dargestellt wird als das **SDR**-Signal, da es sich beim zweiten um eine relative Codierung handelt. Wie an früher Stelle erwähnt, wird die **SDR**-Version für ein Referenzweiß von 100 cd/m^2 erstellt, allerdings an den inzwischen deutlich helleren Heimempfängern mit 200 oder 300 cd/m^2 wiedergegeben. Entsprechend verschiebt sich der Helligkeitseindruck zwischen der **HDR**- und **SDR**-Version.

Positiv ist, dass die **SDR**- generell von der **HDR**-Version profitieren kann. Dies gilt aber nur gesetzt den Fall, dass die **HDR**-Version eine Postproduktion durchlaufen hat. Durch die sekundäre Farbkorrektur ist es möglich, helle und dunkle Partien im Bild anzugleichen und so eine lokale Kontrastkompression durchzuführen. Für eine Live-Produktion muss auf die global operierende **CMU** zurückgegriffen werden. Wie sich in Kapitel 3 noch zeigen wird, sind die Möglichkeiten einer solchen Kontrastkompression weitaus beschränkter als die von lokalen Verfahren.

Das technologische Hauptproblem bildeten allerdings die zwei Layer und die Metadaten und die damit verbundene Komplexität sowie Schwierigkeiten in der Interoperabilität. Dem Feedback von Rundfunkanstalten konnte entnommen werden, dass sie ein solches System für ihre täglichen Anwendungen als ungeeignet ansahen. Es wäre nicht möglich, die korrekte Erzeugung und Beibehaltung der Metadaten sicherzustellen. So könnte es zu ihrem Verlust über die verschiedensten Wegstrecken und Geräte hinweg kommen. Problematisch wäre auch die Verwendung von altem (Archiv-)Material in diesem Kontext oder Zu- oder Verkauf von fremdem Material. Deshalb findet sich der Dual-Layer-Ansatz heute nur noch auf der UHD-Bluray⁶. Der für das Broadcast angedachte Dual-Layer-Ansatz⁷ wird seit Ende 2017 [24] nicht mehr unterstützt. Das bedeutet auch, dass bei Dolby Vision heute nur eine Abwärtskompatibilität über das Versenden zweier unabhängiger Signale erreicht werden kann.

Für den **HDR**-Kanal ist durch die Definition von dynamischen Metadaten in der SMPTE ST 2094 [117] eine bessere Anpassung des Bildes an die jeweiligen Spezifikationen des Endgerätes hinzugekommen. Die ursprünglich definierten statischen Metadaten bezogen sich auf das gesamte Video. Mit dynamischen Metadaten hingegen können einzelne Szenen oder noch kürzere Sequenzen separat beschrieben werden. Diese sind allerdings unter

⁶wird auch Dolby Vision Profile 7 genannt [24]

⁷wird auch Dolby Vision Profile 4 genannt [24]

den vier Subgruppen der SMPTE ST 2094 nicht kompatibel. Dazu müsste Dolby Vision wieder auf statische Metadaten zurückgreifen, welches dem Profile 8.1 [24] entsprechen würde.

Für die Rundfunkanstalten ist die proprietäre Formatvielfalt ein erschwerender wirtschaftlicher Aspekt. Zum einen spielt sie für die Archivierung eine zentrale Rolle. Bei diesen proprietären Formaten ist aber unklar, welche Folgekosten zu erwarten sind und ob die Daten noch Jahrzehnte später gelesen werden können. Zum anderen schlägt sich der proprietäre Gedanke auch in der Beschränkung potenzieller Produktions- und Consumerdisplays nieder und wirft die Frage auf, wie groß der Markt ist, den man mit dem neuen Format erreichen kann.

Grundsätzlich bringt die Einführung von HDR noch einen Nebenaspekt mit sich, der näher in Kapitel 7 betrachtet werden soll. Bei der oben beschriebenen Erzeugung des zusätzlichen Layers wird zunächst das SDR-Bild wieder decodiert und mit dem HDR-Bild verglichen. Es wird versucht, eine möglichst gute Rekonstruktionsfunktion zu finden. Diese, sowie die noch wahrnehmbaren Differenzen, werden schließlich codiert. Ein Overhead von 20% bis 30% im Vergleich zum klassischen SDR-Signal ist die Folge. Je lokaler allerdings die Kontrastkompression durchgeführt wird, desto schwieriger ist es, eine geeignete Rekonstruktionsfunktion zu finden mit entsprechendem Einfluss auf die Datenrate. [83] Dabei bleibt allerdings unberücksichtigt, dass der Basis-Layer nicht dem heutigen Fernsehbild entspricht; wurde er doch mithilfe des Tonemappings aus dem HDR-Bild generiert. In der Folge ist mehr Zeichnung in Schatten und Lichter vorhanden und somit auch mehr Informationen, die es zu codieren gilt. Somit ist dieser Effekt nicht alleine auf den Dual-Layer-Ansatz beschränkt. Gerade für die Auswirkung auf den Basis-Layer gibt es keine Studie.

2.5.4 Beispiel HLG

Aufgrund der im vorangegangenen Abschnitt ausgeführten Schwierigkeiten bezüglich der Abwärtskompatibilität verfolgt die BBC ein vollkommen anderes Konzept. Die Kernidee ist dabei das sogenannte Hybrid-Log-Gamma (HLG) [58] aus nachstehender Gleichung.

$$L = \begin{cases} \frac{\sqrt{V}}{2} & \text{wenn } 0 \leq V \leq 1 \\ a * \ln(V - 0,28466892) + 0,55991073 & \text{wenn } 1 < V \end{cases} \quad (2.6)$$

$$V = \begin{cases} 4 * L^2 & \text{wenn } 0 \leq L \leq 0,5 \\ e^{\frac{L-c}{a}+b} & \text{wenn } 0,5 < L \end{cases} \quad (2.7)$$

Für eine konsistente Darstellung wurden die gleichen Symbole genutzt, wie schon bei der **PQ**. Allerdings ist L innerhalb eines Bereichs von 0 bis 12 skaliert. Diese **OETF** kann sowohl mit der entsprechenden **EOTF** als auch mit der Gamma-Kurve nach ITU-R BT.1886 [55] entzerrt werden. Ermöglicht wird das durch einen annähernd ähnlichen Kurvenverlauf von **HLG** und der Gamma-Vorentzerrung nach ITU-R BT.709 [54]. Erst bei ca. 50% des ursprünglichen Referenzweißes geht die **HLG** in eine logarithmische Form über und ist so in der Lage, auch Werte deutlich über dem Referenzweiß ($L > 1$) zu codieren. Es handelt sich somit um eine sehr rudimentäre Kontrastkompression. Sie könnte auch als extreme Knee-Funktion bezeichnet werden. Bei der **HLG** liegt entsprechend keine klassische **CSF** vor, obwohl die Grundidee identisch ist. Durch den hybriden Aufbau kommt das System ohne einen zweiten Layer und ohne Metadaten aus.

Aufgrund seiner hohen Kompatibilität und Analogie zu derzeitigen Systemen sind Broadcaster diesem Vorschlag deutlich zugewandter, auch wenn er einige Nachteile birgt. So kann das **SDR**-Bild kaum von einer **HDR**-Aufzeichnung profitieren. Durch die extreme Stauchung der Lichter aufgrund der nicht korrekten Entzerrung und der damit verbundenen flachen Gradation, könnte das **HLG**-Bild unter Umständen von den Zuschauern weniger präferiert werden als das derzeitige **SDR**-Bild. Auch ist der Creative Intent nicht sichergestellt. Die Darstellung an verschiedenen **HDR**-Displays kann genauso unterschiedlich ausfallen, wie an verschiedenen **SDR**-Geräten. Da es aber das Ziel ist, das **SDR**-Bild signifikant zu verbessern sowie einen konstanten Bildeindruck zu schaffen, scheint **HLG** zunächst weniger geeignet. Allerdings sind beide Nachteile auf die unzureichende Kontrastkompression zurückzuführen. Wäre eine Modifikation auf diesem Gebiet möglich, könnte dies Abhilfe schaffen.

2.5.5 Zwischenfazit bezüglich der Broadcast-Workflows

Zusammengenommen haben die Beispiel-Workflows gezeigt, wie komplex die Einführung von **HDR** im Broadcast ist und wie viele Aspekte eine Rolle spielen. Allerdings wird auch deutlich, dass die Kontrastkompression, gerade im Live-Betrieb, das entscheidende Element darstellt und dieses erst einmal unabhängig vom eigentlichen **HDR**-Standard ist. Ein synergetischer Workflow, so wie in Abbildung 2.4 dargestellt, bei dem die Blendeneinstellung anhand des **HDR**-Bildes vorgenommen wird, wird faktisch nicht eingesetzt. Die Varianz der Belichtung wäre für die heute verfügbaren Kontrastkompressionssysteme zu groß. Damit ist gemeint, dass bei einer echten **HDR**-Produktion die Blendeneinstellung kaum verändert werden muss, da der Leuchtdichtebereich so weit ist, dass es nur sehr selten zum Clipping kommt. Helligkeitsänderungen der Szene werden somit an den Zuschauer weitergereicht, welches maßgeblich zur Immersion beiträgt. Aufgrund des engen Korridors bei **SDR** müssen dort aber solche Schwankungen mithilfe der Blende

ausgeglichen werden. Soll nun das **SDR**-Signal von einem weiten **HDR**-Signal abgeleitet werden, kann dies nicht funktionieren, indem immer ein definierter Bereich des Signals herausgeschnitten wird. Deshalb werden auch **HDR**-Produktionen nach dem **SDR**-Signal angesteuert. So bleiben aber sowohl das **HDR**-Bild als auch das **SDR**-Bild hinter ihren Möglichkeiten zurück. Das **HDR**-Bild, da es in seiner maximalen Leuchtdichte sowie Varianz beschnitten wird, und das **SDR**-Bild, da es kaum vom erweiterten Kontrastumfang profitieren kann. Selbst in der BT.2408 (*Guidance for operational practices in HDR television production*) [59] wird eingeräumt:

„If the SDR production must not be compromised, both HDR and SDR cameras should be shaded using an SDR monitor fed via a down-converter. Whilst the HDR signals may not always exploit the full potential of the HDR production formats, the HDR pictures can still show significant improvement over SDR.“

Damit also die Blendeneinstellung in **HDR** vorgenommen werden könnte, würde es ein System benötigen, das für **SDR** quasi die Blende nachzieht bzw. deutlich mehr Kontrastumfang innerhalb eines **SDR**-Bildes vereinigt, so dass die Blende an Relevanz verliert. Es würde sich dadurch abheben, dass es das erste System wäre, welches dynamisch arbeitet. Eben jenes erste dynamische System soll in dieser Arbeit entwickelt werden. Indirekt wird in der BT.2408 bereits darauf verwiesen:

„As the exposure latitude of HDR images is far greater than SDR, a dynamic HDR to SDR converter may be required to deliver a satisfactory SDR output. A dynamic converter is designed to optimise the HDR to SDR tone mapping curve for any scene, thereby accommodating a wider range of exposures than might be possible with a fixed (or static) tone mapping curve.“

Das Zitat liefert bereits ein Stichwort zur Herangehensweise - Tonemapping Operatoren (TMOs). Im nächsten Kapitel sollen die bisher bekannten Möglichkeiten auf dem Gebiet der Kontrastkompression genauer beleuchtet werden.

Kapitel 3

Tonemapping und Broadcast-Anforderungen

3.1 Funktionsweise von Tonemapping-Operatoren

3.1.1 Einleitung

Über Jahrzehnte hinweg galten 24-Bit-Bilder mit 256 Werten pro Farbkanal als ausreichend akkurat in der Erstellung und Wiedergabe von fotorealistischen Bildern. [67] Nicht zuletzt dank der in der Fernsehtechnik verwendeten Gamma-Vorentzerrung, welche die Abbildung eines vergleichsweise großen Kontrastumfangs auf einen relativ kleinen Dynamikumfang ermöglicht. Aber bei schwierigen Lichtsituationen, beispielsweise bei Gegenlicht oder starkem Schattenwurf, stößt dieses Format an seine Grenzen. Der Fotograf oder Kameramann muss sich entscheiden, auf welchem Bildbereich der Fokus liegen soll, und diesen zuungunsten von anderen Bildbereichen erhalten. In der Folge kommt es in den sehr hellen bzw. sehr dunklen Bereichen zum Über- oder Untersteuern des Signals.

Durch die mit der Computergrafik aufkommende technologische Möglichkeit der synthetischen Erzeugung von Bildern, beispielsweise durch Radiosity oder das Monte-Carlo-Raytracing, mussten neue Dateiformate gefunden werden, welche in der Lage waren, die errechneten Leuchtdichtewerte abzuspeichern. [67] Auch in der Fotografie wurde es möglich, mit Hilfe von mehreren Aufnahmen (Belichtungsreihe) den Kontrastumfang der Kamera um ein Vielfaches zu steigern und das Ergebnis in so genannten Radiance-Maps zu hinterlegen.

Das Ausgabemedium, sei es ein Display oder ein Druckerzeugnis, war und wird auf absehbare Zeit nicht in der Lage sein, die aufgezeichneten Leuchtdichtewerte direkt eins zu eins wiederzugeben. Aufgrund der unter [2.2.2](#) dargelegten Erkenntnisse über die menschliche Wahrnehmung ist es auch durchaus fraglich, ob es überhaupt ein sinnvolles Ziel wäre, derart große Leuchtdichteunterschiede in so kleiner räumlicher Nähe zu reproduzieren. Gleichwohl gilt die Prämisse, dass trotz dieser Abweichung bei der Betrachtung auf dem Bildschirm dieselben visuellen Empfindungen hervorrufen werden sollen, wie bei der äquivalenten realen Szene.

Um die errechneten Leuchtdichtewerte auf einem Display anzeigen zu können, stellten Tumblin und Rushmeier 1993 [\[124\]](#) den ersten sogenannten **TMO** zur Kontrastkompression vor. Seitdem wurde eine Vielzahl diverser TMOs entwickelt. TMOs versuchen die Charakteristika der menschlichen Wahrnehmung nachzubilden. So steht häufig das Verhalten der Fotorezeptoren im Vordergrund, welche einen annähernd logarithmischen Zusammenhang zwischen Leuchtdichte und empfundener Helligkeit aufweisen. TMOs werden dabei anhand ihres Vorgehens in zwei große Klassen unterteilt:

- Bei den globalen Verfahren wird jedes Pixel mit derselben Übertragungskurve gewichtet. Diese basiert meist auf der mittleren und maximalen Leuchtdichte des Gesamtbildes.
- Bei den lokalen Verfahren wird versucht, den regionalen Kontrast zu erhalten, indem die Umgebung jedes Pixels mitberücksichtigt wird.

Obwohl TMOs im Bereich der Standbilder bereits eine über zwanzigjährige Entwicklungsgeschichte haben und es eine Vielzahl von Vorschlägen gibt, darunter auch mehrere Patente z. B. [\[26\]](#) oder [\[49\]](#), ist keines der Systeme in seiner heutigen Form für den Einsatz beim Broadcast geeignet. Eine großangelegte Evaluationsstudie von TMOs mit temporaler Komponente von Eilertsen et al. [\[31\]](#) kommt zu dem Ergebnis: „From the discussion we draw the conclusion that the problem of tone mapping of HDR-Video sequences is far from being solved.“ Obwohl die untersuchten Kandidaten auf Videos ausgelegt waren, führten die zeitlichen Änderungen zu derart großen Herausforderungen, dass diese nicht für jede Situation zufriedenstellend gelöst werden konnten. Durch das inzwischen reichhaltige und abwechslungsreiche zur Verfügung stehende Testmaterial werden solche Schwächen sichtbar. Diese starke Korrelation zwischen Bildvorlage und Ergebnis machen pauschale Aussagen schwer, wie an späterer Stelle noch erörtert werden wird.

Doch durch den enormen Leistungssprung der Kamerasensoren in den letzten Jahren, welche nun in der Lage sind, nativ einen sehr hohen Kontrastumfang einzufangen, wird es

deutlich wichtiger, ja geradezu unerlässlich, sich dem Problem der Kontrastkompression bei Videos verstärkt anzunehmen. Dabei muss berücksichtigt werden, dass sich die Anforderungen im Vergleich zum Einsatz bei der Fotografie maßgeblich unterscheiden, was der Grund für die unzufriedenstellenden Ergebnisse sein dürfte. Das Hinzukommen der zeitlichen Komponente sowie die mit dem Broadcast einhergehenden Ansprüche an Echtzeit, Automation, universeller Einsetzbarkeit und Zuschauerpräferenz stellen bis heute eine nicht überwundene Hürde dar. Eine genaue Betrachtung der Leistungsfähigkeit von TMOs sowie der angesprochenen Aspekte folgt im nächsten Abschnitt. Begonnen werden soll zunächst mit der prinzipiellen Vorgehensweise von globalen und lokalen Operatoren. Aus diesem Verständnis heraus können die Schwierigkeiten bei der Verwendung im Broadcast analysiert werden.

3.1.2 Globale Operatoren

Die Anzahl an verfügbaren globalen TMOs ist immens. Eine Übersicht in chronologischer Reihenfolge über die wichtigsten Vertreter findet sich in Tabelle 3.1. Anhand des Reinhard- und des Drago-Operators [103] [25] soll ihre prinzipielle Vorgehensweise beschrieben werden.

Person	Name	Jahr	Referenz
Schlick	quantisation technique TMO	1994	[108]
Ferwerda et al.	visual adaption TMO	1996	[38]
Tumblin et al.	revised Tumblin and Rushmeier TMO	1999	[123]
Pattanaik et al.	time-dependent visual adaptation TMO	2000	[93]
Reinhard et al.	photographic TMO	2002	[103]
Drago et al.	adaptive logarithmic TMO	2002	[25]
Irawan et al.	mal-adaption TMO	2005	[60]
Mantiuk et al.	display adaptive TMO	2008	[82]

TABELLE 3.1: Übersicht über die globalen TMOs

Zunächst wird bei beiden ein logarithmisches Mittel \bar{L}_w aus allen Leuchtdichtewerten gebildet.

$$\bar{L}_w = \frac{1}{N} \exp \left(\sum_{x,y} \log(\delta + L_w(x,y)) \right) \quad (3.1)$$

N ist dabei die Anzahl an Pixeln und δ ein minimaler Offset, für den Fall, dass L_w gleich 0 sein sollte. w steht dabei für *world* und impliziert, dass es sich um die Eingangswerte (Werte der Szene) handelt.

Anschließend werden alle Leuchtdichtewerte mithilfe des logarithmischen Mittels und dem so genannten *Key-Value* a vorskaliert. Der Key-Value gibt an, auf welchen Wert das logarithmische Mittel abgebildet werden soll. Der Standardwert 0,18 steht dabei für 18%, was in der Fotografie als mittleres Grau gilt. Kleinere Werte lassen das endgültige Bild dunkler wirken (Nachtszenen), wohingegen größere Werte zu helleren Bildern führen (High-Key-Szenen).

$$L(x,y) = \frac{a}{\bar{L}_w} L_w(x,y) \quad (3.2)$$

Das eigentliche Tonemapping besteht aus Abwandlungen der Naka-Rushton-Gleichung (1.1). L_{wmax} ist dabei der größte Leuchtdichtewert vor dem Tonemapping, oder ein entsprechend gewählter Wert. $L_d(x,y)$ sind die Werte am Ausgang (d wie Display). b ist ein Skalierungsfaktor, der standardmäßig den Wert 0,85 besitzt.¹

Reinhard:

$$L_d(x,y) = \frac{L(x,y) \left(1 + \frac{L(x,y)}{L_{wmax}^2}\right)}{1 + L(x,y)} \quad (3.3)$$

Drago:

$$L_d(x,y) = \frac{1}{\log_{10}(L_{wmax} + 1)} * \frac{\log(L_w + 1)}{\log\left(2 + \left(\frac{L_w}{L_{wmax}}\right)^{\frac{\log(b)}{\log(0,5)}} * 8\right)} \quad (3.4)$$

Auch wenn die Berechnungsvorschriften auf den ersten Blick sehr unterschiedlich anmuten, so ist doch das optische Resultat in den meisten Fällen ausgesprochen ähnlich. Abweichungen sind hauptsächlich in den hellen Bereichen des Bildes zu finden, welche bei Drago etwas kontrastärmer wirken, da die Kurve flacher ausläuft.

3.1.3 Lokale Operatoren

Lokale Operatoren versuchen der Tatsache Rechnung zu tragen, dass das menschliche Auge eine Szene nicht durch eine absolute Helligkeit erfasst, sondern sich regionsweise adaptiert. Der im vorherigen Abschnitt vorgestellte globale Reinhard-Operator existiert auch in einer lokalen Version, welche in Gleichung 3.5 dargestellt ist. Dabei wird $L_{surround}(x,y)$ verwendet, bei welchem es sich um eine weichgezeichnete Variante des Originalbildes handelt. Der Algorithmus sieht vor, dass die dazu verwendete Filtergröße schrittweise optimiert wird, bis der gewünschte lokale Eindruck entsteht.

¹Der Wert 0,85 wird von Drago et al. empfohlen, nachdem er in einer kleinen subjektiven Versuchsreihe bestimmt wurde. Über den Hintergrund der konstanten Zahlenwerte 2, 0,5 und 8 ist keine Information verfügbar.

Reinhard:

$$L_d(x,y) = \frac{L(x,y)}{1 + L_{surround}(x,y)} \quad (3.5)$$

Viele Operatoren arbeiten darüber hinaus mit einer bilateralen Filterung, um das Bild in hoch- und niederfrequente Anteile aufzusplitten. Das eigentliche Tonemapping wird anschließend nur auf den niederfrequenten Anteilen durchgeführt. Dadurch können die Kanten und somit auch die Details besser erhalten werden.

Insgesamt ist auch die Anzahl der aus der Literatur bekannten lokalen Operatoren enorm. Tabelle 3.2 zeigt eine Auswahl bekannter Vertreter in chronologischer Reihenfolge. Eine Vielzahl dieser findet sich auch in den im Anschluss referierten Evaluationsstudien wieder.

Person	Name	Jahr	Referenz
Pattanaik	adpativ gain control TMO	2002	[92]
Reinhard et al.	photographic TMO	2002	[103]
Fattal et al.	gradient domain TMO	2002	[36]
Durand et al.	fast bilateral filtering TMO	2002	[28]
Ledda et al.	local adaption TMO	2004	[68]
Bennett et al.	virtual exposures TMO	2005	[7]
Kuang et al.	iCAM06 TMO	2007	[63]
Benoit et al.	retina model TMO	2009	[8]
Reinhard et al.	color appearance TMO	2012	[102]
Boitard et al.	temporal coherence TMO	2012	[11]

TABELLE 3.2: Übersicht über die lokalen TMOs

3.1.4 Temporaler Aspekt

Die in den vorangegangenen beiden Abschnitten vorgestellten Operatoren führen zur Kontrastkompression eine Intraframe-Analyse durch. Bei einer Abfolge von Bildern kann es aufgrund von Schwankungen der Leuchtdichtewerte zu veränderten Berechnungsergebnissen kommen, woraus Flackern resultieren kann. Um dies zu vermeiden, wird eine Interpolation über mehrere Bilder notwendig. Kiser et al. [61] nutzen dazu folgende Gleichungen,

$$A_n = (1 - \alpha_A)A_{n-1} + \alpha_A A \quad (3.6)$$

$$B_n = (1 - \alpha_B)B_{n-1} + \alpha_B B \quad (3.7)$$

$$a_n = (1 - \alpha_a)a_{n-1} + \alpha_a a \quad (3.8)$$

wobei A sowie B Leuchtdichtewerte der Szene beschreiben, α den Key-Value angibt und n für die Bildnummer steht. Das neue Bild geht zu einem Anteil α und das vorherige zu einem Anteil $(1 - \alpha)$ in das Gesamtergebnis ein. Das vorherige Bild wurde seinerseits ebenfalls aus seinen Vorgängern gebildet. Auf diese Weise kann darauf verzichtet werden, mehrere Bilder zu speichern. Der Wert für α liegt in etwa in einem Bereich zwischen 0,02 und 0,05. Somit findet die Adaption bei sich verändernden Lichtverhältnissen langsam und kontinuierlich statt. Kurzzeitige Schwankungen der Helligkeit werden geglättet.

Auch bei Durand et al. [27] wird ähnlich vorgegangen. Hier wird die mittlere Leuchtdichte des Bildes, also die Leuchtdichte, an die das Auge adaptiert ist, ausgeglichen, um eben jenen Adaptionprozess zu simulieren.

$$Y_a = Y_a + (Y - Y_a) * (1 - e^{-\frac{T}{\tau}}) \quad (3.9)$$

Dabei ist T die Dauer eines Frames und τ eine zeitliche Konstante, welche die Adaptionprozesse von den Stäbchen und Zapfen wiedergibt. Bei [62] werden für τ 0,4 s bzw. 0,1 s für Stäbchen bzw. für Zapfen angenommen.

Das beschriebene Vorgehen wird *Parameter-Smoothing* genannt. Es ist allerdings nur beim globalen Tonemapping einsetzbar, da es darauf abzielt, die Analyseergebnisse der vorherigen Bilder (z. B. die maximale und mittlere Leuchtdichte) zu speichern und mit den Ergebnissen des aktuellen Bildes zu verrechnen. Im lokalen Fall wird hingegen ein Adaptionniveau pro Pixel berechnet. Theoretisch wäre auch dies speicherbar, allerdings passieren die räumlichen Änderungen zu schnell, um eine zufriedenstellende Interpolation zu ermöglichen. Sprich, bereits kleinste Bewegungen würden große Nachzieheffekte verursachen. Würde man die Interpolationszeit allerdings stark verkürzen, könnte die Bildhelligkeit nicht stabil gehalten werden. Entsprechend muss beim lokalen Tonemapping auf der Bilddomäne mit z. B. Blockmatching oder Verfahren zur Bewegungskompensation gearbeitet werden. [4] Solche Verfahren sind in der Folge viel komplexer und erfordern Kenntnis über zukünftige Bilder.

Bei Live-Anwendungen ist es allerdings nicht möglich, den Bearbeitungszeitpunkt bzw. das zu bearbeitende Bild in die Vergangenheit zu verlegen, da sonst das Delay zu groß werden würde. Auch die Bearbeitungszeit ist beim Live-Videosignal durch die Bildwechselfrequenz begrenzt. Für 60 Hz liegt diese bei 16,6 ms. Entsprechend bilden die letzten beiden Punkte die ideale Überleitung zum nächsten Abschnitt, in welchen nun die Broadcast-Anforderungen gesammelt und mit den TMOs verglichen werden sollen.

3.2 Tonemapping und Broadcast-Anforderungen

Damit ein Workflow, wie in Abbildung 2.6, für Broadcast-Anwendungen in Frage kommt, müssen bei den TMOs prinzipielle Voraussetzungen erfüllt sein. Diese sollen im Folgenden erarbeitet und direkt mit dem Stand der Technik verglichen werden.

3.2.1 Echtzeit

Zunächst einmal muss die Rechengeschwindigkeit des TMO eine Echtzeitanwendung erlauben. Ob dies gegeben ist, hängt nicht allein von seiner Komplexität ab. Entscheidend sind darüber hinaus beispielsweise die Möglichkeiten zur Parallelisierbarkeit, die Art der Implementierung, die Bildgröße und die geforderte Bildwechselfrequenz. Das resultierende Ergebnis spiegelt dabei eine Momentaufnahme wider und muss vor dem zeitlichen Hintergrund gesehen werden. Laut Moorschem Gesetz verdoppelt sich die Rechenleistung in etwa alle 20 Monate.² Was bis vor wenigen Jahren noch als zu komplex galt, kann bald schon echtzeitfähig sein.

Zhao et al. haben 2008 [132] die Rechenzeit von acht TMOs untersucht und dafür für zwei von ihnen (Drago [25] und Pattanaik [93]) eine neue GPU-Implementierung realisiert. Die Berechnungszeit für die fünf globalen Operatoren waren sehr ähnlich. Bei der größten Bildvorlage³ kamen Reinhard (global), Ashikhmin (global) und Drago jeweils auf 6 Hz. Eine einfache logarithmische Kurve erreicht nur ein Bild mehr. Unter Berücksichtigung, der für UHD angestrebten spatialen Auflösung und der Tatsache, dass sich die Rechenleistung bis heute mehr als sechsmal verdoppelt haben dürfte, ergibt sich eine Bildwechselfrequenz von ca. 20 Hz. Eine Bearbeitung von Videosignalen in Echtzeit mit einer für das Fernsehen ausreichenden Bildwechselfrequenz erscheint somit im Bereich des Möglichen.

Die lokalen Operatoren sind dagegen etwas langsamer. Reinhard (lokal) kommt auf 3 Hz; Ashikhmin (lokal) ebenfalls und Durand auf 1 Hz. Hier eine Echtzeitfähigkeit herzustellen dürfte schwieriger, aber nicht unmöglich sein. Die Ausnahme bildet Zhaos neue Implementierung des Pattaniak Operators, welcher auf 24 Hz kommt, sprich viermal so schnell war wie die globale Konkurrenz. Bereits 2005 konnten auch Krawczyk et al. [62] zeigen, dass unter geschickter Ausnutzung von Synergien bezüglich der benötigten Gauß-Kurven der lokale Reinhard Operator deutlich effizienter implementiert werden kann. Für das größte zur Verfügung stehende Bild⁴ wurden 80 ms benötigt, was ca. 12

²Genaugenommen muss zwischen der Rechenleistung bei der zentrale Verarbeitungseinheit (CPU) und der Grafikprozessor (GPU) unterschieden werden. Hier soll lediglich darauf aufmerksam gemacht werden, dass sich Aussagen auf diesem Gebiet schnell überholen.

³Price Western mit einer Auflösung von 3272 x 1280 Pixel

⁴1024 x 768 Pixel

Hz entspricht. Berücksichtigt man erneut Anzahl der Pixel und Entwicklung der Rechenleistung, ergibt sich so eine erzielbare Bildwechselfrequenz von 15 Hz.

Dass aber noch deutlich höhere Werte möglich sind, zeigen Kiser et al. 2012 [61]. Ihr leichtgewichtiger, auf dem globalen Reinhard basierender Algorithmus mit einer einfachen zeitlichen Interpolation kann ein 1080p-Bild in der Zeit einer vertikalen Abtastlücke (1,6 ms) berechnen. Übertragen auf die heutigen Anforderungen ergäben sich so 300 Hz.

Es zeigt sich, dass globale Operatoren im Allgemeinen einen Geschwindigkeitsvorteil besitzen, aber nicht alle lokalen prinzipiell aufgrund der Rechenzeit ausscheiden. Unter Umständen kann dies eine Einschränkung darstellen. Generell gilt es, in dem maximal für Echtzeit zulässigen 16,6 ms ein bestmögliches optisches Ergebnis zu erzielen.

3.2.2 Szenenwechserkennung und temporale Interpolation

Ein Verfahren zur Kontrastkompression ist im Hinblick auf Broadcast an mehreren Stellen der Verarbeitungskette denkbar. So könnte es direkt zu Anfang in der Kamera zum Einsatz kommen, wie die heutige Kamerakennlinie auch. Eine weitere Option wäre, das volle HDR-Signal zum Übertragungswagen bzw. zur Regie zu führen und dort das Verfahren vor der Distribution anzusetzen. Dies entspräche auch dem eingangs in Abbildung 2.6 gezeigten Workflow.

Im zweiten Fall würde es sich um den so genannten Programmausgang handeln. In der Folge würde die Videosequenz Schnitte enthalten. Essenzen von mehreren Kameras wären in einem konstanten Strom vereint. Es kommt zum Umschnitt zwischen verschiedenen Szenen. Dies führt aber zu Problemen im Bezug auf die zeitliche Verschleifung, wie sie in Abschnitt 3.1.4 dargestellt wurde.

In diesem Fall wäre es wünschenswert, wenn das erste Bild einer neuen Szene nicht von den Werten vorheriger Bilder der alten Szene beeinflusst würde. Es wäre also nötig, die zeitliche Interpolation für ein Bild zu deaktivieren bzw. die gespeicherten Gesamtwerte mit den Werten des aktuellen Bildes zu überschreiben. Hierfür müsste die Szene als solche erkannt werden. Die Berücksichtigung dieser Problematik findet sich in keinem vorhandenen TMO. Entsprechend ist es im Hinblick auf die zeitliche Interpolation nötig, Parameter-Smoothing für lokales Tonemapping zu ermöglichen, um Blockmatching oder ähnliche Verfahren zur Bewegungskompensation zu umgehen (s. 3.1.4) und jenes darüber hinaus um eine Szenenwechserkennung zu erweitern.

3.2.3 Universale Verwendbarkeit und Robustheit gegen Artefakte

Das Fernsehbild zeichnet sich durch eine Vielfalt von Bildinhalten aus. Hierzu zählen beispielsweise Tag- und Nachtszenen, schnelle und ruhige Bewegungen, Menschen und Tiere, Städte und Landschaften, Sport und Live-Events, usw.. Hieraus ergibt sich als dritte Grundvoraussetzung die universelle Einsetzbarkeit bei all diesen Szenarien. Da das Grundwesen einer Evaluation die Testsequenzen auf eine endliche Menge beschränkt, können allerdings nie garantierte Aussagen bezüglich jeglichen denkbaren Fällen getroffen werden. Durch eine gezielte Auswahl der Inhalte kann allerdings versucht werden, eine möglichst hohe Aussagekraft zu erreichen.

Ein Hauptaugenmerk liegt darauf, ob der **TMO** bei der Kontrastkompression von Videosequenzen Artefakte erzeugt, also für Effekte, Objekte oder Erscheinungen sorgt, welche in der natürlichen Szene nicht vorkommen. Die drei wichtigsten Artefakte in diesem Zusammenhang wären:

Halo (halo): Als Halo wird der Schein um Konturen bezeichnet. Er kommt zustande, wenn in einem sehr kleinen Bereich des Bildes unterschiedliche Leuchtdichtewerte zusammentreffen - gut zu erkennen an den Kanten der Skyline von Frankfurt in Abbildung **3.1**. Dort kommt es auf der dunklen Seite zu einer positiven und auf der hellen Seite zu einer negativen Polarisierung. **[104]** In der Fotografie wird dies zuweilen als gewollter künstlerischer Effekt eingesetzt. Für den Broadcast-Bereich gilt es, dieses Artefakt zu vermeiden.



ABBILDUNG 3.1: Die Skyline von Frankfurt mit starken Halo-Artefakten

Flackern (flicker): Als Flackern werden schnelle Helligkeitsänderungen zwischen den einzelnen Bildern einer Sequenz bezeichnet. Ein wahrnehmbares Pulsieren der Helligkeit ist die Folge. Der Effekt tritt auf, wenn kohärente Bilder allein für sich betrachtet und bearbeitet werden. Durch zeitliche Veränderungen in der Szene, sei es durch Bewegung

oder Rauschen, werden die aufeinander folgenden Bilder unterschiedlich prozessiert. Eine wahrnehmbare Inkonsistenz ist die Folge.

Ghosting (ghosting): Ghosting, also eine geisterhafte Erscheinung, kann mehrere Ursachen haben. Zum einen tritt sie auf, wenn zeitlich sequenziell aufgenommene Bilder genutzt werden, um daraus eine einzige hochkontrastreiche Aufnahme zu generieren. Die Veränderungen über die Zeit lassen die Bilder nicht mehr ideal übereinander liegen. Zum anderen, und hier von Relevanz, kann eine Rauschunterdrückung oder Bewegungskompensation zu solchen Artefakten führen.

Für die Einschätzung bekannter TMOs soll an dieser Stelle auf zwei Studien [31] [96], von welchen die erste bereits zuvor erwähnt wurde, zurückgegriffen werden. In der Studie von Eilersten et al. wurden 11 (6 globale und 5 lokale) für Videomaterial ausgelegte TMOs miteinander verglichen. Bevor es allerdings zu einer jeweils paarweisen Gegenüberstellung aller Kandidaten durch einen Probandentest kam, wurde eine Vorselektion durchgeführt. Die TMOs, welche inakzeptable Artefakte produzierten und so zu keiner sinnvollen Bewertung geführt hätten, wurden eliminiert. Dies traf auf 4 lokale Operatoren zu, da das Flackern und das Ghosting eine Verwendung in der Praxis unmöglich machten. Einer dieser Operatoren hatte zudem mit erheblichen Übersättigungen zu kämpfen, wie auch ein globaler, welcher ebenfalls aussortiert wurde. Im direkten Vergleich der verbleibenden 7 Operatoren blieb der letzte lokale Operator hinter anderen globalen zurück. Positiv hervor taten sich der *mal-adaption TMO* [60] von Irawan und der *display adaptive TMO* [82] von Mantiuk. Den globalen Operatoren wurde in der Folge eine Robustheit gegen Flackern, Ghosting und Halos attestiert.

Diese beiden zuvor erwähnten TMOs kamen auch bei der Studie von Petit et al. [96] zum Einsatz. Sie zeigten über die 7 getesteten Sequenzen eine konstant gute Leistung. Neben den beiden wurde noch ein **TMO**, welcher eine klassische Kamerakennlinie nachahmen sollte, und der lokale *iCAM06*-Operator untersucht. Der *iCAM06* konnte dabei dreimal, teils sehr knapp, den Spitzenrang belegen. Er lag dafür aber bei allen 4 Nachtszenen abgeschlagen auf dem letzten Platz. Der Grund war ein deutlich zu helles und somit inakzeptables Erscheinungsbild.

Die starke Ambivalenz von lokalen Operatoren mit teils hervorragenden, aber auch in diskutablen Ergebnissen, kann aus eigenen Voruntersuchungen mit dem *iCAM06* in der MatLab-Implementierung von [53] und Testszenen von Fröhlich [41] bestätigt werden. Zwar war die generelle Unzulänglichkeit bei Nachtszenen in den eigenen Untersuchungen nicht im vollen Umfang gegeben, aber auch hier gab es Licht und Schatten. So liefert der Operator bei der Szene *carousel_06* ein überzeugendes Ergebnis ab, scheitert aber gleichzeitig bei dem Versuch, blauen Himmel, Feuerwerk oder ein Dämmerlicht abzubilden, wie in Abbildung 3.2 ersichtlich ist.



ABBILDUNG 3.2: 4 Beispiele des iCAM06-Operators mit dem Testmaterial von Fröhlich
[41]

Aber auch die globalen Operatoren führten teilweise zu leichten Überbelichtungen, deren Ausprägungen in den referierten Studien jedoch noch in einem verträglichen Rahmen lagen. Zudem ist nicht nur die Stärke der Ausprägung, sondern auch die Art des Artefakts dafür entscheidend, wie störend es vom Zuschauer wahrgenommen wird. So gilt Rauschen eher als akzeptabel als Flackern, Ghosting oder Übersättigungen. [31]

Ein prinzipieller Nachteil von globalen Operatoren war der vergleichsweise geringe lokale Kontrast. [31] Feine lokale Unterschiede gingen verloren, wodurch auch die Schärfeleistung subjektiv abnahm, da ein kontrastreiches Bild gleichzeitig auch schärfer wirkt. In diesem Punkt verhielten sich die lokalen Operatoren deutlich besser. Das in Abbildung 3.3 gezeigte Bild wirkt flau, nicht obwohl, sondern gerade weil der abgebildete Kontrastumfang deutlich größer ist als in Abbildung 2.5. Dort wurde auf die gleiche Vorlage ein Kamerakennlinien ähnlicher TMO angewendet. Durch das Clipping und die damit verbundene steilere Gradation wirkt es kontrastreicher.

Abschließend bleibt deshalb festzuhalten, dass nicht mit letzter Sicherheit gesagt werden kann, wie sich derzeitige lokale Operatoren in einem Broadcast-Workflow verhalten. Das Risiko für die Ausprägung von Artefakten und anderen optischen Phänomenen für eine Echtzeitanwendung, bei der es keine nachträgliche Möglichkeit der Kontrolle und Korrektur gibt, scheint dennoch recht hoch zu sein. Gleichzeitig deutet sich aber auch an, dass nur durch die Berücksichtigung lokaler Aspekte eine kontrastreiche Darstellung hoher Kontrastumfänge gelingen kann und rein globales Vorgehen nicht ausreichend wäre. Diesen Zielkonflikt gilt es Laufe der Arbeit aufzulösen.



ABBILDUNG 3.3: Szene aus einem Fußballspiel. Der Kontrastumfang wurde mit dem Drago **TMO** komprimiert. Das Ergebnis wirkt subjektiv kontrastarm.

3.2.4 Akzeptanz durch den Zuschauer

In der bereits zitierten Studie von Eilertsen et al. [31] werden die untersuchten TMOs in drei Kategorien eingeteilt:

- *visual system simulators (VSS)*: Sie simulieren die Limitationen und Eigenschaften des visuellen Systems. Zum Beispiel können sie Glanz, Nachtsehen oder die Reduktion von Sättigung und Kontrast in dunklen Bildbereichen nachstellen.
- *scene reproduction (SRP) operators*: Sie versuchen, den Eindruck der Szene im Original bezüglich Kontrast, Schärfe und Farben zu erhalten, wenn das Bild auf einem Display mit kleinerem Farbraum, geringerem Kontrast und geringerer Leuchtdichte gezeigt wird.
- *best subjective quality (BSQ) operators*: Sie produzieren das am meisten präferierte Bild oder Video hinsichtlich der subjektiven und/oder künstlerischen Vorlieben.

Es wird deutlich, dass eine Bewertung der Operatoren nur unter einer definierten Zielsetzung erfolgen kann, da ihnen unterschiedliche Intentionen zu Grunde liegen. Somit muss die Frage gestellt werden, was im Falle des Broadcasts erreicht werden soll. In der Zielsetzung in Abschnitt 2.4 wurde definiert, bei der Kontrastkompression ein möglichst von den Zuschauern subjektiv präferiertes Ergebnis zu schaffen. Denn letztlich bilden

sie die Instanz, an die sich das Fernsehprogramm richtet und die somit auch über Erfolg oder Niederlage entscheiden.

Dieses Ziel findet sich in der dritten der oben genannten Kategorien (*best subjective quality*) wieder. Allerdings weist diese Kategorie gerade einmal zwei Vertreter auf - den *camera TMO* und den *virtual exposures TMO*. Der *camera TMO* ist nichts anderes als eine Nachbildung einer heutigen Kamerakennlinie, mit den damit verbundenen Vor- und Nachteilen. Sprich, er ist sehr robust und subjektiv kontrastreich, besitzt dafür aber nur einen sehr eingeschränkten Kontrastumfang. Der *virtual exposures TMO* ist ein Tonemapping-Ansatz, um **SDR**-Videos zu verbessern. Beide vermögen somit nicht eine Kontrastkompression unter Beibehaltung eines möglichst großen Kontrastumfangs der Szene durchzuführen und entsprechen folglich aus anderen Gründen nicht der Zieldefinition.

Trotzdem ist es sinnvoll den *camera TMO* näher zu betrachten. In der Studie von Petit et al. [96] schnitt dieser beispielsweise nicht signifikant schlechter ab, als die anderen drei getesteten TMOs, obwohl die Schwächen des klassischen **SDR**-Signals bekannt sind. Die hierfür ausschlaggebenden Gründe könnten einen Schlüssel zur Lösung bilden. Dass eine Kamerakennlinie überhaupt mit „richtigen“ TMOs verglichen wurde, liegt an einer Untersuchung von Akyuzetal et al. aus dem Jahre 2007 [2]. Als jene überprüfen wollten, ob **HDR**-Displays auch **SDR**-Material unterstützen, kamen sie zu folgendem Fazit:

„We also find, perhaps surprisingly, that tone-mapped HDR images are often no better than the best single **LDR**⁵ exposure from a bracketed sequence. In other words, to truly benefit from the new technology you will probably need an **HDR** display, and not just an **HDR** camera.“

Davon inspiriert kam der *camera TMO* auch bei Eilersten et al. [31] und, wie bereits erwähnt, Petit et al. [96] zum Einsatz. In beiden Fällen könnte der **TMO** durch Konstanz und Robustheit überzeugen. Des Weiteren teilten Petit et al. die Probanden in zwei Gruppen ein. Die erste Gruppe war dazu angehalten die Videos hinsichtlich der Gesamtqualität (*overall quality*) zu bewerten. Die zweite Gruppe sollte ihr Augenwerk auf die Natürlichkeit bzw. Realitätsnähe (*fidelity with the memorised reality*) richten. Obwohl es sich bei der ersten Zielsetzung um ein sehr subjektives Kriterium handelt und bei der zweiten um ein vermeintlich eher objektives, konnten keine signifikanten Unterschiede zwischen den Gruppen festgestellt werden.

Natürlich besteht zwischen diesen beiden Zielsetzungen eine starke Korrelation, aber je nach Bildvorlage kann der *camera TMO* aufgrund seiner Einschränkungen (Informationsverlust in Lichtern und Schatten) kein realistisches Abbild schaffen. Letztlich

⁵**LDR** ist hier als Synonym mit **SDR** zu verstehen.

bleibt fraglich, ob der Zuschauer einschätzen kann, wie die Szene im Original aussehen müsste. Denkt man beispielsweise an Zeugenbefragungen, ist bekannt, wie massiv sich das Gedächtnis täuschen lässt.

Folgend sei ein Beispiel angeführt. Das erste Bild in Abbildung 3.4 zeigt eine möglichst realistische Reproduktion des Sinneseindrucks der Szene; so tendiert beispielsweise in den sehr dunklen Bildbereichen die Sättigung gegen null, um das Nachtsehen zu simulieren. Auch ist die Gradation eher flach gewählt. Das zweite Bild hingegen wurde nach künstlerischen Gesichtspunkten optimiert. Die deutliche Mehrheit der Zuschauer würde sich mutmaßlich für dieses entscheiden ohne es dabei für unnatürlich zu halten.



ABBILDUNG 3.4: Zwei Bilder der selben Szene. Das Linke wurde an die menschliche Wahrnehmung angelehnt, das Rechte nach künstlerischen Gesichtspunkten erstellt.

Ein Punkt der dabei nicht zu verachten ist, ist die Tatsache, dass immer eine Vorprägung durch Erfahrungen mit bestehenden Medien besteht, welche unterbewusst ein ästhetisches Verständnis formt. Allen voran seien - neben dem heutigen Fernsehen - Film, Fotografie und Malerei zu nennen. In den allermeisten Fällen wird dabei eine künstlerische Intention verfolgt und versucht, auf diesem Wege eine Botschaft zu übermitteln. Da wir in der Folge diesen Bildeindruck gewohnt sind, verschwimmt die Grenze zwischen natürlich und künstlerisch. Weist das Bild keine über- oder unterbelichteten Bereiche auf, werden diese unter Umständen sogar vermisst. (s.a. [2]) Deshalb ist die Frage essenziell, wie es gelingen kann, Präferenz für diesen definierten Fall quantitativ messbar und somit abbildbar und nutzbar zu machen. Denn der Autor ist nicht wie Akyuz et al. [2] der Meinung, dass es unbedingt ein HDR-Display benötigt, um von einer HDR-Kamera zu profitieren, sondern der Überzeugung, dass es möglich ist den dargestellten Kontrastumfang auf SDR-Displays sinnvoll zu steigern, wenn es gelingen würde, die für den Zuschauer wichtigen Qualitätskriterien weiterhin beizubehalten.

Es bleibt festzuhalten, dass TMOs, bis auf die erwähnten Ausnahmen, das Ziel verfolgen, ein möglichst realistisches Abbild der Szene zu erzeugen. Dank der relativ hohen Korrelation zwischen Realismus und Präferenz lassen sich so mit TMOs, welche eigentlich einer anderen Intention dienen, bereits teilweise sehr gute Ergebnisse erreichen. Meistens werden jedoch nicht alle Aspekte erfüllt, welche für den Zuschauer eine hohe subjektive

Bildqualität bedeuten. Gerade die Helligkeit und Steilheit eines Bildes scheinen hierfür verantwortlich zu sein. Hinzu kommen die Probleme hinsichtlich der Artefakte.

3.2.5 Zusammenfassung der Broadcast-Anforderungen

Durch eine stärkere Berücksichtigung der ästhetischen Vorprägung, die dafür Sorge trägt, dass der Kontrastumfang nur gesteigert wird, wenn eben jene Aspekte davon nicht betroffen sind, ist ein deutliches Steigerungspotenzial gegeben. So kommen Mantiuk et al. [82] bei ihrem *display adaptiv TMO*⁶ zu dem Schluss:

„Many recent studies on tone-mapping undertake the difficult task of producing images that will be subjectively preferred. We avoid this challenging problem from the area of computational aesthetics by hiding it in the image enhancement step. Although many attempts to model subjective preference or quality of images have been made [Keelan 2002], there exist no reliable models that would measure subjective image preference. For this reason we do not claim an operator that produces the best looking images (although in our opinion in many cases it does), but rather the operator that objectively solves the problem of reproducing large dynamic range on a displays of low contrast with the least visible contrast distortions.“

Das Ziel der nun folgenden Arbeit ist es deshalb, zunächst ein Modell zu entwickeln, das die sich verändernden visuellen Präferenzen bezüglich der Darstellung bei unterschiedlichen Leuchtdichtewerten des Displays berücksichtigt. Auch soll in diesem Zuge der Einfluss der Umgebungsleuchtdichte mit untersucht werden. Dazu werden in Kapitel 4 weitere Studien angeführt und Vortests durchgeführt, um die entscheidenden Parameter besser definieren zu können. Hierauf aufbauend wird ein so genanntes Probanden-Grading durchgeführt, mit dessen Hilfe sich quantitative Aussagen über die Parameter treffen lassen. In Kapitel 5 liegt dann der Fokus auf der Sättigung und der genaueren Betrachtung der Probleme bezüglich der Farberscheinung bei TMOs. Zusammengenommen sollen die Erkenntnisse anschließend in Kapitel 6 genutzt werden, um TMOs hinsichtlich der Broadcast-Anforderungen zu modifizieren. So kann es gelingen, den Zuschauer dort abzuholen, wo er sich aktuell befindet, und auf diesem Wege eine Akzeptanz für die neue Technik zu schaffen.

Zusammenfassend kann festgehalten werden, dass in den vorangegangenen Kapiteln folgende Anforderungen an ein zukünftiges HDR-System ermittelt werden konnten, für welche im weiteren Verlauf der Arbeit Lösungen gefunden werden sollen:

⁶Dieser TMO schnitt bei Eilersten am Besten ab.

- Berücksichtigung der Zuschauerpräferenz bezüglich der Kontrastkompression von Leuchtdichtewerten (nichtlineare Reproduktion von Leuchtdichtewerten)
- Berücksichtigung der Zuschauerpräferenz bezüglich der Farberscheinung (Sättigung) bei einer Kontrastkompression
- Algorithmus basierend auf den Erkenntnissen der ersten beiden Punkte zur Modifikation bestehender TMOs, welcher ferner folgende Rahmenbedingungen erfüllt:
 - Vermeidung von Artefakten (Halo, Flackern, Ghosting)
 - Echtzeitfähigkeit (<16,6 ms bei 4k)
 - Parameter-Smoothing für lokales Tonemapping mit Szenenwechselerkennung
- Erarbeitung eines zukunftsfähigen Workflows unter Berücksichtigung existierender Infrastrukturen, **EOTEs** und Farbräume

Kapitel 4

Visuelle Präferenz bezüglich der Helligkeitswahrnehmung

Wie im vorangegangenen Kapitel beschrieben soll die Präferenz des Zuschauers bezüglich einer Kontrastkompression quantifiziert werden. Dafür ist es zunächst nötig, die unterschiedlichen Parameter für das Gelingen einer Modifikation der Helligkeitskomponente zu identifizieren. Gleichzeitig ist zu beachten, ob diese ggf. mit äußeren Gegebenheiten korrelieren.

4.1 Genauere Untersuchung zur simultanen Kontrastwahrnehmung

Zunächst soll allerdings die simultane Kontrastwahrnehmung des Menschen näher betrachtet werden. Die Intention ist, dass die hierdurch gewonnen Erkenntnissen bei der Konzeption der Probandentests sowie bei der Erklärung der etwaigen Beobachtungen genutzt werden können. Das Wissen aus den Abschnitten [2.2.2](#) und [2.2.3](#) ist hierfür nicht ausreichend und substanziell genug. Im nun folgenden Abschnitt wird deshalb der Stand der Forschung kritisch betrachtet und um ein eigenes Experiment erweitert.

4.1.1 Kritische Betrachtungen zum Stand der Forschung

Wie bereits im Abschnitt [2.3.2](#) erwähnt begann zu Beginn des neuen Jahrtausends mit den Prototypen von Seetzen [\[112\]](#) ein neues Zeitalter in der Displaytechnologie. Aus diesem Grund versuchten Kunkel et al. 2010 [\[64\]](#) den simultanen Kontrastumfang des [HVS](#) genauer zu bestimmen, um so eine Art Guideline für zukünftige Displaydesigns

zu schaffen. In [2.2.3](#) wurde hierzu bereits erläutert, dass das menschliche Wahrnehmungssystem zwar einen viel größeren Kontrastumfang verarbeiten kann, dies aber nur aufgrund von Adaptionsprozessen möglich ist, die Zeit benötigen, welche beim Fernsehen so nicht zur Verfügung steht. Entsprechend war es nötig, eben jenen simultanen Kontrastumfang genauer zu quantifizieren. Die zu diesem Zeitpunkt aus der Literatur zu entnehmenden Werte wiesen eine extreme Spannweite von 2 bis hin zu 3,5 Dekaden auf, was einer Differenz von bis zu 1500% entspricht. Laut Kunkel et al. kann dies auf die Tatsache zurückgeführt werden, dass die Resultate nicht aus direkten Messungen, sondern auf Kalkulationen der Reaktionskurven der Photorezeptoren basieren. Aufgrund der Komplexität des [HVS](#) sollte der Prozess des Sehens aber eher als eine Einheit gesehen werden, wodurch die losgelöste Betrachtung einer Komponente, in diesem Fall der Photorezeptoren, nicht ausreichend ist.

Deshalb entwickelten Kunkel et al. ein psychophysikalisches Experiment unter Zuhilfenahme eines Gabor-Rasters, das eine kleine Modulation der Leuchtdichte (Stimulus) auf einem mit Rauschen belegten Hintergrund zeigte. Die Idee dahinter war, dass gleiche prozentuale Änderungen in der Leuchtdichte unterschiedlich großen Zuwachs in der Reaktion des [HVS](#) auslösen. Weist der Stimulus die gleiche Leuchtdichte wie der Hintergrund auf, fällt der Zuwachs am größten aus und nimmt mit steigender Differenz ab. Entsprechend wurde die Leuchtdichte des Stimulus so lange erhöht und anschließend herabgesetzt, bis der Zuwachs unter die visuelle Wahrnehmungsgrenze fiel. Der Bereich zwischen den beiden Werten entspricht dem simultanen Kontrastumfang und wurde auf 3,7 Dekaden taxiert.

Allerdings räumen Kunkel et al. ein, dass mehrere Freiheitsgrade im Testdesign das Resultat beeinflussen können. Wird ein stärker modulierter Stimulus für einen längeren Zeitraum bei einer größeren Hintergrundleuchtdichte gezeigt, werden Werte über 4 Dekaden, also über 10.000:1 erreicht. Diese zweite Zahl wurde in den letzten Jahren oft zitiert und als Argument für bestimmte Sachverhalte angeführt, ohne deren Entstehung genau zu kennen. Denn das Experiment weist noch eine weitere Limitation auf.

In der Masterarbeit von Bychkov [\[17\]](#) [\[75\]](#) konnte gezeigt werden, dass der simultane Kontrastumfang unter realen Bedingungen deutlich reduziert sein kann. Bei Kunkel et al. wurde die Messung bezüglich des hellen und des dunklen Wahrnehmungsgrenzwertes nacheinander durchgeführt. Eine natürliche Vorlage, weist allerdings je nach Bildinhalt, im Gegensatz zum verwendeten künstlichen Testbild, Lichter und Schatten zur selben Zeit auf. Man könnte dies also mit der Messung des Absolutkontrastes (*On-Off-Contrast*) anstelle des Simultankontrastes bei Displays und Projektoren vergleichen. Aufgrund der von Bychkov durchgeführten Voruntersuchungen ist zu vermuten, dass helle Bildbereiche – gerade in [HDR](#) – die Sensitivität des [HVS](#) bezüglich leichter Abstufungen in den

angrenzenden dunklen Bildbereichen herabsetzen. Ähnliche Maskierungseffekte (auch Verdeckung genannt) werden beispielsweise bei der Audiocodierung genutzt um die Datenrate zu reduzieren, da leise Töne nicht gehört werden können, wenn diese sich zeitlich und von der Frequenz zu dicht an laute Töne befinden.

Bei Bychkov lag der Kontrastumfang teilweise unter 300:1, was einen enormen Unterschied zu den idealen Bedingungen bei Kunkel et al. bedeutet. Allerdings war der Test auf eine maximale Leuchtdichte von 100 cd/m^2 limitiert und berücksichtigte nur Situationen, bei denen Highlight und Schatten direkt aneinander angrenzten. Entsprechend sind die quantitativen Aussagen für diese Arbeit nicht ausreichend. Im Folgenden soll nun ein Probandentest entworfen werden, der erstmals den simultanen Kontrastumfang unter realen Sehbedingungen in Abhängigkeit von mehreren Parametern abbilden kann.

Dass der Test für verschiedene Umgebungsleuchtdichten wiederholt werden sollte, zeigen die Ergebnisse von Kusakabe et al. [65]. Diese untersuchten, inwieweit der kleinste noch wahrnehmbare Schwarzwert mit der Umgebungsleuchtdichte zusammenhängt. Hierfür kamen zwei verschiedene Testcharts zum Einsatz. Zum einen wurde ein Bild genutzt, welches nur sehr dunkle Grauabstufungen enthielt. Hier stieg der Schwarzwert nahezu proportional mit der Umgebungsleuchtdichte an. Andererseits wurde ein Testchart genutzt, das auch ein helles Quadrat in der Mitte des Bildes aufwies. Hier blieb der Schwarzwert bei zunehmendem Umgebungslicht beim hellsten Display (4000 cd/m^2) unverändert, während er beim dunkelsten Display (100 cd/m^2) weiterhin leicht anstieg.

4.1.2 Testdesign und -durchführung

Um den simultanen Kontrastumfang in Abhängigkeit mehrerer Parameter bestimmen zu können, wurde ein Probandentest entworfen, dessen Schwerpunkt auf der Messung des Effekts der Blendung auf die Wahrnehmbarkeit der dunklen Bildbereiche lag. Ursprünglich war geplant sowohl die Limitationen in den Lichtern, als auch in den Schatten zu messen. In den Vortests hatte sich allerdings gezeigt, dass es für die Probanden nie ein Problem darstellte, zwischen den beiden hellsten darstellbaren Patches zu unterscheiden. Grund hierfür war die Limitation der maximalen Leuchtdichte des LCD-LED-Displays auf gut 1.000 cd/m^2 .

Die Versuche wurden im Rahmen der Bachelorarbeiten von Pascal Kutschbach [66] und Mark Benyamin [9] durchgeführt. Die Ergebnisse wurden auf der IBC-Konferenz [77] präsentiert und finden sich im **SMPTE** Motion Imaging Journal [70] wieder.

Um die Wahrnehmbarkeit von dunklen Bildelementen zu messen, wurde ein synthetisches Testbildset erzeugt. In der Mitte jedes Bildes wurde eine Grautreppe platziert,

auf welche der Proband sich fokussieren sollte. Die Abstufungen in der Leuchtdichte wurden möglichst klein und wahrnehmungsgleichabständig gewählt. Die korrespondierenden Werte sind in Tabelle 4.1 gegeben. Buchstaben in absolutem Schwarz wurden in die einzelnen Stufen der Grautreppe platziert, die im Folgenden Patches genannt werden. Die Probanden sollten den Buchstaben nennen, welchen sie gerade noch identifizieren konnten.

Patch	Leuchtdichte	Patch	Leuchtdichte	Patch	Leuchtdichte
1	0,38 cd/m ²	2	0,32 cd/m ²	3	0,28 cd/m ²
4	0,25 cd/m ²	5	0,21 cd/m ²	6	0,18 cd/m ²
7	0,16 cd/m ²	8	0,14 cd/m ²	9	0,13 cd/m ²
10	0,12 cd/m ²	11	0,11 cd/m ²	12	0,10 cd/m ²
13	0,09 cd/m ²	14	0,08 cd/m ²	15	0,07 cd/m ²
16	0,06 cd/m ²				

TABELLE 4.1: Übersicht über die Leuchtdichtewerte der Grautreppe

Auf beiden Seiten der Grautreppe wurde ein Highlight in Form eines weißen Balkens platziert. Anschließend wurden 36 (3x4x3) verschiedene Testbilder aus diesem Template erstellt, indem die Freiheitsgrade „Distanz zwischen Highlight und Grautreppe“, „Größe des Highlights“ und „Leuchtdichte des Highlights“ variiert wurden, wie es in Tabelle 4.2 gezeigt wird. Im Folgenden wird die Distanz als Schwinkel in Grad angegeben, wodurch der Parameter unabhängig von der Displaygröße wird. Dazu wird der Strecke zwischen den beiden Außenseiten der Highlights herangezogen.

Parameter	Anzahl an Varianten	Einheit	Namen der Varianten			
Entfernung zum Highlight	3	Schwinkel in °	8 - 28 (close)	34 (medium)	56 (far)	
Größe des Highlights	4	% des Gesamtbildes	2,5	10	20	30
Leuchtdichte des Highlight	3	cd/m ²	1.000	300	100	

TABELLE 4.2: Übersicht über die Variationen der drei Testparameter

Ein optischer Eindruck der verschiedenen Testbilder für eine feste Leuchtdichte ist in Abbildung 4.1 gezeigt. Es ist zu berücksichtigen, dass die Leuchtdichtewerte der Grautreppe modifiziert wurden, sodass sie in der gedruckten Version sichtbar sind.

Damit sich die Probanden die Reihenfolge der Buchstaben auf der Grautreppe nicht merken konnten, wurden diese zufällig in den einzelnen Testbildern vertauscht. Des Weiteren wurden die Probanden dazu aufgerufen, eine intuitive Antwort innerhalb weniger Sekunden zu geben, um zu starke Adaptionsprozesse zu vermeiden. Der Test wurde unter drei

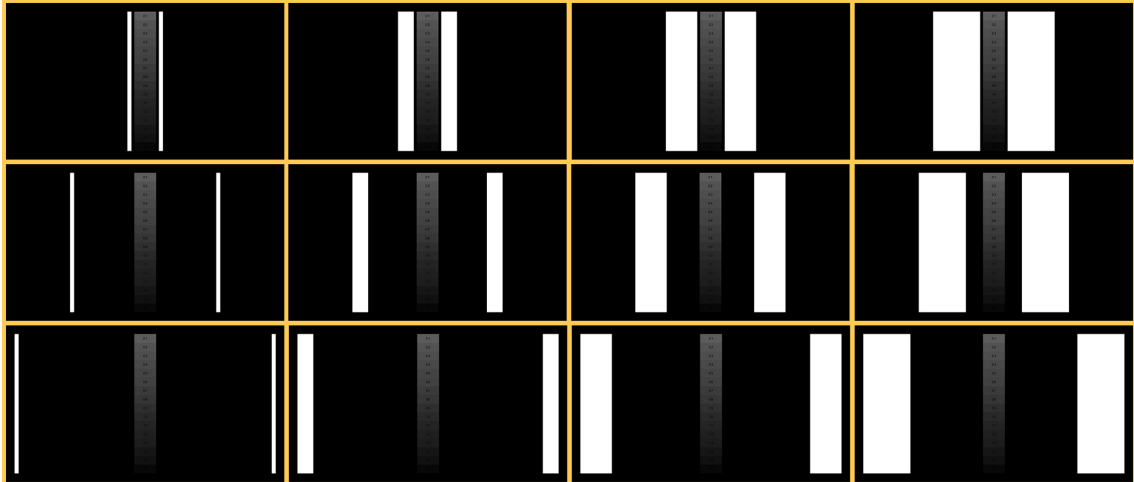


ABBILDUNG 4.1: Testbilder zur Evaluierung der menschlichen simultanen Kontrastwahrnehmung

verschiedenen Umgebungsleuchtdichten (10 lx, 50 lx, 100 lx)¹ wiederholt. Insgesamt nahmen 41 Probanden am Test teil. 66% von ihnen waren männlich und 34% weiblich. Das durchschnittliche Alter lag bei 31 Jahren.

Es muss festgehalten werden, dass das LED-Backlight nicht ganz ideal für diesen Test war. Im Fall, dass sich das Highlight sehr nah an der Grautreppe befand, konnte Übersprechen zu einer ungewollten Aufhellung führen. Im Fall, dass sich das Highlight am Rand des Displays befand, wurde es schwierig die 1.000 cd/m^2 zu erreichen. Durch Optimierungen der Testbilder konnten die Effekte stark reduziert werden, indem beispielsweise die Highlightgrenzen genau auf die Segmentgrenzen der Hintergrundbeleuchtung gelegt wurden. In der nun folgenden Auswertung wird deshalb auch immer mit den nominalen Werten gerechnet.

4.1.3 Resultate

In Abbildung 4.2 sind die Resultate in verschiedenen Repräsentationsformen dargestellt, um bestimmte Aspekte zu verdeutlichen. Zunächst sei das Diagramm in der linken oberen Ecke (a) genauer betrachtet. Hier ist der Sehwinkel auf der Abszissenachse abgebildet, während die Ordinatenachse die Anzahl an wahrgenommenen Grauwerten repräsentiert. Die Farbe jeder Kurve gibt die Leuchtdichte und das Symbol die Größe des Highlights an. Alle Kurven zeigen dieselbe monoton steigende Tendenz, denen zufolge

¹Mit der Einheit Lux (lx) wird die Beleuchtungsstärke und mit der Einheit cd/m^2 die Leuchtdichte gemessen. Erstes ist somit eine Empfängergröße, während zweites eine Sendergröße ist. Da der Raum sowohl Licht von Lampen empfängt, als auch Licht auf das Display reflektiert, finden sich in der Literatur verschiedene Angaben. Für die folgende Arbeit kann vereinfacht davon ausgegangen werden, dass 3,5 lx zu einer Aufhellung der Displayoberfläche von $1,0 \text{ cd/m}^2$ führt.

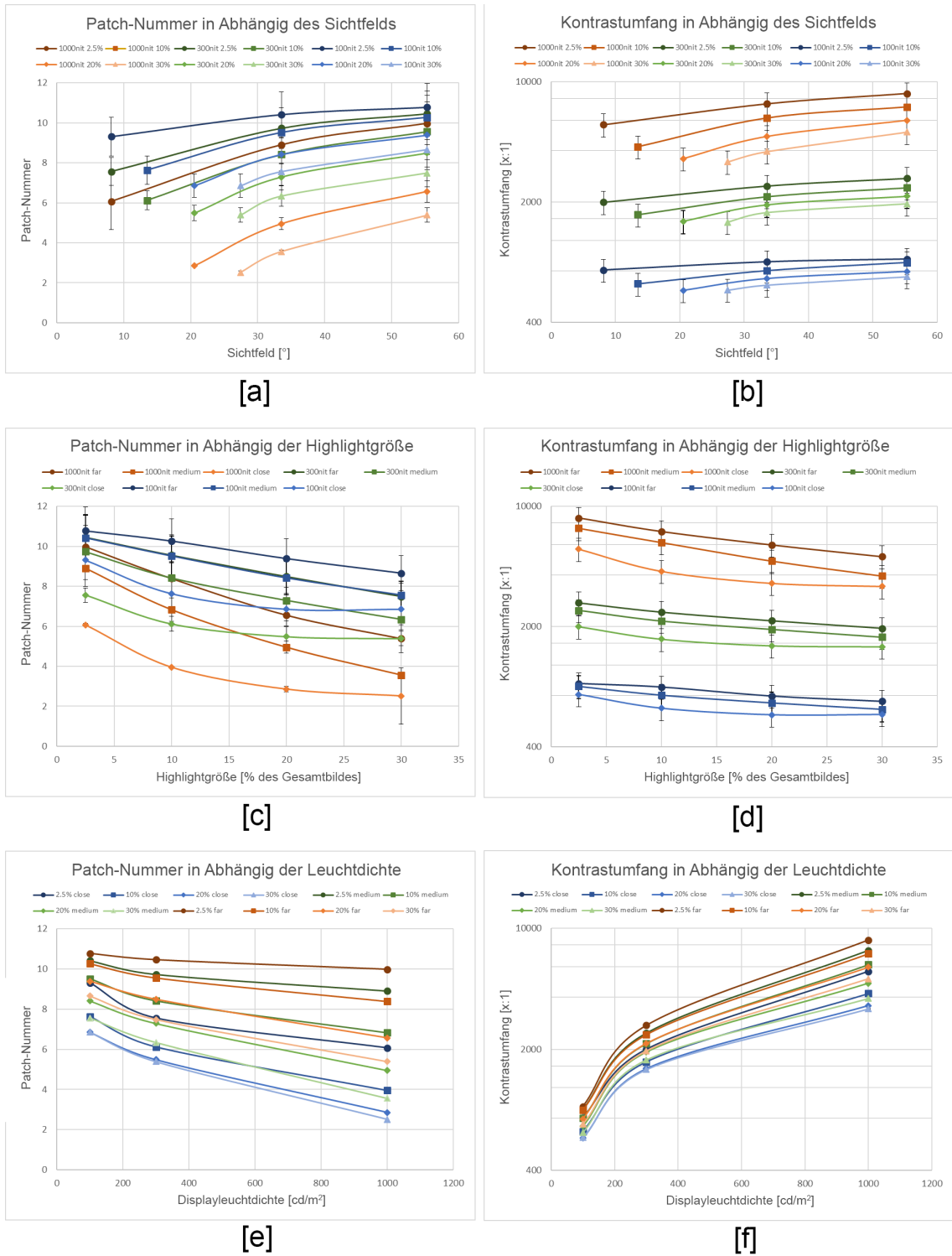


ABBILDUNG 4.2: Ergebnisse des Probandentests bezüglich des simultanen Kontrastumfangs. Auf der linken Seite ist die Anzahl an Patches aufgetragen, welche unterschieden werden konnte. Auf der rechten Seite findet sich der daraus resultierende Kontrastumfang.

eine größere Distanz zwischen Grautreppe und Highlight die Wahrnehmbarkeit der Patches verbessert. Des Weiteren zeigt sich auch, dass dieser Effekt für höhere Leuchtdichtewerte stärker ausgeprägt ist und auch bei der maximalen Entwertung wahrscheinlich noch nicht die Sättigung erreicht.

Die 95%-Vertrauensintervalle zeigen an, dass die gemessenen Differenzen für einige Testbilder nicht signifikant sind.² Bei genauere Analyse konnte festgestellt werden, dass die Anzahl an wahrgenommenen Grauwerten mit dem Alter der Probanden korrelierte. Älteren Teilnehmern fiel es schwerer zwischen den einzelnen Patches zu unterscheiden. Entsprechend bedeutet die fehlende Signifikanz nicht, dass die Verdeckung für den einzelnen Teilnehmer irrelevant wäre. Die Kurvenverläufe waren für jeden Probanden sehr ähnlich, allerdings war das Grundniveau zwischen den Probanden leicht verschieden.

In Diagramm (c) ist die Anzahl an Grauwerten gegen die Größe des Highlights aufgetragen. Wie bereits vermutet, wird eine Unterscheidung der Patches bei einer Steigerung dieses Parameters schwieriger. Dieser Effekt ist für den **SDR**-Bereich weniger von Bedeutung als für den **HDR**-Bereich. Bei 100 cd/m^2 schwankt der Wert für die Schwarzwahrnehmung bei einer mittleren Entfernung zwischen $0,115$ und $0,155 \text{ cd/m}^2$. Für den selben Fall bei 1.000 cd/m^2 bewegen sich die Werte zwischen $0,135$ und $0,265 \text{ cd/m}^2$.

Diagramm (e) verifiziert noch einmal die obigen Beobachtungen bezüglich des Einflusses der Leuchtdichte des Highlights indem diese nun auf der Abszissenachse dargestellt wird. Die Farbe der Kurve zeigt den Sehwinkel. Es ist offensichtlich, dass sich die Kurven mit zunehmender Leuchtdichte auseinander bewegen. Bei 100 cd/m^2 schwankt der Wert zwischen $0,11$ und $0,165 \text{ cd/m}^2$ wohingegen er sich bei 1.000 cd/m^2 zwischen $0,12$ und $0,30 \text{ cd/m}^2$ bewegt. Der Maskierungseffekt ist im zweiten Fall deutlicher ausgeprägt. Es kann somit festgehalten werden, dass bei **HDR**-Displays der wahrnehmbare Kontrastumfang deutlich stärker vom Bildinhalt abhängt und entsprechend höchst variabel ist.

Ein weiterer interessanter Aspekt kann bei den Diagrammen auf der rechten Seite von Abbildung 4.2 ((b), (d) und (f)) beobachtet werden. Verglichen mit (a), (c) und (e) zeigt die Abszissenachse den Kontrastumfang in logarithmischer Skalierung und nicht die Anzahl an Grauwerten. Dazu wurde die Patch-Nummer in die korrespondierende Leuchtdichte umgerechnet und mit der Leuchtdichte des Highlights zusammengeführt. Dadurch verändert sich das Erscheinungsbild der Kurven deutlich. Auch wenn die hohen Leuchtdichten dazu führen, dass die Sensitivität der Probanden bezüglich des Schwarzwertes

²Da sich bei einer Anzahl von 41 Messwerten die Kurve der t-Verteilung stark der Kurve der z-Verteilung (Gleichverteilung) annähert, wurde bei der Berechnung der Standardabweichung und des Konfidenzintervalls von einer Normalverteilung ausgegangen. Allerdings muss berücksichtigt werden, dass der Grenzwert von 0,21 beim Kolmogorov-Smirnov-Tests bei α gleich 0,05 in zwei Fällen überschritten wurde (0,26 und 0,28). Das ist darauf zurückzuführen, dass nur eine sehr geringe Anzahl an diskreten Antwortmöglichkeiten bestand und es in den beiden Fällen zu einer starken Ballung auf einen einzelnen Wert kam.

herabgesetzt wird, steigern sie den Weißwert deutlich stärker, wodurch ein insgesamt vergrößerter Kontrastumfang resultiert. Aus der Perspektive eines Displayherstellers kann also klar gesagt werden, dass es zielführender ist, die maximale Leuchtdichte zu steigern, als die minimale Leuchtdichte zu senken.

Des Weiteren kann aufgrund des Kurvenverlaufs angenommen werden, dass dieser Effekt bei 1.000 cd/m^2 noch nicht in der Sättigung angelangt ist. Höhere Werte für den Kontrastumfang wären entsprechend bei 2.000 cd/m^2 oder gar 4.000 cd/m^2 denkbar. Im vorgestellten Experiment wurde im besten Fall (kleines, weit entferntes 1.000 cd/m^2 - Highlight) ein Kontrastumfang von ca. 8.500:1 gemessen. Entsprechend wird vermutet, dass bei höheren Leuchtdichten Werte über 10.000:1 möglich sind. Diese Resultate passen sehr gut zu denen von Kunkel et al..

Um für einen beliebigen Fall den Kontrastumfang schätzen zu können, wurde, basierend auf dem psychophysikalischen Experiment, eine Näherungsformel ermittelt. Diese ist in Gleichung 4.1 gegeben. L_{max} ist dabei die Leuchtdichte des Highlights in cd/m^2 , α der Sehwinkel in Grad und s die Größe des Highlights in % der Bildschirmfläche. Folgende Werte wurden iterativ für die drei Konstanten k ermittelt: $k_1 = 15,7$, $k_2 = 0,22$ und $k_3 = 0,18$. Ein Vergleich der gemessenen Werte und der Näherungsformel ist in Abbildung 4.3 gezeigt. Eine hohe Übereinstimmung ist sichtbar. Die durchschnittliche Abweichung beträgt 5,7%.

$$\text{Kontrastumfang} = L_{max} * k_1 * \left(\frac{\left(\frac{\alpha}{100} \right) k_2}{(1 + s) k_3} \right)^{\log_{10}(L_{max})} \tag{4.1}$$

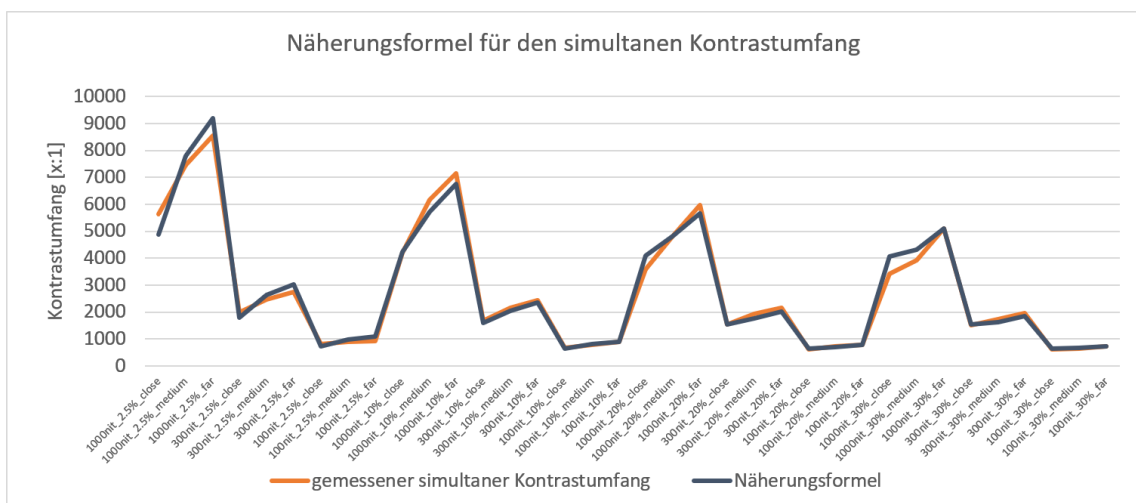


ABBILDUNG 4.3: Vergleich zwischen den Messergebnissen und der Näherungsformel bezüglich des wahrnehmbaren Kontrastumfangs.

Die bis zu diesem Zeitpunkt vorgestellten Ergebnisse beziehen sich auf eine Umgebungsleuchtdichte von 10 lx. Für diesen Fall lag der Kontrastumfang, je nach Testbild, zwischen 616:1 bis 8561:1. Für 50 lx bzw. 100 lx war der Bereich auf 570:1 bis 7245:1 bzw. 557:1 bis 6534:1 reduziert. Somit kommt es unter realen Sehbedingungen zu zwei verschiedenen Maskierungseffekten. Der Erste, welcher vom Bildinhalt abhängig ist, und der Zweite, welcher sich auf die Umgebungsleuchtdichte bezieht. Der zweite Effekt war besonders immer dann zu beobachten, wenn der erste eher gering ausfiel. War also die Sensitivität bereits durch ein nahes, helles und großes Highlight stark herabgesetzt, hatte die Umgebung nur noch einen sehr geringen bis keinen Effekt auf den Kontrastumfang. War der erste Maskierungseffekt aber gering, gewann die Umgebung deutlich an Bedeutung.

Der durchgeführte Probandentest hat gezeigt, dass der simultan wahrnehmbare Kontrastumfang signifikant durch Blendungseffekte beeinflusst wird. Highlights können, je nach Größe und Entfernung, die Sensitivität des **HVS** bezüglich kleiner Unterschiede in den dunklen Bildbereichen deutlich herabsetzen. Entsprechend sind kleinere Displays stärker betroffen. Auch besteht eine Korrelation mit der Leuchtdichte, welche die Tatsache besonders relevant für **HDR** macht. Das bedeutet, dass es durch die neuen technologischen Entwicklungen zwar möglich ist, ein deutlich größeres Spektrum an Leuchtdichtewerten einzufangen und auch wiederzugeben, der Zuschauer aber aufgrund der Limitation der menschlichen Wahrnehmung nicht zwangsläufig davon profitieren kann.

4.1.4 Zwischenfazit bezüglich des simultanen Kontrastumfangs

Bisher war bekannt, dass der wahrnehmbare simultane Kontrastumfang des Menschen deutlich beschränkter ist, als jener, der die Adaptionsmöglichkeiten miteinbezieht. Die Untersuchungen [\[17\]](#) und [\[65\]](#) legten darüber hinaus nahe, dass der reale simultane Kontrastumfang in natürlichen Vorlagen durch Maskierungseffekte noch weitaus geringer ausfallen kann. Allerdings waren diese unzureichend quantifiziert. Die dargelegten neuen Messergebnisse stellen die Sinnhaftigkeit eines fundamentalen Konzepts der Fernsehtechnik in Frage. Bis heute werden die Relationen der Leuchtdichtewerte bei der Reproduktion im Allgemeinen unverändert belassen.³ Konnte der geringe Kontrastumfang bei **SDR** noch ziemlich gut vom simultanen Kontrastumfang des **HVS** abgedeckt werden, so hat sich gezeigt, dass dies bei **HDR** nicht mehr der Fall ist. Denn nicht einzig die Betrachtungsdauer, sondern auch das schmalere Sichtfeld stellt eine Beschränkung der Adaptionsfähigkeiten des Menschen im Vergleich zu einer realen Szene dar. Damit trotzdem genauso viele Differenzierungen wie in der natürlichen Vorlage wahrgenommen werden können, müsste diese Limitation mithilfe von technischen Mitteln ausgeglichen

³Natürlich führt auch z. B. das Over-All-Gamma zu einer leichten Abweichung der linearen Verhältnisse. Allerdings entspricht das längst nicht der hier angedachten Größenordnung.

werden, um Maskierungseffekten vorzubeugen. Dies ist bisher auch für **HDR** nicht in Aussicht.

Die zu Beginn ausgegebene Richtlinie **HDR**-Material so anzupassen, dass es auf einem **SDR**-Display wiedergeben werden kann, scheint somit nicht allein auf diesen Anwendungsfall beschränkt zu sein. Auch eine Anpassung von **HDR**-Material für **HDR**-Displays könnte unter dem Gesichtspunkt der Sichtbarkeit von dunklen Bildinhalten sinnvoll sein, da es sonst zum Verlust von Informationen kommen kann. Die Helligkeit der Schatten müsste angehoben werden, wohingegen die Lichter abgesenkt werden müssten. Eine Angleichung verschiedener Bildregionen und somit eine Homogenisierung des Eindrucks würde automatisch auch zu einer Kontrastkompression führen, wie sie in den vorherigen Abschnitten gefordert wurde.

4.2 Stand der Forschung bezüglich der Präferenz

Im Folgenden sollen einige ausgewählte, für diese Arbeit relevante, Studien vorgestellt werden, um aus deren Erkenntnissen ein Testdesign für die quantifizierenden Untersuchungen zu entwickeln. Vorrangig beschäftigen sich die Untersuchungen mit der präferierten Helligkeit bei unterschiedlichen Display- und Umgebungsleuchtdichten.

Bereits 1969 hat Novick [91] gezeigt, dass bei steigender Umgebungsleuchtdichte das vom Zuschauer präferierte Gamma merkbar abnimmt. Diese Beobachtung war auch unabhängig davon, ob ein farbiges oder ein monochrom grünes Bild gezeigt wurde. Auf der Basis dieser Resultate schätzte De Marsh [21] den idealen Gamma-Wert für verschiedene Situationen: 1,0 für eine helle Umgebung (69 cd/m^2), 1,2 für eine gedimmte Umgebung (14 cd/m^2) und 1,5 für eine komplett dunkle Umgebung. Noch heute trägt das Systemgamma⁴ von 1,2 beim Broadcast diesen Angaben Rechnung.

Die **BBC** untersuchte im Rahmen ihres Vorschlages für ein Display unabhängiges **HDR**-TV-System [13], der die Basis für **HLG** bildete, inwieweit sich dieser Effekt bei **HDR**-Displays reproduzieren lässt. Als Ergebnis steht eine Funktion zur Berechnung einer **OOTE** in Abhängigkeit der Display- und der Umgebungsleuchtdichte.

$$\gamma = 1 + \frac{1}{5} \log_{10} \left(\frac{L_{dmax}}{L_{surround}} \right) \quad (4.2)$$

⁴Das Systemgamma wird auch Over-All-Gamma oder **OOTE** genannt. Dieses ist immer dann ungleich 1,0, wenn die **EOTE** keine hundertprozentige Inverse der **OETF** ist.

In einem gedimmten Raum und mit einem $1000\text{cd}/\text{m}^2$ -Display wäre somit ein Gamma von 1,4 ideal, also 0,2 Punkte größer als heute. Dieser vergleichsweise klein wirkende Unterschied hat durchaus deutliche Auswirkungen auf den Bildeindruck.

Indirekt setzt die Formel voraus, dass zwischen den beiden Parametern keine Korrelation besteht. Dabei kam Rempel et al. [105], dem als erster für eine derartige Untersuchung ein HDR-Display-Prototyp mit $4000\text{cd}/\text{m}^2$ zur Verfügung stand, bereits 2009 zu dem Schluss, dass dies nicht zwangsläufig gegeben sein muss. In seinem Versuch sollten die Probanden mit Hilfe einer Fernbedienung das Display immer wieder auf die für sie subjektiv beste Spitzenleuchtdichte einstellen. Dazu wurden ihnen fünf Spielfilme an fünf aufeinanderfolgenden Tagen gezeigt. Da die Spielfilme allerdings nur in SDR vorlagen, wurden die Helligkeitswerte linear hochgerechnet. Die Beleuchtungsstärke der Umgebung wurde im Versuch in den drei Stufen 7 lx, 70 lx und 700 lx variiert. Zwischen 7 und 70 lx konnte kein Unterschied beobachtet werden. Der eingestellte Median der Leuchtdichte des Bildes lag in beiden Fällen bei etwa $37\text{cd}/\text{m}^2$. Von 70 auf 700 lx kam es näherungsweise zu einer Verdopplung.

Auch die 2015 von der NHK auf der International Broadcasting Conference (IBC) vorgestellten Untersuchungen [65] bezüglich dieser Problematik, deuten darauf hin, dass die beiden Faktoren sich für extreme Situationen gegenseitig bedingen. Im vorliegenden Beispiel bekamen 120 Probanden zehn Testbilder mit unterschiedlicher durchschnittlicher Leuchtdichte bei fünf verschiedenen Display- und vier verschiedenen Umgebungsleuchtdichten gezeigt. Sie mussten entscheiden, ob sie das gezeigte Motiv für *zu hell* (+1 Score), *passend* (0 Score) oder *zu dunkel* (-1 Score) empfanden. Je heller das Display und je dunkler der Raum, desto eher wurde das Motiv als *zu hell* (>0 Score) eingestuft. Die Verdopplung der Displayleuchtdichte ließ den Score durchschnittlich um 0,09 steigen. Bei der geringsten Umgebungsleuchtdichte ($0,01\text{cd}/\text{m}^2$) betrug diese Steigerung aber eher 0,06, wohingegen sie für 1 und $10\text{cd}/\text{m}^2$ bei 0,11 lag. Diese Ergebnisse passen zu den in Kapitel 4.1 gewonnenen Erkenntnissen bezüglich des Maskierungseffektes. Auch hier war es so, dass, wenn die Leuchtdichtewerte für Display und Umgebung sehr weit auseinander lagen, der geringere Maskierungseffekt nahezu vernachlässigbar klein wurde.

Neben der Problematik bezüglich der Korrelation stellt sich die Frage, ob eine Gammafunktion die Bildaussteuerung unter verschiedenen Bedingungen am besten beschreibt, da die Gammafunktion die Mitten überproportional stark abdunkelt. Dazu ließ die NHK [65] einen Coloristen ein Grading bei drei verschiedenen Displayleuchtdichten durchführen. Es wurde untersucht, wie sich der Anteil der Partien im Bild verändert, die mehr als $100\text{cd}/\text{m}^2$ Leuchtdichte aufweisen (hier als Highlight bezeichnet), wenn die Displayleuchtdichte variiert wird. Bestand das Bild bei $100\text{cd}/\text{m}^2$ noch zu 9% aus Highlights, waren es bei $500\text{cd}/\text{m}^2$ bereits 23% und bei $3000\text{cd}/\text{m}^2$ 40%. Die Aussagekraft

dieser Werte ist allerdings sehr begrenzt, da nur ein Colorist nach seiner persönlichen Einschätzung zwei Testbilder bearbeitet hat. Einzig eine Tendenz ist zu erkennen.

Die für die geplante Untersuchung relevanteste Studie ist von Stokkermans et al. [121]. 20 Probanden wurden gebeten, den Key-Value⁵ für drei verschiedene Testbilder einzustellen. Dabei wurde die Displayleuchtdichte von 100 cd/m^2 auf 550 cd/m^2 variiert und die Umgebungsleuchtdichte von 20 cd/m^2 auf 275 cd/m^2 . Für beide Freiheitsgrade konnte eine Korrelation zum eingestellten Key-Value nachgewiesen werden. Während ein helles Display zu einem geringeren Key-Value führte, stieg dieser mit einer heller werdenden Umgebung an. Diese Effekte konnten unabhängig von der Einstellung des jeweils anderen Freiheitsgrades beobachtet werden.

Neben der Umgebungs- und der Displayleuchtdichte stellt sich die Frage, ob auch die Displaygröße einen Einfluss auf die visuelle Präferenz ausübt. Dazu soll zunächst vereinfacht angenommen werden, dass der Sehwinkel sich mit der Displaygröße nicht ändert, sprich sich der Betrachtungsabstand proportional zur Displaygröße verhält.

Daly et al. [20] versuchten 2013 für Dolby den idealen Kontrastumfang eines HDR-Displays zu quantifizieren, um 90% der Betrachter zufrieden zu stellen. Dieser Test wurde 2015 [35] mit einem Projektor und einer 4 m großen Leinwand wiederholt. Die Resultate wiesen eine hohe Übereinstimmung auf. So lag der Wert für das diffuse Weiß in beiden Fällen bei etwa 3000 cd/m^2 . Allerdings muss dabei berücksichtigt werden, dass es dem Projektor nur möglich war eine Spitzenleuchtdichte von 2100 cd/m^2 zu erzeugen. Um noch hellere Bilder zu simulieren, die Leuchtdichten über 2100 cd/m^2 möglichst nahe kommen, wurde Clipping zugelassen. So konnten die erwähnten 3000 cd/m^2 gemessen werden. Indirekt wurde somit auch eine Toleranz für Clipping nachgewiesen. Die Akzeptanz hierfür führten die Autoren darauf zurück, dass die Kante eines Spitzlichtes deutlich später zu clippen anfängt als das Spitzlicht selber. In der Folge kann die Leuchtdichte weiter erhöht werden, um so die Differenz von Spitzlicht und Umgebung weiter zu vergrößern. Die Resultate bezüglich des Schwarzwerts waren im Vergleich zu denen des diffusen Weiß leicht unterschiedlich. So sank der gemessene Wert von $0,005 \text{ cd/m}^2$ in der ersten Studie auf $0,002 \text{ cd/m}^2$ in der zweiten Studie. In Anbetracht der Abweichung der Displaygrößen um den Faktor 13 (0,3 m auf 4 m), liegen die Ergebnisse aber nah beieinander. Im Bereich der TV-Geräten ist mit einer weitaus geringeren Varianz bei den Displaygrößen zu rechnen.

Alle bisher vorgestellten Studien evaluieren wie hell eine Szene letztendlich auf dem Zieldisplay dargestellt werden soll. Dabei wird aber nicht drauf eingegangen, wie die linearen Verhältnisse des Eingangssignals modifiziert werden sollen. Von den TMOs ist bekannt, dass diese unterschiedlich starke Gradationen aufweisen. Die mittlere Helligkeit kann

⁵Die Funktion des Key-Value wurde in 3.1.2 beschrieben.

trotzdem identisch sein. Es handelt sich somit um die Abwägungsentscheidung zwischen einem hohen subjektiven Kontrast inklusive Clipping und einer Repräsentation eines hohen Szenenkontrastumfangs. Dieser, für das angestrebte Ziel der Arbeit wahrscheinlich wichtigste Aspekt, fehlt in der Literatur gänzlich. Es gibt keinerlei Informationen, wie sich die Display- bzw. Umgebungsleuchtdichte auf die Größe des reproduzierten Kontrastumfangs der Szene auswirkt. Entsprechend ist auch nicht bekannt, ob die Präferenz bezüglich lokalem oder globalen Tonemapping mit den äußeren Gegebenheiten zusammenhängt.

Der Autor stellt die Hypothese auf, dass bei einem Display mit höherer Leuchtdichte auch ein größerer Kontrastumfang dargestellt werden kann, ohne dass das Bild *flau* aussieht. Gesetzt diesem Fall würde sich die durchschnittliche Helligkeit auch indirekt über die Stärke der Kontrastkompression steuern lassen. Je größer die Begrenzung, also das Clipping der Lichter ist, desto heller wird das Bild. Fällt das Clipping geringer aus, wird das Bild dunkler. Im Vergleich zur reinen Steuerung der Helligkeit würden auf diesem Weg, abhängig von den äußeren Gegebenheiten, zusätzliche Informationen in den Schatten bzw. Lichtern sichtbar.

An dieser Stelle schließt sich der Kreis mit den in Kapitel 3 vorgestellten TMOs. Diese nutzen die maximale Leuchtdichte der Szene zur Berechnung der Transferfunktion. Aufbauend auf den Erkenntnissen über die Zuschauerpräferenz könnte dieser Wert gegen eine Art *maximale relevante Leuchtdichte* ausgetauscht werden, die eine bessere Gradation und Helligkeit bezüglich der gegebenen Bedingungen liefern würde.

Ebenfalls aus Kapitel 3 ist bereits mit der Sättigung ein weiterer Parameter bekannt, der einen hohen Einfluss auf die subjektive Qualität einer Kontrastkompression ausübt. Dieser Aspekt wird allerdings zunächst zurückgestellt und in Kapitel 5 genauer behandelt. An dieser Stelle soll der Fokus allein auf der Helligkeitskomponente liegen.

Zusammenfassend lässt sich festhalten, dass im Laufe der nächsten beiden Kapitel folgende Fragen beantwortet werden sollen?

- Wie hell soll das Bild sein? ((normierte) Helligkeit)
- Wie steil soll der Bildeindruck sein? (Kontrast und Clipping)
- Wie stark soll es lokal modelliert sein? (Homogenisierung)
- Wie stark sollen die Farben gesättigt sein? (Sättigung)

Des Weiteren hat sich gezeigt, dass die Display- und die Umgebungsleuchtdichte die entscheidenden Faktoren für die äußeren Bedingungen darstellen. Auch ist es möglich diese

unabhängig voneinander zu variieren, zumindest solange sie keine zu große Diskrepanz aufweisen. In diesem Fall wären die beiden Parameter nicht ausreichend dekorreliert.

Auf das Erheben verschiedener Displaygrößen soll dagegen verzichtet werden. Die bisherigen Ergebnisse deuten darauf hin, dass bei einem konstanten Sehwinkel keine nennenswerten Änderungen zu erwarten sind. Allerdings muss eingestanden werden, dass eine von der [BBC](#) im Jahre 2015 durchgeführte Studie [\[90\]](#) bezüglich der Sehsituation Zuhause zeigt, dass der Betrachtungsabstand (gemessen in Vielfachen der Bildschirmhöhe) durchaus variiert. Andererseits kann dies auch bereits bei zwei Betrachtern vor dem selben Bildschirm der Fall sein. Hier stellt sich letztendlich die Frage der Praktikabilität. Sprich, es ist nicht möglich, auf einem Bildschirm verschiedene Bilder für verschiedene Zuschauer darzustellen. Aus den genannten Gründen fiel die Abwägungsentscheidung zu Ungunsten der Displaygröße aus.

Darüber hinaus hat die [BBC](#) auch die Verbreitung der verschiedenen Displaytechnologien erfasst. Theoretisch könnte dieser Punkt auch ein Parameter für die äußeren Bedienungen darstellen. Allerdings sieht der Autor hier erstens die Pflicht bei den Displayherstellern, für eine konsistente Darstellung eines identischen Eingangssignals auf verschiedenen Displaytechnologien zu sorgen und zweitens sind die Variationen innerhalb einer Displaytechnologie bereits je nach Preisklasse so immens, dass es nicht möglich wäre, generelle Aussagen abzuleiten.

Es wurde dargelegt, dass der referierte Stand der Forschung wichtige Anhaltspunkte für das visuelle Präferenzmodell liefert. Allerdings bedarf es für ein finales Testdesign noch bessere Kenntnis über die Auswirkungen und Effekte, die mit hochkontrastreichen Videoinhalten einhergehen, um die benötigten Parameter zu verifizieren, sinnvolle Wertebereiche für diese festzulegen und geeignetes Testmaterial zu selektieren. Deshalb wurden zunächst zwei vorbereitende Probandentests durchgeführt.

4.3 Qualitätsgewinn von HDR im Vergleich zum heutigen SDR

Der erste Probandentest wurde im Rahmen der Masterarbeit von Martin Schultz [\[111\]](#) durchgeführt. Hier lag der Fokus darauf, den Qualitätsgewinn von [HDR](#) im Vergleich zum heutigen [SDR](#) (linear ohne Tonemapping) zu messen. Es handelt sich also eher um eine Frage, welche Aufschluss über den Mehrwert der Einführung von [HDR](#) gegenüber dem Status-Quo geben soll. Trotzdem lassen sich hieraus indirekte Erkenntnisse für diese Arbeit ableiten, die im Folgenden kurz zusammengefasst werden sollen.

Insgesamt nahmen 42 Probanden an dem Versuch teil. Es handelte sich um Mitarbeiter von sky Deutschland. Trotzdem bezeichneten sich nur 23 als Experten, da unterschiedlichste Abteilungen, wie beispielsweise die Buchhaltung oder die Personalverwaltung, teilnehmen konnten. 33 Probanden waren männlich, neun weiblich. Der Altersdurchschnitt lag bei 37,4 Jahren. Beim Test kam die Methode des Double-Stimulus Continuous Quality-Scale (DSCQS) [56] zum Einsatz, die Ergebnisse nach dem fünfstufigen Mean Opinion Score (MOS) liefert.⁶ Der Test wurde bei zwei verschiedenen Umgebungsleuchtdichten wiederholt (16 und 76 cd/m^2).

Die fünf unterschiedlichen Testsequenzen, alle aus dem Bereich Fußball, wurden von den Probanden in der SDR-Variante im dunkleren Raum durchschnittlich mit 2,3 MOS-Punkten bewertet. Die obere Grenze des 95%-Konfidenzintervalls⁷ wies dabei den Wert 2,5, die untere Grenze den Wert 2,1 auf. Die Bewertung der HDR-Variante fiel im Durchschnitt um 1,4 MOS-Punkte höher aus. Das Konfidenzintervall war in diesem Fall um 0,05 kleiner (3,55 bis 3,85). Bei der Wiederholung des Versuchs im helleren Raum sank die durchschnittliche Bewertung für die SDR-Variante um 0,4 Punkte. Bei der HDR-Variante waren es etwa 0,2 Punkte. Die Konfidenzintervalle blieben dabei vergleichbar.

Ferner konnte beobachtet werden, dass bei den Testsequenzen 1 und 5 der Unterschied mit 1,85 MOS-Punkten überdurchschnittlich groß ausfiel. Hier waren sowohl ausgedehnte schattige und somit dunkle Bildanteile als auch sonnige und somit helle Bildanteile vorhanden. Es zeigt sich, dass bei großen Kontrastumfängen der realen Szene ein größerer Kontrastumfang besonders gut zum Tragen kommt. Bei den Testsequenzen 2 und 4 lagen die Bewertungen beider Versionen näher beieinander (1,25 MOS-Punkte Unterschied). Hier war der aufgenommene Kontrastumfang geringer.

Beide Beobachtungen deuten darauf hin, dass der Qualitätsgewinn bei HDR nicht nur durch die absolute Leuchtdichte, sondern auch durch den höheren Informationsgehalt zustande kommt. So scheint eine Verbindung zwischen dem Kontrastumfang der Szene und der Differenz in der Bewertung zwischen HDR und SDR zu bestehen. Dies wäre nicht der Fall, wenn die absolute Leuchtdichte das einzige Maß darstellen würde. Die Erhöhung der Umgebungsleuchtdichte führte ebenfalls zu einer Steigerung der Differenz zwischen HDR und SDR. Der Grund hierfür könnte die verbesserte Sichtbarkeit durch

⁶Diese Methode wird auch später in Kapitel 7 verwendet, um den Qualitätsgewinn des in der vorliegenden Arbeit gewonnenen Verfahrens gegenüber dem Stand der Technik aufzuzeigen. An besagter Stelle wird es auch genauer erläutert.

⁷Da sich bei einer Anzahl von 42 Messwerten die Kurve der t-Verteilung stark der Kurve der z-Verteilung (Gleichverteilung) annähert, wurde bei der Berechnung der Standardabweichung und des Konfidenzintervalls von einer Normalverteilung ausgegangen. Dass eine ausreichend große Übereinstimmung besteht, wurde mit Hilfe des Kolmogorov-Smirnov-Tests nachgewiesen. Wird α gleich 0,05 angenommen, darf bei 42 Messwerten die Abweichung nicht größer als 0,21 ausfallen. Die Bedingung konnte für alle Vergleiche im vorliegenden Test erfüllt werden. In den meisten Fällen wurde sogar ein Wert unter 0,1 gemessen.

die gesteigerte Leuchtdichte sein, das anhand der in Abschnitt 4.1 nachgewiesenen Verdeckungsproblematik mit einhergehendem Informationsverlust, gerade bei geringeren Displayleuchtdichten, erklärt werden könnte. Entsprechend wäre aber auch hier die absolute Leuchtdichte nicht per se qualitätsfördernd, sondern nur im Zusammenhang mit entsprechenden Bildinformationen. Dies lässt vermuten, dass ein kontrastkomprimiertes SDR weitaus bessere Ergebnisse erzielen könnte, als der Status-Quo, wenn jenem Informationsverlust vorgebeugt würde. Es bleibt die Frage, wie diese genau zu bewerkstelligen ist?

4.4 Coloristen-Test

4.4.1 Aufbau und Durchführung Coloristen-Test

Wie im vorangegangenen Kapitel deutlich geworden ist, handelt es sich bei der Kontrastkompression keineswegs um ein rein technisches Feld. Da diese derzeit oft manuell durchgeführt wird, besteht ein hoch subjektiver Einfluss des einzelnen Bildingenieurs oder Coloristen. Gerade die zweit Genannten verstehen sich eher als Künstler und prägen damit immer noch das ästhetische Verständnis des Zuschauers.

Um sich dem Thema weiter zu nähern, wurde deshalb zunächst eben jener manuelle Weg des Gratings eingeschlagen, in der Hoffnung von diesem Vorgehen lernen zu können. Dazu wurden fünf professionelle Coloristen aus Film und Fernsehen eingeladen, sieben ausgewählte hochkontrastreiche Testsequenzen nach ihren Vorstellungen für die Anzeige auf einem 100-cd/m^2 -Display zu bearbeiten.⁸ Das Material enthielt unterschiedliche Motive zu unterschiedlichen Tageszeiten. Die Bedingungen waren für alle Coloristen identisch. Sie erhielten nicht nur das gleiche Ausgangsmaterial sondern arbeiteten auch (nacheinander) in den gleichen Räumlichkeiten, die von ARRI in München bereitgestellt wurden. Die Organisation sowie Durchführung dieses Aspekts oblag der Bachelorarbeit von Daniel Schmitt [110].

Bereits vor Beginn herrschte die Vermutung, dass sich die Resultate der einzelnen Coloristen unterscheiden würden. Trotzdem bestand die Hoffnung eine gemeinsame Basis ausmachen zu können. Dem war allerdings nicht so. Diese Verschiedenartigkeit bot aber die Grundlage für einen Probandentest. Dazu wurden alle fünf Varianten einer Sequenz jeweils direkt miteinander verglichen. Pro Sequenz ergaben sich somit zehn Gegenüberstellungen, bei denen sich die Probanden für das eine bzw. gegen das andere Video entscheiden mussten. Dabei wurden die Varianten nacheinander gezeigt - getrennt

⁸Folgende Testsequenzen wurden aus dem Material von Jan Fröhlich [41] ausgewählt: *bistro01*, *bistro02*, *carouselfireworks01*, *carouselfireworks08*, *carouselfireworks09*, *carsfullshot*, *fireplace01*

von einem 5 s dauernden Graubild. Die Gegenüberstellung endete ebenfalls mit einem 5 s dauernden Graubild und dem Schriftzug „Bitte bewerten“. Um den Versuch in einem angemessenen zeitlichen Rahmen zu halten und gleichzeitig das intuitive Moment zu fördern, wurden die Varianten bei der Gegenüberstellung nicht wiederholt.

Jede Variante wurde in ihren vier Vergleichen zwei Mal als Erstes und zwei Mal als Zweites gezeigt. Allerdings wurde die Reihenfolge innerhalb einer Gegenüberstellung nicht variiert. Das heißt, Colorist 1 wurde beispielsweise gegen Colorist 2 und 3 immer als erstes und gegen Colorist 4 und 5 immer als zweites gezeigt. Antworttendenzen aufgrund der Reihenfolge sollten sich gleichwohl im Hinblick auf das bei dieser Untersuchung entscheidende Gesamtergebnis, bei welchem über alle Vergleiche gemittelt wird, ausgleichen. Aber selbst für den in der späteren Auswertung dargestellten Einzelvergleich (Abbildung 4.5) scheint solch ein Effekt nicht beobachtet werden zu können. So schwankt die Zustimmung zwischen Colorist 2 und 4 trotz gleichbleibender Reihenfolge bei den einzelnen Sequenzen stark hin und her.

Der Versuch wurde unter kontrollierten Laborbedingungen (10 cd/m^2) durchgeführt. Das Display⁹ wurde auf 100 cd/m^2 begrenzt. Insgesamt nahmen 19 Personen an dem Versuch teil. Der Altersdurchschnitt betrug 28,6 Jahre, wobei der jüngste Teilnehmer 16 und der älteste 58 Jahre alt waren. Die Verteilung von männlich zu weiblich lag bei 53% zu 47%.

4.4.2 Ergebnisse Coloristen-Test

Bei der Auswertung wurde jede Bevorzugung in einem direkten Vergleich mit einem Punkt honoriert. In der Folge spiegelt die Höhe der Balken in Abbildung 4.4 die Anzahl der Probandenentscheidungen für das Grading des entsprechenden Coloristen wider. Es ist ersichtlich, dass das Grading von Colorist 2, 4 und 5 dem ihrer Kollegen 1 und 3 deutlich vorgezogen wurde. Die genaue Reihenfolge variiert aber innerhalb der Sequenzen. Abbildung 4.5 schlüsselt den direkten Vergleich zwischen den Coloristen noch etwas genauer auf.

Für diese Arbeit ist nicht entscheidend, welchen Coloristen bzw. welches Grading die Zuschauer am meisten bevorzugen, sondern ob die Ergebnisse mit den in vorherigen Abschnitten genannten Parametern verknüpft werden können. Deshalb soll im Folgenden untersucht werden, inwiefern zwischen der Bewertung der Probanden und den Größen Sättigung, Kontrast, Helligkeit und Clipping eine Korrelation besteht. Da in diesem Kapitel zunächst der Fokus auf der Helligkeitsdarstellung liegen soll, werden die Ergebnisse bezüglich der Sättigung zurückgestellt und im folgenden Kapitel betrachtet.

⁹HP Dreamcolor Z27x

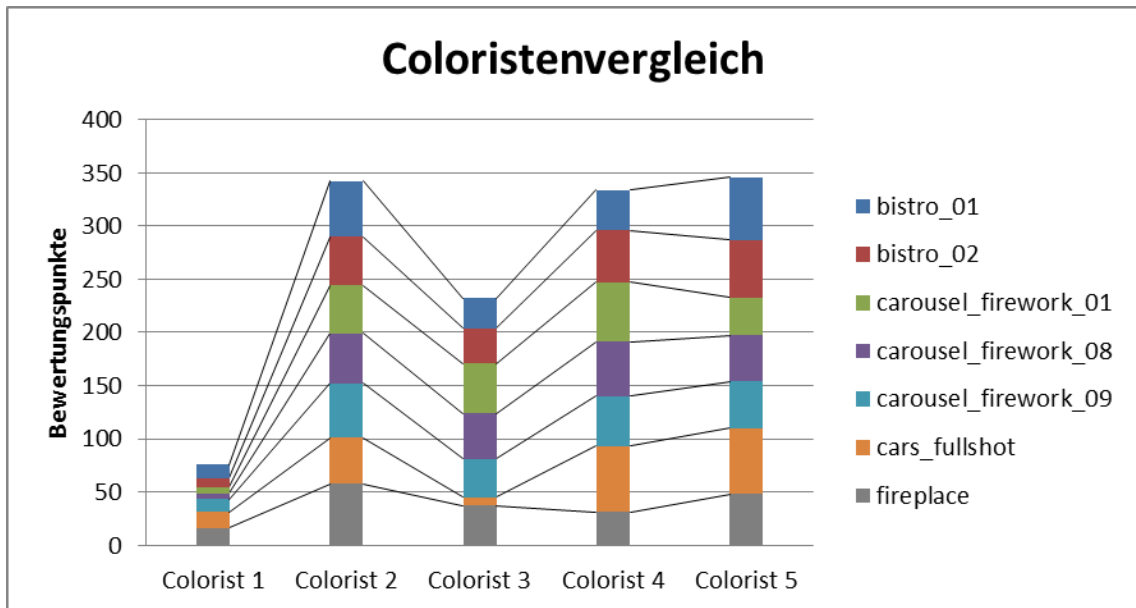


ABBILDUNG 4.4: Die Gradings von Colorist 2,4 und 5 wurden von den Probanden deutlich bevorzugt. Die genau Reihenfolge variiert aber innerhalb der Sequenzen.

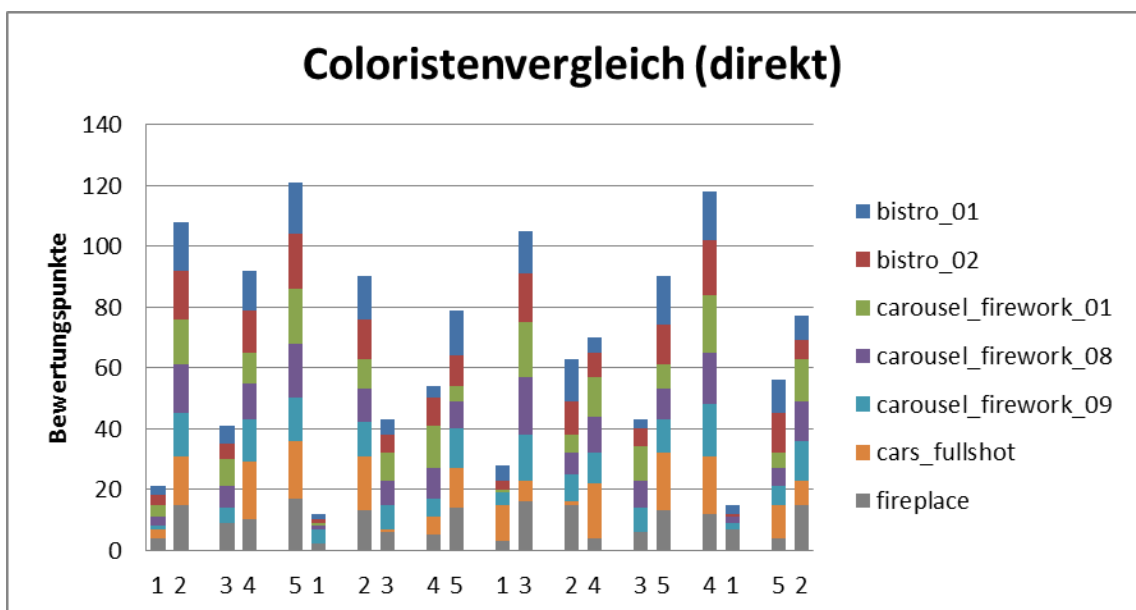


ABBILDUNG 4.5: Die Bewertungen der einzelnen Coloristen Gradings im direkten Vergleich. So war die Gegenüberstellung von Grading 1 und 5 immer eine klare Angelegenheit, wohingegen es zwischen Grading 2 und 4 von der Sequenz abhing.

Für Kontrast, Helligkeit und Clipping wird nun die betreffende Größe jeweils auf der Abszissenachse aufgetragen, wohingegen die Punkte der Probandenbewertung die Ordinatenachse bilden. Jedes Grading wird mithilfe eines repräsentativen Einzelbildes als Punkt im Koordinatensystem dargestellt. Dabei wird zwischen den einzelnen Testsequenzen unterschieden, indem diese durch verschiedene Farben und Formen repräsentiert werden. Die Diagramme finden sich in Abbildung 4.6.

Kontrast: Der Kontrast wurde als durchschnittlicher Abstand zum mittleren Grau, sprich zum Codewort 128 (bei einer 8-Bit-Quantisierung), ermittelt. In der Abbildung ist zu erkennen, dass sich die Messpunkte tendenziell im oberen rechten Bereich befinden. Sequenzen mit einem niedrigeren Kontrast erhielten eher eine geringere Bewertung. Allerdings gibt es mehrere Ausreißer. Hier fällt besonders der orange Punkt unten rechts ins Auge. Dieser lässt sich aber über die Sättigung des dahinterstehenden Gradings erklären. Der Colorist hatte sich für eine schwarz-weiß Darstellung entschieden, die den Probanden nicht zusagte. Bei der Sequenz *fireplace* (graue Quadrate) gibt es keine klare Tendenz. Dies ist darauf zurückzuführen, dass die hier verwendete Definition von Kontrast für diese Sequenz ungeeignet ist. So führt eine deutlich dunklere Darstellung der Szene, durch die damit verbundene Abweichung vom Codewort 128, zu einem hohen Kontrastwert, ohne dass dafür helle Bildbereiche vorhanden sein müssen. Auch bei den Sequenzen *bistro01* (rote Quadrate) und *carouselfirework09* (blaue Sterne) erhält jeweils die Variante mit dem höchsten Kontrastwert nur eine mäßige Bewertung. Es ist denkbar, dass die gewählte Gradation als zu steil empfunden wird. Generell scheint bei einem Kontrastwert um 100 (gemessen in Codewörtern) das Maximum erreicht zu sein.

Helligkeit: Die Helligkeit wird mithilfe des arithmetischen Mittels der 8-Bit-Werte dargestellt. Genaugenommen handelt es sich somit um eine Art Luma. Da die Gradings des Coloristen 1 aufgrund der schwachen Sättigung und des geringen Kontrastes stark abgewertet wurden, werden sie bei der Helligkeit nicht berücksichtigt, um falsche Rückschlüsse zu vermeiden. Insgesamt fällt es bei der Helligkeit schwerer, eine eindeutige Tendenz abzulesen. Für die drei Tagszenen (*bistro01* (rote Quadrate), *bistro02* (blaue Rauten) und *carsfullshot* (orange Punkte)) führte eine hellere Darstellung zu besseren Bewertungen. Gleiches gilt auch für die Szene *fireplace* (graue Quadrate), welche im Bereich der Dämmerung zu verorten ist. Bei den Nachtszenen (*carouselfirework01* (grünes Dreieck), *carouselfirework08* (violetttes Kreuz) und *carouselfirework09* (blauer Stern)) scheint eher ein gegensätzlicher Zusammenhang zu bestehen, so dass eine dunklere Darstellung eine bessere Bewertung nach sich zieht. Für eine gesicherte Aussage sind die Daten allerdings nicht aussagekräftig genug. Erschwerend kommt hinzu, dass es sich bei den Nachtszenen um Szenen mit relativ künstlichen Inhalt handelt. Inwieweit dies auch für natürliche Vorlagen gilt, bleibt unklar.

Clipping: Das Clipping wird durch die Anzahl der Pixel mit dem Codewort 255 repräsentiert. Um eine bessere Darstellung zu ermöglichen, wurde die Abszissenachse logarithmisch skaliert und alle Gradings ohne Clipping erhielten den Wert 1. Auch hier gilt es zwischen den Szenen bei Tag, Dämmerung und Nacht zu unterscheiden. Erstere profitieren vom Clipping, was durch die damit verbundene Steigerung von Kontrast und Helligkeit zusammenhängen dürfte. Für die Nachtszenen ist hingegen keine klare Tendenz auszumachen. Bei den beiden Ausreißern, welche bei relativ starkem Clipping eine

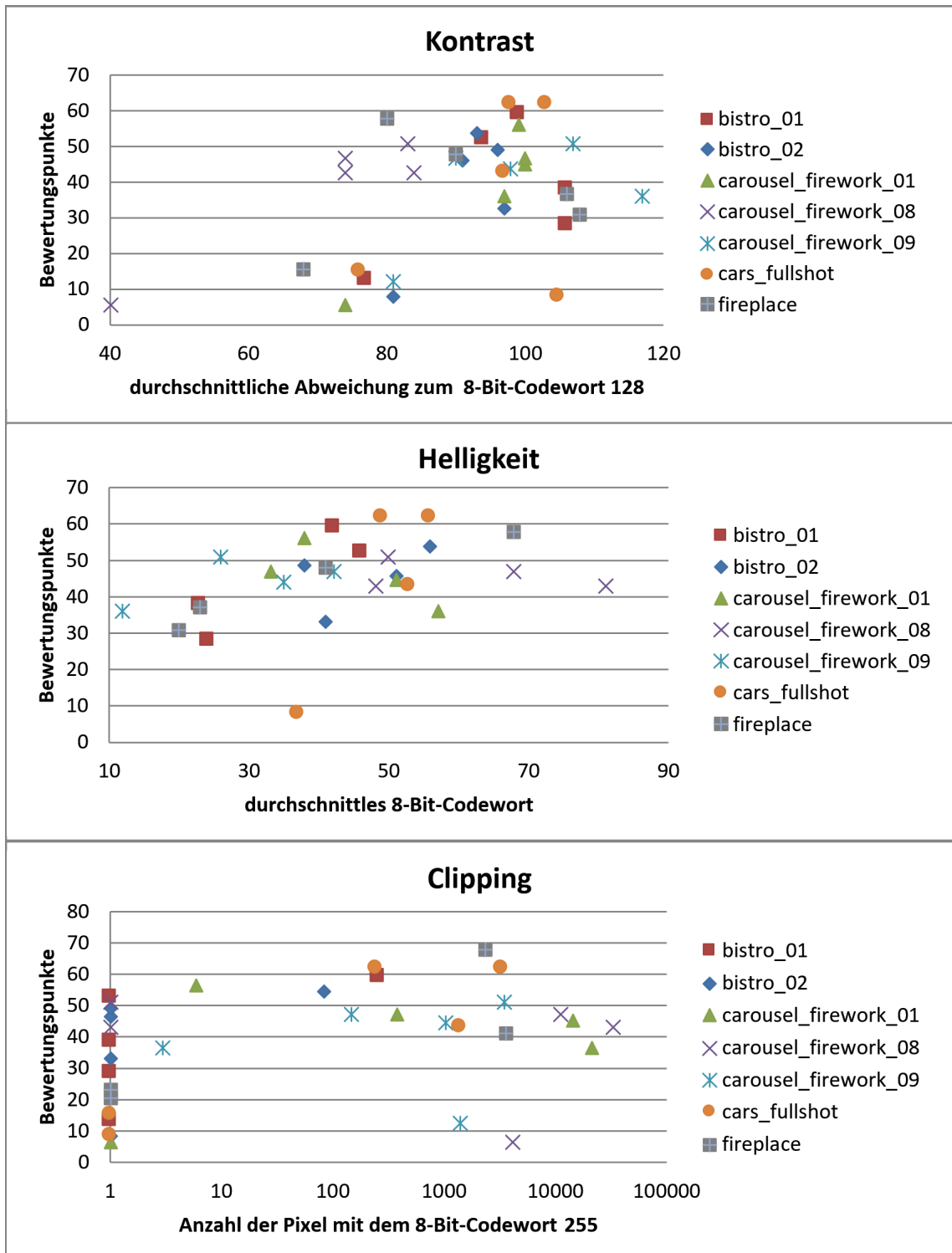


ABBILDUNG 4.6: Die drei Diagramme zeigen die Korrelation zwischen Probandenbewertung und Kontrast, Helligkeit oder Clipping.

sehr geringe Bewertung erhalten haben, handelt es sich um Gradings von Colorist 1 mit der bereits besprochenen Problematik bezüglich der schwachen Sättigung und des geringen Kontrastes. Auf Grundlage dieser Ergebnisse kann keine pauschale Aussage getroffen werden, dass sich Clipping grundsätzlich negativ auf die wahrgenommene Bildqualität auswirken würde. Dieser Sachverhalt benötigt eine weitaus differenziertere Betrachtung.

Zusammenfassend lässt sich festhalten, dass sich die Vermutung, Kontrast und Helligkeit seien entscheidende Einflussparameter bei der Konvertierung von **HDR** nach **SDR** bestätigt. Prinzipiell führt ein höherer Kontrast und ein helleres Bild zu einer höheren Bewertung. Das Clipping kann hierbei indirekt hilfreich sein, da es diese beiden Faktoren entsprechend bedingt. In diesem Fall wird Clipping nicht unbedingt als negativ empfunden, obwohl es zum Verlust von Informationen kommt. An dieser Stelle sei auch an den *cameraTMO* erinnert, der trotz dieser Einschränkung sehr gute Bewertungen erfuhr.

Deshalb stellt der Autor folgende Hypothese auf: Der Zuschauer bevorzugt ein eher helles und ausreichend steiles Bild. Dies ermöglicht ihm, ohne größeren Aufwand alle wichtigen Informationen der Szene wahrzunehmen. Wenn dies nur erreicht werden kann, indem Extremwerte abgeschnitten werden, nimmt er das als notwendiges Übel in Kauf. Wäre es aber möglich, den hellen und steilen Eindruck ohne größeren Informationsverlust zu erreichen, z. B. durch ein dahingehendes modifiziertes lokales Tonemapping, würde dieses bevorzugt werden.

Bis zu diesem Zeitpunkt der Arbeit sind die präsentierten Ergebnisse eher qualitativer Natur. Um sie aber nutzbar zu machen und auch die oben stehende Hypothese zu überprüfen, wird es nötig, aussagekräftigere Werte zu produzieren. Deshalb wird im folgenden Abschnitt, basierend auf den bisherigen Erkenntnissen, ein Probandentest entwickelt, der in der Lage ist, die Eindrücke besser zu quantifizieren.

4.5 Probanden-Grading

4.5.1 Aufbau und Durchführung Probanden-Grading

Aufbauend auf den referierten Studien und durchgeführten Voruntersuchungen wurde ein Experiment entwickelt, das den Probanden erlaubte, die fraglichen Parameter direkt einzustellen (Grading) und somit die Basis für das visuelle Präferenzmodell zu bilden. Die Studie wurde auf der Electronic Imaging Konferenz 2017 in Burlingame vorgestellt [\[75\]](#).

Da die meisten Teilnehmer des Probanden-Gradings bis zu diesem Zeitpunkt noch keine Erfahrung mit Bild- bzw. Videobearbeitung hatten, mussten die Parameter sehr intuitiv

und leicht zu bedienen sein. Des Weiteren wurde der Test unter verschiedenen äußeren Bedingungen (Display- und Umgebungsleuchtdichte) wiederholt, so dass die Anzahl der Parameter sowie die der Testbilder entsprechend beschränkt und sorgfältig ausgewählt war. Die meisten Teilnehmer benötigten etwa eine Stunde (inklusive Pausen) für den gesamten Test.

Die Teilnehmer wurden gebeten, die Parameter so einzustellen, wie es ihnen subjektiv am besten gefällt. Zu Beginn sollten sie zunächst alle Parameter ausprobieren, um ein besseres Verständnis für ihre Wirkung auf das Bild sowie ihren Arbeitsbereich zu bekommen. Über den Hintergrund der Evaluation wurden sie nicht in Kenntnis gesetzt.

Die für die Bearbeitung der Bilder genutzten Einstellungsparameter, wurden in einem Overlay, wie in Abbildung 4.7 zu sehen, dargestellt. Sie waren wie folgt geordnet:



ABBILDUNG 4.7: Mithilfe eines Overlays werden die Regler der vier Einstellungsparameter über dem Testbild eingeblendet.

Helligkeit (brightness): Der erste Einstellungsparameter konnte genutzt werden um die Helligkeit anzupassen. Der Arbeitsbereich erstreckte sich von -2 bis 2 (standardmäßig 0). Ein Schritt entsprach der Änderung um eine Blendenstufe, ausgehend von 0 der durchschnittlichen logarithmischen Leuchtdichte (Y_{mean}). Die zugehörige Transferfunktion war eine S-Kurve mit folgender Funktion:

$$Y_{out}(x,y) = \frac{1}{(1 + e^{((-a) * (\ln(Y_{in}) - b))})} \quad (4.3)$$

$$b = \ln(Y_{mean} - brightness * 5) \quad (4.4)$$

Y_{in} steht dabei für die Eingangsleuchtdichte und Y_{out} für die Ausgangsleuchtdichte. Y_{out} kann dabei Werte zwischen 0 und 1 annehmen. a beschreibt den Kontrast, der im folgenden dargestellt wird.

Kontrast (contrast): Der zweite Regler diente zur Kontrolle des Kontrastes, welcher der Steilheit der besagten S-Kurve entsprach. Entsprechend wurde festgelegt, wie viele Blendenstufen im finalen Bild erhalten bleiben. Die Probanden hatten somit die Aufgabe, den ansprechendsten Trade-Off zwischen dem subjektiven Kontrast und dem reproduzierten Kontrastumfang der Szene zu finden. Der Regler besaß einen Arbeitsbereich von -2 bis 2 (standardmäßig 0). Wie aus Formel [4.5](#) ersichtlich, entsprach 0 dabei einem Kontrastumfang von 11 Blendenstufen, welcher etwa sechsmal so groß ist (oder 2,5 EVs größer), als jener von Petits *camera TMO*. Gleichung [4.6](#) wurde genutzt, um die Blendenstufen so umzurechnen, dass sie zu der Skalierung in Gleichung [4.3](#) passten.

$$stops = 6 + (1 - contrast) * 5 \quad (4.5)$$

$$a = 0,0088 * stops^2 - 0,288 * stops + 2,99 \quad (4.6)$$

Sättigung: Die Sättigung wurde wie bereits beim Coloristen-Test mit erhoben, während die Auswertung auch hier erst in Kapitel [5](#) erfolgt.

Homogenisierung: Mit dem letzten Regler hatten die Probanden die Möglichkeit, die Kalkulation von Y_{mean} zu beeinflussen. Der Regler besaß einen Arbeitsbereich von 0 bis 1 (standardmäßig: 0). 0 bedeutete, dass die durchschnittliche Leuchtdichte einzig über das gesamte Bild berechnet wurde; sprich, es handelte sich um globales Tonemapping. 1 hingegen führte zu einer Berechnung, die ausschließlich auf der lokalen Umgebung des Pixels beruhte. Dazu wurde ein vergleichsweise großer Bereich mit einem Durchmesser von ca. 300 Pixeln herangezogen und mit einer Gaußfunktion weich gezeichnet. Werte zwischen 0 und 1 waren entsprechende Mischformen. Der Regler wurde mit dem Begriff *Homogenisierung* bezeichnet, da dies die beste Beschreibung für seine Funktionsweise darstellte. Je stärker von diesem Regler Gebrauch gemacht wurde, desto mehr konnten helle Bereiche des Bildes abgedunkelt und dunkle Bereiche aufgehellt werden. Der gesamte Helligkeitseindruck erschien homogener.

$$Y_{mean}(x,y) = homogenisation * Y_{meanlocal}(x,y) + (1 - homogenisation) * Y_{meanglobal} \quad (4.7)$$

Für die Display-Leuchtdichte wurden drei repräsentative Werte ausgewählt. 100 cd/m^2 entsprechen dem heutigen SDR-Standard [54] [55], 300 cd/m^2 kommen den heutigen TV-Geräten der Konsumenten am nächsten und 1000 cd/m^2 erfüllt die HDR-Anforderungen [57] [58].

Auch für die Umgebungsleuchtdichte wurden drei verschiedene Level ausgewählt. Die Stufen repräsentieren einen dunklen, einen gediminten und einen hellen Raum, wie es auch schon bei ähnlichen Studien in der Vergangenheit der Fall war. Dabei wird gedimmt ebenfalls oft mit 10 cd/m^2 angegeben, was auch den Vorgaben nach [56] entspricht. Der helle Raum wurde mit 50 cd/m^2 und der dunkle Raum mit $0,5 \text{ cd/m}^2$ angesetzt.

Um die Anzahl der Durchläufe und somit auch die Dauer des Tests nicht zu groß werden zu lassen, wurden beide Faktoren unabhängig voneinander variiert. Dazu wurde als Startpunkt die Kombination 300 cd/m^2 (Display) und 10 cd/m^2 (Umgebung), welche die typische Sehsituation zu Hause am nächsten abbildet, gewählt. Zunächst wurde die Umgebungsleuchtdichte auf 10 cd/m^2 fixiert und die Displayleuchtdichte variiert. Anschließend wurde die Displayleuchtdichte auf 300 cd/m^2 fixiert, wohingegen die Umgebungsleuchtdichte verändert wurde. Dieses Vorgehen setzt voraus, dass die beiden Leuchtdichtewerte dekorreliert sind, was aber für den angestrebten Bereich bereits nachgewiesen werden konnte [121].

Der gesamte Test wurde auf einem Panasonic VIERA TX-58DXW904 durchgeführt, welches das erste Display war, das die Ultra-HDR-Premium-Bedingungen erfüllen konnte und somit eine sehr hohe Farbtreue und eine maximale Leuchtdichte von über 1000 cd/m^2 aufwies.

Das Bildmaterial, welches in diesem Experiment genutzt wurde, deckte einen weiten Bereich von möglichen, für den Broadcast relevanten Szenen und Genres ab. Insgesamt wurden 12 Standbilder genutzt. Sechs von ihnen wurden im Rahmen der bereits erwähnten Masterarbeit von Nico Schultz [111] aufgenommen; die anderen sechs stammen von Jan Fröhlich [41].¹⁰ Die ersten sechs zeigen Situationen aus einem Fußballspiel, die anderen sechs teils sehr unterschiedliche Motive. Die zugrundeliegende Überlegung war, zunächst eine Präferenz für eines der wichtigsten Genres im Fernsehen (Live-Sport) zu finden und im Anschluss zu analysieren, inwieweit dies mit anderen Motiven korreliert und ob es möglich ist, die Ergebnisse zu generalisieren. Mit dieser Vorgehensweise könnte herausgefunden werden, ob es notwendig ist, je nach Genre, Veränderungen oder Transformationen vorzunehmen.

Insgesamt nahmen 34 Probanden an dem subjektiven Test teil. Alle von ihnen absolvierten die zusammengenommenen fünf Testrunden. Nur fünf gaben an, bereits Erfahrung

¹⁰ *Anstoß, Stadion Torwart, Stadion Schwenk 2, Zuschauer, Eckball, Stadion Torwart 2, carousell fireworks 08, cars closeshot, cars longshot, fishing longshot, poker travelling slowmotion, smith hammering*

mit Bildbearbeitung und Qualitätsanalysen zu haben. Das durchschnittliche Alter betrug 28,9 Jahre. Nur fünf waren älter als 50 Jahre. Es konnten keine signifikanten Unterschiede aufgrund des Alters festgestellt werden. 77% der Teilnehmer waren männlich, 23% waren weiblich.

4.5.2 Ergebnisse Probanden-Grading

Zunächst wurden die Korrelation mit den äußeren Bedingungen analysiert. Insgesamt wurden so 1800 Datenreihen, bestehend aus Helligkeit, Kontrast, Sättigung, Homogenisierung, Display- und Umgebungsleuchtdichte, erzeugt. Abbildung 4.8 zeigt einen beispielhaften 3D-Plot eines der 12 Bilder. Helligkeit, Display- und Umgebungsleuchtdichte bilden dabei die drei Achsen. Somit repräsentieren die vertikalen Linien die 34 verschiedenen Einstellungen der jeweiligen Probanden. Sie weisen eine Streuung im Bereich von etwa einer Blendenstufe auf. Die Standardabweichung beträgt 0,33. Die Darstellungen der anderen Bilder oder der anderen Parameter sehen ähnlich aus.

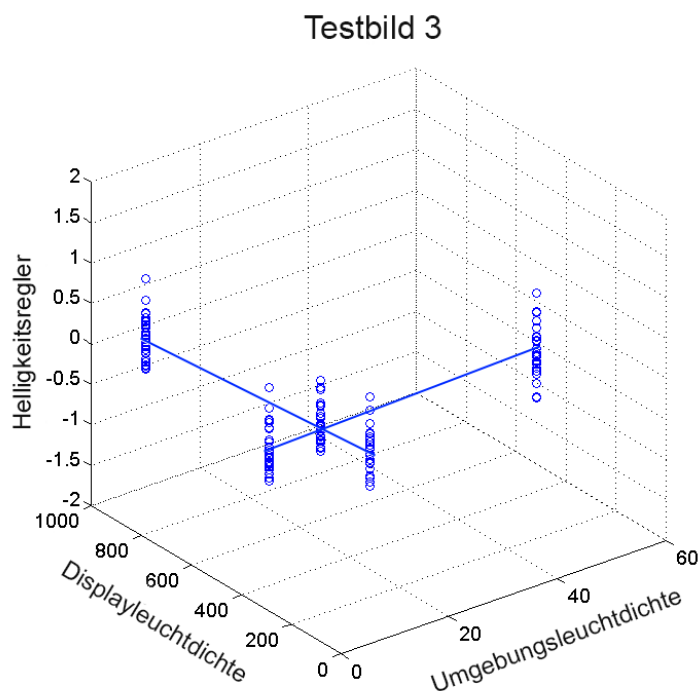


ABBILDUNG 4.8: 3D-Plot von einem Bild mit Helligkeit, Display- und Umgebungsleuchtdichte

Um einen Gesamteindruck der Korrelationen zwischen den äußeren Bedingungen, den Einstellungsparametern und den Bildern zu bekommen, ist es deshalb zielführender, den Mittelwert aller 34 Probanden zu verwenden, wie in Abbildung 4.9 dargestellt. Der

Einstellungsparameter wird dazu auf die Ordinatenachse abgebildet und die Abszissenachse in die fünf verschiedenen äußeren Bedingungen eingeteilt. Pro Parameter ergeben sich zwei Grafiken, da die Auswertung in die ersten sechs und die zweiten sechs Bilder aufgeteilt wurde.

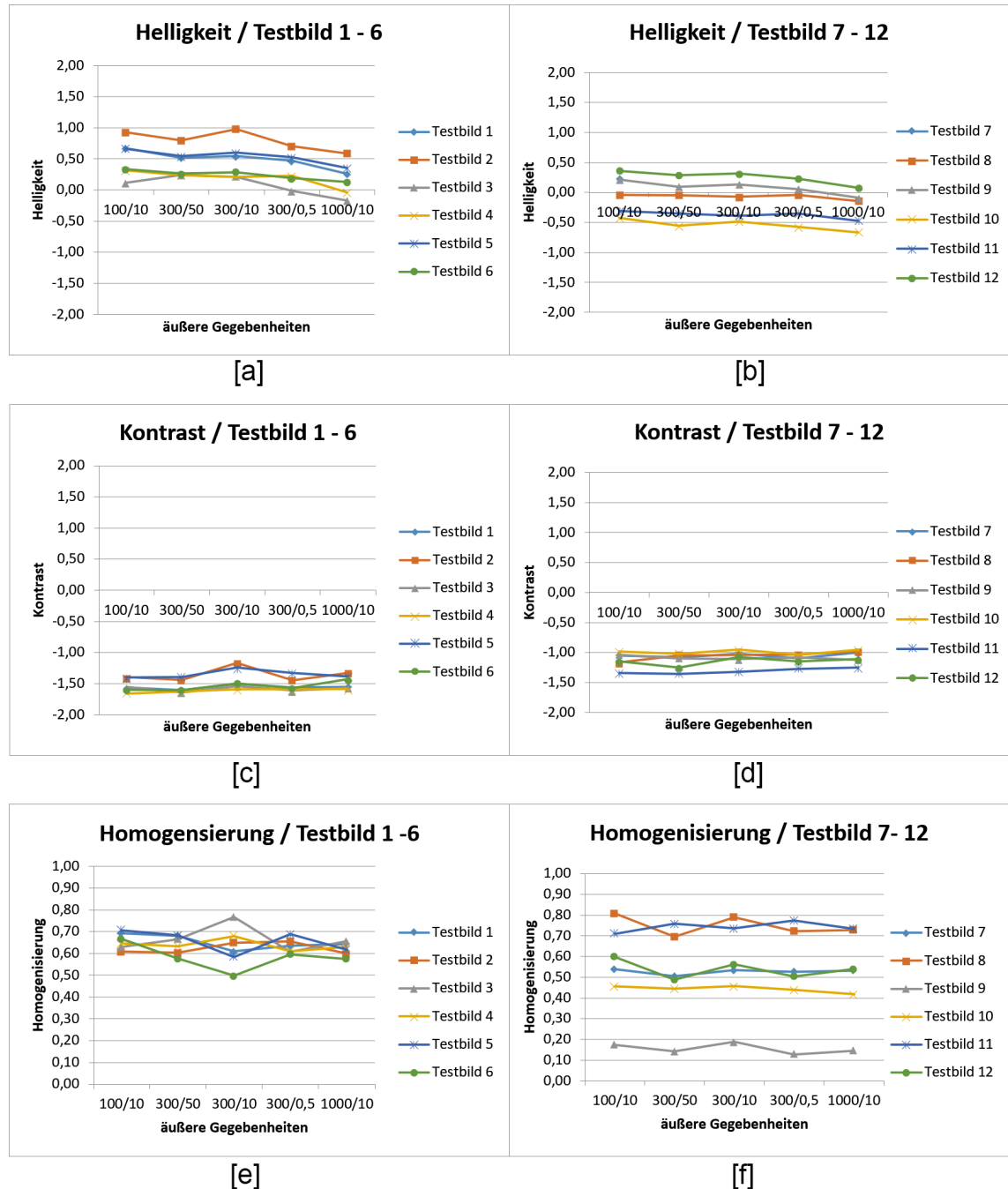


ABBILDUNG 4.9: Das Diagramm zeigt den durchschnittlichen Wert aller 30 Probanden für alle Einstellungsparameter bei den unterschiedlichen äußeren Gegebenheiten und Bildern.

Die beiden oben stehenden Diagramme ((a) und (b)) in Abbildung [4.9](#) stellen die Resultate bezüglich der Helligkeit dar. Es wird deutlich, dass eine ansteigende Displayleuchtdichte zu einer absinkenden Helligkeit führt. Die Differenz zwischen 100 cd/m^2 und 1000 cd/m^2 beträgt durchschnittlich 0,4 EVs. Die Umgebungsleuchtdichte scheint nur einen sehr geringen Effekt zu haben. Hier liegt die Differenz zwischen $0,5 \text{ cd/m}^2$ und 50 cd/m^2 bei gerade einmal 0,06 EVs. Zum anderen hängt der Wert für die Helligkeit deutlich vom Bild selber ab. Je nach Szene ist eine andere Ausbelichtung gewünscht. Sowohl der Einfluss von Display- und Umgebungsleuchtdichte sowie der dargestellten Szene decken sich mit den bis zu diesem Zeitpunkt referierten Studien.

Anders stellt sich die Situation beim Kontrast dar. Aus den folgenden beiden Diagrammen ((c) und (d)) kann geschlossen werden, dass der präferierte Kontrast unabhängig von den äußeren Bedingungen ist. Es liegt keine entsprechende Tendenz im Linienverlauf vor. Des Weiteren sind auch nur sehr geringe Schwankungen aufgrund des Bildes zu verzeichnen. In allen Fällen wird ein Kontrastumfang von 12 bis 12,5 Blendenstufen gewählt. Dies ist vergleichsweise viel und deutlich mehr als im heutigen Broadcast. Auch wählten manche Teilnehmer den Wert -2, was einem Blendenumfang von 13 EVs entsprach und den im Test höchstmöglichen Wert darstellte. Für die überwiegende Anzahl der Probanden war damit aber eine zu flache Gradation verbunden. 12 Blendenstufen waren ausreichend, um sowohl die Inhalte in den Lichtern als auch in den Schatten wahrnehmen zu können.

Somit wurde von den Probanden ein höherer Szenenkontrastumfang einem subjektiv hohen Kontrast vorgezogen. Der gewählte Kontrastumfang lag etwa 3-4 Blendenstufen über dem des heutigen Broadcast. Die Zuschauer scheinen also prinzipiell durchaus bereit zu sein, eine flachere Gradation in Kauf zu nehmen, solange ihnen dafür ein deutlicher Mehrwert in Form von mehr Information angeboten wird. Dieses Ergebnis scheint auf den ersten Blick konträr zu den bisherigen Erkenntnissen zu sein, bei welchen ein eher hoher Kontrast bevorzugt wurde.

Die folgenden zwei Aspekte könnten eine mögliche Erklärung liefern. Erstens verfügten alle im Coloristen-Test zur Auswahl stehenden Versionen bereits über einen größeren Kontrastumfang. Im Vergleich zum Broadcast finden in der Postproduktion typischerweise flachere Gradationskurven Anwendung. Es war für die Probanden also gar nicht erst möglich, eine wirklich steile Version auszuwählen. Zweitens, und das ist wahrscheinlich der entscheidende Aspekt, sollte die Einstellung des Kontrastreglers nicht isoliert betrachtet werden, denn die nun folgende Einstellmöglichkeit der Homogenisierung sorgt durch ihre potenzielle Steigerung der lokalen Modulation für ein gesteigertes Kontrastempfinden.

In den beiden letzten Diagrammen ((e) und (f)) können die höchsten Schwankungen gefunden werden. Es kommt zu hohen Abweichungen in der Präferenz aufgrund der Bildinhalte. Für die Fußballszene liegt der Wert durchschnittlich bei 0,64 bei einer Standardabweichung von 0,03. Bei den anderen Szenen im Mittel bei 0,53 bei einer Standardabweichung von 0,2. Allerdings sollte Bild 9 als Ausreißer betrachtet werden, da in dieser Gegenlichtszene die Sonne direkt in die Kamera scheint. Die Sonne bildet somit ein extremes Highlight, das bei einem hohen lokalen Einfluss zu einer überproportionalen Abdunkelung der umgebenden Bereiche und in der Folge zu sichtbaren Artefakten führt. Ein Grenzwert für lokale Maxima wäre zur Vermeidung nötig. Dies erklärt die geringen Werte bei der Homogenisierung.

Erstaunlicherweise ist die Homogenisierung unabhängig von den äußeren Bedingungen. Selbst bei sehr hohen Displayleuchtdichtewerten kann kein signifikanter Unterschied gemessen werden. Oder in anderen Worten: Selbst bei der Darstellung einer **HDR**-Quelle auf einem **HDR**-fähigen Display ist es hiernach sinnvoll, eine Kontrastkompression durchzuführen, um Lichter und Schatten aneinander anzugleichen. Diese Ergebnisse sind mit den durchgeführten Untersuchungen bezüglich der Kontrastwahrnehmung zu Beginn dieses Kapitels zu erklären, welche die negative Wirkung von hellen Bildbereichen auf die Sichtbarkeit der Schattenregionen nachweist. An früherer Stelle hieß es dazu: „Denn nicht einzig die Betrachtungsdauer, sondern auch das schmalere Sichtfeld [eines Fernsehers] stellt eine Beschränkung der Adaptionsfähigkeiten des Menschen im Vergleich zu einer realen Szene dar. Damit trotzdem genauso viele Inhalte wie in der natürlichen Vorlage wahrgenommen werden können, müsste diese Limitation mithilfe von technischen Mitteln ausgeglichen werden, um Maskierungseffekten vorzubeugen.“

Durch den Regler Homogenisierung wurden nun Lichter und Schatten aneinander angeglichen und der Maskierungseffekt reduziert. Somit hat sich die Vermutung bestätigt, dass die Anwendung einer Kontrastkompression z. B. in Form von Tonemapping auch bei **HDR**-Displays sinnvoll ist. Dies ist eine fundamental neue Sichtweise auf das gesamte Thema **HDR** und eine der zentralen Erkenntnisse der vorliegenden Arbeit.

Des Weiteren können durch den konstant hohen Einsatz der Homogenisierung (bis auf die oben erwähnte Ausnahme) die Werte beim Kontrast erklärt werden. Dieser sorgt, wie bereits kurz angesprochen, für eine stärkere lokale Modulation, die von den Zuschauern als kontrastreicher empfunden wird. Entsprechend kann eine flachere Gradation akzeptiert werden. Das bestätigt die im vorherigen Abschnitt aufgestellte Hypothese, dass Clipping in Kauf genommen wird, um ein helles und steiles Bild zu erzeugen, sofern dies nicht über andere Wege erreicht werden kann. Entsprechend kann in diesem Fall durch das lokale Tonemapping weitestgehend auf Clipping verzichtet werden.

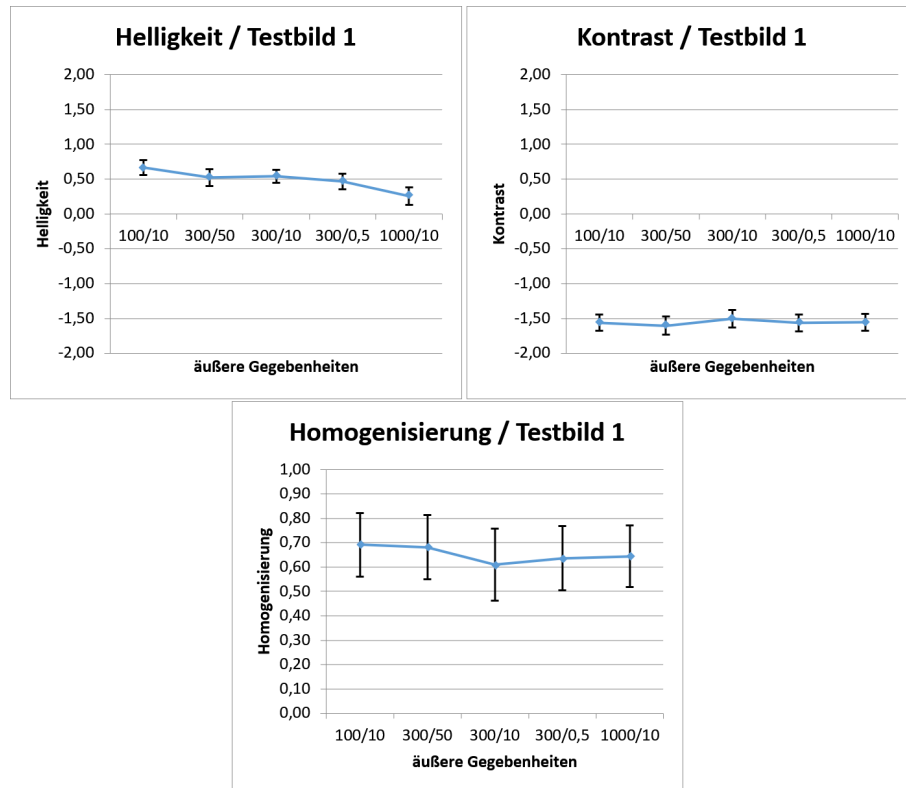


ABBILDUNG 4.10: Die Vertrauensintervalle fallen bei Helligkeit und Kontrast sehr gering aus. Bei der Homogenisierung sind sie deutlich größer.

Zur Verbesserung der Lesbarkeit wurden die 95%-Konfidenzintervalle, in welchen sich der arithmetische Mittelwert befindet, aus der gerade analysierten Abbildung 4.9 entfernt und deshalb an dieser Stelle exemplarisch für Bild 1 nachgereicht¹¹. Ihre Darstellung findet sich in Abbildung 4.10. Es ist zu erkennen, dass die Schwankungsbreite bei Helligkeit und Kontrast sehr gering ausfällt, wohingegen es bei der Homogenisierung zu einer größeren Streuung der Werte kommt. Auch die anderen Bilder weisen hier ein sehr ähnliches Muster auf. Dieser Eindruck relativiert sich, wenn man sich den Einfluss der drei Regler auf den Bildeindruck anschaut. Während sich mit Helligkeit und Kontrast ein stark unterbelichtetes bzw. überbelichtetes und sehr flaches bzw. ziemlich steiles Bild erzeugen lässt, verschiebt die Homogenisierung „nur“ das Verhältnis der Bildbereiche zueinander. Der angegebene Mittelwert scheint zwar nicht für alle Probanden als ideal empfunden zu werden, bildet aber dennoch eine gute Näherung.

Bei der Helligkeit war aufgefallen, dass die Werte bei den verschiedenen Bildvorlagen variierten. Der zweite Teil der Auswertung soll sich deshalb mit den Korrelationen der

¹¹Da sich bei einer Anzahl von 34 Messwerten die Kurve der t-Verteilung stark der Kurve der z-Verteilung (Gleichverteilung) annähert, wurde bei der Berechnung der Standardabweichung und des Konfidenzintervalls von einer Normalverteilung ausgegangen. Dass eine ausreichend große Übereinstimmung besteht, wurde mit Hilfe des Kolmogorov-Smirnov-Tests nachgewiesen. Wird α gleich 0,05 angenommen, darf bei 34 Messwerten die Abweichung nicht größer als 0,23 ausfallen. Im hier vorliegenden Fall wurden folgende Werte ermittelt: 0,13 beim Kontrast, 0,12 bei der Helligkeit und 0,16 bei der Homogenisierung

von den Probanden eingestellten Werten mit den die Szene beschreibenden Parametern beschäftigen. Dazu werden die durchschnittliche Szenenleuchtdichte¹², der Szenenkontrast¹³ und der Szenenkontrastumfang¹⁴ herangezogen.

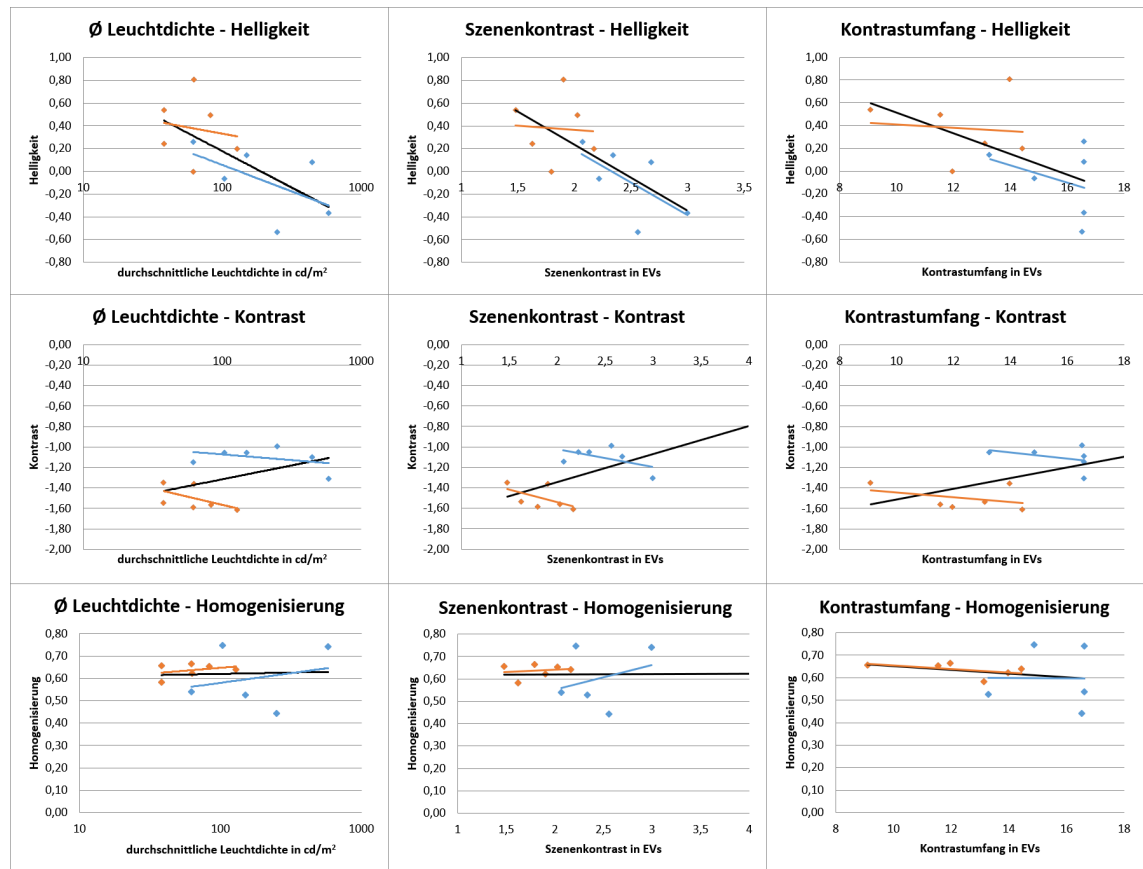


ABBILDUNG 4.11: Korrelationen zwischen den von den Probanden eingestellten Werten und den die Szene beschreibenden Parametern

Die Ergebnisse können Abbildung 4.11 entnommen werden. Die Fußball-Szenen (Bild 1 bis 6) sind jeweils als orange und die Fröhlich-Szenen (Bild 7 bis 12) als blaue Punkte in den Diagrammen verortet. Darüber hinaus ist je eine Trendlinie für die beiden Bildergruppen eingezeichnet, sowie eine Gesamttrendlinie in schwarz.

Die drei Diagramme in der ersten Zeile zeigen die Korrelation des eingestellten Helligkeitswerts mit den drei die Szene beschreibenden Parametern. In allen drei Fällen weist die schwarze Trendlinie eine negative Steigung auf. Die zugehörigen Korrelationskoeffizienten lauten -0,61, -0,68 und -0,58. Die beiden weiteren Trendlinien legen nahe, dass diese Korrelation bei den Fröhlich-Szenen, den kontrastreicherer Szenen, stärker

¹²logarithmischer Mittelwert der Leuchtdichte

¹³durchschnittliche Differenz in Blendenstufen zwischen einem Pixelwert und der durchschnittlichen Szenenleuchtdichte

¹⁴Verhältnis zwischen der höchsten (99,9%) und der niedrigsten (0,1%) Leuchtdichte der Szene in Blendenstufen

ausgeprägt ist als bei den Fußball-Szenen. Insgesamt wurden also bei Szenen mit einer größeren Spreizung der Leuchtdichtewerte eine geringere Helligkeit gewählt. Da eine größere Spreizung der Werte auch die Wahrscheinlichkeit von geclippten Lichtern erhöht, wäre dies eine mögliche Erklärung.

An dieser Stelle sollte angemerkt werden, dass sehr helle Szenen oft einen geringen Kontrastumfang oder Szenenkontrast aufweisen, da es ihnen an Schattenpartien fehlt. In den hier vorliegenden Szenen ist dies allerdings nicht der Fall. Deshalb ergibt sich eine Korrelation zwischen der durchschnittlichen Leuchtdichte der Szene und dem eingestellten Helligkeitswert. Dieses Ergebnis ist entsprechend mit Vorsicht zu betrachten, da es sich aufgrund der dargelegten Szenenauswahl um eine Scheinkorrelation handeln könnte.

Für den eingestellten Kontrastwert zeichnet die schwarze Trendlinie in allen Fällen eine positive Korrelation ($r=0,34$, $0,51$ und $0,53$). Allerdings kann dieser Eindruck bei separater Betrachtung der beiden Bildergruppen nicht bestätigt werden. Die farbigen Trendlinien weisen eine negative Steigung aus. Das Problem dürfte in der Auswahl der Szenen zu suchen sein. Zum einen besitzen alle Fußballszene eine charakteristische Trennung zwischen Lichtern und Schatten, die von der Tribüne verursacht wird. Diese unnatürliche Trennung verleiht dem Bild aber keine Tiefe, wie es sonst oft der Fall ist, sondern löst eine zusammengehörige Handlung in zwei Teile auf. Eine flachere Gradation bringt die Teile wieder näher zusammen. Deshalb dürfte der eingestellte Kontrastwert der Probanden so niedrig ausgefallen sein. Zum anderen fehlen im Bilderpool Aufnahmen mit sehr geringem Kontrastumfang, bei denen eine flache Gradation zu keinem inhaltlichen Mehrwert führt, sondern einzig das Bild kontrastarm wirken lässt.¹⁵ Aufgrund der Ergebnisse beim Coloristen-Tests kann angenommen werden, dass in solch einem Fall ein deutlich höherer Kontrastwert gewählt worden wäre.

Zusammenfassend sollte deshalb festgehalten werden, dass nicht nur der Szenenkontrast für die zu wählende Gradation wichtig ist, sondern auch dessen Verteilung innerhalb der einzelnen Bildpartien. Große Schattenpartien deuten eher drauf hin, dass sich dort eine wichtige Handlung abspielt, die aufgehellt werden sollte. Kleine Schattenpartien sorgen hingegen für eine Tiefenwirkung und Plastizität, welche bei zu starker Aufhellung aufgelöst wird.

Für die Homogenisierung kann, wie schon bei den äußeren Bedienungen, keine Korrelation nachgewiesen werden.¹⁶ Alle Trendlinien verlaufen recht horizontal. Die zugehörigen Korrelationskoeffizienten lauten $0,19$, $0,00$ und $-0,23$. Eine mögliche Erklärung lässt sich in einer Art Selbstregulierung finden. Besitzt eine Szene nur einen geringen Kontrast,

¹⁵Ein prädestiniertes Beispiel für solch eine Szene wäre ein Skirennen bei Sonnenschein. Die meisten Helligkeitswerte befinden sich in einen zwar hellen, aber engen Leuchtdichtekorridor.

¹⁶Der Ausreißer (Bild 9) wurde bei der Auswertung der Homogenisierung nicht mit berücksichtigt.

fällt der Unterschied zwischen globalem und lokalem Adaptionniveau klein aus. Der Homogenisierungseffekt ist entsprechend weniger stark ausgeprägt. Darüber hinaus muss auch hier auf die Einschränkung hingewiesen werden, dass keine Bildvorlagen mit sehr geringem Kontrastumfang ausgewählt wurden.

Generell muss berücksichtigt werden, dass auch wenn versucht wurde, einen möglichst großen Umfang an äußeren Bedingungen, Einstellungsparametern und Bildern zu wählen, dies letztendlich immer nur einen Ausschnitt aller denkbaren Optionen bilden kann. Gerade bei der Messung der Korrelation mit den Szenenparametern, sind 12 Bilder eher eine kleine Testsumme.

Des Weiteren konnten nur Standbilder betrachtet werden. Dies ist ein Mangel vieler Studien in der Vergangenheit und sollte noch einmal mit Videos verifiziert werden, welches im finalen Probandentest in Kapitel 7 sodann erfolgt. Allerdings bilden Standbilder auch den Ausgangspunkt beim Grading in der Postproduktion, indem ein repräsentatives Bild der Sequenz ausgewählt wird. In einem laufenden Video kann die Wirkung der Regler nur schwer beobachtet werden.

4.5.3 Zusammenfassung Probanden-Grading

Mit dem hier beschriebenen Experiment wurde versucht, quantitative Werte für die visuelle Präferenz bei einer Kontrastkompression zu ermitteln. Mithilfe von durchschnittlichen Fernsehzuschauern, die 12 Testbilder selber bearbeiten sollten, wurde analysiert, inwieweit der eingestellte Wert für mehrerer Bildbearbeitungsparameter von der Display- und der Umgebungsleuchtdichte sowie dem Bildinhalt abhängt. Diese Untersuchungen könnten in Zukunft weiter vertieft werden. Die Resultate bilden aber bereits jetzt einen sehr guten Ausgangspunkt für das gesetzte Ziel. Schlussendlich können folgende Aspekte festgehalten werden:

- Eine höhere Reproduktion des Szenenkontrastumfanges als im heutigen Broadcast üblich, wird präferiert, obwohl dies zu einem subjektiv geringeren Kontrast führt. Mindestens 12 Blendenstufen wurden in allen Fällen gewählt.
- Der gewählte Helligkeitswert verhält sich wie in früheren Studien. Für hellere Displays wird ein geringerer Wert gewählt. Bei hellerem Umgebungslicht ist ein höherer Wert präferiert. Allerdings ist der zweite Effekt fast vernachlässigbar klein und nicht statistisch signifikant.
- Die Homogenisierung ist unabhängig von äußeren Bedingungen. Auch bei HDR Displays sollte Tonemapping mit lokalem Einfluss zum Einsatz kommen, würde

man der Zuschauerpräferenz folgen. Das genaue Verhältnis zwischen lokaler und globaler Homogenisierung ist vom Bildinhalt abhängig. Eine Gleichgewichtung führt aber bei fast allen Szenen zu einem ansprechenden Resultat.

4.6 Zwischenfazit bezüglich der Helligkeitspräferenz

Die bisher bekannten Studien zur Helligkeitspräferenz hatten sich lediglich auf die klassische Aussteuerung im Sinne von *Gamma* und *Gain* konzentriert. Deren Ergebnisse konnten, wie oben dargelegt, bestätigt werden. Im Hinblick auf die Gradation und der damit verbundenen Abwägungsentscheidung zwischen Steilheit und Informationsgehalt sowie der Stärke der Homogenisierung, fehlte es vor diesem Versuch an aussagekräftigen Studien.

Insgesamt lassen sich die Einzelbeobachtungen in diesem Kapitel zu einem Gesamtbild zusammenfügen. Übergeordnet steht die zentrale Forderung, dass alle relevanten Informationen im Bild gut sichtbar sein sollen. Dabei wird nach Möglichkeit ein kontrastreiches Bild bevorzugt, solange dies nicht zu einem zu großen Informationsverlust führt. Die Aussteuerung der Helligkeit basiert auf der Bildvorlage und den räumlichen Gegebenheiten. Auch Clipping kann dienlich sein, da es sich positiv auf Helligkeit und Kontrast auswirkt. Besser wäre es aber, diese Aspekte auf anderem Wege und ohne Clipping zu erreichen, um so mehr Informationen erhalten zu können.

Die verbesserte Bildqualität aufgrund von **HDR** basiert nicht in erster Linie auf der größeren Leuchtdichte oder dem besseren Schwarzwert. Vielmehr ist es entscheidend, den Kontrastumfang der Szene bestmöglich zu reproduzieren. Dies kann auch (größtenteils) auf einem **SDR**-Display realisiert werden, wenn die Gradation und die Homogenisierung entsprechend optimiert werden. Selbst bei **HDR** ist eine lineare Repräsentation der Leuchtdichtewerte nicht gewünscht. Dies kann mithilfe der unter **4.1** gemessenen Maskierungseffekte erklärt werden. Zwar ist das **HDR**-Display in der Lage, fast beliebig dunkle Grauabstufungen darzustellen, allerdings wird die Wahrnehmbarkeitsschwelle durch helle Bildelemente in der Umgebung teils so stark herabgesetzt, dass keine Unterscheidung mehr möglich ist.

Der Autor leitet daraus folgendes Fazit ab: „So viel Kontrastumfang reproduzieren wie nötig, aber so wenig wie möglich.“ Sprich, es wäre nötig abzuschätzen, welche Helligkeitswerte von Relevanz sind und erhalten werden sollten und auf welche zugunsten einer steileren Gradation verzichtet werden kann. Solch eine Prüfung wäre sowohl im globalen, also pro Bild denkbar, als auch im sektionalen, also in einzelnen Bildabschnitten. Der

sektionale Aspekt würde die Homogenisierung abbilden und so in der Lage sein, Clipping weiter zu reduzieren, wobei gleichzeitig der lokale Kontrasteindruck gesteigert würde.

In Kapitel [6](#) soll ein Algorithmus erarbeitet werden, der mithilfe von statistischen Auswertungen eine solche Relevanzprüfung vorzunehmen vermag. Im nun folgenden Kapitel [5](#) gilt es aber zunächst die Präferenz bezüglich der Farbdarstellung in analoger Weise zu bestimmen.

Kapitel 5

Visuelle Präferenz bezüglich der Farbwahrnehmung

Der Fokus der Arbeit liegt zwar auf der Kompression der Leuchtdichte, doch wie zu Beginn bereits erwähnt, ist diese untrennbar mit der Farbdarstellung verbunden. Deshalb sollen im folgenden Kapitel die Farbräume und Möglichkeiten zur Erhaltung eines konstanten Farbeindruckes näher erörtert werden.

5.1 RGB-Farbgamut

Die Darstellung von digitalen Bildern erfolgt bei der Aufnahme in den meisten Fällen durch einen additiven RGB-Farbraum. Dabei ergibt sich der Farbeindruck durch die jeweiligen Anteile an den drei Grundfarben Rot, Grün und Blau. Diese Grundfarben werden auch Primärvalenzen genannt. In einer zweidimensionalen Darstellung spannen sie ein Dreieck auf. Hierfür findet sich der Ausdruck Farbgamut. Eine Mischfarbe der Primärvalenzen kann folglich als Punkt in das Farbgamut eingetragen werden. Dabei gilt, dass je geringer der Abstand zu einer Primärvalenz ist, desto höher ist der Anteil der jeweiligen Farbkomponente.

Folglich handelt sich also um kein absolutes System. Die Beschreibung der Farbe findet nur in Relation zu vorher definierten Primärvalenzen statt, welche mit einem Spektrum an Wellenlängen gleichzusetzen sind. Die beim heutigen **HDTV**-Broadcast gültige Vorschrift ist in der ITU-R BT.709 [\[54\]](#) spezifiziert und wurde seinerzeit im Hinblick auf die technischen Möglichkeiten der Kathodenstrahlröhre beschrieben. Die dort angegebenen Primärvalenzen sind identisch mit dem so genannten sRGB-Farbraum, der in der Computergrafik und in **JPEG**-Bilddaten Verwendung findet.

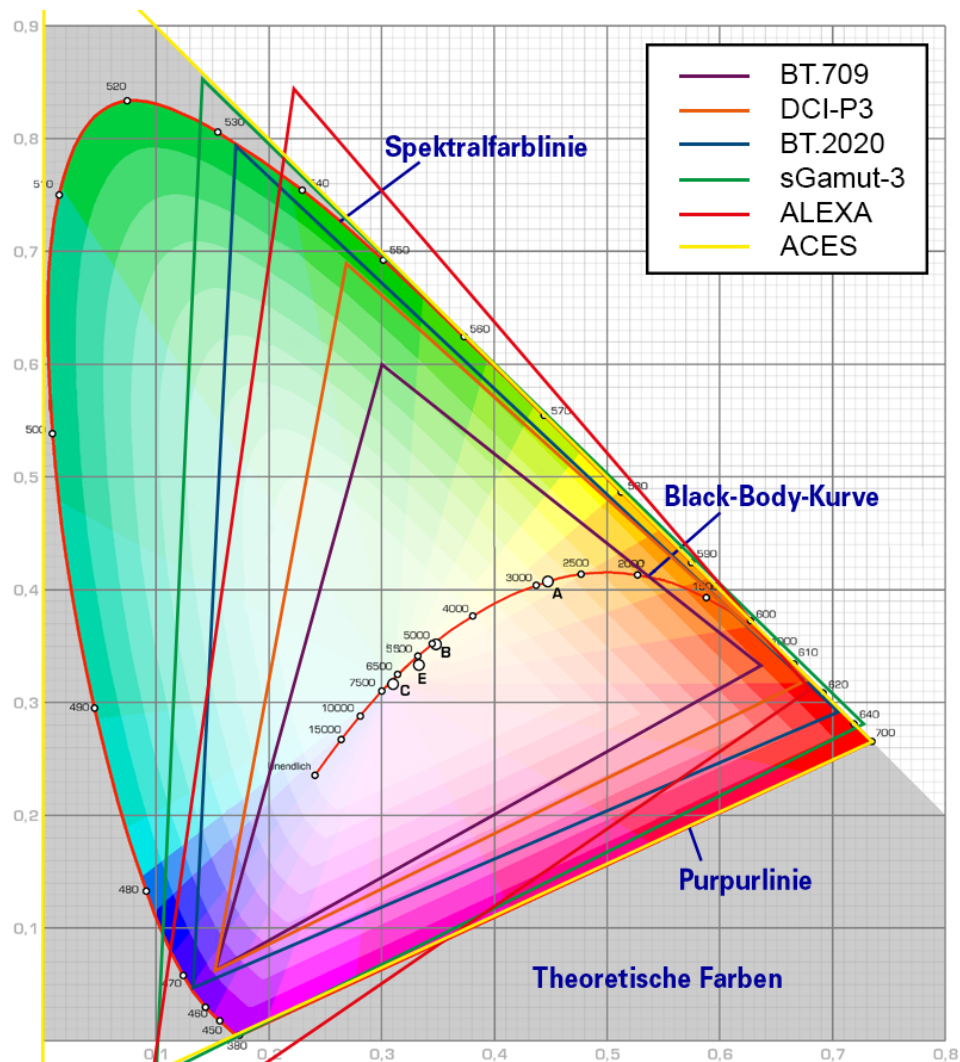


ABBILDUNG 5.1: Vergleich verschiedener Farbgamuts anhand der CIE-Normfarbtafel auf Basis der Grafikvorlage von Torge Anders (<https://de.wikipedia.org/wiki/RGB-Farbraum#/media/Datei:CIE-Normfarbtafel.png>)

Die Sensoren moderner Kameras besitzen in der Regel allerdings andere Primärvalenzen, die ein weitaus größeres Farbgamut aufspannen. Die Betrachtung dieses Aspekts ist deshalb so wichtig, da eine Aufzeichnung mit erhöhtem Kontrastumfang fast immer mit einem erweiterten Farbgamut einhergeht. Wird das RAW-Format zur Aufzeichnung gewählt, entsprechen RGB den Primärvalenzen der Kamera. Bei der Arri ALEXA heißt das hieraus resultierende Gamut beispielsweise ALEXA, bei der Sony F55 sGamut3. Wird das für den **HDR**-Broadcast angeordnete Format gewählt, entsprechen RGB den Primärvalenzen nach ITU-R BT.2020 [57] bzw. ITU-R BT.2100 [58]. Darüber hinaus gibt es noch den Fall, dass digitale Kinodaten vorliegen, welche den Farbgamut DCI-P3 [118] oder den Universalfarbraum Academy Color Encoding Specification (**ACES**) [113] besitzen. Letztgenannter geht deutlich über das sogenannte *Hufeisen* hinaus, das alle für den Menschen sichtbaren Farben umfasst. Ein Vergleich der angesprochenen Farbgamuts findet sich in Abbildung 5.1.

Bei der Darstellung eines größeren in einen kleineren Farbraum muss folglich eine Farbraumtransformation vorgenommen werden.¹ Dabei kann es passieren, dass Werte außerhalb des Zielgamuts liegen. Dies äußert sich als negativer Anteil einer oder mehrerer Primärvalenzen. Auch hierfür gibt es, genauso wie bei der Leuchtdichte, spezielle Mapping-Techniken. Es wird vom so genannten Gamut-Mapping gesprochen. Eine kleine Übersicht findet sich in [42].

In dieser Arbeit soll der Fokus aber auf einem anderen Problem liegen. Im vorherigen Kapitel wurde die Präferenz bezüglich der Helligkeitsdarstellung bei einer Kontrastkompression evaluiert. Es hat sich gezeigt, dass eine durchaus deutliche Kompression sinnvoll sein kann. Allerdings führt die Änderung der Helligkeit im Bild auch immer zu einem anderen Farbeindruck. Besonders kritisch verhält sich dabei die Sättigung, wie in Abschnitt 3.2 deutlich wurde. Deshalb soll in diesem Kapitel geprüft werden, wie es gelingen kann, trotzdem einen konstanten Sättigungseindruck zu erreichen. Natürlich sollte dabei der Farbton so wenig wie möglich beeinflusst werden.

5.2 Klassische Sättigungskompensation auf RGB

Mit einigen wenigen Ausnahmen (z. B. [101]) ist es allgemein anerkannt, die Kontrastkompression einzig auf Basis der Leuchtdichte zu berechnen und somit die Farbkanäle nicht unterschiedlich zu behandeln. [104] Der Farbton wird so im Vergleich weniger beeinflusst. Die Leuchtdichte L wird dazu, falls nötig, aus den linearen RGB-Eingangswerten gebildet. Für ein Bild mit BT.709-Primärvalenzen ergibt sich beispielsweise folgende Gleichung [98]²:

$$L_w = 0,2126 * R + 0,7152 * G + 0,0722 * B \quad (5.1)$$

Nach der Kompression muss die modifizierte Helligkeitskomponente (L_d) mit den unkomprimierten Farbwerten rekombiniert werden, um das finale Farbbild zu erhalten. Um dabei Farbverschiebungen auf ein Minimum zu begrenzen, werden die Verhältnisse zwischen den Farbkanälen erhalten. [104] Klassischerweise kommt dabei folgende Funktion von Schlick [108] zum Einsatz. Dabei steht M für eine der drei Farben und w (world) bzw. d (display) implizieren, ob es sich um die Eingangs- oder Ausgangswerte handelt.

¹Dies gilt natürlich auch umgekehrt.

²Eine sehr ähnliche Gleichung findet sich auch bei Poynton [98] auf Seite 122. Dort wird das Luma mit den identischen Koeffizienten aus den verzerrten RGB-Eingangswerten gebildet. Wie aber in einem der nächsten Abschnitte gezeigt werden wird, ist YC_bC_r nur unzureichend dekorreliert, wodurch sich die Helligkeitsrelationen bei einzelnen Farbtönen zwischen Luma (Y') und Leuchtdichte (L_w) gravierend verschieben können.

$$M_d = \frac{M_w}{L_w} L_d \quad (5.2)$$

Bei einer starken Kontrastkompression ist allerdings zu beobachten, dass die Farben übersättigt wirken. Tumblin et al. [125] und Fattal et al. [36] führen deshalb den Exponenten s ein, der quasi eine Gamma-Funktion für jeden Kanal darstellt. Dabei steht s nicht für die absolute Sättigung, sondern für die durch die Formel herbeigeführte Sättigungsänderung.

$$M_d = \left(\frac{M_w}{L_w} \right)^s L_d \quad (5.3)$$

Eine ähnliche Alternative bieten Mantiuk et al. [81] an:

$$M_d = \left(\left(\frac{M_w}{L_w} - 1 \right) s + 1 \right) L_d \quad (5.4)$$

Die erste Gleichung (5.3) führt zu einem sehr konstanten Farbton, der allerdings mit sehr starken Änderungen in der Helligkeit der Farbwerte einhergeht. Bei der zweiten Gleichung (5.4) ist es genau entgegengesetzt. Hier kann eine Verschiebung bei der Helligkeit vermieden werden. Dafür kommt es zu größeren Abweichungen beim Farbton - besonders für Rot und Blau.

Sowohl für die erste als auch für die zweite Gleichung haben Mantiuk et al. [81] eine Formel entwickelt, um s zu berechnen und somit automatisch setzen zu können. c gibt dabei die Stärke der vorgenommenen Kontrastkompression wieder. Die Zahlenwerte für k_1 und k_2 wurden anschließend empirisch mithilfe von subjektiven Untersuchungen bestimmt.

$$s(c) = \frac{(1 + k_1)c^{k_2}}{1 + k_1c^{k_2}} \quad (5.5)$$

Ein Abgleich mit dem zu Beginn definierten Ziel, sowohl Farbton als auch den Sättigungseindruck konstant zu halten, fällt für Gleichung 5.3 positiv aus. Bei genauerer Betrachtung ist allerdings die daraus resultierende Veränderung der Helligkeitskomponente nicht zu tolerieren. Ihr Einfluss auf das Ergebnis des TMO ist zu gravierend. Auch räumen Mantiuk et al. ein, dass die Stärke der Kontrastkompression und somit ein Wert für c auf die gezeigte Weise nur für globale Operatoren berechnet werden kann. Lokale Operatoren weisen für jeden Pixel andere Verzerrungen auf. [81] [97]

Die Autoren [81] ziehen deshalb in Erwägung, die Kontrastkompression in einem Farbraum durchzuführen, der besser an die menschliche Wahrnehmung angepasst ist und eine

Dekorrelation zwischen der Helligkeits- und der Farbkomponente bewirkt. Die Theorie hinter dem Ansatz ist, dass eine perfekte Entkopplung von Helligkeit und Farbe es erlauben würde, die Helligkeit beliebig zu modifizieren, ohne den Sättigungseindruck zu ändern - bzw. eine bessere Entkopplung würde zumindest das Problem verringern. Klassische Beispiele für wahrnehmungsgerechte Farbräume sind CIELAB oder IPT. Im Folgenden soll eine exemplarische Farbraumtransformation gezeigt werden.

5.3 XYZ-Farbraum

Um ein RGB-Bild in CIELAB oder IPT transformieren zu können, muss ein Umweg über den von der Commission Internationale de l'Éclairage (CIE) aufgestellten Normfarbraum XYZ gegangen werden. X , Y und Z sind dabei virtuelle Primärvalenzen, also rein rechnerische Koordinaten, die auf Basis der Zapfenempfindlichkeit erstellt wurden. Der Farbraum ist nicht begrenzt und im Vergleich zu RGB absolut. Jeder Eingangsfarbraum kann ohne Verluste dorthin transformiert werden. Somit sind die verschiedenen RGB-Farbräume über ihre jeweilige Definition in XYZ miteinander verbunden und können so ineinander überführt werden.³ Die entsprechenden Matrizen im nachstehend gezeigten Muster werden von den Kameraherstellern bereitgestellt.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (5.6)$$

Für ITU-R BT.709 [54] nimmt die Matrix folgende Form an:⁴

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0,4124 & 0,3576 & 0,1805 \\ 0,2126 & 0,7152 & 0,0722 \\ 0,0193 & 0,1192 & 0,9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (5.7)$$

Der Wertebereich bei XYZ ist nicht vorgegeben. Sehr gebräuchlich ist die Darstellung als Relative vom Referenzweiß (100 cd/m^2), wodurch Y beim klassischen SDR¹ auf eins normiert wird. Aber auch die Darstellung als cd/m^2 wird des öfteren genutzt.

³Wenn gewünscht, besteht dabei des Weiteren die Möglichkeit eines nachträglichen Weißabgleiches, indem X und Z verändert werden und so sich die Relationen zu Y ändern. Bei der Konvertierung nach IPT wird D65 als Weißpunkt angenommen.

⁴Weitere Beispiele für andere Farbräume finden sich im Anhang

5.4 IPT-Farbraum

Der IPT-Farbraum ist eine Weiterentwicklung des CIELAB und soll an dieser Stelle näher beleuchtet werden. Ebeners und Fairchilds [29] Ziel war es, eine Beschreibung für den Farbton zu finden, so dass ein konstanter Farbton-Wert einem konstanten Farbton entspricht, unabhängig der Werte für die Helligkeitskomponente und dem Chroma. Dazu werden die XYZ-Werte zunächst mit Hilfe der Hunt-Pointer-Estevéz-Matrix (MHPED65) in LMS-Werte umgerechnet und anschließend nichtlinear verzerrt. LMS steht für long, medium und short, womit die drei Typen von Zapfen beschrieben werden und sollte nicht mit der Leuchtdichte L oder der Sättigung S verwechselt werden.

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0,4002 & 0,7075 & -0,0807 \\ -0,2280 & 1,1500 & 0,0612 \\ 0,0000 & 0,0000 & 0,9184 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (5.8)$$

$$L' = |L|^{0,43} \quad M' = |M|^{0,43} \quad S' = |S|^{0,43} \quad (5.9)$$

Abschließend werden L'M'S' mithilfe einer weiteren Matrix in IPT umgerechnet.

$$\begin{bmatrix} I \\ P \\ T \end{bmatrix} = \begin{bmatrix} 0,4000 & 0,4000 & 0,2000 \\ 4,4550 & -4,8510 & 0,390 \\ 0,8056 & 0,3572 & -1,1628 \end{bmatrix} \begin{bmatrix} L' \\ M' \\ S' \end{bmatrix} \quad (5.10)$$

Alle diese Schritte sind reversibel und werden entsprechend bei der Rücktransformation umgekehrt.

5.5 Sättigungskompensation nach Pouli

Pouli et al. [97] bieten eine Lösung für die Farbkorrektur im IPT-Farbraum an. Dazu wird sowohl das Eingangs- als auch das Ausgangsbild zunächst nach IPT gewandelt. Anschließend findet mit Hilfe der nachstehenden Gleichungen eine Konvertierung nach ICh statt, um Chroma (C), Farbton (h) und Luma (I) zu isolieren.

$$h = \arctan\left(\frac{P}{T}\right) \quad (5.11)$$

$$C = \sqrt{P^2 + T^2} \quad (5.12)$$

Anschließend wird die Sättigung nach folgender modifizierter Formel berechnet:

$$S(C,I) = \frac{C}{\sqrt{C^2 + I^2}} \quad (5.13)$$

Die Lösung der Sättigungsproblematik besteht also darin, die in Kapitel 1 gängige Definition der Sättigung durch eine neue Berechnungsvorschrift zu ersetzen. Diese Formel geht auf Lübke [79] und Richter [106] zurück und wurde von Lübke 2013 [80] durch eine subjektive Untersuchung mit Probanden untermauert, in der eben jene Farbkarten in einer Lichtkabine gefühlt gleichabständig einordnen sollten.

Als Gesamtfunktion postulieren Pouli et al. schließlich folgende Formel, mit der versucht wird, einen konstanten Sättigungseindruck zu erreichen. C_c steht dabei für das kompensierte Chroma. Der Index 0 markiert die vor dem Tonemapping vorliegenden Werte, während t auf das durchgeführte Tonemapping hinweist. Diese Art der Indexierung wird auch im restlichen Kapitel der vorliegenden Arbeit beibehalten.

$$C_c = \frac{S(C_0, I_0)}{S(C_t \frac{I_0}{I_t}, I_t)} \frac{I_0}{I_t} C_t \quad (5.14)$$

5.6 hdrIPT, IC_tC_p und YC_bC_r

Der vorangegangene Abschnitt zeigt, dass selbst im wahrnehmungsgerechten IPT-Farbraum auf eine Sättigungskorrektur bei der Kontrastkompression nicht verzichtet werden kann. Dies könnte daran liegen, dass der IPT-Farbraum eigentlich ein SDR-Farbraum ist. Entsprechend beziehen sich die verwendeten nichtlinearen Funktionen zur Erreichung der gefühlten Gleichabständigkeit auf eben jenen Leuchtdichtebereich. Die Funktionen müssen in der Folge nicht zwangsläufig auch für einen höheren Kontrastumfang bei HDR anwendbar sein.

Als logische Fortsetzung des IPT entwickelten Fairchild et al. deshalb den hdrIPT. Die wahrnehmungsgerechte Verzerrung der drei Komponenten L , M und S wird dabei durch folgende Formel realisiert, wobei ω für eine der Komponenten steht [34]:

$$f(\omega) = 2,47 \frac{\omega^{0,58}}{\omega^{0,58} + 20^{0,58}} + 0,0002 \quad (5.15)$$

Mit IC_tC_p ist in der ITU-R BT.2100 [58] eine Abwandlung dieses Farbraums zu finden, welche auf die Dolby Laboratories [23] zurückgeht. Im Vergleich zum IPT weisen die

Matrizen von XYZ nach LMS und von LMS nach IC_tC_p leichte Unterschiede auf (s. Anhang). Das Koordinatensystem wurde somit leicht rotiert. Die wahrnehmungsgerechte Verzerrung wird über die PQ-Kurve aus Kapitel 2 realisiert. Sie ist so ausgelegt, dass der Leuchtdichtewert 10.000 cd/m^2 auf 1 abgebildet wird. Dolby verspricht eine nahezu perfekte Dekorrelation von Farb- und Helligkeitskomponente - besonders bei hohen Leuchtdichtewerten. Wäre dies der Fall, würde es die Anwendung einer Sättigungskompensation obsolet machen. Im Verlauf des Kapitels wird sich allerdings zeigen, dass dies nicht zutrifft. Hintergrund ist die unterschiedliche Interpretation von Dekorrelation der Farb- und Helligkeitskomponente. Für die Zielsetzung der Arbeit ist es nötig, dass die Helligkeit beliebig verändert werden kann, ohne sich dabei auf den Farbeindruck auszuwirken. Die Helligkeit darf sich nicht in den Farbinformationen widerspiegeln. Dolby [23] meint allerdings damit, dass der Farbton keinen Einfluss auf die Berechnung der Helligkeit haben darf. Zwar gelingt durch die PQ-Kurve auch eine relativ hohe Unabhängigkeit von Änderungen der Helligkeit wie Abbildung 5.2 zeigt, eine vollständige Dekorrelation kann aber nicht erreicht werden.

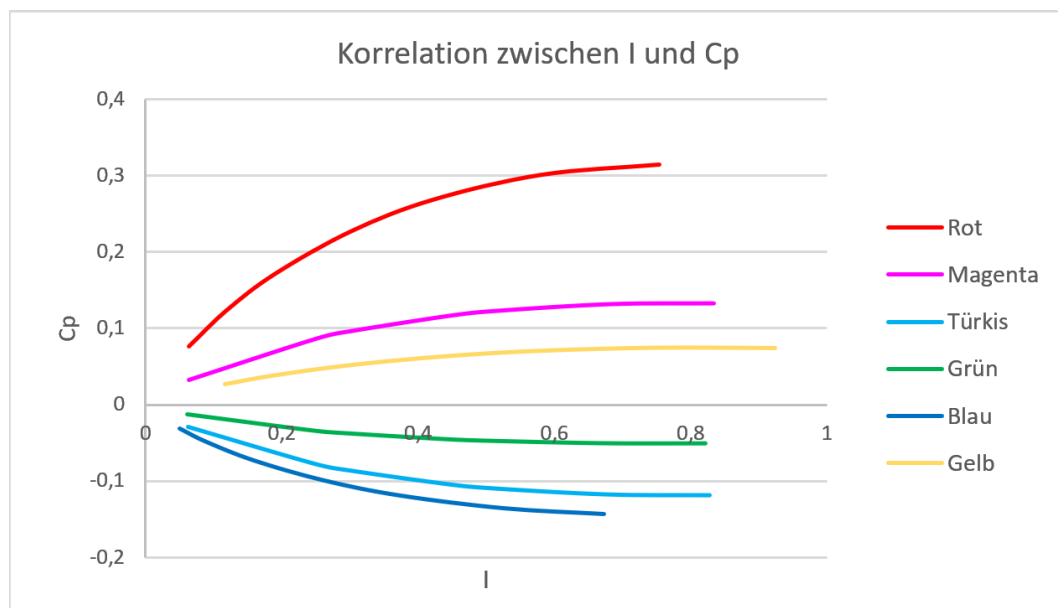


ABBILDUNG 5.2: Konstanter Farbort (xy) bei veränderter Leuchtdichte dargestellt in I und C_p

Die derzeit beim SDR-Broadcast gebräuchlichste Darstellung der Helligkeits- und Farbinformation ist allerdings YC_bC_r . Auch hier wird versucht, Luma von Chroma zu trennen. C_b und C_r sind dabei Farbdifferenzsignale, die sich direkt aus den vorverzerrten Primärvalenzen $R'G'B'$ bilden. Die genauen Transformationsmatrizen finden sich im Anhang. Für HDR wurde spezifiziert [58], dass anstelle der klassischen Gamma-Vorverzerrung auch die PQ-Kurve oder die HLG-Kurve genutzt werden kann. In der Folge findet $Y'C_b'C_r'$ auch bei HDR Anwendung. Aus der Tradition heraus tendiert die

Branche derzeit dazu, weiterhin mit $Y'C_b'C_r'$ zu arbeiten. Die Dekorrelation bezüglich des Farbtons fällt allerdings schlechter aus, als es bei IC_tC_p der Fall ist. [23] [100]

Strittig ist, ob $Y'C_b'C_r'$ und IC_tC_p als Farbräume oder nur als Farbdifferenzsignale bezeichnet werden können. Bei der Betrachtung der Konstruktionsweisen von IPT und IC_tC_p wird allerdings deutlich, dass diese identisch sind und Abweichungen nur bei den Zahlenwerten sowie der geänderten Verzerrungsformel zu finden sind. Da es sich beim IPT um einen Farbraum handelt, muss folglich, nach Ansicht des Autors, auch IC_tC_p als Farbraum bezeichnet werden können. Für $Y'C_b'C_r'$ werden zwar andere Transformationen verwendet, trotzdem ist das Resultat eine dreidimensionale Farbdarstellung. Deshalb wird $Y'C_b'C_r'$ für die folgende Arbeit ebenfalls unter Farbraum geführt.

5.7 Farberscheinungsmodelle

Neben den klassischen Farbräumen gibt es noch Farberscheinungsmodelle (engl. color appearance models), welche Erscheinungsphänomene mitberücksichtigen. Ein klassisches Beispiel für solch ein Phänomen ist der Hunt-Effekt [51]. Dieser besagt, dass eine Erhöhung der Leuchtdichte eines Reizes zu einer Erhöhung der Buntheit führt. Ähnliche Beobachtungen finden sich auch beim Stevens-Effekt⁵ oder der Bartleson-Breneman-Gleichung⁶ wieder. Darüber hinaus werden teilweise spatiale Wechselwirkungen betrachtet, wie z. B. beim iCAM [63]. In der Folge sind die Modelle im Vergleich zu einem einfachen Farbraum deutlich komplexer und benötigen darüber hinaus mehr Informationen für die Simulation der Erscheinungsphänomene. Deshalb rät Fairchild [33] dazu, Farberscheinungsmodelle nur dann zu nutzen, wenn sie wirklich absolut notwendig („absolutely necessary“) sind.

Aus gleich drei Gründen sieht der Autor Farberscheinungsmodelle für die in dieser Arbeit definierte Problemstellung als nicht zielführend an. Erstens werden, wie gerade erwähnt, für die Simulation der Erscheinungsphänomene bestimmte Informationen benötigt. Dies beinhaltet z. B. die absolute Leuchtdichten der Szene (scene referred values), welche aber durch die gewählte Blendeneinstellung bei der Aufnahme unbekannt sind oder Informationen über die Wiedergabe, welche aufgrund der Diversität bei den Empfängern nicht vereinheitlicht angegeben werden können. Entsprechend sind die Modelle entweder nicht in der Lage ihr volles Potenzial auszuschöpfen oder kommen gar aufgrund von inkorrekten Angaben zu falschen Ergebnissen.

⁵Eine Erhöhung der Leuchtdichte eines Reizes führt zu einer Erhöhung des Kontrastes. In der Folge wirkt die Szene bei Tageslicht kontrastreicher als in der Dämmerung. [120]

⁶Eine Erhöhung der Umgebungsleuchtdichte führt zu einer Erhöhung des Kontrastes. In der Folge wirkt die Reproduktion einer Szene als Print in einem Raum bei Tageslicht kontrastreicher als in einem völlig abgedunkelten Raum. Grund hierfür ist, dass dunkle Bereiche der Reproduktion im dunklen Umfeld heller wirken, wohingegen sich die Wahrnehmung der hellen Anteile kaum verändert. [6]

Zweitens wurde bereits in Abschnitt [3.2.4](#) aufgezeigt, dass es nicht das Ziel ist, eine möglichst realistische Abbildung der Szene zu schaffen, sondern eher photographische Farben zu produzieren. Diese Farben werden bei einer idealen Belichtung mit der Kamera erreicht. [\[81\]](#) Sie werden beispielsweise bei dunklen Lichtverhältnissen nicht blass und ignorieren somit den Hunt-Effekt. In der Folge sind sie näher an den Gedächtnisfarben eines bestimmten Objekts. [\[81\]](#)

Als dritten Punkt kommt neben den wissenschaftlichen Aspekten die Problematik bei einer späteren Nutzung in der Praxis hinzu. Hohe Komplexität bedeutet auch rechenintensiv. Es ist unklar, inwiefern dies die Verwendung in unterschiedlichen Applikationen eingrenzt.

Fairchild [\[33\]](#) sagt darüber hinaus, dass bereits CIELAB für gewisse Anwendungen als eine Art Farberscheinungsmodell gesehen werden kann und bereits gute Ergebnisse erzielt. Neben dem Nichtvorhandensein der oben beschriebenen Erscheinungsphänomene, welche für die hier angestrebte Nutzung auch unerwünscht sind, kritisiert er einzig die unzureichend farbliche Adaption bei abweichenden Beleuchtungssituationen zwischen Aufnahme und Wiedergabe. In Anbetracht der Verbesserungen von CIELAB zu IPT und von IPT und IC_tC_p kann IC_tC_p als ideale Ausgangsbasis für die gegebene Problemstellung gesehen werden. Auf die Nutzung eines Farberscheinungsmodells im engeren Sinne soll verzichtet werden.

5.8 Erste Messungen der subjektiven Präferenz bezüglich der Sättigungswahrnehmung

Im Folgenden soll die Präferenz des Zuschauers qualitativ und quantitativ gemessen werden. Dabei soll auch geklärt werden, was genau unter „Sättigung“ bei der gegebenen Problemstellung zu verstehen ist und ob „Sättigung“ dafür überhaupt die korrekte Bezeichnung ist. Ferner gilt es zu klären, ob IC_tC_p solche Überlegungen nicht obsolet macht. Falls nicht, soll eine Methode zur Kompensation aufgezeigt werden, die nach Möglichkeit ohne aufwendige Transformationen in andere Farbräume auskommt, sondern auf Basis der beiden in ITU-R BT.2100 [\[58\]](#) definierten Farbräume ($Y'C_b'C_r'$ und IC_tC_p) arbeitet.

5.8.1 Sättigung beim Coloristentest

Zunächst sei kurz der Eindruck bezüglich der Sättigung beim Coloristentest wiedergegeben. Der Aufbau des Experimentes wurde bereits in Kapitel [4](#) beschrieben. Einzig

auf die Darstellung bezüglich der Sättigung war zu diesem Zeitpunkt verzichtet worden, da der inhaltliche Bezug nicht gegeben war. Wie auch schon bei den anderen Parametern sind in Abbildung 4.6 die einzelnen Gradings als Punkte eingetragen. Die Ordinate wird durch die durchschnittliche Wertung der Probanden gebildet. Die Sättigung wird in diesem Fall als Quotient aus Chroma und normierter Helligkeit verstanden.

Betrachtet man die Verteilung der einzelnen Punkte pro Testsequenzen, ist meist ein ansteigendes Muster auszumachen. Die am wenigsten gesättigte Variante weist jeweils die geringste Bewertung auf. Die höchste Bewertung entfällt in allen bis auf einen Fall auf die am zweit stärksten gesättigte Variante. Allerdings konnten drei Gradings mit den höchsten Sättigungswerten nur bedingt überzeugen. Dies könnte die Theorie unterstützen, dass sehr stark gesättigte Bilder ab einem gewissen Punkt als nicht mehr präferiert angesehen werden. Eine mögliche Erklärung neben einer zu hohen Sättigung wäre aber auch, dass alle drei Gradings eine verhältnismäßig geringe Helligkeit aufwiesen, was zu einer generellen Abwertung geführt haben könnte.

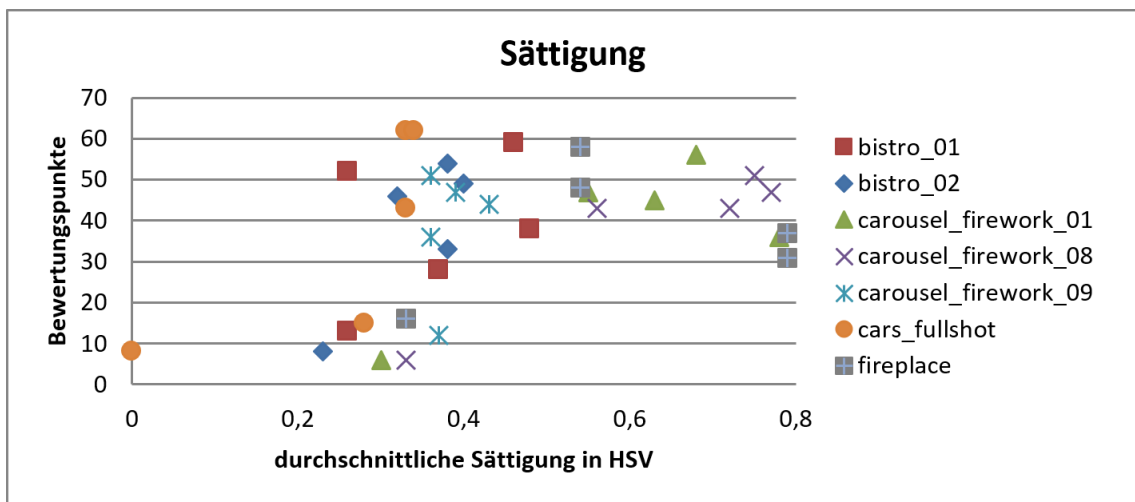


ABBILDUNG 5.3: Das Diagramm zeigt die Korrelation zwischen Probandenbewertung und Sättigung.

Wie auch schon bei den Parametern Helligkeit, Kontrast und Clipping konnten durch diesen Test nur einzelne Indizien gesammelt werden. Analog der Untersuchungen zur Helligkeitspräferenz sollte deshalb das Probanden-Grading auch für die Sättigung quantitative Ergebnisse liefern.

5.8.2 Sättigung oder Chroma

Dazu musste zunächst die Wirkweise des Sättigungsreglers definiert werden. Da zu diesem Zeitpunkt der IC_tC_p -Farbraum noch nicht zur Verfügung stand, wurde dies im IPT-Farbraum getan. Die bislang gewonnenen Erkenntnisse können auf eine Kernfrage

reduziert werden: Zu welchem Anteil muss die durch die Kontrastkompression herbeigeführte Helligkeitsänderung auch auf das Chroma übertragen werden, um den subjektiven Eindruck einer Sättigungsänderung zu kompensieren? Oder anders ausgedrückt: Zu welchen Anteilen soll die Zielsetzung konstante Sättigung und zu welchen Anteilen die Zielsetzung konstantes Chroma verfolgt werden?⁷

Ein konstantes Chroma könnte einfach erreicht werden, indem I um einen Faktor x verändert werden würde, C hingegen nicht. Die Indizes orientieren sich dabei an der Notation von Pouli. Es handelt sich somit bei C_t um eine der beiden Chroma-Komponenten nach der Chroma-Kompensation und nicht um C_t von IC_tC_p .

$$\begin{aligned} I_t &= I_0 x \\ C_t &= C_0 \end{aligned} \tag{5.16}$$

Würden, wie bei Schlicks-Formel, I und C , identisch behandelt, entspräche dies einer konstanten Sättigung. Allerdings konnte bereits gezeigt werden, dass diese Form einer „konstanten Sättigung“ oft eher einer Übersättigung entspricht.

$$\begin{aligned} I_t &= I_0 x \\ C_t &= C_0 x \end{aligned} \tag{5.17}$$

Als Gleichung kann diese Abwägungsentscheidung folgendermaßen ausgedrückt werden:

$$C_t = C_0((1 - k) + k * x) \tag{5.18}$$

Äquivalent kann der Faktor auch direkt auf beide Farbdifferenzsignale angewendet werden:

$$\begin{aligned} P_t &= P_0((1 - k) + k * x) \\ T_t &= T_0((1 - k) + k * x) \end{aligned} \tag{5.19}$$

Dabei spiegelt k das Mischungsverhältnis zwischen einer konstanten Sättigung ($k = 1$) und einem konstanten Chroma ($k = 0$) wider. Je besser der Farbraum Farbe und Helligkeit zu entkoppeln vermag, je geringer ist der erwartete Wert für k , da die Farbe nicht

⁷Jeweils in den ursprünglichen Definitionen von Chroma und Sättigung.

durch die geänderte Helligkeit beeinflusst wird. Die Vorgehensweise, den subjektiven Sättigungseindruck durch direkte Modifikation der beiden Farbdifferenzsignale zu erreichen, soll im folgenden als Chroma-Korrektur oder Chroma-Kompensation bezeichnet werden und die Methode als Koeffizienten-Methode.

Im Vergleich zu Pouli et al. [97] ist es bei der Chroma-Korrektur im allgemeinen und bei der Koeffizienten-Methode im speziellen (Gleichung 5.19) nicht nötig, eine erste Version des Ausgangsbildes zu erstellen, das nachfolgend noch korrigiert werden muss, sondern die finale Version wird ohne Umwege erreicht. Ferner kann auf die Berechnung von C , h oder s verzichtet werden. Die Anwendung erfolgt direkt auf P und T .

Durch die drei Gleichungen wird auch deutlich, warum der Zielkonflikt im SDR-Bereich nicht in gleichem Umfang besteht. Wird ein SDR-Bild in der Bearbeitung modifiziert, variiert x nur leicht um den Wert eins. Beide Zielsetzungen führen zu einem nahezu identischen Ergebnis. Bei der Kontrastkompression eines HDR-Bildes wird die Varianz von x um ein Vielfaches größer.

Der Sättigungsregler beim Probanden-Grading wies folglich einen Arbeitsbereich von 0 bis 1 (standardmäßig: 0,5) auf. 0 bedeutete, dass das Chroma von der Änderung der Leuchtdichte durch das Tonemapping unberührt blieb, während 1 bedeutete, dass die Änderung in gleicher Höhe hier ebenfalls Anwendung fand. Der Startwert setzte sich somit aus 50% Chroma und 50% Sättigung zusammen.

$$P_t = P_0 * \left((1 - saturation) + saturation * \left(\frac{I_t}{I_0} \right) \right) \quad (5.20)$$

5.8.3 Sättigung beim Probanden-Grading

Beim Probanden-Grading mussten die Teilnehmer 12 Bilder nach ihren Vorstellungen unter variierenden äußeren Bedingungen (Variation von Display- und Umgebungsleuchtdichte) einstellen. Hierfür standen ihnen vier Parameter zur Verfügung. Das genaue Vorgehen wurde bereits in Kapitel 4 beschrieben. An dieser Stelle sollen nun die Ergebnisse bezüglich des Sättigungsreglers vorgestellt werden. Die durchschnittlich pro Bild und äußerer Bedingung gewählten Werte werden in Abbildung 5.4 dargestellt. Sie liegen in einem recht engen Korridor von 0,63 bis 0,85. Das arithmetische Mittel liegt bei 0,73 und die Standardabweichung bei 0,05.

Wie auch schon bei den weiteren drei Einstellungsparametern des Probanden-Gradings wurden zur Verbesserung der Lesbarkeit die 95%-Konfidenzintervalle exemplarisch für Bild 1 in einem weiteren Diagramm dargestellt (Abbildung 5.6). Sie fallen ähnlich klein

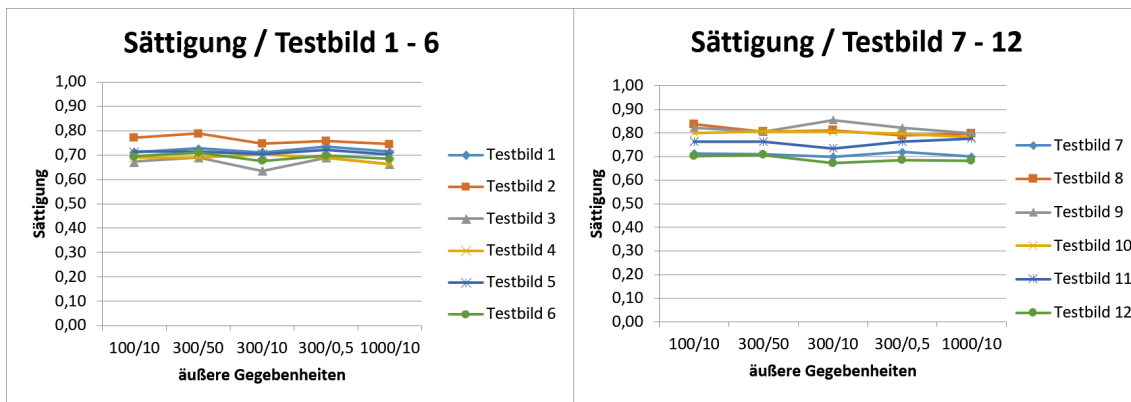


ABBILDUNG 5.4: Das Diagramm zeigt den durchschnittlichen Wert aller 34 Probanden für die Sättigung bei den verschiedenen äußeren Gegebenheiten und Bildern.

aus wie bei Kontrast und Helligkeit. Der Mittelwert der 34 Probanden besitzt entsprechend eine hohe Aussagekraft.

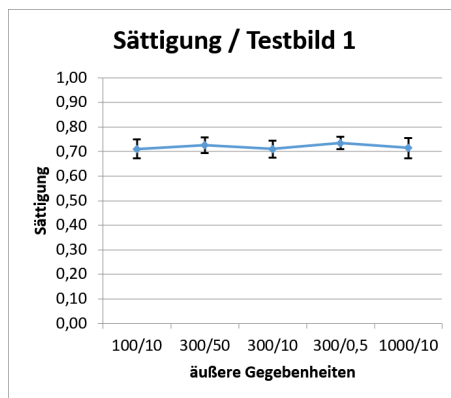


ABBILDUNG 5.5: Die Vertrauensintervalle fallen bei der Sättigung sehr klein aus.

Ebenfalls analog zu den weiteren drei Einstellungsparametern soll nachfolgend die Einstellung des Sättigungsreglers auf mögliche Korrelationen untersucht werden. In Abbildung 5.4 war bereits ersichtlich, dass die Schwankungen bei den unterschiedlichen äußeren Bedingungen minimal sind. Diese Größenordnung ist für die Praxis irrelevant, da sie einen kaum sichtbaren Effekt auf das Erscheinungsbild ausübt.

In der ersten Zeile von Abbildung 5.6 wird die Sättigung gegen die drei die Szene beschreibenden Parameter aufgetragen. Die zugehörigen Korrelationskoeffizienten lauten 0,54, 0,59 und 0,5. Allerdings stellt sich auch hier die Frage nach der praktischen Relevanz. Alle drei Parameter überspannen einen extrem großen Wertebereich, wohingegen die Sättigung nur minimal zunimmt.

Interessanter als die zwei vorher genannten Korrelationsgruppen (Korrelationen mit den äußeren Bedingungen und mit den die Szene beschreibenden Parametern) ist aber eine

dritte Gruppe. In den vorangegangenen Abschnitten wurde dargelegt, dass das Tonemapping eine Korrektur der Farbkomponente nötig macht. Entsprechend müssten sich Korrelationen zwischen der Sättigung und den drei weiteren Einstellungsparametern ergeben, da jene die Tonemapping-Funktion definieren. Dargestellt ist dies in der zweiten Zeile von Abbildung 5.6. Während Helligkeit und Homogenisierung eher indifferente Erscheinungsbilder liefern, lässt sich beim Kontrast eine klare Tendenz ausmachen. Bereits eine geringe Modifikation der Steilheit (etwa 1 EV) sorgt für eine merkbliche Änderung der Sättigungseinstellung. Je nach Tonemapping-Funktion kann die Gradation noch deutlich stärker beeinflusst werden. Der Korrelationskoeffizient liegt bei 0,69.

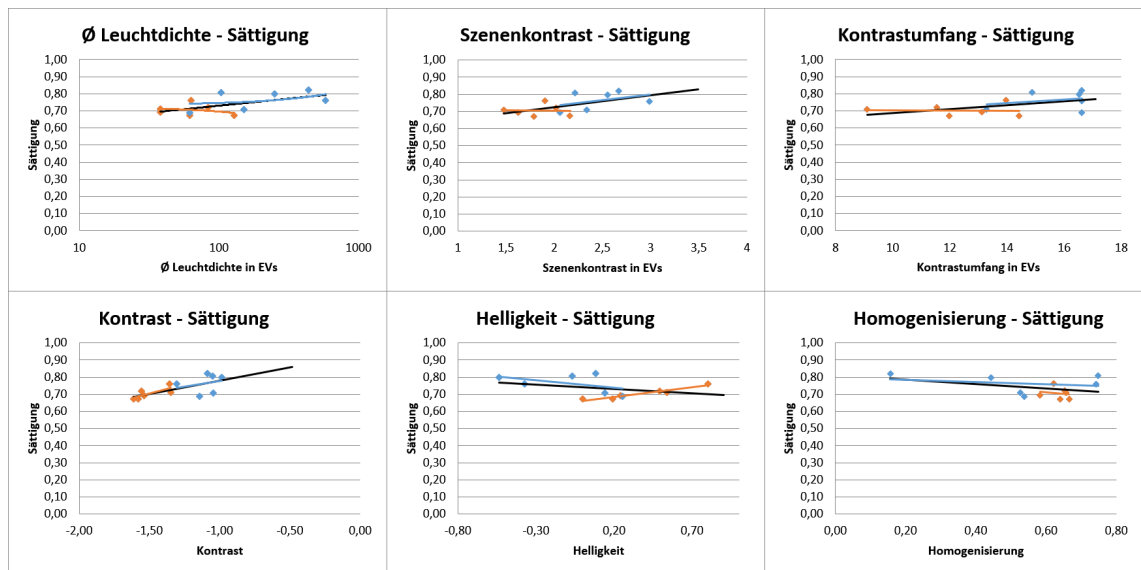


ABBILDUNG 5.6: Korrelation der Sättigungseinstellung mit den Szenenparametern

Zusammenfassend kann aus den Ergebnissen geschlossen werden, dass bereits mit einem fixen Koeffizienten um 0,73 eine gute Chroma-Kompensation für die durchgeführte Kontrastkompression bei einer Vielzahl an Bildern erreicht werden kann. Allerdings sollte beachtet werden, dass die Änderung der Steilheit wahrscheinlich nur unzureichend abgebildet wird. Deutliche Abweichungen von der im Experiment verwendeten Gradation könnten einen anderen Koeffizienten benötigen.

In Abbildung 5.7 sind die Korrekturfunktionen von Pouli und der Koeffizienten-Methode bei sich verändernder Helligkeit beispielhaft für den Fall C_0 entspricht $I_0/2$ dargestellt. Für die Koeffizienten-Methode wurde der Wert von 0,73 aus dem Probanden-Grading angenommen. Es wird deutlich, dass die Kurvenverläufe besonders zwischen 0,5 und 2 eine hohe Analogie aufweisen. Dieser Bereich ist maßgeblich, da Reduktionen der Helligkeitskomponente von mehr als 50% bzw. Erhöhungen von mehr als 100% seltener auftreten.

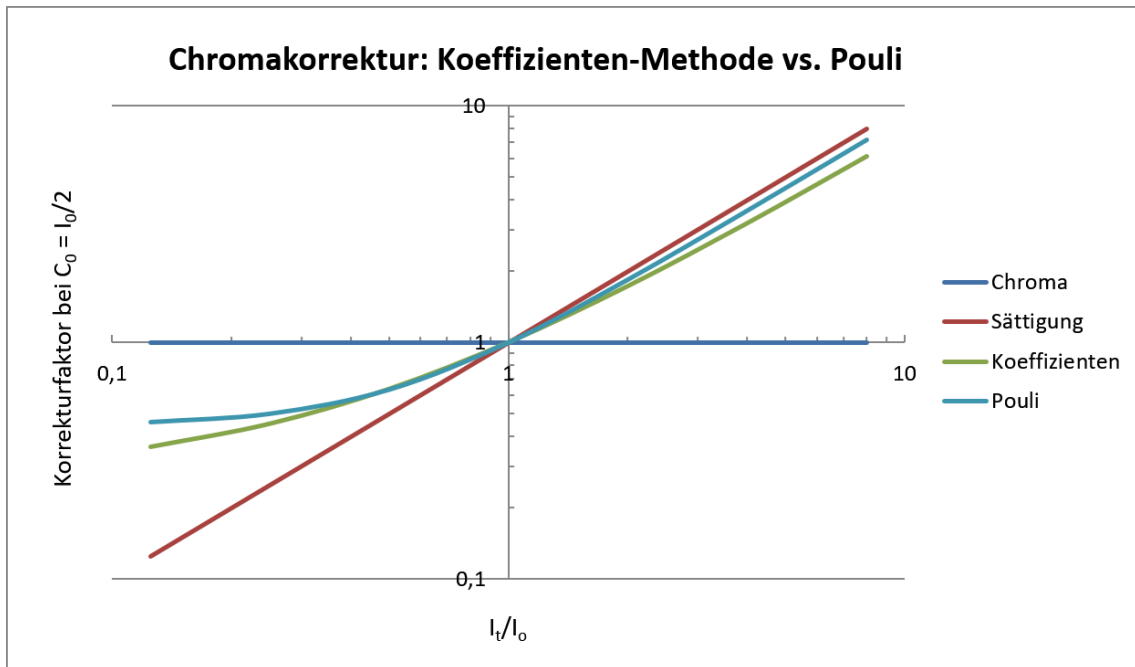


ABBILDUNG 5.7: Vergleich der Chroma-Korrektur von Pouli mit der Koeffizienten-Methode

5.9 Probandentest Farbpatches

5.9.1 Aufbau und Durchführung

Im letzten Abschnitt konnte gezeigt werden, dass mit der Koeffizienten-Methode unter Zuhilfenahme eines festen Wertes bereits für eine Vielzahl von Bildern ein sehr ansprechendes Resultat erzielt werden konnte, auch wenn die Gradation nicht ausreichend berücksichtigt zu sein scheint. Es schloss sich die Frage an, ob dieser Koeffizient für alle Farbtöne und Sättigungslevel gleich gut funktioniert, da die Qualität der Dekorrelation jedes Farbraumes, auch von IC_tC_p , ungleich verteilt ist. Hierfür wurde ein dritter Probandentest entwickelt, der sich rein auf die Farbwahrnehmung konzentrierte. Dessen Ziel war es, eine Funktion zur Berechnung eben jenes Koeffizienten auf Basis von Farbton, Sättigung und Helligkeitsniveau zu ermitteln.

Dazu wurden auf einem **HDR**-Display zwei große Farbpatches gezeigt, die in IC_tC_p definiert wurden. Der rechte Farbpatch bildete dabei die Referenz. Die Referenz wies einen um 0,05 Punkte höheren I-Wert auf als der linke Patch. Zusätzlich wurde auch der graue Hintergrund beim linken Patch um 0,05 Punkte abgesenkt. Die C_t - und C_p -Werte der beiden Patches waren zunächst identisch. Die Aufgabe des Probanden bestand nun darin, C_t und C_p des linken Patches gemeinsam anzuheben oder abzusenken, so dass sich der subjektive Eindruck der beiden Patches, trotz Helligkeitsunterschied, bestmöglich an

einander annäherte. Die Steuerung erfolgte über Tastendruck. Der aktuell eingestellte Wert war, im Gegensatz zum Probanden-Grading, für die Probanden nicht einzusehen.

Abbildung 5.8 zeigt den grünen Farbton in zwei verschiedenen Sättigungsstufen und diese jeweils in drei unterschiedlichen Helligkeitsniveaus. Im CIE-xy-Diagramm in Abbildung 5.9 sind alle 54 Referenzpatches als Punkte eingetragen. Sie ergeben sich aus den sechs Farben (Rot, Magenta, Blau, Türkis, Grün und Gelb), welche jeweils in drei Sättigungs- und drei Helligkeitsstufen unterteilt wurden. Bei der Durchführung wurde jeweils mit der geringsten Helligkeitsstufe begonnen, da eine Adaption bezüglich geringer Leuchtdichten langsamer vonstatten geht, als eine Adaption in Richtung höherer Leuchtdichten. Die maximale Leuchtdichte war dabei an den Farbton gekoppelt. Der hellste Farbton (gelb) erreichte 209 cd/m^2 , der dunkelste Farbton (blau) erreichte $16,4 \text{ cd/m}^2$. 41 der 54 Patches lagen innerhalb des Farbgamuts von ITU-R BT.709 [54]. Die restlichen 13 Patches lagen innerhalb des Farbgamuts von ITU-R BT.2020 [57].

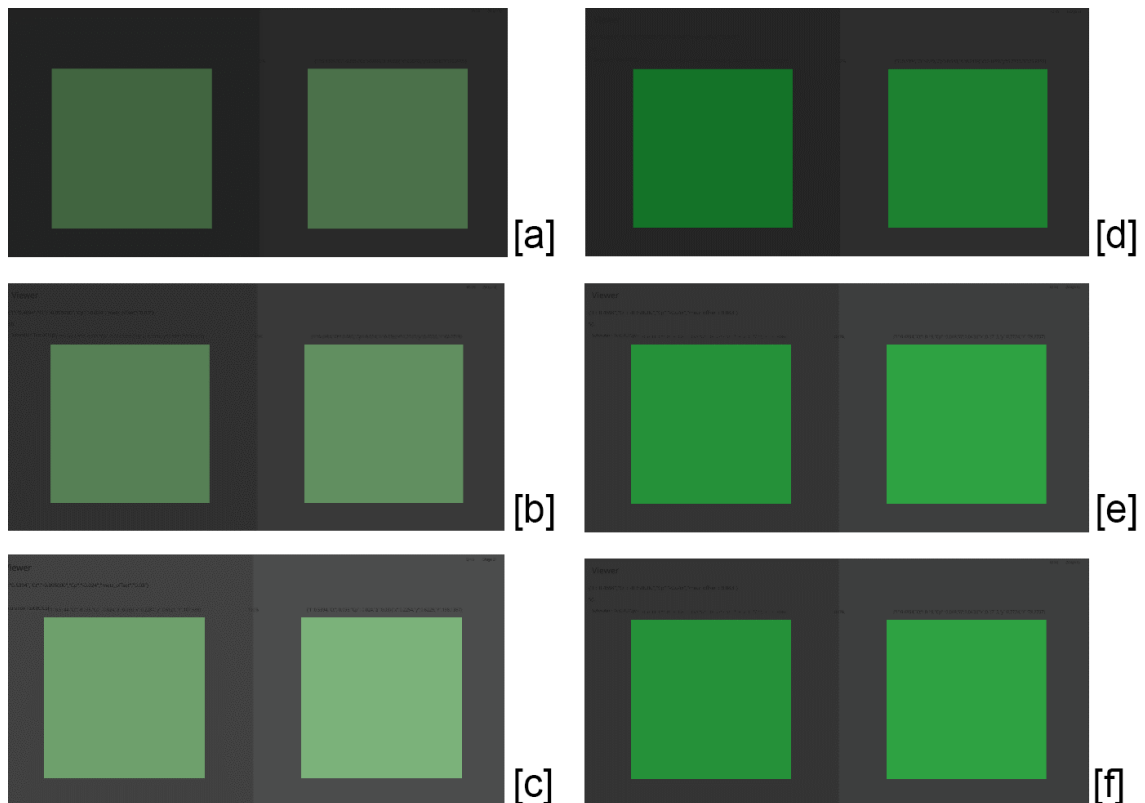


ABBILDUNG 5.8: Grüne Farbpatches bei unterschiedlichen Helligkeits- und Sättigungsstufen (a) Sättigung Stufe 1, Helligkeit Stufe 1, (b) Sättigung Stufe 1, Helligkeit Stufe 2, (c) Sättigung Stufe 1, Helligkeit Stufe 3, (d) Sättigung Stufe 2, Helligkeit Stufe 1, (e) Sättigung Stufe 2, Helligkeit Stufe 2, (f) Sättigung Stufe 2, Helligkeit Stufe 3

3

Für die Darstellung auf dem Display wurden die $IC_t C_p$ -Werte nach R'G'B' umgerechnet. Das Display, ein Metz **OLED**, wurde dazu in den **PQ**-Modus versetzt. Die Schnittstelle war dabei auf 10 Bit begrenzt.

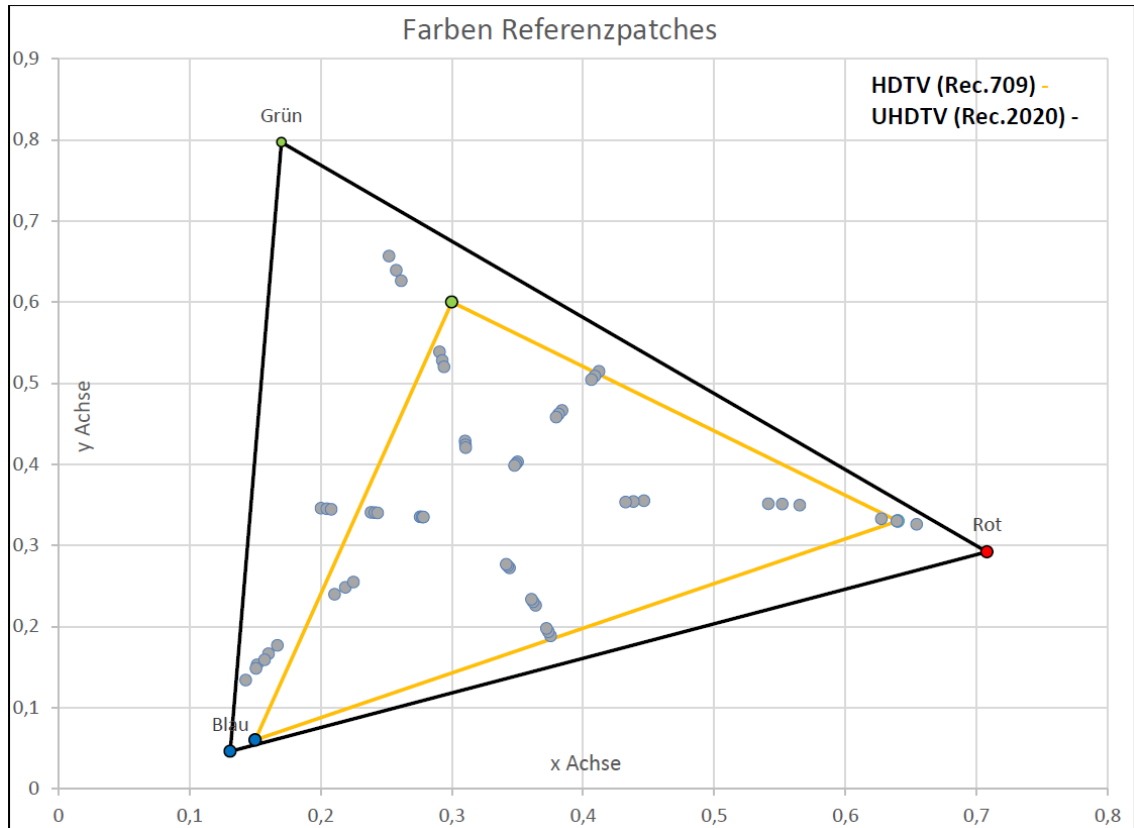


ABBILDUNG 5.9: Farbpatches zur Messung der Chroma-Kompensation im CIE-xy-Diagramm

Am Test nahmen insgesamt 6 Frauen und 25 Männer, im Alter zwischen 23 und 67 Jahren teil.

5.9.2 Ergebnisse und Fehlerbetrachtung

Die Hypothese war, dass bei kleineren Werten für I auch C_t und C_p abgesenkt werden würden. Dieses Ergebnis legt auch die Grafik aus Abschnitt 5.6 nahe, die zeigt, dass Helligkeit- und Farbkomponenten nicht perfekt von einander entkoppelt sind. Die Frage war, bei welchem Farbton und Sättigungsniveau sich dieser Effekt wie stark bemerkbar machen würde. Allerdings entsprechen die gemessenen Werte nicht der dargelegten Vermutung.

Abbildung 5.10 zeigt die ermittelten Werte unterteilt in die sechs Farbtöne. Der Wert 50 gibt dabei die Nullstellung des Reglers wieder. Der Wert 51 steht für eine Erhöhung um 1%, der Wert 49 für eine Erniedrigung um 1%. Einzig bei blau befinden sich alle Durchschnittswerte unter der Nullstellung. Bei den anderen Farbtönen schwanken die Werte teilweise stark und gefühlt willkürlich um den Wert 50. Hinzukommen große Abweichungen in den vorgenommenen Einstellungen der einzelnen Probanden und damit

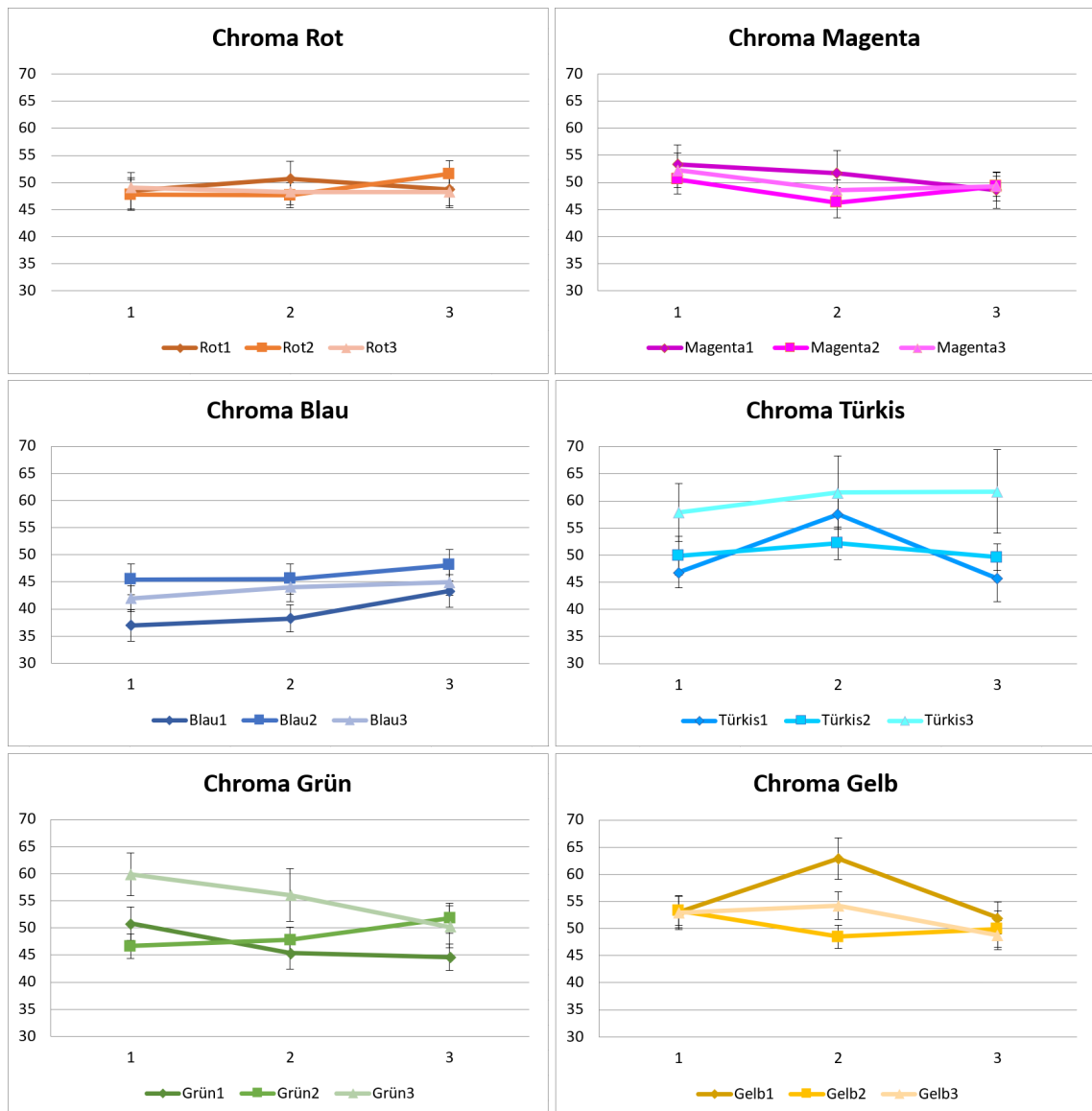


ABBILDUNG 5.10: Ergebnisse der subjektiven Untersuchung mit Farbpatches. Es sind die durchschnittlich eingestellten Werte mit den dazugehörigen Konfidenzintervallen angegeben

verbundene große 95%-Konfidenzintervalle. Entsprechend besitzen die Messwerte keine Aussagekraft bezüglich der Chroma-Kompensation, sondern lassen eher auf Fehler beim Testdesign schließen.

Viele Probanden gaben im Nachgang an, oft keine für sie zufriedenstellende Einstellung gefunden zu haben. Es war ihnen nicht möglich, den Helligkeitsunterschied auszublenden und eine bestmögliche farbliche Übereinstimmung zu finden. Dies zeigt, wie problematisch es ist, verschiedene Helligkeitsniveaus in einen gemeinsamen Test einzubeziehen. Weiterhin zeigt sich, dass der direkte Vergleich von **SDR** mit **HDR**, wie in Abschnitt **4.3**, äußerst fragwürdig sind. Für den vorliegende Fall wurde allerdings angenommen,

dass durch die vergleichsweise kleinen Abstufungen trotzdem mit einer helleren Referenz gearbeitet werden könne. Dem war allerdings nicht so. Die vergleichsweise geringen Helligkeitsabstufungen führten darüber hinaus dazu, dass der durch eine nicht perfekte Dekorrelation verursachte Fehler sehr klein und somit schwer messbar wurde.

Des Weiteren wurde nicht bedacht, dass eine Sättigungswahrnehmung, so wie sie für eine Chroma-Kompensation entscheidend ist, ohne Kontrasteindruck deutlich erschwert wird. Bereits bei Mantiuk [81] war es so, dass die Stärke der Kontrastkompression und daher die Steilheit des Bildes den Faktor für die Sättigungsanpassung bestimmte (Abschnitt 5.2). Auch im Probanden-Grading hatte sich der gewählte Kontrast als Einflussfaktor herauskristallisiert (s. Korrelation zwischen Sättigung und Kontrast in Abschnitt 5.8.3). Eine isolierte Messung, wie mit den Farbpatches, konnte in der Folge nicht sinnvoll sein. Diese Ableitung aus eigentlich bereits bekannten Informationen wurde aber zunächst übersehen und erst durch den Versuch mit den Farbpatches augenscheinlich.

Letztendlich bedeuten diese Ergebnisse auch, dass selbst bei einem perfekt dekorrelierten Farbraum immer noch die Notwendigkeit einer Chroma-Kompensation bestünde, auch wenn diese deutlich geringer ausfallen würde, da zumindest die Dimension der Helligkeit bereits inkorporiert ist. Die Dimension der Steilheit wäre allerdings noch nicht berücksichtigt.

5.10 Konstante Buntheit

5.10.1 Theorie zur konstanten Buntheit

In den vorherigen Abschnitten konnte gezeigt werden, dass weder die Zielsetzung „konstante Sättigung“ noch die Zielsetzung „konstantes Chroma“ zu einem konstanten Farbeindruck führt. Mithilfe eines Koeffizienten, der das Mischverhältnis beider Zielsetzungen angibt, kann unkompliziert ein in den meisten Fällen ansprechendes Ergebnis erzielt werden (s. 5.8.3). Allerdings bleibt unklar, ob unterschiedliche Farbtöne nicht divergente Koeffizienten benötigen. Darüber hinaus wird die Stärke der Kontrastkompression nur unzureichend abgebildet⁸, so dass es bei sehr flachen Repräsentationen einer Szene weiterhin zu Übersättigungen bzw. bei sehr steilen Repräsentationen zu Untersättigungen kommen kann. Aus diesen Gründen soll nachfolgend ein Ansatz gefunden werden, der die genannten Limitationen zu überwinden vermag. Dazu sind nochmals die theoretischen Zusammenhänge zwischen Sättigung und Chroma zu betrachten. Zu Beginn der Arbeit wurde definiert:

⁸faktisch wird ein Mittelwert angenommen.

$$\text{Chroma} = \frac{\text{Buntheit}}{\text{absolute Helligkeit weiß}} \quad (5.21)$$

$$\text{Sättigung} = \frac{\text{Buntheit}}{\text{absolute Helligkeit}} \quad (5.22)$$

$$\text{relative Helligkeit} = \frac{\text{absolute Helligkeit}}{\text{absolute Helligkeit weiß}} \quad (5.23)$$

Somit muss auch gelten:

$$\text{absolute Helligkeit weiß} = \frac{\text{absolute Helligkeit}}{\text{relative Helligkeit}} \quad (5.24)$$

$$\text{Sättigung} = \frac{\text{Chroma}}{\text{relative Helligkeit}} \quad (5.25)$$

Durch das Tonemapping wird nun die relative Helligkeit geändert. Damit die letzte Gleichung weiterhin Bestand hat, muss sich folglich auch die Sättigung und / oder das Chroma verändern. Welche der beiden Größen wie stark angepasst werden muss, könnte von der relativen Helligkeit abhängig sein. Es wäre denkbar, dass dabei nicht der Wert der relativen Helligkeit entscheidend ist, sondern wie sich dessen Änderung auf seine beiden Bestandteile (absolute Helligkeit und absolute Helligkeit von weiß) aufteilt. Der Autor stellt hierzu folgende Hypothese auf:

Wenn die Änderung der relativen Helligkeit überproportional durch eine Änderung der absoluten Helligkeit ausgelöst wird, ist die Sättigung stärker zu korrigieren. Wenn die Änderung der relativen Helligkeit überproportional durch eine Änderung der absoluten Helligkeit von *weiß* ausgelöst wird, ist das Chroma stärker zu korrigieren. Sprich, die Hypothese wäre, dass ein subjektiv konstanter Farbeindruck einer konstanten Buntheit entspricht.

Bis zu diesem Punkt ist in der vorliegenden Arbeit nicht immer zwischen relativer und absoluter Helligkeit unterschieden worden. Aus den Gleichungen [5.21](#) und [5.22](#) gehen ganz klar die absoluten Helligkeiten hervor. Dies impliziert, dass auch Buntheit und Chroma entsprechend absolut und linear skaliert sind. In den beiden angestrebten Farbräumen IC_tC_p und $Y'C_b'C_r'$, werden allerdings die Farbdifferenzsignale aus den normierten und verzerrten Farbkanälen gebildet. Deren Äquivalent in der Helligkeitsdomäne wäre entsprechend das Luma (Y'). Aus diesem Grund soll bei der Chroma-Kompensation, die auf einem [HDR](#)-Farbraum angewendet wird, für die Hypothese der

konstanten Buntheit die Annahme gelten, dass die absolute Helligkeit mithilfe des Lumas quantifiziert werden kann.

Unter dieser Prämisse gilt:

$$C_0 = \frac{B}{Y'_{white-0}} \quad (5.26)$$

$$C_t = \frac{B}{Y'_{white-t}} \quad (5.27)$$

Somit ergibt sich:

$$C_t = C_0 \frac{Y'_{white-t}}{Y'_{white-0}} \quad (5.28)$$

B steht dabei für die Buntheit, C für das Chroma und Y' für das Luma. Die Indizes zeigen wieder an, ob es sich um den Wert vor (0) oder nach dem Tonemapping (t) handelt. Der zusätzliche Index (*white*) gibt an, dass es sich nicht um das Luma des Pixels, sondern um das Luma eines weißen Pixels unter den identischen Beleuchtungsbedingungen handelt.

Im Falle der konstanten Buntheit würde aber gleichzeitig für Sättigung und Chroma gelten:

$$S_t = S_0 \frac{Y'_t}{Y'_0} \quad (5.29)$$

$$C_t = S_t \frac{Y'_t}{Y'_{white-t}} \quad (5.30)$$

Durch Einsetzen, Umformen und unter Zuhilfenahme der folgenden bereits bekannten Gleichung

$$C_0 = S_0 \frac{Y'_0}{Y'_{white-0}} \quad (5.31)$$

erhält man in diesem Fall:

$$C_t = C_0 \frac{Y'_{white-0}}{Y'_{white-t}} \left(\frac{Y'_t}{Y'_0} \right)^2 \quad (5.32)$$

Gleichung [5.28](#) und [5.32](#) wären somit gleichzeitig nur für den Fall gültig, wenn folgender Zusammenhang gilt:

$$\frac{Y'_{white-t}}{Y'_{white-0}} = \frac{Y'_t}{Y'_0} \quad (5.33)$$

Das die Änderung des Lumas von Weiß, der Änderung des Lumas des Pixels entspricht, ist nur erfüllt, wenn die Helligkeit gleichabständig skaliert ist, der Farbton die Helligkeit nicht beeinflusst und keine Modifikation der Steilheit vorliegt. Die Veränderung der Helligkeit eines bestimmten Pixels ist aber von bis zu allen dieser drei Faktoren abhängig. Zwar ist im Falle von [PQ](#) die Helligkeit sehr gleichabständig skaliert und im Falle von IC_tC_p der Farbton sogar fast perfekt dekoriert (im Gegensatz zu $Y'C_b'C_r'$), allerdings wird die Kontrastkompression nicht berücksichtigt. Deshalb sieht der Autor hierbei die Ursache für das beim Tonemapping beobachtete Phänomen der Übersättigung. Würde hingegen ein Referenzweiß, wie in Gleichung [5.28](#) zur Chroma-Kompensation verwendet, ergäbe sich eine absolute Unabhängigkeit vom Farbton des Pixels und es wäre möglich, die Änderung der lokalen Steilheit des Bildes pro Pixel indirekt mit zu berücksichtigen. Damit würde die Problematik von Mantiuk [\[81\]](#) umgangen, bei der nur eine globale Kontrastkompression berücksichtigt werden konnte.

Warum die Anwendung der Referenzweiß-Methode die Steilheit des Bildes bzw. die Stärke der Kontrastkompression widerspiegelt, soll anhand der folgenden Abbildung [5.11](#) verdeutlicht werden. Die Graphen zeigen beispielhafte Verläufe sehr flacher lokaler Tonemapping-Kurven, jeweils getrennt für Lichter und Schatten, anhand von zwei Beispielbildern. Die korrespondierenden optischen Resultate folgen an späterer Stelle in Abbildung [5.12](#).

In jedem der Diagramme finden sich zwei Kurvenverläufe. Die graue Linie zeigt durch ihre Diagonale an, dass die Helligkeitskomponente I von Eingang zu Ausgang unverändert belassen wurde, die blaue Linie hingegen gibt den Verlauf einer flachen lokalen Tonemapping-Funktion wieder. Es ist zu erkennen, dass beim Bild *Moderatorin* die dunklen Schattenpartien massiv verstärkt werden. Für die Methode der konstanten Sättigung würde dies bedeuten, dass die Änderung von I (Verstärkung um mehr als 300%) auch auf C_t und C_p übertragen werden müsste. Unberücksichtigt bleibt dabei, dass sich die Steilheit des Bildes extrem verflacht hat. So überspannt I bei der Tonemapping-Funktion nur einen Bereich von etwa 0,1 Einheiten, während es im Originalbild deutlich über 0,4 sind. Die Auswirkung dieser Kurvenverläufe kann beim Referenzweiß (ca. 0,508 I) gut nachvollzogen werden. Anstelle einer Erhöhung von über 300% legt das Referenzweiß eine Reduzierung um 26% nahe. Ein ähnliches Ergebnis zeichnet sich für die Schatten im Bild *Testchart* ab. Allerdings würde das Referenzweiß hier zu einer Verstärkung um 7%

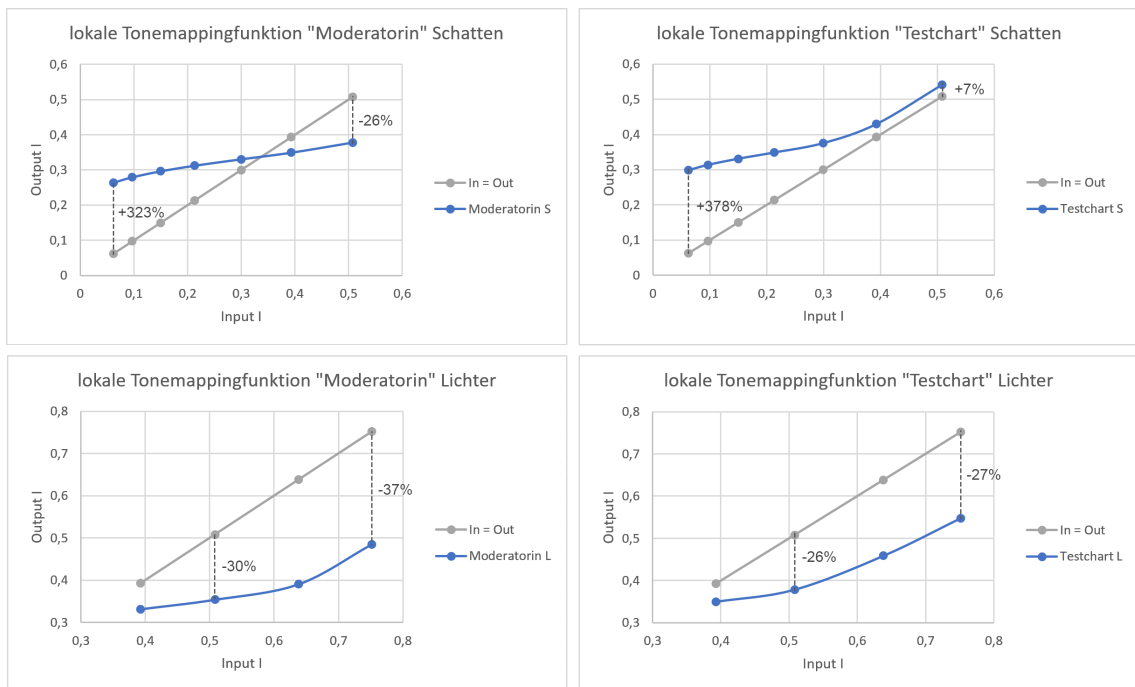


ABBILDUNG 5.11: Konstante Buntheit beim lokalen Tonemapping

führen. Zu erklären ist dies mit einer noch ausgeprägteren Erhöhung der I -Komponente sowie einem etwas steileren Kurvenverlauf des Tonemappings.

Für den Bereich der Lichter weist die Tonemapping-Funktion ein gänzlich anderes Erscheinungsbild auf. Im Vergleich mit der Diagonalen ist der Verlauf der blauen Linie nur wenig flacher, dafür aber deutlich geringer. Es wäre also anzunehmen, dass die Reduzierung der Chroma-Komponenten stärker ausfällt, als es bei der Methode der konstanten Sättigung der Fall wäre. Den Diagrammen ist allerdings zu entnehmen, dass dies nicht zutrifft, sondern die Reduzierung sogar minimal geringer ausfällt. Zu erklären ist dies mit der fortschreitenden Dekorrelation zwischen I und $C_t C_p$ bei hohen Werten für I , wie an früherer Stelle in Abbildung 5.2 gezeigt wurde. So bleibt das Chroma trotz Helligkeitsänderungen nahezu konstant. Zu ausgeprägte Entsättigungen müssen in der Folge unterbunden werden.

Abbildung 5.12 zeigt anhand von zwei Szenen, wie sich die einzelnen Chromakompensations-Verfahren auf den Bildeindruck auswirken. Um den Effekt besser zu illustrieren, wurde jeweils die Helligkeitskomponente im Vergleich zur Aufnahme (jeweils oben links) stark verändert. Die Darstellung ist heller und deutlich flacher. Bei der Methode der konstanten Sättigung (oben Mitte) führt dies zu einem sehr übersättigten Bildeindruck. Weniger Sättigung, wenn auch immer noch relativ viel, zeigt das Bild rechts daneben. Hier wurden die Chroma-Komponenten unverändert belassen. Das erste Bild in der zweiten Zeile zeigt die Koeffizienten-Methode. Es handelt sich um eine Mischung der beiden



ABBILDUNG 5.12: Die verschiedenen Chroma-Kompensationen im Vergleich. Pro Szene von oben links nach unten rechts: Originalaufnahme dargestellt bis 100 cd/m^2 , konstante Sättigung, konstantes Chroma, Koeffizienten-Methode, Pouli und konstante Buntheit

genannten Vorgänger. Der optische Eindruck ist aber deutlich näher am konstanten Chroma.

Die letzten zwei Darstellungen gehen auf Pouli und die konstante Buntheit zurück. Diese Verfahren sind als einzige in der Lage, die Hauttöne trotz starker Modifikation der Helligkeit - vor allem der Steilheit - natürlich zu reproduzieren. Auch wenn der Gesamteindruck ähnlich ist, sind doch Unterschiede zwischen den beiden Verfahren auszumachen. So nehmen im Hintergrund der Moderatorin Zelt und Springhindernis bei Pouli einen leuchtenden Rotton an und der dunkelblaue Farbpatch auf dem MacBeth-Chart der zweiten Szene wird eher hellblau dargestellt. Der größte Unterschied ist aber im Umgang mit chromatischem Rauschen zu erkennen. Tiefe Schatten werden beim Tonemapping meist am stärksten angehoben, gleichzeitig weisen sie aber auch das größte Rauschen auf. In der Folge kommt es zu einer enormen Verstärkung und somit zu einem sichtbaren chromatischen Rauschen. Im vorliegenden Bild äußert sich dies dadurch, dass das schwarze Hemd grünlich erscheint. Bei der konstanten Buntheit hingegen wird nicht die Helligkeit des Pixels, sondern die des Referenzweiß herangezogen. Die Verstärkung bleibt somit aus und das Hemd schwarz.

Theoretisch scheint die konstante Buntheit einen guten Lösungsansatz für die Sättigungsproblematik darzustellen. Im Hinblick auf die Praxis bedarf es aber noch der Klärung zweier Aspekte. Zum einen sind solch starke Modifikationen der Helligkeitskomponente eher unüblich, was die Unterschiede zwischen den verglichenen Methoden verringert. Zum anderen stellt sich die Frage, ob diese Lösung auch jene ist, die von den Zuschauern präferiert wird bzw. ob von den Zuschauern überhaupt große Unterschiede wahrgenommen werden? In einem der vorangegangenen Probandentests hatte sich die Koeffizienten-Methode trotz nun aufgezeigter Limitationen beispielsweise bereits bewährt. Deshalb sollen im nachfolgenden Abschnitt die gezeigten fünf Verfahren in einem weiteren Probandentest verglichen werden.

5.10.2 Testaufbau zur konstanten Buntheit

In den vorangegangenen Abschnitten hat sich gezeigt, dass der Sättigungseindruck einer Farbe nicht entkoppelt von der Kontrastwahrnehmung betrachtet werden kann. Deshalb sollten im finalen Präferenztest wieder reale Bildvorlagen zum Einsatz kommen. Zum ersten Mal wurden hierfür keine Einzelbilder, sondern kurze **HDR**-Videosequenzen genutzt. Insgesamt wurden sechs Sequenzen⁹ ausgewählt. Dabei handelte es sich vornehmlich um Szenen, die ein breites Farbspektrum aufweisen und in den meisten Fällen auch Hauttöne zeigen. Gleichzeitig besitzen sie unterschiedliche Kontrastumfänge, was zu Schwankungen bei der zugehörigen Tonemapping-Funktion führen sollte.

Jede der ausgewählten Sequenzen wurde mithilfe der in Kapitel 4 gewonnenen Erkenntnisse präferenzbasiert kontrastkomprimiert. Ziel war es, praxisnahe Vorlagen in typischer Gradation zu erhalten. Die sich anschließende Chroma-Kompensation wurde, wie bereits im vorherigen Abschnitt angekündigt, jeweils mit den fünf verschiedenen Methoden (konstante Sättigung, konstantes Chroma, Koeffizienten-Methode, Pouli und konstante Buntheit) durchgeführt, so dass entsprechend von jeder Videosequenz fünf **SDR**-Varianten entstanden. Als Farbraum wurde dazu IC_tC_p eingesetzt.

Um den Probandentest auf eine Dauer von knapp 20 Minuten zu begrenzen, wurde kein paarweiser Vergleich durchgeführt. Anstelle dessen wurde die Koeffizienten-Methode als Referenz angesehen, so dass immer eine der anderen vier Varianten gegen die Koeffizienten-Methode verglichen wurde. Im Gegensatz zum Coloristen-Test wurden bei jedem Vergleich beide Varianten doppelt gezeigt, um einer Adaption an das Sättigungsniveau der zuerst gezeigten Variante vorzubeugen. Auch die Verwendung der

⁹*FKT-Stilleben, guitarist, Anstoß, festival crowd, Moderatorin und hdr test image*

Koeffizienten-Methode als Referenz unterstützte dieses Ansinnen, da diese sich, hinsichtlich des Sättigungsniveaus, im Mittelfeld der fünf Kandidaten bewegt und hierdurch Adaptionen in Richtung sehr stark gesättigter bzw. sehr gering gesättigter Bilder vermieden werden.

Auch die Bewertung wurde in Relation zur Referenz vorgenommen. Dazu stand eine stufenlose Skala bereit, welche mithilfe von sieben Begriffen (*much worse*, *worse*, *slightly worse*, *the same*, *slightly better*, *better* und *much better*) in mehrere Bereiche aufgeteilt wurde und so den Teilnehmern als Orientierung dienen sollte. Insgesamt war das Testdesign somit an das Vorgehen bei [56] angelehnt.

Der Test wurde an einem SDR-Referenzmonitor der Klasse 1 unter kontrollierten gedimmten (10 cd/m^2) räumlichen Bedingungen durchgeführt. Insgesamt nahmen 14 Teilnehmer im Alter zwischen 19 und 67 Jahren an dem Test teil. Das durchschnittliche Alter betrug 25,5 Jahre. Bei 14% der Teilnehmer handelte es sich um Frauen. 86% waren Männer. 79% bezeichneten sich selbst als Expert-Viewer.

5.10.3 Ergebnisse zur konstanten Buntheit

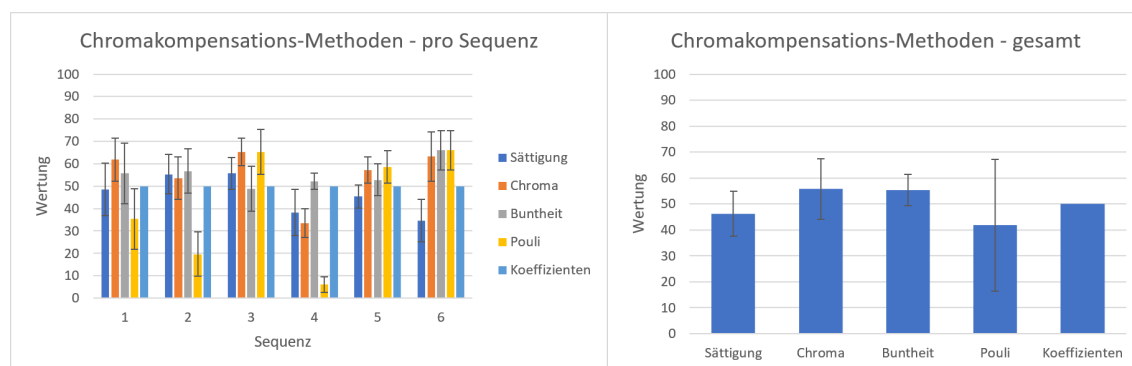


ABBILDUNG 5.13: Ergebnisse der Probandenuntersuchung für die fünf verschiedenen Verfahren zur Chroma-Kompensation. Die linke Seite zeigt die Auswertung unterteilt in die sechs verwendeten Testsequenzen inklusive der 95%-Konfidenzintervalle. Die rechte Seite zeigt die Gesamtbewertung für die jeweilige Methode inklusive der Standardabweichung der arithmetischen Mittelwerte der Einzelsequenzen.

Abbildung 5.13 zeigt die Ergebnisse inklusive der 95%-Konfidenzintervalle¹⁰. Auf der linken Seite sind die Ergebnisse pro Sequenz dargestellt. Die Balken der Koeffizienten-Methode weisen bei jeder Sequenz den Wert 50 auf, da sie als Referenz auf diesen Wert festgelegt wurde. Überschreitet die Wertung die Referenz, wurde das dazugehörige Verfahren bei der entsprechenden Videosequenz von den Probanden der Referenz vorgezogen.

¹⁰Es muss berücksichtigt werden, dass der Grenzwert von 0,35 beim Kolmogorov-Smirnov-Test bei α gleich 0,05 in zwei Sequenzen bei der Buntheit überschritten wurde (0,42 und 0,44). Das ist darauf zurückzuführen, dass in diesen Fällen fast ausschließlich der Wert 50 gewählt wurde.

Zunächst ist zu beobachten, dass die Differenzen zwischen den fünf Verfahren teilweise relativ klein ausfallen und es häufig zu Überlappungen der 95%-Konfidenzintervalle kommt. Dies scheint die im vorangegangenen Abschnitt geäußerte Vermutung zu bestätigen, dass die Unterschiede in der Praxis nicht so deutlich hervorstechen, wie es beispielsweise in Abbildung 5.12 der Fall ist. Als möglicher Grund wird dort die meist moderate Steilheitsänderung angeführt. Hinzukommt die gute Dekorrelation zwischen Helligkeit und Farbe, durch den in diesem Test verwendete IC_tC_p -Farbraum. Das wird besonders auf der rechten Seite von Abbildung 5.13 deutlich, wo die Gesamtergebnisse für die Chroma-Kompensationsmethoden dargestellt sind. Die höchste Wertung entfällt hier auf das Chroma; also auf die Nichtanwendung einer Chroma-Kompensation.

Somit scheint es auf den ersten Blick, als ob die Anwendung einer Chroma-Kompensation in der Praxis durch die Einsatz von IC_tC_p obsolet geworden ist und die in den vorherigen Abschnitten diskutierten Probleme, hauptsächlich im Bezug auf die Steilheit, nicht relevant genug sein würden. Bei genauerer Betrachtung des Abschneidens der Verfahren in bestimmten Sequenzen werden allerdings doch signifikante Unterschiede sichtbar.

Auffällig ist beispielsweise Sequenz 4, bei der die konstante Buntheit und die Koeffizienten-Methode deutlich besser abschneiden als die weiteren drei Verfahren. Dabei handelt es sich um eine Szene, die im HDR sehr hell aufgezeichnet wurde und fast ausschließlich Hauttöne enthält, da ein Publikum bei einem Open-Air-Konzert gezeigt wird. Auch bei Sequenz 2 handelt es sich um eine sehr helle Szene. Dieses Mal schneidet aber nur die Methode nach Pouli signifikant schlechter ab als die anderen Verfahren. Ob die modifizierte Sättigungsfunktion, welche von Pouli genutzt wird, generell Probleme mit hohen Helligkeiten hat, oder ob die Adaption von IPT auf IC_tC_p nicht ohne weiteres möglich ist, kann an dieser Stelle nicht beantwortet werden.

Die Buntheit erhält ihre geringste Wertung für Sequenz 3. Die Szene zeigt ein Fußballstadion von innen. Insbesondere das Grün des Spielfeldes ist bei der konstanten Buntheit in dieser Sequenz geringer gesättigt als bei den anderen Methoden. Dies scheint hierbei als nachteilig empfunden zu werden. Die Differenz zwischen den Verfahren fällt allerdings eher gering aus und ist lediglich zum konstanten Chroma signifikant (s. die 95%-Konfidenzintervalle bei Sequenz 3). Dies ist der einzige Fall, wo eine andere Methode als signifikant besser empfunden wird als die konstante Buntheit. Im Gegensatz dazu bietet die konstante Buntheit, im Vergleich zu den anderen Verfahren (mit Ausnahme der Referenz), bei durchschnittlich zwei Sequenzen einen signifikanten Qualitätsgewinn.

Der Vorteil der Buntheit besteht also nicht darin, bei einzelnen Bildern besonders hohe Wertungen zu erzielen, sondern das Verfahren bietet ein konstant gutes Niveau ohne starke Ausschläge. Im besten Fall wurden im Test 66 Punkte erreicht. Im schlechtesten Fall nur 17 Punkte weniger. Beim Chroma ist diese Differenz immerhin schon 32 Punkte

groß. Die Beobachtung wird auch durch die Gesamtauswertung gestützt. Zwar erhält, wie bereits erwähnt, das konstante Chroma eine minimal höhere durchschnittliche Bewertung (56 zu 55), dafür weist die konstante Buntheit die mit Abstand kleinste Varianz¹¹ auf (6 zu 12). Als einziges Verfahren bietet es somit die Möglichkeit einer potenziellen Qualitätsverbesserung gegenüber der Koeffizienten-Methode ohne dabei gleichzeitig ein nennenswertes Risiko für eine Verschlechterung aufzuweisen. Im Hinblick auf das angestrebte Ziel, den Einsatz im Broadcast-Umfeld, ist diese Zuverlässigkeit das entscheidende Kriterium.

5.11 Zwischenfazit bezüglich der Präferenz bei der Farbwahrnehmung

In diesem Kapitel wurde das Problem von subjektiv zu gesättigten Bildern in Folge einer Kontrastkompression betrachtet. Es konnte gezeigt werden, dass klassische Methoden, die auf den RGB-Komponenten arbeiten, entweder deutliche Auswirkungen auf den Farbton oder auf die Helligkeit nach sich ziehen, was für den angestrebten Verwendungszweck nicht gewünscht ist. Des Weiteren sind sie in ihrer ursprünglichen Form nur bei globalen Tonemapping-Verfahren einsetzbar, da nur für diese eine einheitlich starke Kontrastkompression vorliegt, die zur Berechnung benötigt wird.

Deshalb nutzt Pouli mit dem IPT einen Farbraum, der visuell gleichabständiger konstruiert ist. Durch seine höhere Dekorrelation zwischen Farbe und Helligkeit soll der Farbton durch die Kontrastkompression nicht beeinflusst werden. Die Chroma-Kompensation selbst wird durch eine Neudefinition des Sättigungsbegriffs erreicht, der wiederum auf Lübbe zurückgeht und auch beim lokalen Tonemapping eingesetzt werden kann. Für eine Anwendung im Broadcast wäre es allerdings wünschenswert direkt mit dem verwendeten Farbdifferenzsignal arbeiten zu können. In der abschließenden subjektiven Untersuchung in diesem Kapitel konnte darüber hinaus nachgewiesen werden, dass die Bildqualität hinter anderen Verfahren zurücksteht.

Bei dem im Rahmen dieser Arbeit durchgeführten Probanden-Grading wurde deshalb versucht, die Sättigungsproblematik auf eine Abwägungsentscheidung zwischen konstanter Sättigung und konstantem Chroma zu reduzieren (Koeffizienten-Methode), die durch die Teilnehmer nach ihrer subjektiven Präferenz getroffen werden sollte. Mit dem Koeffizient 0,73 konnte ein Wert gefunden werden, der für alle getesteten Bildvorlagen und äußeren Gegebenheiten einen guten Mittelwert mit nur geringer Standardabweichung bildete. Dabei konnte allerdings nicht ermittelt werden, ob der Faktor für alle Farbtöne

¹¹In diesem Fall als Standardabweichung eingezeichnet.

und Sättigungsniveaus identisch ausfällt. Die hierfür im Anschluss durchgeführte Untersuchung lieferte aufgrund von mangelhaftem Testdesign keine verwertbaren Resultate. Vielmehr bestätigte sie die Zweidimensionalität, der in diesem Kapitel behandelten Problematik. Zum einen muss die Chroma-Kompensation die nicht perfekte Dekorrelation des Farbraumes ausgleichen und zum anderen muss sie die veränderte Steilheit des Bildes berücksichtigen. Das der zweite Aspekt von der Koeffizienten-Methode nur unzureichend adressiert wird, konnte an sehr flachen Bildern im Abschnitt [5.10.1](#) gezeigt werden.

Auf Basis der bis zu diesem Zeitpunkt dargelegten Erkenntnisse, wurde deshalb das Konzept der konstanten Buntheit postuliert. Die Neudefinition des Sättigungsbegriffs bei Lübbe legte bereits nahe, dass das, was als konstanter Farbeindruck empfunden wird, nicht der Sättigung im klassischen Sinne entspricht. Gleiches gilt auch für das Chroma. Als letzter der drei Begriffe, welcher eine Art von Farbigkeit beschreibt, blieb entsprechend nur die Buntheit übrig. Für die vom Autor vorgeschlagene Berechnung wird neben der Helligkeit des Pixels noch ein Referenzweiß benötigt. Beide Werte durchlaufen in der Folge die identische Tonemapping-Funktion, wodurch eine Relation abgebildet werden kann. Wie hierdurch die Steilheitsänderungen kompensiert werden können, wurde in Abschnitt [5.10.1](#) dargelegt.

Abschließend galt es, die Theorie anhand eines Probandentest zu überprüfen. Es hat sich gezeigt, dass der Unterschied zwischen der Koeffizienten-Methode und der konstanten Buntheit unter realitätsnahen Bedingungen eher gering ausfällt, zumindest für die getesteten sechs Videosequenzen mit der gewählten Kontrastkompression. Da die konstante Buntheit doppelt so viele Tonemapping-Operationen notwendig macht, und deshalb ggf. nicht in allen denkbaren Applikationen eingesetzt werden kann, sollte in solchen Fällen auf die Koeffizienten-Methode zurückgegriffen werden. Die weiteren drei Verfahren konnten aus nicht näher bekannten Gründen aufgrund ihrer hohen Varianz bei den Bewertungen nicht überzeugen. Das bedeutet, dass aus wissenschaftlicher Sicht nicht alle Aspekte der Sättigungswahrnehmung bei der Kontrastkompression im Rahmen dieser Arbeit abschließend geklärt werden konnten. Allerdings war es möglich neue Zusammenhänge aufzuzeigen und hieraus eine praxistaugliche Lösung abzuleiten.

Kapitel 6

EVI: Theoretisches Modell zum TMO-Tuning

6.1 Virtuelle Blende

Basierend auf den Erkenntnissen der visuellen Präferenz soll im Folgenden ein Verfahren entwickelt werden, welches diese in Form eines Algorithmus abbildet. Es soll dazu dienen, die Lehren für den (Live-)Broadcast anwendbar zu machen. Das Verfahren wurde im Rahmen der **IBC** 2016 vorgestellt [72] und mit dem **IEI** *Journal Young Professional Award* ausgezeichnet [73]. Es ist gemeinhin unter dem Namen Enhanced Video Imaging (**EVI**) bekannt.¹ Des Weiteren findet es sich in der Patentschrift [74] wieder und wird dort aufgrund der Wirkungsweise „virtuelle Blende“ genannt.

Beim Coloristen-Test zeigte sich, dass die Probanden einen höheren subjektiven Kontrast bevorzugen, allerdings nur dann, wenn dies nicht zu einer deutlichen Abnahme des wahrnehmbaren Kontrastumfangs der Szene führt, wie später aus dem Probanden-Grading hervorging. Die Sichtbarkeit der relevanten Bildelemente wird demnach über eine hohe Gradation gestellt. Entsprechend muss es die Intention des Algorithmus sein, den Kontrastumfang zu erhalten, so lange er einen inhaltlichen Mehrwert bietet und Werte darüber hinaus zu verwerfen, um so ausreichend viele Abstufungen für den zu erhaltenden Helligkeitsbereich bereitzustellen.

Um eine automatische Ausbelichtung anhand der vorliegenden Leuchtdichtewerte vornehmen zu können und somit die beschriebene Abwägungsentscheidung zu treffen, bedarf es einer Bildanalyse. Dabei gilt es für den hier definierten Fall in erster Linie herauszufinden, wie groß der Kontrastumfang ausfällt und wie sich die Leuchtdichtewerte auf

¹**EVI** ist der Name des Forschungsprojektes. Produkte, die auf der **EVI**-Technologie beruhen, tragen eigene Handelsmarken.

diesen verteilen. So kann beispielsweise eine Szene mit einem großen Kontrastumfang, bei welchem sich allerdings die meisten Leuchtdichtewerte in einem engen Bereich zentrieren, anders behandelt werden, als es bei einer gleichmäßigen Verteilung der Fall wäre. Bei der Zentrierung wäre es sinnvoll, den Kontrastumfang stärker einzuschränken, um so den Einfluss der Extremwerte zu verringern und mehr Abstufungen für die dicht gedrängten Leuchtdichtewerte bereit zu stellen, so dass die Sichtbarkeit im betreffenden Bereich erhöht wird.

Das im Gebiet von Fotografie und Video oft verwendete Histogramm stellt hier ein geeignetes Werkzeug dar. Anhand dessen kann die Verteilung der Leuchtdichtewerte abgelesen werden. Der Abschnitt der größten Akkumulation soll nach der Bearbeitung die stärkste Gradation erhalten, um die Sichtbarkeit zu erhöhen. Gleichzeitig lässt sich erkennen, ob es sich um eine Tag- oder Nachtszene handelt, was einen Vorschlag für die präferierte Belichtung ermöglicht. Um der menschlichen Wahrnehmung Rechnung zu tragen, soll das Histogramm logarithmisch aufgetragen werden. Dies hat darüber hinaus den Vorteil, dass gleichgroße Abschnitte auf dem Histogramm den gleichen Kontrastumfang aufweisen.

Gegeben seien nun die Leuchtdichtewerte Y_0 eines Bildes am Eingang. Mithilfe der nachstehenden Gleichung wird die Logarithmierung durchgeführt. Die Konstante k_1 dient dabei dazu, dass Y_{log} keine negativen Werte annimmt. Standardmäßig weist k_1 den Wert 5 auf. Entsprechend können Werte bis minimal $0,00001 \text{ cd/m}^2$ verarbeitet werden. In der Testimplementierung wird der Kontrastumfang auf 16 Bit aufgetragen, wobei jeweils 5.000 Werte pro Dekade zur Verfügung stehen. Der maximal mögliche Wert läge somit über $128.000.000 \text{ cd/m}^2$. Jeder theoretisch in der Natur vorkommende Wert könnte also im vorliegenden Histogramm abgebildet werden.

$$Y_{log}(Y_0) = \begin{cases} \log_{10}(Y_0) + k_1 & \text{wenn } Y_0 > 10^{-k_1} \\ 0 & \text{wenn } Y_0 \leq 10^{-k_1} \end{cases} \quad (6.1)$$

Um den Abschnitt der größten Akkumulation auf dem Histogramm zu ermitteln, empfiehlt es sich, das kumulative Histogramm zu bilden und die Position der maximalen Steigung auf einer definierten Strecke W zu bestimmen. Die Strecke sollte eine ausreichende Breite besitzen, um nicht anfällig gegen Extremwerte oder Schwankungen zu sein. Gleichzeitig sollte sie allerdings auch nicht zu groß ausfallen, damit die Gradation nicht unnötig verflacht wird. Ein Wert um 1 Dekade ist dabei als sinnvoll zu erachten. Dazu wurde folgende Berechnungsvorschrift definiert:

$$W = k_2 + k_3 L_{dmax} \quad (6.2)$$

k_2 und k_3 sind Konstanten und weisen in der erwähnten Testimplementierung die Werte 4200 und 4 auf. Bei einer Leuchtdichte des Displays (L_{dmax}) von 200 cd/m^2 ergibt sich für W ein Wert von 5000, was einer Dekade an Kontrastumfang entsprechen würde. Die leichte Kopplung an L_{dmax} ist darin begründet, dass bei einem größeren Wert für W die Helligkeit des finalen Bildes abnimmt², was dem Ergebnis des Probanden-Grading entspricht.

$H(Y_{log})$ sei das kumulative Histogramm von $h(Y_{log})$. In der Folge ergibt sich für die Steigung folgende Bedingung:

$$H(Y_{log}) = \sum_{Y_{log}=0} h(Y_{log}) \quad (6.3)$$

$$H'(Y_{log}) \approx H\left(Y_{log} + \frac{1}{2}W\right) - H\left(Y_{log} - \frac{1}{2}W\right) = \max \rightarrow \mu \quad (6.4)$$

μ entspricht dabei dem Mittelpunkt der Strecke W , an welchem die Ableitung den höchsten Wert erreicht. Der Anfang g_l sowie das Ende g_r ergeben sich somit indirekt:

$$\begin{aligned} g_l &= \mu - \frac{1}{2}W \\ g_r &= \mu + \frac{1}{2}W \end{aligned} \quad (6.5)$$

Um eine Auswertung hinsichtlich der Zentriertheit des Histogramms durchführen zu können, bedarf es der Information über die Anzahl der Leuchtdichtewerte innerhalb des nun ermittelten Bereichs g_l bis g_r . Mithilfe der Normierung durch W ergibt sich eine durchschnittliche Höhe H_a :

$$H_a = \frac{\sum_{g_l}^{g_r} h(Y_{log})}{W} = \frac{H(g_r) - H(g_l)}{W} \quad (6.6)$$

Die durchschnittlicher Höhe H_a und Strecke g_l bis g_r können auch als die zwei Seitenlängen eines Rechtecks interpretiert werden. Dieses Rechteck wird nachfolgend als *Kontrastkasten* bezeichnet. Der Kontrastkasten gibt zum einen eine visuelle Auskunft über die Stärke der Akkumulation, zum anderen kann er genutzt werden, um eine Normalverteilung zu bilden, die das Histogramm bestmöglich abbildet. Dazu wird unter Zuhilfenahme einer zuvor definierten durchschnittlichen mittleren Höhe³ (H_{av}) zunächst

²Dieser Zusammenhang wird erst im weiteren Verlauf deutlich.

³Erfahrungswert gebildet aus der Analyse einer großen Anzahl an Testbildern

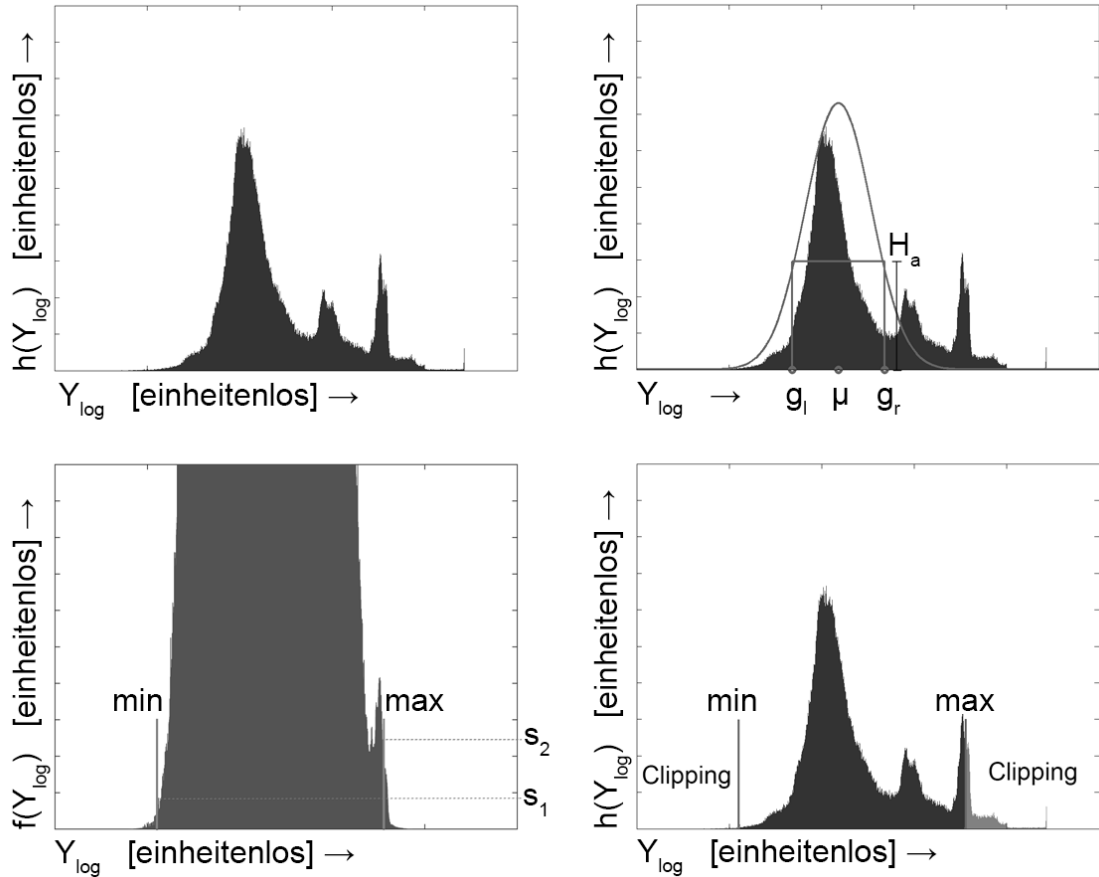


ABBILDUNG 6.1: Abbildung 1: von links oben nach rechts unten: (1) logarithmisches Histogramm der Eingangswerte, (2) Histogramm mit Kontrastkasten und angenäherter Gauß-Funktion, (3) gewichtetes Histogramm durch Multiplikation mit der Gauß-Funktion und Schnittpunkte mit den Grenzwerten, (4) Vergleich von absoluten und relevanten Extremwerten.

sigma (σ) bestimmt, welches so konstruiert ist, dass die anschließend aufgestellte Gauß-Funktion ungefähr durch die oberen beiden Ecken des Kontrastkastens verläuft.

$$\sigma = \left(\frac{H_{av}}{H_a} \right)^{k_5} H_{av} k_6 \quad (6.7)$$

$$g(Y_{log}) = \frac{H_a k_4 W}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_{log}-\mu)^2}{2\sigma^2}} \quad (6.8)$$

H_{av} , k_4 , k_5 und k_6 entsprechen in der Testimplementierung den heuristisch ermittelten Werten 250, 2,2, 1,5 und 10. Bei hoher Akkumulation bildet die Gauß-Funktion steile Flanken und fällt schnell ab. Bei einer niedrigen Akkumulation sind die Flanken entsprechend flach und laufen lange aus. Die Gauß-Funktion wird anschließend genutzt, um das eigentliche Histogramm zu gewichten.

$$f(Y_{log}) = h(Y_{log})g(Y_{log}) \quad (6.9)$$

Dieses nun sehr überhöhte Histogramm $f(Y_{log})$ wird abschließend mit den Schwellwerten s_1 und s_2 verglichen, welche beispielsweise beide den Wert 1 aufweisen können. Bei Bedarf können diese zusätzlich an die Umgebungsleuchtdichte $L_{surround}$ gekoppelt werden. Ein größerer Wert für $L_{surround}$ würde zu einem helleren Bild führen. Allerdings müsste für den Broadcast-Bereich zwangsläufig eine typische Betrachtungssituation angenommen werden. Auch hat sich beim Probaden-Grading gezeigt, dass die Umgebungsleuchtdichte in der untersuchten Größenordnung nicht relevant ist.

Durch den Abgleich der Funktion $f(Y_{log})$ mit s_1 und s_2 und anschließender Delogarithmierung ergeben sich das relevante Minimum Y_{min} und Maximum Y_{max} für das vorliegende Bild. Für s_1 muss dazu der kleinstmögliche Wert, für s_2 der größtmögliche Wert gefunden werden, der zu einer Überschreitung des Grenzwertes führt.

$$\begin{aligned} f(Y_{log}) &\geq s_1 L_{surround} \\ f(Y_{log}) &\geq s_2 L_{surround} \end{aligned} \quad (6.10)$$

In der Folge können die nicht relevanten Leuchtdichtewerte geclippt werden, um eine an das Bild besser angepasste Gradation zu ermöglichen.

$$f(Y_0) = \begin{cases} Y_{min} & \text{wenn } L_w < Y_{min} \\ Y_{max} & \text{wenn } L_w > Y_{max} \\ L_w & \text{sonst} \end{cases} \quad (6.11)$$

In Abbildung [6.1](#) sind die einzelnen Schritte anhand eines beispielhaften Histogramms dargestellt.

In Abbildung [6.2](#) wird anhand von vier Testbilder die unterschiedliche Ausprägung des Kontrastkastens bzw. der Gauß-Kurve und somit die angepasste Behandlung an die Bildvorlage sichtbar. Testbild 1 und Testbild 4 spiegeln den Fall einer hohen Akkumulation wider, wohingegen bei Testbild 2 eine hohe Streuung vorliegt. Bei Testbild 3 kann das Zusammenspiel von Histogramm und Gauß-Kurve am besten beobachtet werden. Das Zentrum des Interesses, das Auto, wird von niedrigen Leuchtdichten gebildet. Entsprechend weist die Gauß-Funktion bei hohen Leuchtdichten nur noch sehr kleine Werte auf. Da es aber beim Histogramm dort zu einem größeren Ausschlag kommt, überschreitet das Produkt aus Gauß und Histogramm den gesetzten Grenzwert auch bei einer sehr

hohen Leuchtdichte noch einmal. Sprich, obwohl dieser helle Leuchtdichtenbereich sich weit entfernt vom eigentlichen Zentrum des Interesses befindet, ist er doch so groß, dass ein vollständiges Clipping nicht sinnvoll wäre.

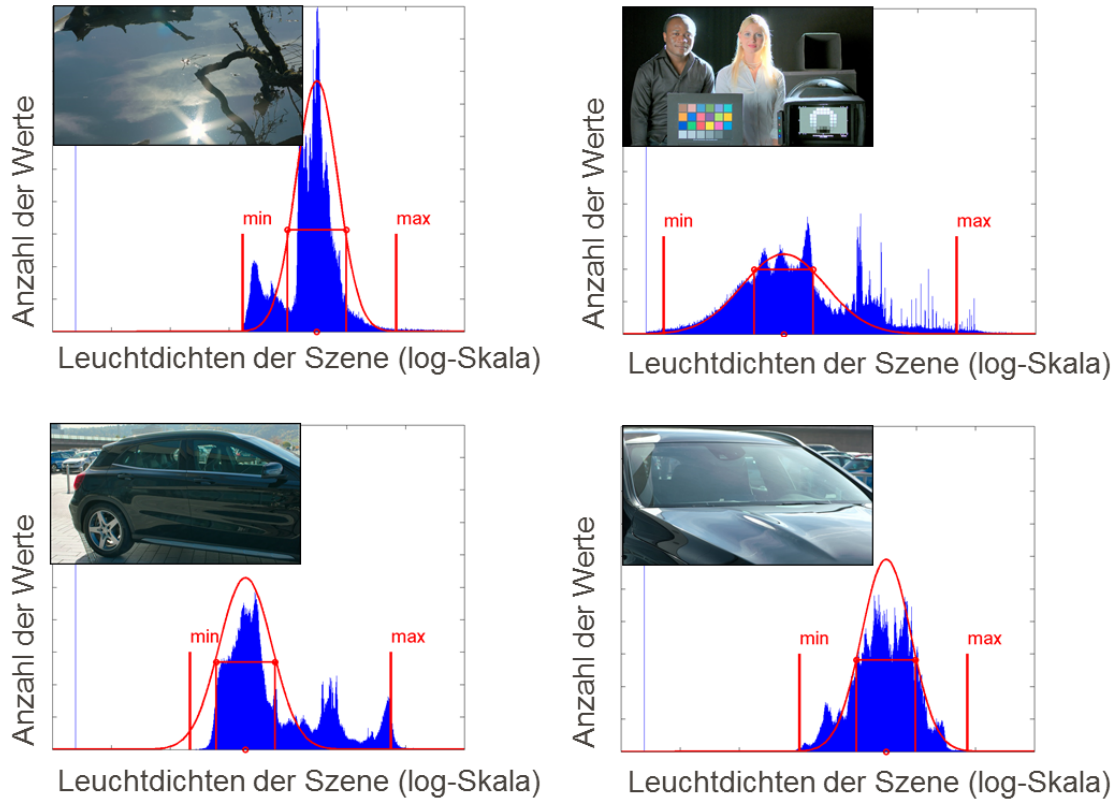


ABBILDUNG 6.2: Die virtuelle Blende reagiert je nach Histogramm sehr unterschiedlich auf die vier Testbilder

In Abbildung [6.3](#) sind die Auswirkungen der virtuellen Blende im Vergleich zum klassischen globalen Tonemapping sichtbar. Das gezeigte Bild korrespondiert mit dem in Abbildung [6.1](#) gezeigten Histogramm. Das mit der virtuellen Blende behandelte Bild auf der rechten Seite besitzt eine steilere Gradation, was die Helligkeit und die Sichtbarkeit sowie den Kontrast erhöht. Die negativen Auswirkungen sind in den für den Betrachter eher irrelevanten Bildpartien im Himmel und am Dach zu erkennen. Hier kommt es zu Clipping und Farbtonverschiebungen, Details sowie Zeichnung gehen verloren. Es wird somit eine Abwägungsentscheidung zu Gunsten der Geschehnisse auf dem Spielfeld getroffen und den Ergebnissen bezüglich der Parameter Helligkeit und Kontrast aus den Probandenversuchen Rechnung getragen.

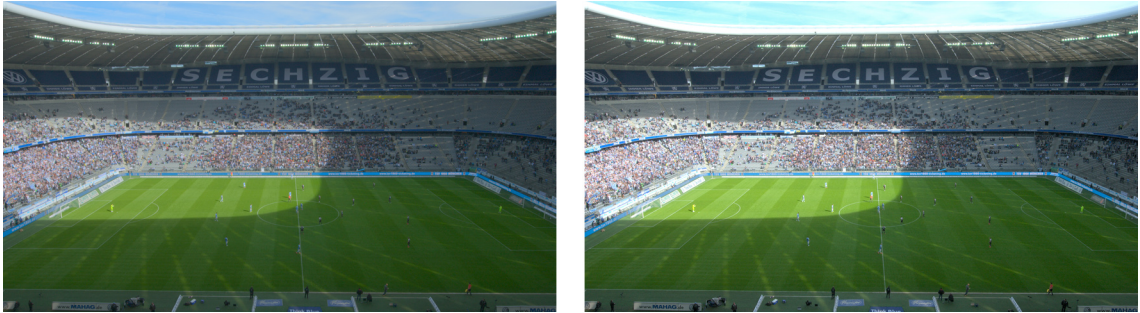


ABBILDUNG 6.3: Auswirkung der virtuellen Blende (rechts) im Vergleich zum klassischen globalen Tonemapping (links)

6.2 Sektionales Tonemapping und bilaterale Filterung

In Abbildung 6.3 wird gezeigt, wie die virtuelle Blende TMOs hinsichtlich deren Helligkeits- und Kontrastdarstellung in Richtung der visuellen Präferenz modifizieren kann. Gleichzeitig werden aber erneut die Limitationen eines globalen Vorgehens deutlich, weswegen die Homogenisierung beim Probanden-Grading in den meisten Fällen einen Wert um 0,5 aufwies. Auf der anderen Seite hat sich in Kapitel 2 gezeigt, dass klassische lokale Operatoren schwer zu kontrollieren sind.

Deshalb soll im Folgenden das Verfahren aus Abschnitt 6.1 um eine lokale Komponente erweitert werden. Im Vergleich zum klassischen lokalen Tonemapping ist das Vorgehen aber ein gänzlich anderes. Während dort ein Pixel betrachtet und seine Nachbarschaft zur Bestimmung des Adaptionniveaus genutzt wird, soll für die virtuelle Blende das Bild gitterförmig aufgeteilt werden und die Auswertung bezüglich der relevanten Leuchtdichtewerte auf den einzelnen Blöcken erfolgen. Abbildung 6.4 zeigt eine solche Auswertung. Die grüne Linie repräsentiert dabei den min- und die rote Linie den max-Wert. In Anlehnung an die Blöcke soll dieses Vorgehen sektionales Tonemapping genannt werden.

In Abbildung 6.4 sind insgesamt 144 Blöcke zu sehen. Diese Zahl hat sich als günstig erwiesen, da sie zum einen ausreichend groß ist, um lokale Unterschiede aufzuzeigen, aber auch noch nicht zu groß, so dass zu wenig Werte für eine statistische Auswertung vorhanden wären. Versuche mit 64 Blöcken hatten gezeigt, dass eine derart große Unterabtastung die Stärken eines lokalen Tonemappings nicht mehr gänzlich ausspielen kann und Helligkeitsgrenzen zwischen Objekten oder Lichtern und Schatten zu stark verschliffen werden. Bei 256 Blöcken hingegen konnte keine aussagekräftige Statistik mehr gebildet werden und die Bereiche waren so klein, dass es bei stärkeren Bewegungen in der Szene, wie beim klassischen lokalen Tonemapping, wieder zum Flackern oder Pumpen kam. Da 144 auch der Multiplikation aus 16 und neun entspricht, ergibt sich der Vorteil von ideal quadratischen Blöcken⁴

⁴Die Argumentation gilt maßgeblich für HD und UHD-1. Größere Bildauflösungen könnten zukünftig mit weiteren Bildausschnitten einhergehen, wodurch die Anzahl der Blöcke beeinflusst werden könnte.

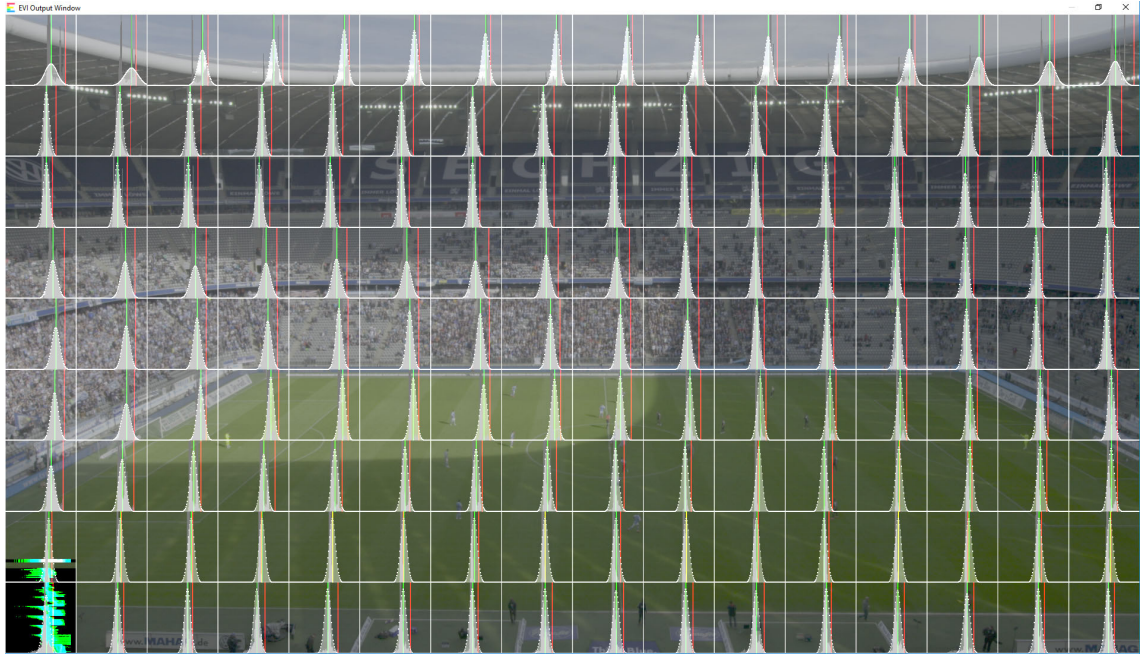


ABBILDUNG 6.4: Sektorenweise Berechnung der virtuellen Blende

Um letztlich jedem Pixel einen Wert zuweisen zu können und die Blockgrenzen aufzulösen, muss das Ergebnis der Sektorenanalyse anschließend gefiltert werden. Auf der linken Seite von Abbildung [6.5](#) kann beobachtet werden, wie das Bild ohne solche eine Filterung aussähe. Die rechte Seite nimmt das Ergebnis vorweg. Im Vergleich zu Abbildung [6.3](#) können nun der Himmel und die in der Sonne befindlichen Zuschauerränge besser erhalten und gleichzeitig der Schatten weiter angehoben werden. Das Bild wirkt schärfer und plastischer.

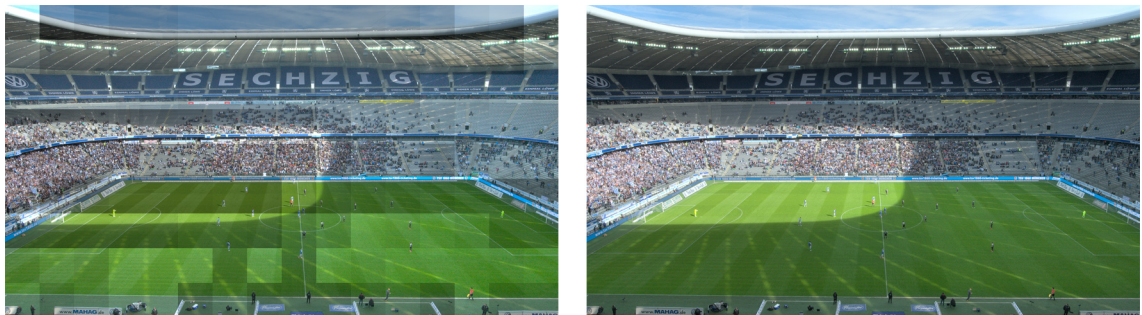


ABBILDUNG 6.5: Auswirkung der sektionalen Blende mit und ohne Filterung der Blöcke

Um ein Ergebnis, wie in Abbildung [6.5](#) gezeigt, zu erreichen, bedarf es eines Weichzeichnungsfilters. Dieser Filter wird allerdings nicht auf die Bildvorlage angewendet, sondern einzig auf die Resultate der Sektorenanalyse. Es muss eine Korrekturmaske gebildet werden, die letztendlich die selbe Größe aufweist wie die Bildvorlage. Es handelt sich somit um eine starke Vergrößerung mit gleichzeitiger Weichzeichnung.

Klassische Weichzeichenfilter (Bilinear, Bikubisch, Gauß, Butterworth) gewichten dazu den Einfluss der Umgebung des Pixels basierend auf ihrer räumlichen Distanz. Problematisch in Bezug auf den vorliegenden Fall ist dabei die Tatsache, dass Objektkanten nicht erhalten werden können. Die Folge ist das bereits erwähnte Übersprechen, das sich in Halo-Artefakten äußert. Da durch die Sektoren die Filterung deutlich großflächiger geschieht, als es beim lokalen Tonemapping der Fall wäre, ist das Risiko maßgeblich reduziert. Ein Filterkernel, auch Faltungsmatrix genannt, der Größe 5x5 wirkt normalerweise auf eine Fläche von 25 Pixeln. Im Falle des sektionalen Vorgehens sind es entsprechend 25 Blöcke. Bei einem FullHD-Bild, das in 144 Blöcken unterteilt wurde, ergibt sich somit indirekt eine Fläche von 90.000 Pixeln⁵. Dabei muss beachtet werden, dass die Koeffizienten des Filterkerns sich je nach Lage des Zielpixels innerhalb des Blocks unterscheiden und deshalb zunächst pro Pixel berechnet werden müssen, während sie bei klassischen Weichzeichenfiltern für jeden Pixel gleich sind. Befindet sich das Pixel am Rand des Blocks, gehen die Nachbarblöcke zu einem größeren Teil ein, als es bei einem Pixel im Blockzentrum der Fall wäre. Trotz des großen Filterradius können aber auch beim sektionalen Tonemapping mit beispielsweise einem bikubischen Filter Halo-Artefakten nicht gänzlich vermieden werden.

Um Flächen weichzuzeichnen, Kanten aber trotzdem zu erhalten, wurde ab 1990 auf die Anisotrope Diffusion [95] zurückgegriffen, die eine so genannte Kanten-Stopp-Funktion verwendet, die sich nach den Bildgradienten richtet. Ab 1998 wurde sie durch die bilaterale Filterung [122] abgelöst. Dabei wird neben der räumlichen Distanz noch die inhaltliche Distanz als Maßstab zur Berechnung des Filterkerns herangezogen. Nachbarpixel, die eine ähnliche Helligkeit (oder Farbe) aufweisen, wie das ausgewählte Pixel, gehen stärker in die Berechnung des lokalen Adaptionsniveaus ein, als solche, die sich deutlich unterscheiden. Objektgrenzen bleiben in der Folge erhalten. Die Tonemapping-Verfahren von Durand und Dorsey [28] oder Pattanaik und Yee [92] basieren auf dieser Idee. Mathematisch lässt sich die Vorschrift folgendermaßen ausdrücken:

$$\hat{I}_p = \frac{1}{W_p} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I_p - I_q|) I_q \quad (6.12)$$

Der Normalisierungsterm $1/W_p$ ergibt sich folglich durch:

$$W_p = \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I_p - I_q|) \quad (6.13)$$

⁵Natürlich werden gewöhnlich deutlich größere Filterkernel eingesetzt. Allerdings würde die Betrachtung einer Umgebung von 90.000 Pixeln mindestens 180.001 Rechenoperationen pro Pixel bedeuten, wenn der Filterkernel aus unterschiedlichen Gewichtungen besteht.

Dabei ist p die Position und I_p die Intensität des Ausgabepixels, während q und I_q die Position bzw. die Intensität eines Pixels in der Nachbarschaft S sind. Die Gaußfunktionen G_{σ_s} und G_{σ_r} gewichten die Pixel aufgrund ihrer räumlichen bzw. inhaltlichen Distanz.

Für die Anwendbarkeit auf Blöcke wird die bilaterale Filterung neu interpretiert. Anstelle einer Faltung des Bildes im Ortsbereich mit sich selbst, wird hier die stark unterabgetastete Maske mit lediglich 144 Werten unter Zuhilfenahme des Originalbilds gefaltet. Damit wird es möglich, die Objektkanten aus der aktuellen Bildvorlage in die Maske zu transferieren, obwohl die Bildanalyse nur eine sehr begrenzte Anzahl an Werten liefert. Sprich, im Vergleich zum klassischen bilateralen Tonemapping waren hier die Objektkanten bereits verloren, können aber durch die Filterung zurückgewonnen werden.

Diese Tatsache ermöglicht in der Folge erstmals das im folgenden Abschnitt [6.3](#) erklärte Parameter-Smoothing auf lokaler Ebene, da 144 Werte leicht verschliffen werden können, aber trotzdem nach der Filterung scharfe und pixelgenaue Kanten liefern. Ferner ist sichergestellt, dass die Analyseergebnisse weiterhin den passenden Pixeln zugeordnet werden, auch wenn sich diese durch Bewegungen verschoben haben. Der 5×5 -Kernel K , der mit Hilfe des modifizierten bilateralen Filters errechnet wird und die Gewichtung der Blöcke angibt, lässt sich wie folgt ausdrücken:

$$K = \begin{bmatrix} sd(2bs, 2bs, x, y) id(2bs, 2bs, x, y) & \dots & sd(-2bs, 2bs, x, y) id(-2bs, 2bs, x, y) \\ \vdots & sd(0, 0, x, y) id(0, 0, x, y) & \vdots \\ sd(2bs, -2bs, x, y) id(2bs, -2bs, x, y) & \dots & sd(-2bs, -2bs, x, y) id(-2bs, -2bs, x, y) \end{bmatrix} \quad (6.14)$$

bs ist dabei die Blockweite in Pixeln. x und y sind die Koordinaten des Eingangsbildes bzw. der gefilterten Maske. Die spatiale Distanz $sd()$ ergibt sich aus:

$$sd(m, n, x, y) = 2,5bs - \sqrt{(x \bmod bs - m - 0,5bs)^2 + (y \bmod bs - n - 0,5bs)^2} \quad (6.15)$$

Die inhaltliche Distanz $id()$ ergibt sich anschließend wie folgt:

$$id(m, n, x, y) = id_{max} - \left| I_b \left(\lfloor \left(\frac{m+x}{bs} \right) \rfloor, \lfloor \left(\frac{n+y}{bs} \right) \rfloor \right) - I_o(x, y) \right| \quad (6.16)$$

I_b liefert den Intensitätswert der Blockmaske (Ergebnis der sektionalen virtuellen Blende) an der entsprechenden Stelle. I_o bildet das Äquivalent für das Eingangsbild. id_{max} ist

der Wert für die theoretisch maximale Distanz. Alle drei Werte sollten als logarithmisch verzerrte Leuchtdichten vorliegen.

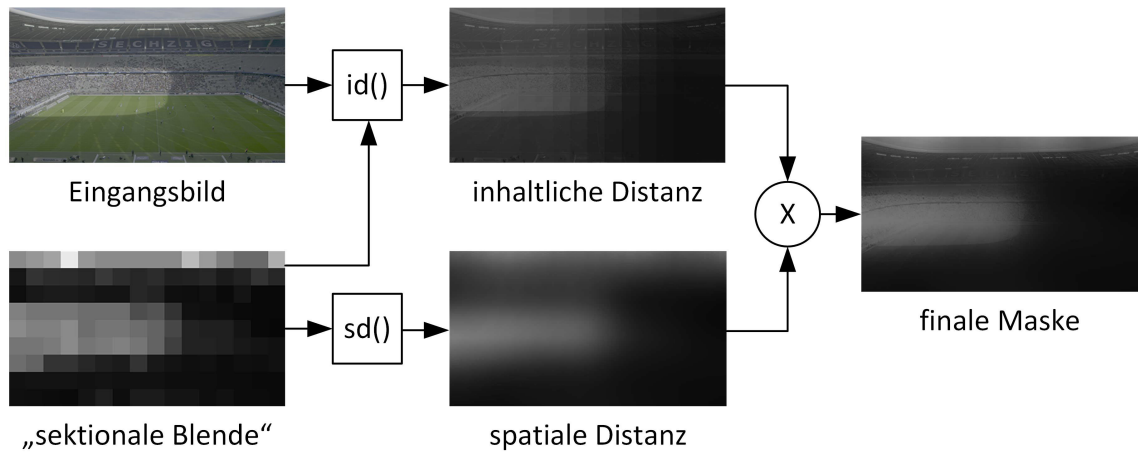


ABBILDUNG 6.6: Einzelschritte der bilateralen Filterung

Abbildung 6.6 zeigt die einzelnen Schritte der Filterung, während in Abbildung 6.7 der Unterschied zwischen einem klassischen bikubischen Filter und dem implementierten bilateralen Filter anhand von zwei Bildern illustriert wird.



ABBILDUNG 6.7: Vergleich zwischen der bikubischen (links) und der bilateralen Filterung (rechts): Die bilaterale Filterung ist in der Lage die Kanten aus dem eingehenden Bild in die Maske zu transferieren.

Das nun vorliegende Ergebnis der gefilterten sektionalen Auswertung wird im Anschluss, wie schon in der Probanden-Studie, nach unterstehender Formel (6.17) mit dem globalen Ergebnis verrechnet, um den Gesamteindruck der Szene zu erhalten. Zu Gunsten einer

klaren Trennung von Skalaren werden die Matrizen im folgenden mit (x,y) gekennzeichnet. Die reine Verwendung der sektionalen Werte würde einer Homogenisierung von eins entsprechen. In diesem Fall könnte, durch den betonten lokalen Kontrast und die starke Angleichung von Lichtern und Schatten, eine leicht surreale Bildanmutung entstehen.

$$Y_{max}(x,y) = \text{homogenisation} Y_{max}(x,y) + (1 - \text{homogenisation}) Y_{max} \quad (6.17)$$

In der obigen Gleichung wurde die Verwendung für die maximale Leuchtdichte gezeigt. Entsprechend können auch die minimale Leuchtdichte sowie das logarithmische Mittel anhand der Blöcke berechnet werden, je nachdem, welche Anforderungen der gewählte **TMO** besitzt. Nachstehend sollen die für das sektionale Tonemapping modifizierten Gleichungen beispielhaft am Drago Operator [25] gezeigt werden:

$$Y_{t1}(x,y) = \frac{a}{Y_{mean}(x,y)} Y_0 \quad (6.18)$$

$$Y_t(x,y) = \frac{1}{\log_{10}(Y_{max}(x,y) + 1)} \frac{\ln(Y_{t1}(x,y) + 1)}{\ln \left(2 + \left(\left(\frac{Y_{t1}(x,y)}{Y_{max}(x,y)} \right)^{\frac{\ln(b)}{\ln(0,5)}} \right) 8 \right)} \quad (6.19)$$

Zusammenfassend kann gesagt werden, dass beim lokalen Tonemapping die Betrachtung vom Kleinen ins Große erfolgt, wohingegen beim sektionalen Tonemapping von Groß nach Klein geschaut wird. Im zweiten Fall kann eine deutlich größere Nachbarschaft pro Pixel berücksichtigt werden. Entsprechend fällt es zwar schwerer, feine lokale Adaptioneniveaus zu erhalten, dafür können aber Probleme wie Halos und ein surrealer Bildeindruck umgangen werden. Gleichzeitig entspricht die großflächigere Betrachtungsweise eher der menschlichen Wahrnehmung. Darüber hinaus ermöglicht das sektionale Vorgehen in Kombination mit dem bilateralen Filter erstmals den Einsatz von Parameter-Smoothing auf lokale Bereiche und damit das Erreichen einer temporalen Stabilität ohne komplexe Prädikationsmodelle. Der folgende Abschnitt widmet sich diesem Aspekt.

6.3 Temporale Glättung und Szenenwechsel

In Kapitel 3 wurde definiert, dass eine temporale Glättung nötig ist, um Flackern zu vermeiden. Hierfür konnten brauchbare Beispiele in der Literatur gefunden werden. Gleichzeitig wurde allerdings eine Lücke bezüglich der Szenenerkennung deutlich, die eine Interpolation über Umschnitte hinweg unterbinden soll. Mithilfe von μ und σ aus

Gleichung [6.4](#) bzw. [6.7](#) kann im Folgenden bei der virtuellen Blende festgestellt werden, ob ein Szenewechsel vorliegt.

Zunächst sei eine Interpolation ähnlich der bei Kiser et al. [\[61\]](#) angenommen. Anstelle für μ und σ die Werte des aktuellen Bildes (μ_n und σ_n) bei der Generierung der Gauß-Funktion zu verwenden, wird die temporal gefilterte Variante (μ_{gn} und σ_{gn}) genutzt. Diese wiederum bildet sich aus den Werten des aktuellen Bildes und aus den gefilterten Vorgängern ($\mu_{g(n-1)}$ und $\sigma_{g(n-1)}$). Bei solch einem Konzept ist es in der Folge nicht nötig, die Werte der vorherigen Bilder zu speichern. Einzig der direkte gefilterte Vorgänger muss bekannt sein.

z_μ und z_σ geben an, zu welchem Anteil der Wert des aktuellen Bildes und zu welchem Anteil der gefilterte Vorgänger in die Berechnung eingehen. In der Praxis können z_μ bzw. z_σ identisch sein. Typisch wäre ein Wert um 0,95. Dies würde bedeuten, dass das neue Bild zu einem Anteil von 5% in das Gesamtergebnis eingeht. Es ist sinnvoll diesen Faktor an die Bildwechselfrequenz zu koppeln, um eine gleichbleibend schnelle Adaption unabhängig von der Hz-Zahl zu gewährleisten.

$$\begin{aligned}\mu_{gn} &= \mu_{g(n-1)} * z_\mu + \mu_n * (1 - z_\mu) \\ \sigma_{gn} &= \sigma_{g(n-1)} * z_\sigma + \sigma_n * (1 - z_\sigma)\end{aligned}\tag{6.20}$$

$$z_\mu = z * \frac{50}{\text{framerate}}\tag{6.21}$$

Formel [6.20](#) kann für den sektionalen Fall einfach für alle Blöcke wiederholt werden. Dadurch wird erstmals das Parameter-Smoothing für ein lokales Verfahren einsetzbar. i und j bilden die Nummerierung der einzelnen Blöcke.

$$\begin{aligned}\mu_{gn}(i,j) &= \mu_{g(n-1)}(i,j) * z_\mu + \mu(i,j) * (1 - z_\mu) \\ \sigma_{gn}(i,j) &= \sigma_{g(n-1)}(i,j) * z_\sigma + \sigma(i,j) * (1 - z_\sigma)\end{aligned}\tag{6.22}$$

Gleichzeitig können μ_n und $\mu_{g(n-1)}$ bzw. σ_n und $\sigma_{g(n-1)}$ genutzt werden, um starke Schwankungen im Histogramm und somit auch im Bildinhalt zu detektieren. Übersteigt die Differenz den zulässigen Grenzwert s_μ bzw. s_σ , wird von einem Szenenwechsel ausgegangen. Die Grenzwerte befinden sich in einer Größenordnung von 15%. Auch diese

Gleichungen können auf den sektionalen Fall übertragen werden. Dort wird solche eine Auswertung auf jedem Block durchgeführt. Werden in einer definierten Anzahl von Blöcken Grenzwertüberschreitungen festgestellt, kann von einem Szenenwechsel ausgegangen werden.

$$\begin{aligned} \left| \frac{\mu_n - \mu_{g(n-1)}}{\mu_{g(n-1)}} \right| &\leq s_\mu \\ \left| \frac{\sigma_n - \sigma_{g(n-1)}}{\sigma_{g(n-1)}} \right| &\leq s_\sigma \end{aligned} \quad (6.23)$$

In der Praxis wird mit den sektorenbasierter Szenenwechsellerkennung gearbeitet. Statistisch gesehen ist es mäßig wahrscheinlich, dass zwei Bilder mit unterschiedlichem Inhalt ein ähnliches Histogramm aufweisen. Dass diese Verteilung aber auch räumlich ähnlich ausfällt, ist hingegen sehr unwahrscheinlich. Gesetzt dem Fall, eine solche Verteilung würde trotzdem vorliegen, ist ein nicht erkannter Szenenwechsel für den Zuschauer nicht sichtbar, da beide Szenen zu einem nahezu identischen Analyseergebnis kommen würden. Genau genommen ist somit der Terminus Szenenwechsellerkennung nicht ganz korrekt, da es vielmehr darum geht, starke Schwankungen in der Helligkeitsverteilung zwischen zwei Bildern zu detektieren, welche dann einer Neuausrichtung der dynamischen Kontrastkompression bedürfen. Dies kann auch innerhalb einer Szene bzw. muss nicht unbedingt bei jedem Umschnitt der Fall sein. Folglich kann nicht ohne Weiteres auf konventionelle Techniken der Szenenwechsellerkennung zurückgegriffen werden. Verfahren, die Kenntnis über zukünftige Bilder benötigen, scheiden wegen des Delays grundsätzlich aus (s. [3.1.4](#)). μ und σ sind bereits als Nebenprodukt der Kontrastkompression bekannt und spiegeln gleichzeitig die Relevanz der Helligkeitsänderung für die Kontrastkompression wider.

6.4 Chroma-Korrektur

Nach den Ergebnissen von Kapitel [5](#) bestehen zwei Möglichkeiten, um die Chroma-Kompensation durchzuführen. Zum einen kann auf den Ansatz zur konstanten Buntheit zurückgegriffen werden:

$$C_t = C_0 \frac{I_{white-t}}{I_{white-0}} \quad (6.24)$$

C_0 steht für eine der beiden Chroma-Komponenten vor und C_t nach der Chroma-Kompensation. Bei $I_{white-0}$ handelt es sich um das Luma des Referenzweiß vor und bei $I_{white-t}$ nach dem Tonemapping.

Zum anderen, kann alternativ wie bereits erörtert, die weniger rechenintensive Formel der Koeffizienten-Methode eingesetzt werden. Hier wird das Luma des Pixels vor (I_0) und nach dem Tonemapping (I_t) herangezogen. Der Faktor f_c entspricht dem Wert 0,73, welcher beim Probanden-Grading ermittelt wurde.

$$C_t = C_0 * \left((1 - f_c) + f_c * \left(\frac{I_t}{I_0} \right) \right) \quad (6.25)$$

Beide Formeln können sowohl auf ein IC_tC_p - als auch auf ein $Y'C_b'C_r$ -Signal angewendet werden, wobei sich aufgrund der besseren Dekorrelation IC_tC_p empfiehlt.

6.5 Zusammenfassung EVI

In vorliegenden Kapitel wurde ein Verfahren zur Modifikation von TMOs bezüglich der subjektiven Präferenz des Zuschauers entwickelt. Im Zentrum steht dabei die automatisierte Abwägungsentscheidung zwischen einem hohen Szenenkontrastumfang und einem hohen subjektiven Kontrast, welche anhand eines logarithmischen Histogramms und einer Gauß-Funktion durchgeführt wird. Dabei wird bewusst auf weniger relevante Leuchtdichtewerte zu Gunsten einer besseren Gradation in den entscheidenden Leuchtdichtekorridoren verzichtet. Durch die Anwendung dieser Funktion auf Blockebene ist es möglich einen höheren Szenenkontrastumfang zu erhalten, ohne dabei einen flachen Bildeindruck zu verursachen. Des Weiteren wird ein Angleichen von Lichtern und Schatten ermöglicht, das zu einem vom Zuschauer präferierten homogeneren Erscheinungsbild führt.

Im Vergleich zum klassischen lokalen Tonemapping können die Analyseergebnisse der einzelnen Blöcke zwischengespeichert und zeitlich interpoliert werden, so dass das sogenannte Parameter-Smoothing eingesetzt werden kann. Auch dienen sie dazu einen möglichen Szenenwechsel zu detektieren, bei dem die Interpolation ausgesetzt werden sollte. Die geringe Auflösung, die mit dem blockbasierten Vorgehen einhergeht, kann über einen modifizierten bilateralen Filter ausgeglichen werden. Dieser zeichnet die Übergänge zwischen den einzelnen Blöcken weich und stellt gleichzeitig die wichtigen Kanten der Szenenobjekte wieder her, wodurch Halos umgangen werden.

Da die Helligkeitsänderung auch die Sättigungswahrnehmung einer Farbe beeinflusst, wurden in der Folge zwei Funktionen zur Chroma-Kompensation vorgestellt, die sowohl auf IC_tC_p - als auch auf ein $Y'C_b'C_r$ -Signale angewendet werden können.

Kapitel 7

Nachweis der Qualitätssteigerung und praxistauglicher Workflow

Im vorherigen Kapitel wurde die Theorie des **EVI**-Verfahrens erklärt. In diesem Kapitel soll der Übergang zur Praxis geschaffen werden. Dabei geht es darum, den Qualitätsunterschied zwischen dem Verfahren und dem Status-Quo zu belegen sowie potenzielle Workflows und die Bedienung darzustellen.

In den folgenden zwei Abschnitten soll es zunächst um die Frage der Bildqualität gehen. Dabei sollen sowohl objektive (Peak Signal-to-Noise Ratio (**PSNR**) und Datenrate) als auch subjektive Messmethoden (Double-Stimulus Continuous Quality-Scale (**DSCQS**) [56]) zum Einsatz kommen. Dabei ist es entscheidend, mit jeweils codierten Sequenzen zu arbeiten, wie sie auch bei der Distribution vorliegen würden. Die bisherigen Erkenntnisse legen nahe, dass der Informationsgehalt einer **EVI**-Sequenz deutlich höher ist als beim klassischen **SDR**. Dies könnte sich auf die benötigte Datenrate bzw. den **PSNR** auswirken. Eine höhere Datenrate bzw. ein niedrigerer **PSNR** wären ein Beleg für den höheren Informationsgehalt. Allerdings stellt sich damit die Frage, ob die Qualitätssteigerung durch die bessere Kontrastdarstellung nicht durch die höhere Anzahl an Codierartefakten egalisiert wird.

7.1 Auswirkungen auf den PSNR

Zunächst seien die Auswirkungen auf die Datenrate bzw. den **PSNR** genauer analysiert. Die dargestellten Ergebnisse finden sich auch im IBC-Paper [76] und im **SMPTE** Motion Imaging Journal [69] wieder.

Ein **SDR**-Bild, das mithilfe von Tonemapping aus einem **HDR**-Bild gewonnen wurde, könnte eine höhere Datenrate benötigen, da im Vergleich zum klassischen **SDR** mehr Details in den Lichtern und Schatten vorhanden sind. Die Datenrate könnte sogar höher liegen als bei **HDR**-Signalen, die mit der **PQ**- oder **HLG**-Kurve [58] quantisiert sind, da jene eine weitaus bessere Bitverteilung aufweisen.¹ Eine Betrachtung dieses Sachverhaltes ist allerdings nicht alleine für die **SDR**, sondern auch für **HDR**-Auspielungen relevant, falls letztere auf den 2-Layer-Ansatz zurückgreifen. In diesem Fall soll der Base-Layer das klassische **SDR**-Bild ablösen. Aus den oben genannten Gründen sind diese jedoch nicht identisch.

Um quantitative Aussagen machen zu können, sollen daher detaillierte Untersuchungen sowohl mit objektiven als auch mit subjektiven Methoden durchgeführt werden. Zunächst gilt es den **PSNR** zu bestimmen. Bezüglich der Auswirkung des Tonemappings auf eine Interframe-Codierung wurden vor Beginn des Experiments zwei Hypothesen aufgestellt:

- Zum einen könnte die Datenrate der I-Frames signifikant steigen, da die zusätzlichen Informationen des erhöhten Kontrastumfangs abgebildet werden müssen.
- Bei P- oder B-Frames könnte allerdings ein umgekehrter Effekt zu beobachten sein, da das Tonemapping die Konvergenz in der zeitlichen Richtung verbessert. So bedarf es beispielsweise keiner Blendenänderungen bei sich moderat ändernden Lichtverhältnissen. Entsprechend kann die Bewegungsschätzung effektiver arbeiten.

7.1.1 Stand der Forschung bezüglich des PSNR

Es sind nur wenige relevante Studien bekannt, die den Einfluss einer **HDR**-Produktion auf die Datenrate untersuchen. Allen voran sind hier die Ergebnisse von Litwic et al. [78] anzuführen. Dabei wurden vier Testsequenzen sowohl für **SDR** als auch für **HDR** aufbereitet. Anschließend erfolgte eine Codierung mit dem HEVC-Verfahren bei Datenraten von 1, sowie 3 und 5 Mbit/s. Die Bittiefe betrug durchgängig 10 Bit. In einem Probandentest wurde die jeweilige Referenz mit der codierten Version verglichen, womit die Problematik des direkten Vergleichs von **SDR** mit **HDR** umgangen wurde. Die Ergebnisse wurden in **MOS** im Sinne des Double-Stimulus Continuous Quality-Scale (**DSCQS**) bestimmt.

Das Ergebnis dieser subjektiven Tests war nicht einheitlich: Bei der niedrigsten Datenrate (1 Mbit/s) wurde die Bildqualität der **SDR**-Sequenzen als besser beurteilt, als die

¹Die Quantisierungsstufen sind besser an die menschliche Wahrnehmung angepasst (s. [2.2.5]).

bei **HDR**, wohingegen die **HDR**-Bilder bei der höchsten Datenrate (5 Mbit/s) als besser bewertet wurden. Eine wichtige Erkenntnis dabei war, dass das Ergebnis stark vom Bildinhalt abhängig ist. Bei den vier untersuchten Testsequenzen konnte, über alle Datenraten gesehen, **HDR** bei zwei Sequenzen einen höheren Wert erzielen. Einmal war das Ergebnis identisch und einmal war der Wert geringer als bei **SDR**. Die Autoren folgerten daraus, dass wegen der höheren Nichtlinearität der **PQ**-Transferfunktion bei **HDR** dies in eher dunkleren Sequenzen einen positiven Einfluss auf die Bildqualität hat und das Banding reduziert, da hierbei mehr Werte zur Verfügung stehen (also feiner quantisiert wird) als dies bei **SDR** mit einer Gamma-Kurve der Fall wäre. Für sehr helle Sequenzen ist ein entgegengesetzter Effekt zu beobachten. Aus diesen Untersuchungen kann allerdings nicht interpretiert werden, dass **HDR**-Sequenzen generell für die im Broadcast-Betrieb typischen Datenraten effizienter codiert werden können, da das **SDR**-Bild aus der **HDR**-Sequenz durch eine manuelle Bildbearbeitung gewonnen wurde. Entsprechend besitzt dieses **SDR**-Bild eine höhere Bildqualität als es bei einer klassischen **SDR**-Aufnahme der Fall wäre.

Mantiuk et al. [83] haben bereits 2006 untersucht, wie groß der Mehraufwand bei der 2-Layer-Methode wäre. Im Vergleich zur Datenrate des Base-Layers, der das **SDR**-Signal enthält, benötigt der Enhancement-Layer für die zusätzliche **HDR**-Übertragung weitere ca. 20-25% Datenrate. Allerdings entspricht auch hier der Base-Layer nicht dem heutigen TV-Signal.

7.1.2 Versuchsaufbau und -durchführung (PSNR)

In der folgenden Untersuchung wurden verschiedene Testsequenzen genutzt, die sich hinsichtlich ihrer zeitlichen Länge, des Kontrastumfangs der Szene, der Bewegung innerhalb des Bildes und der Kamerabewegung unterscheiden. Eine genaue Erläuterung der gerade genannten Charakteristika folgt an späterer Stelle, auch wenn sie bereits in der nachstehenden Übersicht der Testsequenzen in Tabelle 7.1 mit aufgeführt sind.

Testsequenz	fps	Dauer	Kontrastumfang	Bewegung	Leuchtdichte
Anstoß [111]	50	10,00 s	9,10 EVs	0	80,6 cd/m ²
Allianz Arena [111]	50	10,00 s	7,00 EVs	4	739,9 cd/m ²
Stadion Schwenk [111]	50	27,76 s	7,49 EVs	8	173,5 cd/m ²
Zuschauer Schwenk [111]	50	10,00 s	9,61 EVs	7	198,7 cd/m ²
bistro 2 [41]	25	25,92 s	13,96 EVs	3	5,4 cd/m ²
bistro 3 [41]	25	6,80 s	11,26 EVs	4	19,7 cd/m ²
fishing longshot [41]	25	33,96 s	10,55 EVs s	6	7,7 cd/m ²
cars fullshot [41]	25	17,68 s	11,08 EVs s	5	17,1 cd/m ²

TABELLE 7.1: Übersicht der Testsequenzen für die Datenratenanalyse

Alle Testsequenzen lagen im OpenEXR-Dateiformat vor, welches die Speicherung des größtmöglichen Kontrastumfangs bietet. Aus diesen linearen Leuchtdichtewerten wurden vier Versionen generiert:

- **TV-Simulation**: Wie der Name suggeriert, soll diese Version das klassische **SDR**-Bild mit einem Kontrastumfang von bis zu ca. 8,5 Blendenstufen simulieren. Dies entspricht dem maximal möglichen, darstellbaren Kontrastumfang beim heutigen Fernsehsystem nach ITU-R BT.709 [54]. In Szenen mit starken Schwankungen in der Belichtung, beispielsweise ausgelöst durch einen Kameraschwenk, wird die Blende kaum sichtbar nachgeregelt.
- **EVI-Gamma**: Wie die TV-Simulation, basiert die **EVI**-Gamma-Version auf der Gamma-Kurve und ist somit geeignet auf **SDR**-Displays wiedergegeben zu werden. Allerdings kam das in den vorangegangenen Kapiteln entwickelte Verfahren zum **TMO**-Tuning zum Einsatz.
- **HLG-Simulation**: Die dritte Version repräsentiert den live **HDR**-Broadcast-Betrieb unter Verwendung von **HLG**. Im Vergleich zur TV-Simulation bleiben über 12 Blendenstufen erhalten. In Szenen mit starken Schwankungen in der Belichtung, beispielsweise ausgelöst durch einen Kameraschwenk, wird die Blende kaum sichtbar nachgeregelt.
- **EVI-HLG**: Obwohl **HLG** entwickelt wurde, um eine Abwärtskompatibilität mit dem **SDR**-Fernsehen zu gewährleisten, ist der optische Eindruck des Bildes nicht ideal. Verglichen mit dem klassischen **SDR**-Bild, ist hier der lokale Kontrast deutlich schwächer ausgeprägt. Deshalb wird das **EVI**-Verfahren mit der **HLG**-Kurve kombiniert. Damit wird der lokale Kontrast verbessert und es kommt zu einer leichten Homogenisierung (Angleichung von Licht und Schatten). Als Konsequenz weist das Bild eine höhere Kompatibilität bei der Darstellung auf einem **SDR**-Fernsehgerät auf, zeigt aber gleichzeitig eine bessere Bildqualität bei der **HDR**-Darstellung. Verglichen zur reinen **HLG**-Darstellung besitzt diese Version mehr spatiale Informationen, die zu einer steigenden Datenrate führen könnten.

Ein Beispiel für einen visuellen Vergleich dieser vier Versionen findet sich in Abbildung 7.1. Mithilfe der HM 16.3 Software [127] wird das High Efficiency Video Coding (**HEVC**) durchgeführt. Für eine möglichst realitätsnahe Aussage wurden für die beiden Gamma-Versionen 8 Bit und für die beiden **HLG**-Versionen 10 Bit genutzt. In einem ersten Test wurde eine feste Datenrate verwendet und der **PSNR**-Wert gemessen. Da es sich bei allen Szenen um Aufnahmen mit sehr viel Detailreichtum und bei den Aufnahmen 1 bis 4 zusätzlich noch um 50p-Sequenzen handelt, wurde die vergleichsweise hohe Datenrate

von 10 Mbit/s gewählt. Wie bei Litwic et.al. [78] wurden anschließend die codierten Sequenzen nur mit ihrer jeweiligen Referenz verglichen.

7.1.3 Auswertung (PSNR)

In Abbildung 7.2 sind die Unterschiede zwischen den vier Versionen sowie den verschiedenen Testsequenzen dargestellt. Es wird deutlich, dass die EVI-Versionen einen niedrigeren PSNR-Wert aufweisen, als die TV- bzw. die HLG-Simulation. Bei den beiden Gamma-Versionen beträgt der Unterschied durchschnittlich 2,5 dB. Die größte Abweichung liegt bei 4,6 dB und die geringste bei 0,9 dB. Aber nur wenige EVI-Versionen fallen dabei unter die wichtige 42dB-Marke, die im Allgemeinen als Schwelle für sichtbare Artefakte bei derartigen subjektiven Tests angenommen wird. Bei HLG ist die Differenz im Durchschnitt 1,5 dB größer. Erstaunlicherweise ist der PSNR-Wert für HDR höher als für SDR. Dies ist jedoch erklärbar, da die effektivere Verteilung der Bits bei der Codierung den höheren Informationsgehalt bei der 10-Bit-Codierung für HDR gegenüber der 8-Bit-Codierung für SDR wieder kompensiert.²

Aus den Ergebnissen kann allerdings nicht abgeleitet werden, dass die TV-Simulation eine bessere Gesamtbildqualität besitzt, da die codierten Sequenzen jeweils nur mit den zugehörigen Referenzen verglichen wurden. Daher kann zunächst nur festgehalten werden, dass die EVI-Versionen aufgrund ihres höheren lokalen Kontrastes und ihrer besseren Szenenkontrastreproduktion schwieriger zu komprimieren sind. Daraus ergibt sich die Frage, ob die positiven Effekte des Tonemappings auch ausreichend sind, wenn die Datenrate nicht entsprechend erhöht wird. Bei den Messungen des Tests wäre eine Steigerung um ca. 40% nötig, um die oben erwähnten 2,5 dB zu kompensieren. Eine ähnliche Fragestellung gab es auch bei den Untersuchungen zu HFR, bei der die höhere Bildwechselfrequenz für eine flüssigere Bewerbungsdarstellung und vor allem für schärfere Einzelbilder durch die kürzere Belichtung und die damit verbundene geringere Bewegungsunschärfe sorgt. In Probandentests konnte gezeigt werden, dass der reduzierte PSNR subjektiv nicht signifikant messbar war. [71] [44] Die gravierende Änderung (Verbesserung) des Bildeindrucks bei HFR hatten die leicht stärkeren Codierfehler derart überdeckt, dass sie vom Zuschauer nicht wahrgenommen wurden. Ein entsprechender Test existierte bislang für HDR nicht und wurde daher im Rahmen dieser Arbeit durchgeführt.

Daher wurde in einem zweiten Schritt der Einfluss des Tonemappings auf die Videocodierung tiefergehend analysiert. Dazu wurde die Gesamtdatenrate, die weiterhin 10 Mbit/s

²Die Quantisierungsstufen sind besser an die menschliche Wahrnehmung angepasst (s. 2.2.5)

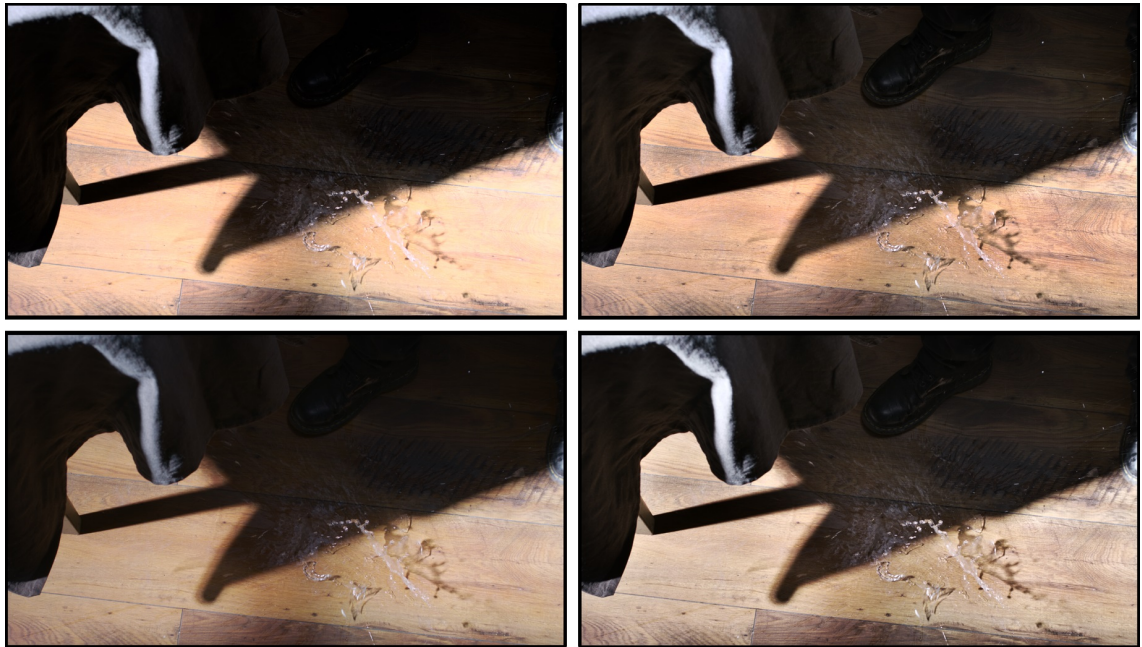


ABBILDUNG 7.1: von oben links nach unten rechts: TV-Simulation, EVI-Gamma, HLG-Simulation, EVI-HLG (bei Sequenz *bistro 3*)

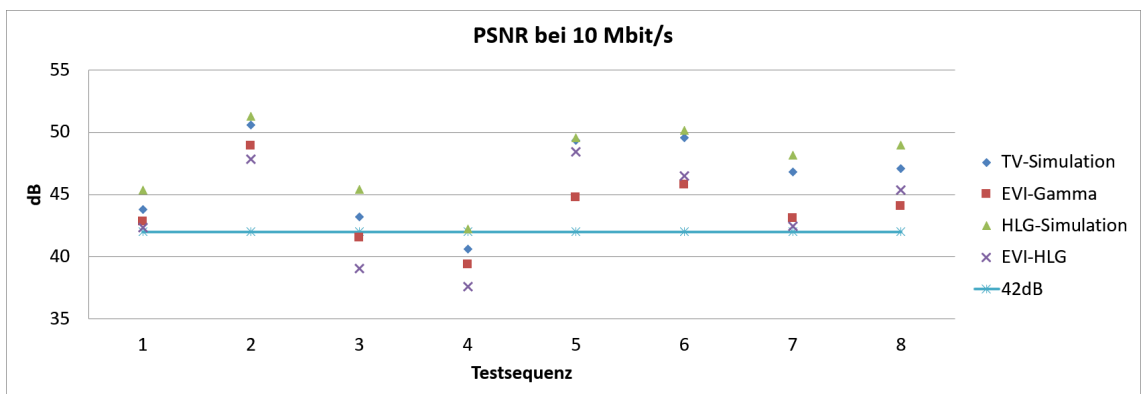


ABBILDUNG 7.2: Auswirkung des Tonemappings auf den PSNR-Wert bei konstanter Datenrate

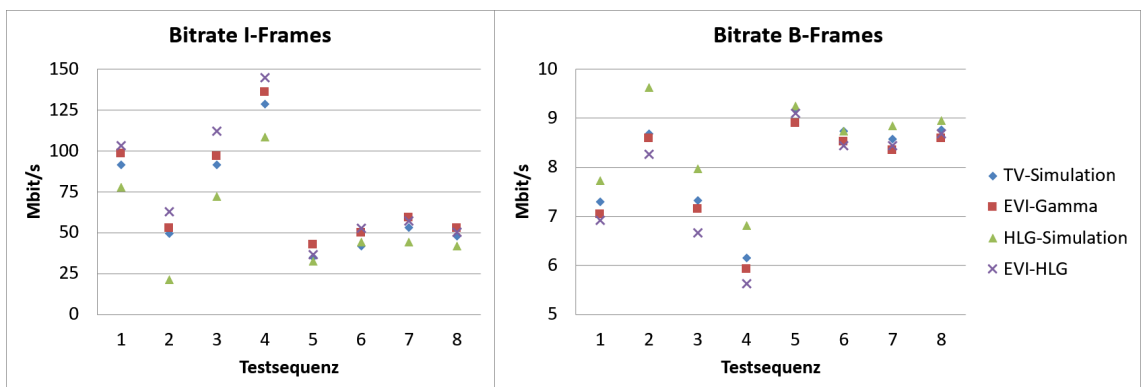


ABBILDUNG 7.3: Das Verhältnis der I-Frame- zur B-Frame-Datenrate ist beim Tone-mapping in Richtung I-Frames verschoben.

betrug, auf die I- und B-Frames aufgeteilt.³ Die Resultate finden sich in Abbildung 7.3. Für alle Sequenzen kann beobachtet werden, dass die Datenrate bei den I-Frames beim EVI-Verfahren stark ansteigt, während sie bei den B-Frames leicht sinkt. Dies stimmt mit der Hypothese überein, dass Einzelbilder nach der Kontrastkompression mit EVI schwieriger zu komprimieren sind, die Korrelation in zeitlicher Richtung zwischen ihnen aber zunimmt und damit die Bewegungskompensation einfacher wird. Der erste Effekt hat allerdings einen größeren Einfluss auf die Gesamtdatenrate.

Im dritten Schritt wurde untersucht, ob es Korrelationen zwischen der Änderung der Datenrate und den die Szene beschreibenden Parametern gibt. Entsprechend wurden die Szenen hinsichtlich ihres Kontrastumfanges, ihrer Bewegung und ihrer durchschnittlichen Leuchtdichte quantifiziert. Die Werte finden sich in Tabelle 7.1

Für den Kontrastumfang wurde das Analyseergebnis von EVI genutzt, da dort eine Bestimmung des höchsten und niedrigsten Leuchtdichtewertes, der für den Zuschauer noch Relevanz hat, durchgeführt wird. Würde hingegen einfach der höchste und niedrigste Leuchtdichtewerte verwendet, würde das Ergebnis durch Rauschen und Ausreißer verzerrt werden, wie in 50 gezeigt wurde. Der Kontrastumfang wurde für jedes Frame einzeln bestimmt und anschließend gemittelt. Für die Bewegung wurden vier Faktoren - nämlich Schwenkgeschwindigkeit, Schwenkdauer, Objektgeschwindigkeit und Objektgröße definiert, die Werte von 0 (keine Bewegung) bis 10 (sehr starke Bewegung) annehmen konnten. Für die Helligkeit wurde die durchschnittliche Leuchtdichte der Szene herangezogen, indem der logarithmische Mittelwert über alle Pixel gebildet wurde.

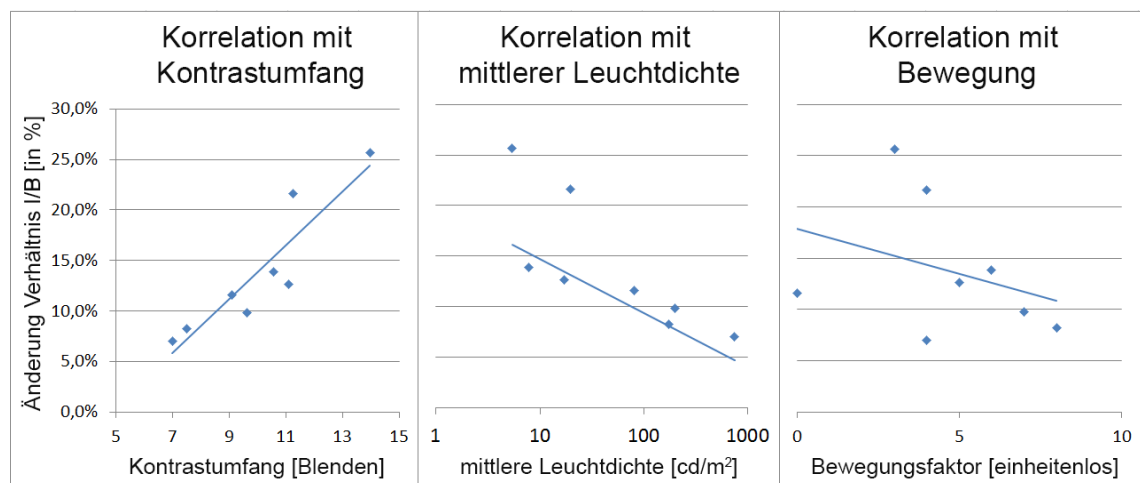


ABBILDUNG 7.4: Korrelation der Änderung des Verhältnisses der Datenrate von I- zu B-Frames mit den die Szene beschreibenden Parametern

³Da die Datenrate immer in Mbit/s angegeben wird, liegt die Datenrate der I-Frames deutlich über den 10 Mbit/s. Es gilt zu beachten, dass ein 10 s dauerndes Video aber nur zu etwa 0,3 Sekunden aus I-Frames besteht.

In Abbildung 7.4 repräsentiert jeder Punkt eine Sequenz. Auf der Ordinatenachse ist die Änderung des Verhältnisses in der Datenrate von I- zu B-Frames in Prozent aufgetragen. Die Skalierung der Ordinatenachse erfolgt anhand der die Szene beschreibenden Parametern. Im Falle einer idealen Korrelation müssten sich die Punkte auf einer diagonalen Linie befinden. Der Korrelationskoeffizient für den Kontrastumfang lag bei 0,91, der für die durchschnittliche Leuchtdichte bei -0,59 und der für den Bewegungsfaktor bei -0,33. Demzufolge ist der Kontrastumfang der Szene der wichtigste Faktor für die gesteigerte Komplexität der EVI-I-Frames und auch für den verringerten PSNR. Dies zeigt aber auch, dass mit Tonemapping-Verfahren ein sehr großer Szenenkontrastumfang gut auf einem klassischen SDR-Display wiedergegeben werden kann und die Diskrepanz zwischen klassischem SDR und EVI-SDR mit zunehmendem Szenenkontrastumfang steigt.

7.2 Auswirkungen auf den MOS

Im vorherigen Abschnitt konnte anhand der Datenrate bzw. des PSNR der gesteigerte Informationsgehalt beim EVI-SDR bewiesen werden. Allerdings stellt sich nun die Frage, ob der damit verbundene Qualitätsgewinn durch Codierartefakte egalisiert werden könnte. Deshalb wurde ein Probandentest nach der DSCQS-Methode [56] durchgeführt, welcher durch die Bachelorarbeit von Yannick Hoff [46] begleitet wurde.

7.2.1 Versuchsaufbau und -durchführung (MOS)

Beim DSCQS [56] handelt es sich um eine zyklische Methode, um die Qualität eines Systems relativ zu einer Referenz zu messen. Jeder Durchlauf besteht aus vier inhaltlich gleichen Clips, wovon jeweils zwei das mit EVI behandelte Bild und zwei das klassische SDR-Bild zeigen. Die Reihenfolge ist dabei zufällig. Der zeitliche Aufbau eines Durchlaufs ist in Abbildung 7.5 dargestellt. Zunächst wird Variante A für ca. 10 s gezeigt. Anschließend folgt für 3 s ein Graubild bevor Variante B gezeigt wird. Nach einem weiteren Graubild wiederholt sich der Ablauf ein weiteres Mal. Abschließend wird zur Bewertung aufgefordert. Dazu stehen den Probanden pro Durchlauf zwei Bewertungsskalen zur Verfügung - je eine für die beiden Varianten. Die Bewertungsskala ist in fünf Abschnitte nach dem MOS, namentlich *bad*, *poor*, *fair*, *good* und *excellent*, unterteilt. Die Teilnehmer können aber trotzdem eine stufenlose Bewertung zwischen den Abschnitten vornehmen. Der Bewertungsbogen ist ebenfalls in Abbildung 7.5 dargestellt.

Auch wenn die Stufen namentliche Entsprechungen besitzen, stehen die absoluten Zahlen eher im Hintergrund, da jeder Proband eine persönliche Vorstellung davon hat, welche Bildqualität er als *gut* bezeichnen würde. Viel wichtiger ist die Differenz zwischen den

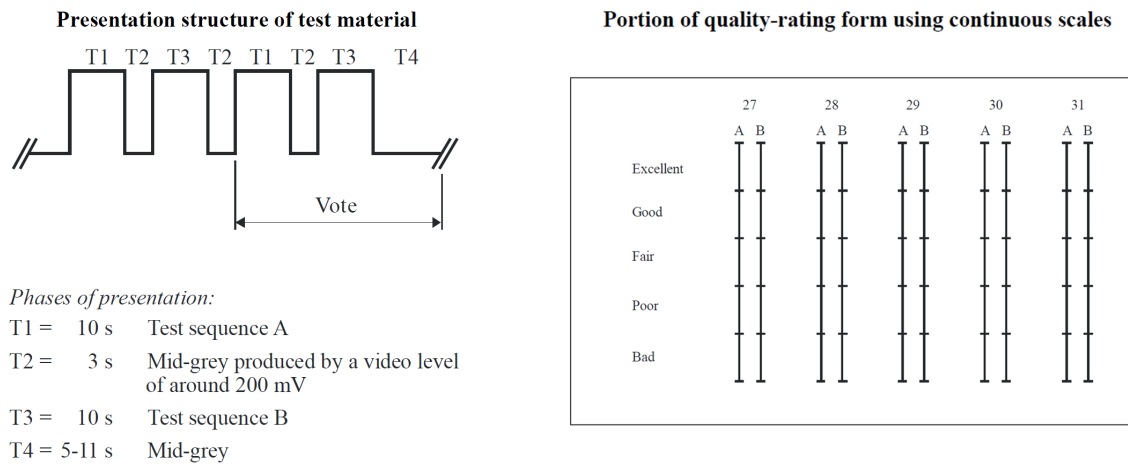


ABBILDUNG 7.5: Ablauf eines Durchlaufs beim DSCQS und Bewertungsbogen [56]

beiden gezeigten Varianten. Diese sollte folglich für die Interpretation der Ergebnisse herangezogen werden.

Insgesamt wurden sieben Testsequenzen⁴ mit unterschiedlichen Kontrastumfängen ausgewählt, da sich gezeigt hat, dass die Steigerung des Informationsgehaltes maßgeblich daran gekoppelt ist. Im Vergleich zur PSNR-Messung wurde die Datenrate auf ca. 3,5 MBit/s beschränkt, da die 10 MBit/s für sehr hohe dB-Werte gesorgt hatten und somit visuell wahrnehmbare Codierartefakte nahezu ausgeschlossen sind.

Die SDR-Varianten wurde aufwendig mit DaVinci Resolve erstellt, indem mit Hilfe von Key-Frames eine bestmögliche Blendensteuerung bei sich veränderten Belichtungsbedingungen simuliert wurde. Es handelte sich somit um ein qualitatives Maximum, dass im heutigen Live-Broadcast nur schwer zu erreichen ist. Auch die EVI-Varianten wurden mit den zur Verfügung stehenden Benutzerparametern, die noch in Abschnitt 7.11 erklärt werden, bestmöglich feinjustiert. Abbildung 7.6 zeigt jeweils ein Einzelbild aus allen 14 Clips.

Der Test wurde unter kontrollierten Laborbedingungen in einem abgedunkelten Raum (17 lx) durchgeführt. Beim verwendeten Display handelte es sich um einen Klasse 1 Referenzmonitor (Sony OLED) nach TECH 3320 [32]. Die maximale Leuchtdichte betrug 100 cd/m². Die Distanz zum Display entsprach dem Normabstand vom dreifachen der Displayhöhe.

Insgesamt nahmen 36 Probanden am Test teil. Von diesen bezeichneten sich 9 selbst als Expert Viewer. Der Anteil der männlichen Probanden betrug 58,3%, der der weiblichen Probanden betrug 41,7%. Der Altersdurchschnitt lag bei 30,3 Jahren.

⁴bistro 02, cars fullshot, fishing longshot, Stadion Schwenk, carousell fireworks 09, Stadion Torwart, Anstoß



ABBILDUNG 7.6: Vergleich der klassischen SDR-Variante (linke Seite) mit EVI (rechte Seite)

7.2.2 Auswertung (MOS)

Die Ergebnisse sind in Abbildung 7.7 dargestellt und in die einzelnen Testsequenzen unterteilt. Die blauen Balken repräsentieren die Bewertungen für die **EVI**-Varianten; die orangen Balken die für die klassischen **SDR**-Varianten. Die eingezeichneten 95%-Konfidenzintervalle basieren auf einer angenommenen Normalverteilung, die bei 36 Probanden bereits stark angenähert ist.⁵

Bei allen sieben Testsequenzen erhält **EVI** eine bessere, über dem angestrebten Signifikanzniveau liegende, Bewertung. Im Durchschnitt beträgt der Unterschied ca. 1,6 **MOS**-Punkte. Dieser variiert aber zwischen den Sequenzen. Besonders deutlich fällt das Ergebnis für *fishing longshot*, *cars fullshot* und *Anstoß* aus. Bei *Stadion Schwenk* und *fireworks* liegen beide Varianten enger beisammen. Hier kann mit einem guten Blendenwert beim klassischen **SDR** entweder eine gute Repräsentation der gesamten Szene erreicht werden oder zumindest für den Bereich des Hauptinteresses. Solange letztgenannter korrekt beleuchtet wird, werden Unterschiede, wie sich durch qualitative Befragungen im Nachgang herausstellt, die sich in weniger wichtigen Bildregionen äußern würden, oft gar nicht wahrgenommen.

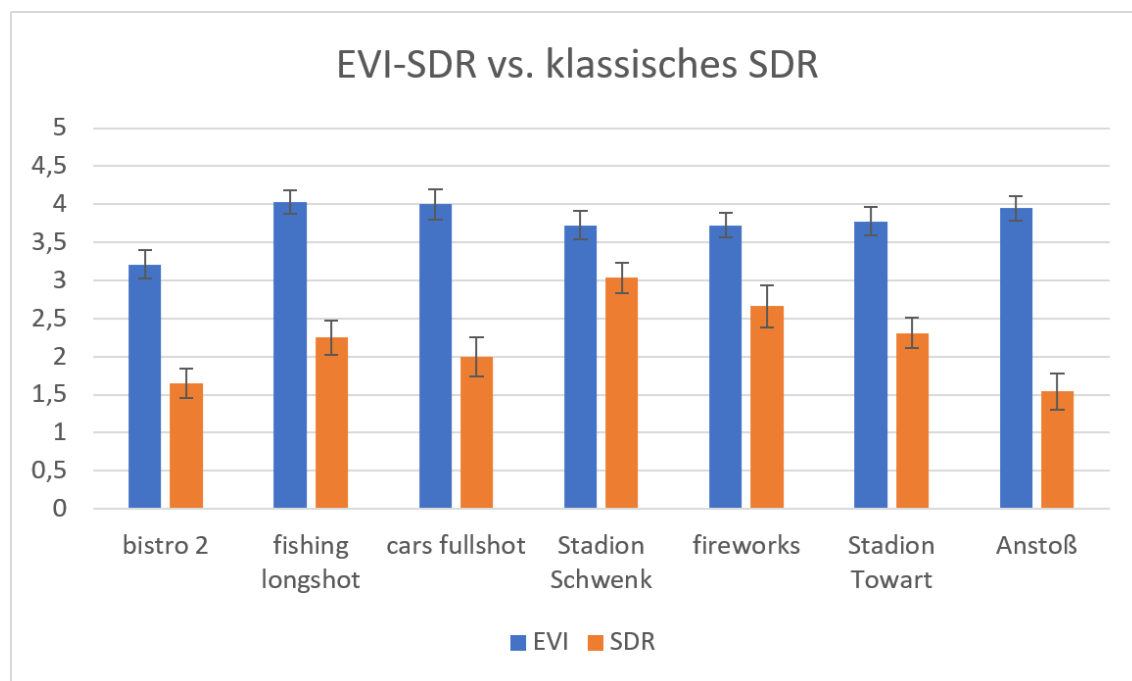


ABBILDUNG 7.7: Vergleich von EVI zu klassischem SDR anhand des MOS

Insgesamt zeichnet sich **EVI** durch eine sehr konstante Wertung über alle Sequenzen hinweg aus und zeigt somit eine gute Robustheit für den Broadcast-Alltag. In sechs von

⁵Das eine ausreichend große Übereinstimmung besteht, wurde mit Hilfe des Kolmogorov-Smirnov-Tests nachgewiesen. Wird α gleich 0,05 angenommen, darf bei 36 Messwerten die Abweichung nicht größer als 0,23 ausfallen. Die Bedingung konnte für alle Vergleiche im vorliegenden Test erfüllt werden. In den meisten Fällen wurde sogar ein Wert unter 0,1 gemessen.

sieben Fällen liegen die **MOS**-Punkte in einem Korridor zwischen 3,7 und 4. Gleichzeitig fallen auch die Konfidenzintervalle geringer aus als bei der **SDR**-Variante. Je nach Vorlage funktioniert die klassische Blendensteuerung besser oder schlechter, was zu Bewertungen zwischen 1,5 und 3 Punkten führt.

Nur bei der Szene *bistro 2* scheint das sektionale Tonemapping an seine Grenzen zu stoßen. Zwar ist auch hier ein Unterschied von ca. 1,5 **MOS**-Punkten gegeben. Absolut werden aber nur 3,2 erreicht. Die Problematik besteht in dieser Vorlage darin, dass sich der unbeleuchtete Kopf des Gastes direkt vor dem sehr hellen Fenster befindet. Der bilaterale Filter ist zwar in der Lage, beide Areale zu trennen, trotzdem fällt, aufgrund der spatialen Komponente, die Aufhellung des Kopfs subjektiv zu gering aus. Letztendlich zeigt das Beispiel, welcher Kompromiss eingegangen werden muss, wenn kein klassisches lokales Tonemapping gewählt wird. Potenzielle Optimierungen werden im Ausblick kurz diskutiert.

Sehr vereinzelt kam es bei der **SDR**-Variante zu einer identischen oder besseren Bewertung als bei **EVI**. Diese Probanden wurden im Nachgang des Experiments zu ihren Gründen befragt. In den meisten Fällen waren jenen keine Unterschiede zwischen den beiden Varianten aufgefallen, da, wie bereits oben angesprochen, der Kontrastgewinn sich in weniger relevanten Bildregionen äußerte und letztendlich nicht bemerkt wurde. In einigen wenigen Fällen wurde der Bildeindruck von **EVI** als zu surreal empfunden. Diese Probanden hätten sich eine leicht geringere Homogenisierung gewünscht. Selten führte die durch **EVI** hervorgerufene Verstärkung dazu, dass Rauschen bemerkt und die Limitation der Kamera sichtbar wurde.

Obwohl beide Varianten in diesem Test mit der gleichen Datenrate encodiert wurden, konnte kein negativer Effekt aufgrund der stärkeren Kompression nachgewiesen werden. Explizit ebenfalls im Nachhinein gefragt, gab kein Proband an, derartige Artefakte oder Effekte wahrgenommen zu haben. Dieser Sachverhalt schien durch den gänzlich anderen Kontrasteindruck komplett überlagert worden zu sein. Zugleich spricht es für die Kompressionseigenschaften von **HEVC**. Auch dem Autor fiel es schwer, bei gezielter Beobachtung Codierartefakte auszumachen.

7.3 Workflow mit Gamma- und / oder HLG-Distribution

In den beiden vorangegangenen Abschnitten konnte gezeigt werden, dass **EVI** in der Lage ist, einen sehr großen Szenenkontrastumfang zu erhalten und trotzdem ein kontrastreiches, ausreichend helles und somit der Zuschauerpräferenz entsprechendes Bild

zu erzeugen. Nachfolgend soll anhand von drei beispielhaften Workflows aufgezeigt werden, wie **EVI** in einer realen Produktionsumgebung eingesetzt werden könnte, um die schrittweise Einführung, Überführung und Etablierung von **HDR** zu ermöglichen.

Option 1: Bei der ersten Option geht es nicht darum, **HDR** zu produzieren, sondern das **SDR**-Bild zu verbessern und den Produktionsprozess höher zu automatisieren. Diese Option ist für viele Broadcaster aktuell das vorrangige Ziel. Auch die in Abschnitt 7.5 durchgeführten Praxistests folgen meist dem in Abbildung 7.8 dargestellten Aufbau. Dabei wird, im Vergleich zum heutigen Workflow, lediglich die Kamera gegen ein **HDR**-fähiges Äquivalent ausgetauscht, oder die vorhandene Kamera entsprechend nachgerüstet. Hinter der Kamera wird die **HDR**-Downkonvertierung durchgeführt, so dass in der folgenden Verarbeitungskette (Mischer, Displays, etc.) weiterhin mit Equipment, welches lediglich ITU-R BT.709 [54] korrekt interpretiert, gearbeitet werden kann.

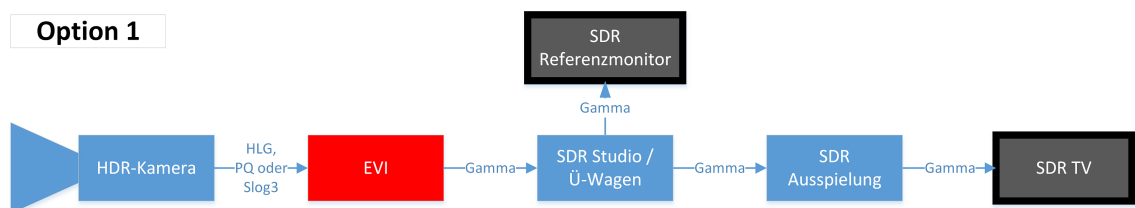


ABBILDUNG 7.8: EVI-Workflow bei reiner SDR-Ausspielung

Dieser Workflow stellt im Vergleich zu einer kompletten Umrüstung aller Komponenten eine finanziell sehr geringe Einstiegshürde dar. Trotzdem ermöglicht er eine hohe Partizipation an den Vorteilen von **HDR**. Zum einen wird es sofort möglich, einen höheren Kontrastumfang und somit ein besseres Bild (s. Abschnitt 7.2) jedem Zuschauer zu senden, zum anderen, und dies ist für viele Fernsehanstalten mindestens von gleichwertiger Relevanz, kann der Produktionsaufwand gegenüber dem Status-Quo gesenkt werden. Dies ist aus zweierlei Gründen möglich. Zum einen führt der gesteigerte Blendenumfang der Aufnahme zu einem größeren Arbeitsbereich. Dies bedeutet, dass die Blende muss weniger nachgeregelt werden. Die eigentliche Ausbelichtung wird stärker automatisiert. Zum anderen kann der Einsatz von Aufhelllicht reduziert werden. Aufhelllicht dient bisher dazu, inhomogene Belichtungsbedingungen auszugleichen, damit der mit **SDR** mögliche, darstellbare Szenenkontrastumfang von etwa 8,5 EVs nicht überschritten wird. Dabei gilt es beispielsweise einen zu gering beleuchteten Protagonisten, der sich vor einem sehr hellen Hintergrund befindet, künstlich aufzuhellen. Letztendlich entspricht dies einer Homogenisierung der Lichtsituation in der Szene. Durch den gesteigerten Kontrastumfang der Sensoren wäre es in diesem Workflow nun möglich, solche Homogenisierungen auch im Digitalen durchzuführen.

Im Hinblick auf zukünftige **HDR**-Produktionen könnte mit einem solchen kontrastkomprimierten **SDR**, unter Zuhilfenahme einer **HDR**-Upkonvertierung, der original **HDR**-Eindruck wieder deutlich stärker angenähert werden, als es mit dem durch Clipping beeinträchtigten **SDR**-Bild denkbar wäre. Somit ermöglicht die erste Option bereits näherungsweise **HDR**-Inhalte in einem **SDR**-Archiv abzulegen.

Option 2: Die zweite Option stellt die Verwendung einer **HDR**-Downkonvertierung im Anschluss an eine **HDR**-Produktion dar (Abbildung 7.9). Im Vergleich zum ersten Workflow sind hier deutlich größere Investitionen nötig, da die **HDR**-Kette bis zur Ausspielung geführt werden muss. Alle eingesetzten Geräte erfüllen die Vorgaben nach [58] bzw. [32]. Die Aussteuerung der Kamera erfolgt in **HDR**. Das **HDR**-Signal kann, sofern ein entsprechender Übertragungskanal zur Verfügung steht, distribuiert werden. Das **SDR**-Signal wird automatisch aus dem **HDR**-Signal abgeleitet und muss in einem separaten Übertragungskanal versendet werden.

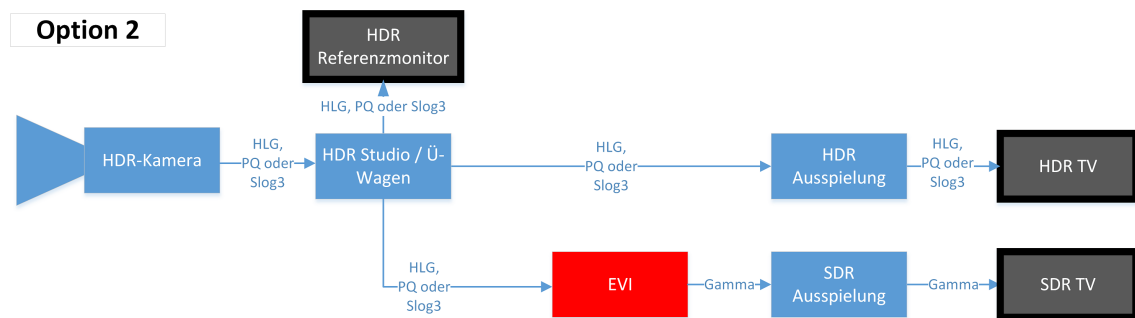


ABBILDUNG 7.9: EVI-Workflow bei Konvertierung eines HDR-Signals am Ende der Kette

Gerade in der Einführungsphase von **HDR**, in welcher der Empfängerkreis für das **HDR**-Signal limitiert ist, oder wo dieses erst gar nicht ausgestrahlt werden kann, ergibt sich die Situation, dass ein Bild kontrolliert und abgenommen wird, welches letztendlich kaum oder keine Verwendung findet. Das nach **SDR** konvertierte Signal, das den Großteil der Zuschauer erreicht, wird hingegen wenig beachtet. Darüber hinaus hat sich im Laufe der Arbeit gezeigt (s. Kapitel 4), dass auch die lineare Repräsentation von **HDR**-Inhalten auf **HDR**-Displays nicht der Wahrnehmung und Präferenz des Zuschauers entspricht.

Option 3: Deshalb stellt die dritte Option (Abbildung 7.10) die Kontrastkompression wieder an den Anfang der Kette, gleich hinter die Kamera. Im Gegensatz zu Option 1, wird allerdings eine **HDR**-Crosskonvertierung durchgeführt, sprich der **HDR**-Eindruck leicht homogenisiert. Die Extreme in der Belichtung werden näher zur Mitte verschoben und der lokale Kontrast wird verstärkt. Diese Tatsache macht zwei verschiedene Ausspielformen denkbar. Zum einen wäre es weiterhin möglich **HDR** und **SDR** über zwei unterschiedliche Kanäle auszuspielen (s. Option 3A in Abbildung 7.10). Dazu müsste der **SDR**-Weg noch eine **HDR**-Downkonvertierung durchlaufen. Da das Bild aber bereits

homogenisiert wurde, sollte in der Folge eine statische Konvertierungsstufe ausreichend sein. Die feste Wirkweise bedeutet eine bessere Prädiktierbarkeit des Ergebnisses. Zum anderen wäre aber auch die Ausspielung mit nur einem Kanal denkbar (s. Option 3B in Abbildung 7.10). Dazu muss als **OETF** zwingend **HLG** gewählt werden. Wie in Abschnitt 2.5.4 dargelegt, ist die **HLG**-Kurve so konzipiert, dass sie sowohl mit einer **EOTF** nach ITU-R BT.1884 [55] als auch nach ITU-R BT.2100 [58] entzerrt werden kann.⁶ Allerdings hatte sich gezeigt, dass die Betrachtung in **SDR** einen zu flauen Bildeindruck erzeugt. Durch die höhere lokale Modulation aufgrund der Homogenisierung, kann diesem Eindruck jedoch entgegengewirkt werden. Der größte Nachteil bei der Verwendung von **HLG** wird dadurch maßgeblich abgemildert. Trotzdem muss beachtet werden, dass die hybride Lösung immer einen Kompromiss darstellt. Das Ergebnis wird sich leicht unterscheiden, je nachdem, ob das Bild auf einem **SDR**- oder **HDR**-Referenzmonitor beurteilt wird. Das bedeutet aber auch, dass in diesem Konstrukt in der Übergangszeit auf die Anschaffung von **HDR**-Displays verzichtet werden kann, wenn es das Ziel ist, ein bestmögliches **SDR**-Bild zu erzeugen.

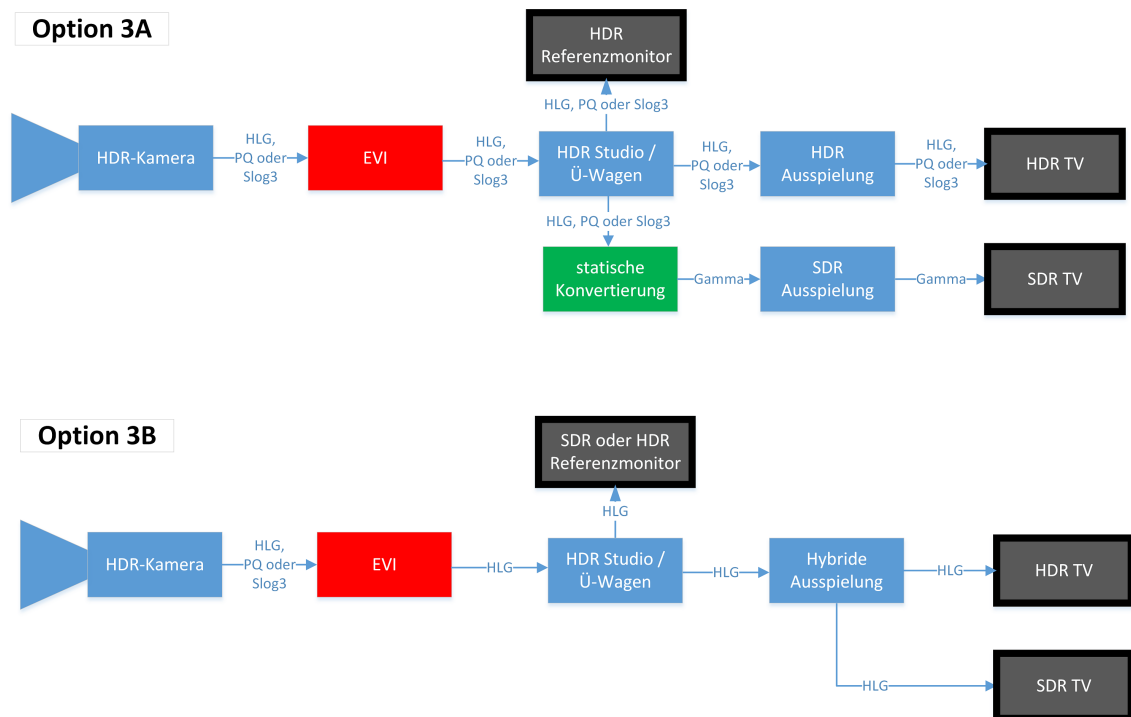


ABBILDUNG 7.10: EVI-Workflow mit Homogenisierung von HDR und SDR

⁶Dies setzt natürlich ein identisches Farbgamut voraus.

7.4 GPU-basierter Echtzeitprototyp mit GUI-Steuerung

Für eine praktische Überprüfung der im vorangegangenen Abschnitt dargestellten Workflows bedarf es einer echtzeitfähigen Implementierungsform der in Kapitel 6 erarbeiteten Algorithmen. Die Hardwarebasis des Prototypen bildet die 1HE hohe Dell Precision 3930 Rack-Workstation, die um eine dezidierte Grafikkarte und eine Serial Digital Interface (SDI)-I/O-Karte erweitert wurde. In der HD-Konfiguration besitzt sie zwei Nvidia Quadro P4000 für die Bildverarbeitung und eine Blackmagic DeckLink Duo 2 für die SDI Ein- und Ausgänge. Diese Einheit ist in der Lage, zwei 3G-SDI-Datenströme parallel zu verarbeiten. Die UHDTV-Variante besitzt eine Nvidia P5000 und eine Blackmagic DeckLink 8k.

Die EVI-Software ist eine Kombination aus C++-Programmierung, die das Grundgerüst bereitstellt und den eigentlichen Bildoperationen als Shader in OpenGL. Sie wird als Windows-Dienst installiert. Die Steuerung erfolgt über ein separates Graphical User Interface (GUI). Dieses kann von einem beliebigen Computer im Netzwerk aufgerufen werden. Zum einen lassen sich über das GUI die benötigten Informationen über das Signal am Ein- und Ausgang (OETF, Gamut, Range, spatiale Auflösung und Bildwechselfrequenz) einstellen, zum anderen bietet es die Möglichkeit, das Bild nach den eigenen Präferenzen zu modifizieren und einen gewissen Look zu kreieren. Insgesamt stehen dafür derzeit sieben Regler zur Verfügung. So ist es beispielsweise möglich, die Steilheit mithilfe von *Kontrast* oder der *Clipping Intensität* den eigenen Vorstellungen anzupassen. Auch die Stärke der Homogenisierung und der Anhebung der dunklen Bildbereiche können justiert werden. Darüber hinaus kann der Nutzer anhand von zwei Reglern auf die Zielhelligkeit Einfluss nehmen und, falls gewünscht, die Sättigung verstellen.

Vorgenommene Einstellungen können in Form von Presets gespeichert und auch wieder geladen werden. Eine Veränderung der Einstellungen während der Produktion ist nur im geringen Maße angedacht. Die Werte sollen im Vorlauf auf die gegebenen Bedingungen und Vorstellungen angepasst werden. Abbildung 7.11 zeigt zwei Screenshots der Oberfläche.

Die Verarbeitungszeit pro Bild hängt von der spatialen Auflösung ab und ist vom Bildinhalt nahezu unabhängig. Sie beträgt bei 1080p ca. 6 ms und bei 2160p ca. 14,5 ms.⁷ Aus dieser Perspektive ist somit die Echtzeitfähigkeit gegeben. Bei 1080p könnten sogar zwar Videoströme auf einer Hardware gleichzeitig prozessiert werden. Ermöglicht wird die schnelle Verarbeitung der hohen Auflösungen durch die Parallelisierbarkeit von

⁷gemessen mit der Nvidia Quadro P5000

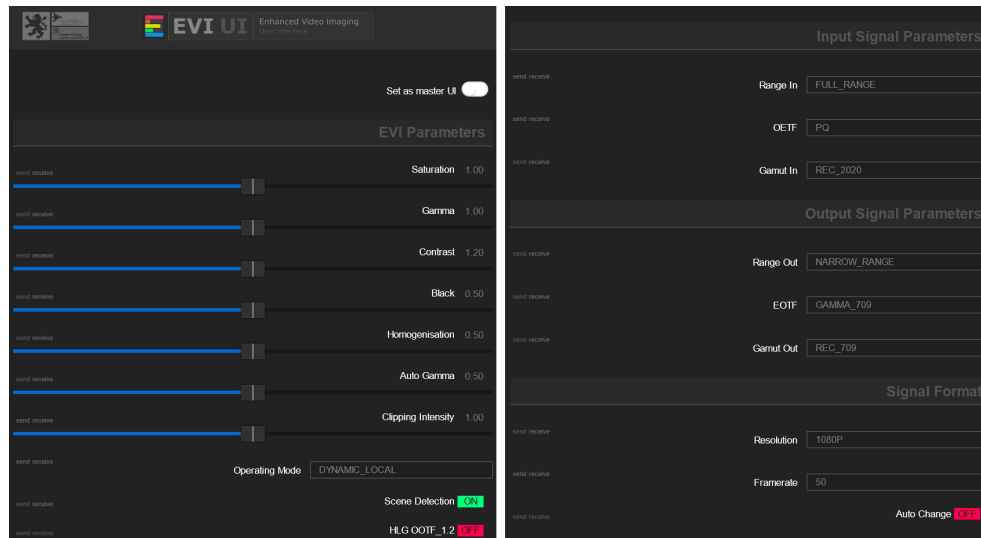


ABBILDUNG 7.11: Das EVI-GUI bietet anhand von sieben Reglern die Möglichkeit, einen gewünschten Look zu erzeugen. Darüber hinaus können die Signalparameter für den Ein- und Ausgang eingestellt werden.

Pixeloperationen. Allerdings benötigt der reine Transport des Videomaterials im Rechner trotz Direct Memory Access (DMA)⁸ mehr Zeit, als die eigentliche Rechenzeit. Für 1080p und 2160p ergeben sich Werte⁹ von ca. 4 ms bzw. ca. 10 ms. Die Herausforderung sind die großen Datenmengen bei unkomprimiertem Videomaterial. Bei 2160p mit 10 Bit Quantisierung, 4:2:2-Farbunterabtastung und 50 Bildern pro Sekunde ergibt sich eine Datenrate von 8,9 Gbit/s. Selbst die neueste Generation an SDI-Karten und DMA-fähigen Grafikkarten unterstützt weiterhin nur Peripheral Component Interconnect Express (PCIe) 3.0, sodass keine schnelleren Transportzeiten zu erreichen sind. Trotz dieser Einschränkung können aber, wie aufgezeigt, Bildwechselfrequenzen oberhalb von 60 Hz erreicht werden.

Entscheidend für den Einsatz im Live-Broadcast ist aber auch die Verarbeitungszeit (Delay). Im vorliegenden Fall beträgt es 2 Vollbilder¹⁰. Dies liegt daran, dass der Algorithmus seine Berechnungen erst beginnen kann, sobald das komplette Bild auf die Grafikkarte übertragen wurde. In der Folge beträgt das Delay ein Vollbild. Die Berechnung selbst dauert, wie aufgezeigt, nur wenige ms. Das fertig berechnete Bild kann aber erst wieder zum Beginn des nächsten Vollbildes ausgegeben werden, um weiterhin synchron zu sein. Dies bedeutet die Verzögerung um ein weiteres Vollbild. Somit ist es konzeptionell nicht möglich einen geringeren Delay zu erreichen. Ein Delay von 2 Vollbildern wird aber für Live-Broadcast-Anwendungen akzeptiert.

⁸Bei diesem Speicherzugriff wird die CPU umgangen. Die Daten werden in der Folge von der SDI-Karte über den Arbeitsspeicher direkt zur GPU kopiert.

⁹gemessen mit der Blackmagic Decklink 8k und DDR4-RAM

¹⁰also 40 ms bei 50 Hz und 33,3 ms bei 60 Hz

7.5 Praxistest und Anwendungsfälle

Sowohl der Bayerischer Rundfunk (**BR**) als auch der Westdeutsche Rundfunk (**WDR**) verfügen seit Dezember 2018 über einen solchen Prototypen. Seit Februar 2019 ist die Workstation vom **BR** in einem ihrer Übertragungswagen eingebaut. Die beiden öffentlich rechtlichen Rundfunkanstalten verfolgen damit Option 1. Sie möchten das **SDR**-Bild verbessern und gleichzeitig in der Produktion einen höheren Automatisierungsgrad erreichen.

Aktuell wird dazu eine separate dedizierte Kamera bei ausgewählten Produktionen in **HDR** betrieben. Dabei handelt es sich sowohl um Freiluftveranstaltungen (*Klassik Open-Air Nürnberg*) als auch um Studioproduktionen (*Schwaben weissblau hurra und helau*, *Wirtshausmusikanten* oder *Ladies Night*). Im Vordergrund steht dabei die Herausforderung, ein algorithmenbasiertes Verfahren in einen bis jetzt komplett manuellen Arbeitsablauf zu integrieren. Solch ein Umbruch besitzt sowohl technische als auch operationell organisatorische Komponenten. Generell ist es entscheidend sicherzustellen, dass die Steuerung von Kamera und **EVI** nicht gegeneinander arbeiten. Als nächster Schritt soll eine Mehrkameraproduktion durchgeführt werden, was den parallel Einsatz mehrerer **EVI**-Instanzen bedeutet.

Auch Option 2 wurde bereits getestet und wird im Vergleich zu Option 1 bereits heute im Fernsehen ausgestrahlt. Zur ersten Mottoshow der 16. Staffel von *Deutschland sucht den Superstar (DSDS)* wurde von Radio Television Luxembourg (**RTL**) eine **HDR/SDR**-Konvertierung auf Basis der **EVI**-Technologie zur Verfügung gestellt.¹¹ Die Sendung wurde komplett in **HDR** produziert und vor der Ausspielung nach **SDR** gewandelt. Bis April 2019 geschah dies mit einem Konverter der Firma AJA (*FS-HDR*). In der ersten Mottoshow wurde nun wie bisher die AJA-Version gesendet, hinter den Kulissen aber zusätzlich eine **EVI**-Version produziert. Der anschließende Vergleich dieser beiden Versionen war scheinbar so eindrucksvoll, dass man sich dazu entschloss, bereits zur nächsten Sendung das Live-Bild mit **EVI** zu produzieren. So erfolgte am 13. April 2019 die erste Live-Produktion auf **EVI**-Basis.

Positiv wurde von den Betatestern sowohl die Darstellung der hellen Bildbereiche, vorzugsweise des Himmels, als auch die bisher unbekannte Durchzeichnung der tiefen Schatten hervorgehoben. Gerade das unbeleuchtete Publikum komme so besser zur Geltung. Allerdings wurde hierdurch auch ein akkurateres Arbeiten beim Studioaufbau nötig,

¹¹Dabei handelte es sich um die greenmachine der Firma LYNX Technik AG, die das Verfahren lizenziert hat.

da Klebestreifen, Kabel und ähnliches nicht mehr mit dem schwarzen Hintergrund verschmelzen. Die nachfolgenden beiden beispielhaften Abbildungen (7.12 und 7.13) zeigen **EVI** im Vergleich mit anderen Konvertierungen bei den angesprochenen Produktionen.



ABBILDUNG 7.12: Die linken Aufnahmen stammen aus der ersten Mottoshow, während die rechten Bilder bei der zweiten Mottoshow aufgenommen wurden. Gerade die Sichtbarkeit bei dunklen Aufnahmen bzw. in dunklen Bildregionen hat sich deutlich verbessert. (Bildquelle: [107](#))

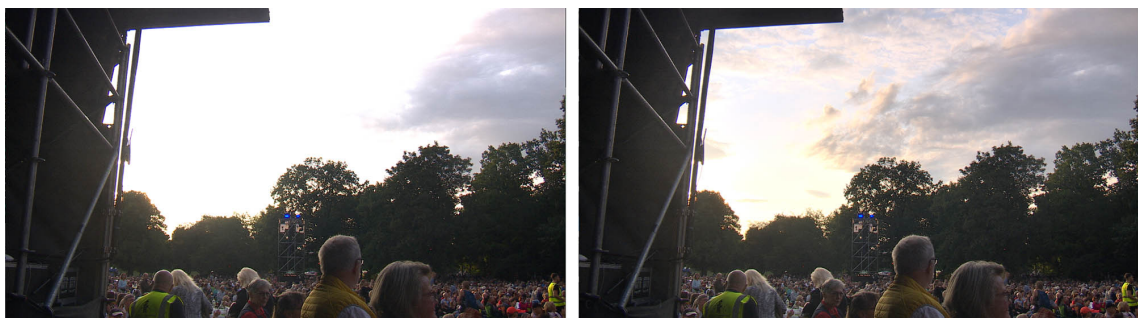


ABBILDUNG 7.13: Die Abbildung zeigt eine Openair-Produktion des **BR**. Links ist klassische **SDR**-Variante zu sehen, rechts die **EVI**-Konvertierung.

Aufgrund der Rückmeldungen der Betatester konnten Optimierungen am Algorithmus sowie an der Bedienung vorgenommen werden. Teilweise wurde die Wirkungsweise von Einstellungsreglern angepasst oder es wurden neue Regler hinzugefügt. Hervorzuheben ist hier die Möglichkeit, einen statischen Tonemapping-Anteil in das dynamische Ergebnis einzurechnen¹². Im Vergleich zum rein dynamischen Ergebnis, ist der erfassbare Kontrastumfang etwas beschränkter, und die Möglichkeiten zur automatischen Aussteuerung des Bildes fallen geringer aus. Dafür ist das Bild in den meisten Fällen

¹²auch als Hybrid-Modus bezeichnet

tendenziell heller und steiler. Außerdem ermöglicht der statische Anteil eine einfachere Nachjustierung per Blendenwert. Wird beispielsweise die Reproduktion des Gesichts des Moderators als zu dunkel empfunden, kann die Blende langsam weiter geöffnet werden, ohne dass das **EVI**-System zu stark nachregeln würde. Dieser Modus empfiehlt sich für Studioproduktionen und ähnliche Situationen mit überschaubarem Kontrastumfang bei gleichzeitig sehr definierten Belichtungswünschen. Gerade bei Produktionen wie *Deutschland sucht den Superstar*, deren Bildeindruck durch die Lichtdesigner bestimmt wird, kann so die kreative Intention gewahrt werden.

Wird hingegen auf einen statischen Anteil verzichtet und rein dynamisch gearbeitet, konnte bei Szenen mit äußerst geringem Kontrastumfang beobachtet werden, dass der Standardwert für die Homogenisierung die kaum vorhandenen Helligkeitsunterschiede im Bild noch weiter angleicht. Da solche Szenen nicht Bestandteil der Probandenversuche waren, wurden sie entsprechend bis zu diesem Zeitpunkt nur unzureichend berücksichtigt. Um den Effekt in Zukunft zu vermeiden, wurde der Algorithmus dahingehend erweitert, dass der Kontrastumfang der Szene auf Basis der Auswertungsergebnisse der Blöcke geschätzt wird. Falls nötig, kann so die Homogenisierung automatisch zurückgenommen werden.

Als essentiell für den Einsatz bei Option 2 hat sich die Szenenwechselerkennung gezeigt, die allerdings noch genauerer Betrachtung und Nachbesserung in der Zukunft bedarf. Über weite Strecken hinweg wurden Umschnitte korrekt detektiert, allerdings war das Ergebnis in Ausnahmesituationen wie sehr schnellen Schwenks oder abrupten Lichtwechsel durch Scheinwerfer, nicht mehr ausreichend zuverlässig. Derzeit wird eine vielversprechende Möglichkeit erprobt, die auf einem flexiblen, an die Bewegung im Bild gekoppelten, Schwellenwert basiert.

Insgesamt haben die Beispiele gezeigt, wie die Idee der automatischen Kontrastkompression in die Praxis überführt werden kann. Es ist naheliegend, dass es im Rahmen der Arbeit nicht möglich ist, alle Optionen unter verschiedenen Rahmenbedingungen zu testen. Das ideale Zusammenspiel von **EVI** mit anderen technischen Geräten und menschlichen Anwendern wird sich erst mit der Zeit ausbilden und letztendlich auch bei jedem Broadcaster andere Formen annehmen. Die sich hieraus ergebenden Lerneffekte werden in Zukunft in die Technologie einfließen müssen, um so einen Reifeprozess zu gestalten. Die bisher gewonnenen praktischen Erkenntnissen zeigen aber, dass die grundlegende Funktionsweise und der bis dato eingeschlagene Weg erfolversprechend zu sein scheinen.

7.6 Zwischenfazit Praxis

Im vorliegenden Kapitel konnte gezeigt werden, wie **EVI** sich auf die Bildqualität auswirkt, neue Workflows ermöglicht und den Übergang zur flächendeckenden Einführung von **HDR** vorantreiben könnte.

Mithilfe einer **PSNR**-Messung konnte nachgewiesen werden, dass die zusätzlichen Informationen, welche bei der Konvertierung mit **EVI** von **HDR** nach **SDR** im Vergleich zum klassischen **SDR** erhalten bleiben, die Komplexität der Bilder für die Codierung erhöhen. Je größer der Kontrastumfang der Szene, desto größer fällt der Unterschied zwischen den beiden **SDR**-Versionen aus. Der niedrigere **PSNR**-Wert bzw. die höhere Datenrate zeigen in diesem Zusammenhang allerdings nur, dass **EVI** in der Lage ist auch sehr große Kontrastumfänge in **SDR** mit hoher Modulation abzubilden. Da der niedrigere **PSNR** keine subjektiv negativen Auswirkungen hat, wurde in einem nachfolgenden Probandentest nachgewiesen. Trotz identischer Datenrate wurde die **EVI**-Version dem klassischen **SDR** in allen Testsequenzen, meist deutlich (> 1 **MOS**-Punkt), vorgezogen.

Für den Praxiseinsatz wurden nachfolgend drei Workflows für verschiedene Zielsetzungen vorgestellt, von denen die ersten beiden Optionen in Feldversuchen erprobt bzw. produktiv eingesetzt werden. Dabei handelt es sich zum einen um die Verwendung einer **HDR**-Kamera in einem klassischen **SDR**-Workflow, welches eine gesteigerte Bildqualität mit einer kostengünstigeren Produktion verbinden soll. Zum anderen wird ein **HDR**-Workflow ermöglicht, bei dem die Blende in **HDR** gesteuert und die Blendendifferenz über die Dynamik des Konverters gelöst wird. Die hierfür benötigte Echtzeitimplementierung wurde in Abschnitt 7.4 vorgestellt. Abschließend wurde dargelegt, wie sich aus den praktischen Anwendungen Lerneffekte und zukünftige Aufgaben ergeben bzw. derzeitige Limitationen des Verfahrens beobachten lassen.

Kapitel 8

Fazit und Ausblick

Die Zielsetzung der vorliegenden Arbeit war es, ein Konzept für die flächendeckende Einführung von **HDR** im Broadcast anzubieten. Dabei galt die Prämisse, dass sowohl Besitzer eines neuen **HDR**-Displays als auch jene mit einem konventionellen **SDR**-Display von der gesteigerten Aufnahmequalität profitieren sollten. Auch sollte sich die Ausstrahlung von zwei verschiedenen Formaten kaum auf die Komplexität des Workflows und damit auf den Arbeitsaufwand auswirken. Es bedurfte also eines synergetischen Ansatzes, bei dem sich das **SDR** automatisch aus dem **HDR** ableitet.

In der Praxis war es bis zu diesem Zeitpunkt aber so, dass die Blendeneinstellung anhand des **SDR**-Bildes erfolgte. Das Problem war, dass die eingesetzten Verfahren zur Kontrastkompression nicht ausreichten, um die starken Schwankungen in einem **HDR**-Bild auszugleichen. Dafür wäre ein dynamisches Verfahren von Nöten, welches für die **SDR**-Variante die Blende nachzieht bzw. deutlich mehr Kontrastumfang innerhalb eines Bildes vereinigt, so dass die Blende an Relevanz verliert. Warum solch ein Verfahren noch nicht am Markt verfügbar war, wurde bei der Betrachtung der sogenannten Tonemapping Operatoren (TMOs) und der Erhebung der Broadcast-Anforderungen deutlich. Als der wahrscheinlich wichtigste Aspekt wurde dabei die menschliche Präferenz identifiziert.

In den darauf folgenden Kapiteln konnte gezeigt werden, wie die Kenntnis über die menschliche Wahrnehmungsgrenze, aber vor allem über die Präferenz dabei helfen kann, die bisherigen Limitationen aufzulösen. Dazu wurden eine Vielzahl von Probandenversuchen durchgeführt. Diese kamen zu dem Ergebnis, dass der Zuschauer zwar kontrastreiche Bilder, also Bilder mit einer steilen Gradation, bevorzugt, gleichzeitig aber große Informationsverluste in den Lichtern oder Schatten gerne vermeiden würde. Ein Schlüssel dieser Arbeit liegt nun darin, diese Abwägungsentscheidung zwischen Gradation und Information pro Szene bestmöglich zu treffen. Ferner wurde deutlich, dass helle Bildbereiche die Wahrnehmbarkeit von dunkleren Bildbereichen in der Umgebung signifikant

reduzieren können. Dieser Effekt ist bei den helleren **HDR**-Displays stärker ausgeprägt. Analog wurde bei den Untersuchungen bezüglich der Präferenz deutlich, dass die lineare Repräsentation einer inhomogen ausgeleuchteten Szene auf dem Display nicht gewollt ist. Die Zuschauer favorisieren ein Angleichen von Lichter und Schatten.

Bei der Überführung der Erkenntnisse in ein einsatzfähiges Verfahren wurden die sich aus den Beobachtungen ergebenden Anforderungen mit den Möglichkeiten der bereits vorhandenen TMOs abgeglichen. Die global operierenden, also das gesamte Bild betrachtenden Vertreter konnten weder für eine ausreichend starke Gradation noch für eine Homogenisierung der Beleuchtungsverhältnisse sorgen. Es bedurfte der Berücksichtigung der Umgebung jedes einzelnen Pixels und somit des lokalen Adaptionsniveaus. Diese lokalen Vertreter wiesen jedoch Nachteile bezüglich ihrer Robustheit auf und waren nicht in der Lage, über verschiedenste Videosequenzen hinweg ein durchgehend ansprechendes Bild zu liefern. Problematisch dabei war ferner die Neigung zu Artefakten wie Halos, Flackern oder Ghosting.

Das in der Folge dargelegte Verfahren führt deshalb die eben angesprochene Abwägungsentscheidung zwischen Gradation und Information sektorenweise durch. Das erstmals auf dieser Basis durchgeführte Verschleifen der Analyseergebnisse sorgt für die nötige Stabilität. Ermöglicht wird das durch die neuartige Anwendung eines bilateralen Filters. Anstelle das Bild mit sich selber zu filtern, wird das unterabgetastete Analyseergebnis mit dem Eingangsbild gefiltert, wodurch eine Objektkanten erhaltende Maske entsteht. Über das potenzielle Aussetzen der Verschleifung am Beginn einer neuen Szene gibt die Szenenwechselerkennung Auskunft.

Neben der Kompression der Helligkeitskomponente ist die Betrachtung des Farbedrucks entscheidend. Es wurde dargelegt, wie die Änderung der Helligkeitskomponente stets Einfluss auf die Farbdarstellung nimmt, was sich vornehmlich in einem geänderten Sättigungsempfinden äußert. Zur Kompensation dieses Effektes wurde gezeigt, wie mit der Betrachtung eines Referenzweiß pro Pixel die Chroma-Komponenten entsprechend modifiziert werden können. Diese Vorgehensweise kann als konstante Buntheit interpretiert werden. Aufgrund der höheren Komplexität des Verfahrens wird es aber unter Umständen nicht für alle Anwendungen zum Einsatz kommen. Für diese Fälle steht mit der Koeffizientenmethode, die nur einen Teil der Änderung der Helligkeitskomponente auch auf die Chroma-Komponenten überträgt, eine meist gute Näherung bereit.

Dass die beschriebenen Maßnahmen letztendlich zu einem Qualitätsgewinn im Vergleich zum klassischen **SDR**-Fernsehen führen, konnte mithilfe einer **PSNR**-Messung, aber vor allem anhand einer subjektiven Untersuchung nach **DSCQS** nachgewiesen werden. Es wurde deutlich, dass die Steigerung maßgeblich vom Bildinhalt beeinflusst wird. Szenen mit einem höheren Kontrastumfang profitierten stärker von der Aufzeichnung in **HDR**

und der anschließenden Konvertierung, als Szenen mit einem eher geringen Kontrastumfang.

Zwar war die Steigerung der Bildqualität das primäre Ziel, trotzdem sollte nicht außer Acht gelassen werden, dass das beschriebene Verfahren auch (gleichzeitig) dazu genutzt werden kann einen höheren Grad an Automatisierung zu erreichen. In einer Zeit die nach mehr und schneller produziertem Content verlangt, ist dies ein wichtiger Aspekt, der auch die Art zu Arbeiten verändern kann. Ob es in Zukunft noch eines Bildingeneurs bedarf, der lediglich die Blende aussteuert, darf nach den Ergebnissen zumindest infrage gestellt werden.

Schlussendlich konnte gezeigt werden, wie **EVI** den Übergang von der derzeitigen **SDR**-Produktionsweise zu zukünftigen Ansätzen maßgeblich gestalten und ermöglichen könnte. Die vorgestellten Workflows bieten eine schrittweise Annäherung an eine komplette **HDR**-Produktion. Dieser fließende Übergang hilft der aktuell bei vielen Broadcastern vorherrschenden Verunsicherung entgegenzuwirken und die Handhabung der Formatvielfalt langsam zu erlernen. Dabei ist das Konzept nicht nur als Übergangsszenario gedacht, sondern auch aufgrund der Homogenisierung und gesteigerten Automatisierung für reine **HDR**-Workflows von Interesse. Der produktive Einsatz bei *Deutschland sucht den Superstar* sowie die Auszeichnung mit dem **NAB** *Show Product of the Year Award* untermauern die gesetzten Ansprüche.

Trotzdem sollte das Bewusstsein vorherrschen, dass es sich erst um den Anfang eines Reifeprozesses handelt, welcher sich in den nächsten Monaten und Jahren über die Erfahrung im praktischen Einsatz ausgestalten wird. Die bis zum heutigen Zeitpunkt durchgeführten Tests und Feldversuche haben bereits für einige Verbesserungen gesorgt, wie beispielsweise die Einführung des hybriden Modus (s. Abschnitt **7.5**). Weitere Punkte, wie die Optimierung der Szenenwechselerkennung oder des bilateralen Filters, werden aktuell untersucht. Der zweite genannte Punkt bezieht sich dabei auf die unter **7.2.2** beschriebene Limitation, dass sehr kleine Bildbereiche, zu welchen es in der näheren räumlichen Nachbarschaft keine korrespondierenden Leuchtdichtewerte gibt, nicht ausreichend aufgehellt bzw. abgedunkelt werden können. In Zukunft soll dies durch eine weitere Verarbeitungsstufe gelöst werden, die in solchen Fällen den Betrachtungsradius erweitern kann und die Korrekturmaske entsprechend modifiziert.

Ferner wird derzeit diskutiert, inwieweit der **EVI**-Prozess invertierbar sein sollte, oder ob eine Downkonvertierung eher als Ablaufende zu verstehen ist. Die statischen Funktionen der im Moment am Markt verfügbaren Lösungen können zwar umgekehrt werden, allerdings kommt es dort im **SDR** zum Clipping, wodurch Details in den Lichtern und Schatten verloren gehen. Bei **EVI** hingegen werden zwar fast alle Informationen erhalten, jedoch scheint aus aktueller Perspektive eine absolute Reversibilität trotzdem

ausgeschlossen, da der bilaterale Filter auf Grundlage des Eingangsbildes errechnet und nicht gespeichert wird. Es soll deshalb getestet werden, wie gut das Ausgangsbild mit Hilfe von Zwischenergebnissen des Algorithmus, welche in Metadaten gespeichert bzw. übertragen werden, angenähert werden kann.

Für den Fortschritt des Kontrastkompressionssystems könnte auch eine technologische Entwicklung relevant werden, welche vielleicht in Zukunft Bestandteil nahezu jeder algorithmischen Lösung sein wird. Die künstliche Intelligenz (KI) hat seit dem Beginn der Dissertation eine hochgradige Ausweitung auf unterschiedlichste Bereiche erlebt. Besonders im Mediumfeld sind viele potenzielle Einsatzfelder gegeben oder sogar bereits erschlossen. Eine denkbare einfache Verbindung von KI und EVI wäre, dass die Einstellung und Steuerung der Benutzerparameter auf Basis von großen Datenmengen aus vorherigen Produktionen erfolgt. Allerdings ist derzeit nur schwer abzusehen wie KI Produktions-Workflows beeinflussen wird. Maßgeblich dafür wird die Bereitschaft der Medienschaffenden sein ihre kreativen Entscheidungen an einen Computer auszulagern. Allein der Wandel hin zu einem dynamisch agierenden Konvertierungsprozess ist eine große Herausforderung und noch nicht abgeschlossen.

Wendet man den Blick vom klassischen Broadcast ab, bietet das in dieser Arbeit vorgestellte Konzept auch Potenzial zur Expansion auf andere Gebiete. Naheliegend sind dabei Nicht-Echtzeit-Anwendungen im Mediumfeld. Der Algorithmus wäre in der Lage, die so genannten Lookup-Tables (LUTs) ablösen, welche heute hauptsächlich in der Postproduktion genutzt werden, um das Akquiseformat statisch in das Zielformat zu überführen. Diese statische Konvertierung weist vergleichbare Nachteile bezüglich des Kontrastumfanges auf, wie sie bei der Blendensteuerung dargelegt wurden. Um trotzdem den hohen Kontrastumfang der Aufnahme zu erhalten, werden Bereiche in der sekundären Farbkorrektur nachträglich gesondert behandelt. Dieser langwierige Prozess könnte durch den Einsatz von EVI in Form eines Plug-Ins für gängige Schnittsysteme deutlich beschleunigt werden. Auch bietet die dynamische Form der Konvertierung eine deutlich bessere Annäherung an das finale Ergebnis und erlaubt so eine realistischere Ansicht für den Schnitt.

Die Technik ist aber nicht auf das Mediumfeld limitiert. Drei weitere Gebiete könnten maßgeblich von dieser Technologie profitieren, da diese mit schwierigen Lichtverhältnissen konfrontiert sind und ohne manuelle Blendensteuerung auskommen müssen. So könnte EVI eine sinnvolle Vorverarbeitung des Kamerasignals für das autonome Fahren darstellen. Durch die verbesserte Kontrastdarstellung dürfte die Erkennung von Objekten durch die zugehörigen Algorithmen vereinfacht werden. Da die Kanten tiefer modelliert sind, heben sich die Objekte in der Folge stärker voneinander ab. Ferner wird eine hohe Durchzeichnung in den Lichtern und Schatten erreicht. Gerade Gegenlicht oder starke

Reflexionen (z. B. auf einer weißen **LKW**-Plane) behindern derzeit noch die zuverlässige Erkennung in den sehr hellen bzw. dunklen Bildarealen. Hierzu wurden erste Versuche mit einem Automobilzulieferer gestartet.

Des Weiteren könnte **HDR** für die Sicherheitstechnik relevant werden. Dies gilt sowohl für klassische Überwachungskameras, aber noch mehr für die so genannten Body-Cams. Diese von Polizisten am Körper getragenen Aufzeichnungsgeräte sind im Feld stark veränderlichen Lichtverhältnissen ausgesetzt, ohne dass eine Korrektur oder Vorschau-funktion bestehen würde. Während eines Einsatzes muss sich der Polizist rein auf seinen Einsatz konzentrieren und der Technik vertrauen können. Ähnlich sieht es in der Medi-zintechnik aus. Hier wäre es beispielsweise denkbar, Endoskope mit **HDR**-Chips auszu-statten und das Signal vor der Anzeige entsprechend zu konvertieren. Die ungleichmäßige Beleuchtung durch die **LED**-Lampe könnte so deutlich angeglichen werden.

In allen drei Fällen würde es somit nicht um ein natives „**HDR**-Erlebnis“ sondern um die Modifikation des derzeitigen Bildes gehen, um somit eine Merkmalsextraktion zu verbessern. Zusammen mit dem Stammgebiet der Medienwelt bieten sich somit der **HDR**-Konvertierung zahlreiche Möglichkeiten, welche gemeinsam mit den verschiedenen Partnern in den nächsten Jahren adressiert werden sollen. Diese Arbeit schafft hierfür die benötigte wissenschaftliche Grundlage, denn die Zeit ist reif für ein Ende von Über- und Unterbelichtung.

Anhang A

Transformationsmatrizen

A.1 Transformation nach XYZ

ALEXA in XYZ [\[16\]](#)

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0,638008 & 0,214704 & 0,097744 \\ 0,291954 & 0,823841 & -0,115795 \\ 0,002798 & -0,067034 & 1,1532994 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (\text{A.1})$$

sGamut3 in XYZ [\[119\]](#)

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0,70648 & 0,128801 & 0,115172 \\ 0,27097 & 0,786606 & -0,0575860 \\ -0,0096 & -0,00460 & 1,094135 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (\text{A.2})$$

ACES in XYZ [\[113\]](#)

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0,9525523959 & 0 & 0,0000936786 \\ 0,3439664498 & 0,7281660966 & -0,0721325464 \\ 0 & 0 & 1,0088251844 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (\text{A.3})$$

A.2 Transformation aus XYZ

XYZ in Rec.709

$$\begin{bmatrix} R709 \\ G709 \\ B709 \end{bmatrix} = \begin{bmatrix} 3,2406 & -1,5372 & -0,4986 \\ -0,9689 & 1,8758 & 0,0415 \\ 0,0557 & -0,2040 & 1,0570 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (\text{A.4})$$

XYZ in DCI-P3

$$\begin{bmatrix} RP3 \\ GP3 \\ BP3 \end{bmatrix} = \begin{bmatrix} 2,7254 & -1,0180 & -0,4402 \\ -0,7952 & 1,6897 & 0,0226 \\ 0,0412 & -0,0876 & 1,1009 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (\text{A.5})$$

XYZ in Rec.2020

$$\begin{bmatrix} R2020 \\ G2020 \\ B2020 \end{bmatrix} = \begin{bmatrix} 1,71666343 & -0,35567332 & -0,25336809 \\ -0,66667384 & 1,61645574 & 0,0157683 \\ 0,01763985 & -0,0427706133 & 0,94210312 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (\text{A.6})$$

A.3 Transformationen Dolby

XYZ in LMS [\[23\]](#)

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0,3592 & 0,6976 & -0,0358 \\ -0,1922 & 1,1004 & 0,0755 \\ 0,0070 & 0,0749 & 0,8434 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (\text{A.7})$$

L'M'S' nach ICtCp [\[23\]](#)

$$\begin{bmatrix} I \\ Ct \\ Cp \end{bmatrix} = \begin{bmatrix} 0,5000 & 0,5000 & 0 \\ 1,6137 & -3,3234 & 1,7097 \\ 4,3780 & -4,2455 & -0,1325 \end{bmatrix} \begin{bmatrix} L' \\ M' \\ S' \end{bmatrix} \quad (\text{A.8})$$

Literaturverzeichnis

- [1] DIE MEDIENANSTALTEN (Hrsg.): *Staatsvertrag für Rundfunk und Telemedien (Rundfunkstaatsvertrag RStV)*. 2019
- [2] AKYÜZ, A. O. ; FLEMING, R. ; RIECKE, B. E. ; REINHARD, E. ; BÜLTHOFF, H. H.: Do HDR Displays Support LDR Content?: A Psychophysical Evaluation. In: *ACM Trans. Graph.* 26 (2007), Jul., Nr. 3. <http://dx.doi.org/10.1145/1276377.1276425>. – DOI 10.1145/1276377.1276425. – ISSN 0730–0301
- [3] ATKINS, R. : Better Pixels: Color Volume and Quantization Errors. In: *HPA 2014*
- [4] AYDIN, T. O. ; STEFANOSKI, N. ; CROCI, S. ; GROSS, M. ; SMOLIC, A. : Temporally Coherent Local Tone Mapping of HDR Video. In: *ACM Trans. Graph.* 33 (2014), Nov., Nr. 6, 196:1–196:13. <http://dx.doi.org/10.1145/2661229.2661268>. – DOI 10.1145/2661229.2661268. – ISSN 0730–0301
- [5] BARTEN, P. G. J.: *Contrast sensitivity of the human eye and its effects on image quality*. Bellingham, Washington : SPIE Press, 1999
- [6] BARTLESON, C. J. ; BRENEMAN, E. J.: Brightness Perception in Complex Fields. In: *J. Opt. Soc. Am.* 57 (1967), Jul, Nr. 7, 953–957. <http://dx.doi.org/10.1364/JOSA.57.000953>. – DOI 10.1364/JOSA.57.000953
- [7] BENNETT, E. P. ; MCMILLAN, L. : Video Enhancement Using Per-pixel Virtual Exposures. In: *ACM Trans. Graph.* 24 (2005), Jul., Nr. 3, 845–852. <http://dx.doi.org/10.1145/1073204.1073272>. – DOI 10.1145/1073204.1073272. – ISSN 0730–0301
- [8] BENOIT, A. ; ALLEYSSON, D. ; HÉRAULT, J. ; CALLET, P. L.: Spatio-temporal Tone Mapping Operator Based on a Retina Model. In: *Computational Color Imaging, Second International Workshop, CCIW 2009, Saint-Etienne, France, March 26-27, 2009. Revised Selected Papers*, 12–22
- [9] BENYAMIN, M. : *Untersuchung zur Wahrnehmung des simultanen Kontrastumfangs in Abhängigkeit der Sichtfeldgröße bei HDR-Systemen*. 2018

- [10] BODMANN, H. : Visibility assessment in lighting engineering. In: *Journal of the Illuminating Engineering Society* 2 (1973), S. 437–443
- [11] BOITARD, R. ; BOUATOUCH, K. ; COZOT, R. ; THOREAU, D. ; GRUSON, A. : Temporal coherency for video tone mapping. In: EIJK, A. M. J. (Hrsg.) ; DAVIS, C. C. (Hrsg.) ; HAMMEL, S. M. (Hrsg.) ; MAJUMDAR, A. K. (Hrsg.): *Proc. SPIE 8499, Applications of Digital Image Processing XXXV*. San Diego, Okt. 2012, 84990D–84990D–10
- [12] BOLEWSKI, N. : Bildwandler-Technologie fuer zukünftige Fernsehformate. In: *FKT* 68 (2014), Dezember, S. 51–52
- [13] BORER, T. ; COTTON, A. : White Paper WHP 309 - A “Display Independent“ High Dynamic Range Television System / BBC. 2015. – White Paper
- [14] BORER, T. : White Paper WHP 283 - Non-linear Opto-Electrical Transfer Functions for High Dynamic Range Television / BBC. 2014. – White Paper
- [15] BOYNTON, R. M. ; WHITTEN, D. N.: Visual Adaptation in Monkey Cones: Recordings of Late Receptor Potentials. In: *Science* 170 (1970), Nr. 3965, S. 1423–1426. <http://dx.doi.org/10.1126/science.170.3965.1423>. – DOI 10.1126/science.170.3965.1423. – ISSN 0036–8075
- [16] BRENDEL, H. ; ARRI (Hrsg.): *ALEXA Log C Curve*. <https://www.arri.com/resource/blob/31918/66f56e6abb6e5b6553929edf9aa7483e/2017-03-alexa-logc-curve-in-vfx-data.pdf>. zuletzt abgerufen am: 02.11.2020
- [17] BYCHKOV, I. : *Evaluierung der menschlichen Kontrastempfindlichkeit bei High Dynamic Range (HDR) Wiedergabe*, Hochschule RheinMain, Masterthesis, 2016
- [18] CENTEN, P. : 4k-HDR-Imagers. In: *HPA Tech Retreat*, 2015
- [19] CIE: ILV: International Lighting Vocabulary. 2011 (S 017 E:2011). – Standard
- [20] DALY, S. ; KUNKEL, T. ; SUN, X. ; FARRELL, S. ; CRUM, P. : Viewer Preferences for Shadow, Diffuse, Specular, and Emissive Luminance Limits of High Dynamic Range Displays. In: *SID Symposium Digest of Technical Papers* 44 (2013), Nr. 1, 563–566. <http://dx.doi.org/10.1002/j.2168-0159.2013.tb06271.x>. – DOI 10.1002/j.2168–0159.2013.tb06271.x. – ISSN 2168–0159
- [21] DE MASH, L. E.: Optimum Telecine Transfer Characteristics. In: *Journal Of The Society Of Motion Picture And Television Engineers* 81 (1972), S. 784–787

- [22] DOLBY LABORATORIES: Dolby Vision White Paper / Dolby Laboratories. zuletzt abgerufen am: 02.11.2020. https://www.diversifiedvideosolutions.com/Dolby_Vision/Dolby_Vision_White_Paper.pdf. – White Paper
- [23] DOLBY LABORATORIES: ICtCp White Paper / Dolby Laboratories. zuletzt abgerufen am: 02.11.2020. <http://www.dolby.com/us/en/technologies/dolby-vision/ICtCp-white-paper.pdf>. – White Paper
- [24] DOLBY LABORATORIES: Dolby Vision profiles and levels / Dolby Laboratories. zuletzt abgerufen am: 23.12.2019. https://www.dolby.com/us/en/technologies/dolby-vision/dolby-vision-profiles-levels_v1.3.1.1.pdf. – White Paper
- [25] DRAGO, F. ; MYSZKOWSKI, K. ; ANNEN, T. ; CHIBA, N. : Adaptive Logarithmic Mapping For Displaying High Contrast Scenes. In: *Computer Graphics Forum* 22 (2003), S. 419–426
- [26] DUAN, J. ; BRESSAN, M. ; DANCE, C. : *Histogram adjustment for high dynamic range image mapping*. Dez. 22 2009. – US Patent 7,636,496
- [27] DURAND, F. ; DORSEY, J. : Interactive Tone Mapping, 2000, S. 219–230
- [28] DURAND, F. ; DORSEY, J. : Fast Bilateral Filtering for the Display of High-dynamic-range Images. In: *ACM Trans. Graph.* 21 (2002), Jul., Nr. 3, 257–266. <http://dx.doi.org/10.1145/566654.566574>. – DOI 10.1145/566654.566574. – ISSN 0730–0301
- [29] EBNER, F. ; FAIRCHILD, M. D.: Development and Testing of a Color Space IPT with Improved Hue Uniformity. In: *6th Color and Imaging Conference, CIC 1998, Scottsdale, Arizona, USA, November 17-20, 1998*, 8–13
- [30] EBU: ETSI TS 103 433-1 High-Performance Single Layer High Dynamic Range (HDR) System for use in Consumer Electronics devices; Part 1: Directly Standard Dynamic Range (SDR) Compatible HDR System (SL-HDR1) . (2017)
- [31] EILERTSEN, G. ; WANAT, R. ; MANTIUK, R. K. ; UNGER, J. : Evaluation of Tone Mapping Operators for HDR-Video. In: *Comput. Graph. Forum* 32 (2013), Nr. 7, S. 275–284
- [32] EUROPEAN BROADCASTING UNION (EBU): TECH 3320 - User requirements for Video Monitors in Television Production. 2017. – Forschungsbericht
- [33] FAIRCHILD, M. D.: *Color Appearance Models*. 3. John Wiley & Sons, 2013. – 85–97, 201–210, 416–417 S.

- [34] FAIRCHILD, M. D. ; CHEN, P.-H. : Brightness, lightness, and specifying color in high-dynamic-range scenes and images. In: *SPIE/IS&T Electronic Imaging Conference* Bd. 7867
- [35] FARRELL, S. ; KUNKEL, T. ; DALY, S. : A Cinema Luminance Range by the People, for the People: Viewer Preferences on Luminance Limits for a Large-Screen Environment. In: *SMPTE Motion Imaging Journal* 124 (2015), July/August, S. 56–61
- [36] FATTAL, R. ; LISCHINSKI, D. ; WERMAN, M. : Gradient Domain High Dynamic Range Compression. In: *ACM Trans. Graph.* 21 (2002), Jul., Nr. 3, 249–256. <http://dx.doi.org/10.1145/566654.566573>. – DOI 10.1145/566654.566573. – ISSN 0730–0301
- [37] FERWERDA, J. A.: Fundamentals of Spatial Vision. In: *V. Interrante (Ed.) Applications of visual perception in computer graphics*, SIGGRAPH '98, 1998, S. 1 – 27
- [38] FERWERDA, J. A. ; PATTANAİK, S. N. ; SHIRLEY, P. ; GREENBERG, D. P.: A Model of Visual Adaptation for Realistic Image Synthesis. In: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. ACM (SIGGRAPH '96). – ISBN 0–89791–746–4, 249–258
- [39] FRANÇOIS, E. ; VAN DE KERKHOF, L. : A Single-Layer HDR Video Coding Framework with SDR Compatibility. In: *SMPTE Motion Imaging Journal* 126 (2017), April, Nr. 3, S. 16–22. <http://dx.doi.org/10.5594/JMI.2017.2660618>. – DOI 10.5594/JMI.2017.2660618
- [40] FROEHLICH, J. ; SU, G. M. ; DALY, S. ; SCHILLING, A. ; EBERHARDT, B. : Content aware quantization: Requantization of high dynamic range baseband signals based on visual masking by noise and texture. In: *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, S. 884–888
- [41] FROEHLICH, J. ; GRANDINETTI, S. ; EBERHARDT, B. ; WALTER, S. ; SCHILLING, A. ; BRENDDEL, H. : Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays. In: SAMPAT, N. (Hrsg.) ; TEZAUR, R. (Hrsg.) ; BATTIATO, S. (Hrsg.) ; FOWLER, B. A. (Hrsg.) ; International Society for Optics and Photonics (Veranst.): *Digital Photography X* Bd. 9023 International Society for Optics and Photonics, SPIE, 279 – 288
- [42] FROEHLICH, J. ; SCHILLING, A. ; EBERHARDT, B. : Gamut Mapping for Digital Cinema. In: *SMPTE Conferences* 2013 (2013), Nr. 10, 1-11. <http://dx.doi.org/10.5594/M001489>. – DOI 10.5594/M001489

- [43] GUTHIER, B. : *Real-Time Algorithms for High Dynamic Range Video*, Universitaet Mannheim, Diss., 2012
- [44] HEDTKE, R. ; LENZEN, L. : HFR für UHD TV - Wie stark muss die Bildwechselfrequenz steigen? In: *FKT 69* (2015), Nr. 8-9, S. 393–398
- [45] HEDTKE, R. ; LENZEN, L. ; CHRISTMANN, M. : HDR - höherer dynamischer Bereich. In: *FKT 70* (2016), Nr. 5, S. 193–201
- [46] HOFF, Y. : *Testreihe zur Ermittlung des subjektiven Eindrucks von Konsumenten bezüglich komprimierten EVI- und SDR-Sequenzen*. 2019
- [47] HOFFMANN, H. : Investigation in an Image Dynamic Range Methodology. In: *NAB Show*, 2014
- [48] HOOD, D. C. ; FINKELSTEIN, M. : Visual sensitivity. In: BOFF, K. (Hrsg.) ; KAUFMAN, L. (Hrsg.) ; THOMAS, J. (Hrsg.): *Handbook of Perception and Human Performance* Bd. 1. 1986, Kapitel 5-1, S. 5–66
- [49] HSU, W. : *Automatic tone mapping method and image processing device*. Jul. 15 2014. – US Patent 8,781,225
- [50] HULUSIC, V. ; VALENZISE, G. ; DEBATTISTA, K. ; DUFAUX, F. : Robust Dynamic Range Computation for High Dynamic Range Content. In: *Proceedings of the IS&T International Symposium on Electronic Imaging 2017*, 2017
- [51] HUNT, R. W. G.: Light and Dark Adaptation and the Perception of Color*. In: *J. Opt. Soc. Am.* 42 (1952), Mar, Nr. 3, 190–199. <http://dx.doi.org/10.1364/JOSA.42.000190>. – DOI 10.1364/JOSA.42.000190
- [52] HUNT, R. : *The Reproduction of Colour*. Wiley, 2005 (The Wiley-IS&T Series in Imaging Science and Technology). – ISBN 9780470024263
- [53] HUSS, A. K.: *Entwicklung eines Workflows zur Anwendung des iCAM06 Tone-mappers auf HDR-Video-Sequenzen*, Fachhochschule Köln, Diplomarbeit, 5 2013
- [54] INTERNATIONAL TELECOMMUNICATION UNION (ITU): *BT.709:1 - Basic parameter values for the HDTV standard for the studio and for international programme exchange*. 11 1993
- [55] INTERNATIONAL TELECOMMUNICATION UNION (ITU): *BT.1886 - Reference electro-optical transfer function for flat panel displays used in HDTV studio production*. 03 2011
- [56] INTERNATIONAL TELECOMMUNICATION UNION (ITU): *BT.500:13 - Methodology for the subjective assessment of the quality of television pictures*. 01 2012

- [57] INTERNATIONAL TELECOMMUNICATION UNION (ITU): *BT.2020:1 - Parameter values for ultra-high definition television systems for production and international programme exchange*. 6 2014
- [58] INTERNATIONAL TELECOMMUNICATION UNION (ITU): *BT.2100:0 - Image parameter values for high dynamic range television for use in production and international programme exchange*. 07 2016
- [59] INTERNATIONAL TELECOMMUNICATION UNION (ITU): *Report ITU-R BT.2408-2 - Guidance for operational practices in HDR television production*. 07 2019
- [60] IRAWAN, P. ; FERWERDA, J. A. ; MARSCHNER, S. R.: Perceptually Based Tone Mapping of High Dynamic Range Image Streams. In: *Proceedings of the Sixteenth Eurographics Conference on Rendering Techniques*. Eurographics Association (EGSR '05). – ISBN 3-905673-23-1, 231–242
- [61] KISER, C. ; REINHARD, E. ; TOCCI, M. ; TOCCI, N. : Real Time Automated Tone Mapping System for HDR Video. In: *Proceedings of the IEEE International Conference on Image Processing*. Orlando, Florida : IEEE, September 2012, S. 2749–2752
- [62] KRAWCZYK, G. ; MYSKOWSKI, K. ; SEIDEL, H. : Perceptual effects in real-time tone mapping. In: *Spring Conference on Computer Graphics, SCCG 2005, Budmerice Castle, Slovakia, May 12-14, 2005*, 195–202
- [63] KUANG, J. ; JOHNSON, G. M. ; FAIRCHILD, M. D.: iCAM06: A Refined Image Appearance Model for HDR Image Rendering. In: *J. Vis. Commun. Image Represent.* 18 (2007), Okt., Nr. 5, 406–414. <http://dx.doi.org/10.1016/j.jvcir.2007.06.003>, – DOI 10.1016/j.jvcir.2007.06.003. – ISSN 1047-3203
- [64] KUNKEL, T. ; REINHARD, E. : A Reassessment of the Simultaneous Dynamic Range of the Human Visual System. In: *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*. ACM (APGV '10). – ISBN 978-1-4503-0248-7, 17–24
- [65] KUSAKABE, Y. ; IKEDA, Y. ; SHIRAI, N. ; MASAOKA, K. ; YAMASHITA, T. ; NISHIDA, Y. ; IKEDA, T. ; SUGAWARA, M. : Extended Image Dynamic Range System For UHDTV Broadcasting. In: *IBC 2015*, 2015
- [66] KUTSCHBACH, P. : *Konzeptionierung und Durchführung einer Untersuchung zum Simultankontrast in Abhängigkeit der Sichtfeldgröße bei HDR Wiedergabe*
- [67] LARSON, G. W. ; RUSHMEIER, H. ; PIATKO, C. : A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes. In: *IEEE Transactions*

- on Visualization and Computer Graphics* 3 (1997), Okt., Nr. 4, 291–306. <http://dx.doi.org/10.1109/2945.646233>. – DOI 10.1109/2945.646233. – ISSN 1077–2626
- [68] LEDDA, P. ; SANTOS, L. P. ; CHALMERS, A. : A Local Model of Eye Adaptation for High Dynamic Range Images. In: *Proceedings of the 3rd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*. ACM (AFRIGRAPH '04). – ISBN 1–58113–863–6, 151–160
- [69] LENZEN, L. ; HEDTKE, R. ; CHRISTMANN, M. : How Tone Mapping Influences the Bit Rate and the Bit Depth of Coded Sequences. In: *SMPTE Motion Imaging Journal* 127 (2018), June, Nr. 5, S. 38–43. <http://dx.doi.org/10.5594/JMI.2018.2810021>. – DOI 10.5594/JMI.2018.2810021. – ISSN 1545–0279
- [70] LENZEN, L. ; HEDTKE, R. ; CHRISTMANN, M. : HDR in Consideration of the Abilities of the Human Visual System. In: *SMPTE Motion Imaging Journal* 128 (2019), June, Nr. 5, S. 40–45. <http://dx.doi.org/10.5594/JMI.2019.2907350>. – DOI 10.5594/JMI.2019.2907350. – ISSN 1545–0279
- [71] LENZEN, L. : *HFR - Analyse des Mehrwerts von höheren Bildwechselfrequenzen bei UHDTV mit Hilfe von subjektiven Untersuchungen*. Wiesbaden, Hochschule RheinMain, Masterthesis, 2014
- [72] LENZEN, L. : HDR for legacy displays using Sectional Tone Mapping. In: *Proceedings of the IBC conference 2016*, 2016
- [73] LENZEN, L. : HDR for legacy displays using Sectional Tone Mapping. In: *The Best of IET and IBC* (2016)
- [74] LENZEN, L. ; CHRISTMANN, M. : *Patent DE102015119137 B3 - Erfindung betreffend Verfahren zur Blendeneinstellung*. 2017
- [75] LENZEN, L. ; CHRISTMANN, M. : Subjective viewer preference model for automatic HDR down conversion. In: *Proceedings of the IS&T International Symposium on Electronic Imaging 2017*, 2017
- [76] LENZEN, L. ; HEDTKE, R. ; CHRISTMANN, M. : How Tone Mapping influences the bit rate and the bit depth of coded sequences. In: *Proceedings of the IBC conference 2017*, 2017
- [77] LENZEN, L. ; HEDTKE, R. ; CHRISTMANN, M. : HDR in consideration of the abilities of the human visual system. In: *Proceedings of the IBC conference 2018*, 2018

- [78] LITWIC, L. ; BAUMANN, O. ; WHITE, P. ; GOLDMAN, M. S.: Bit Rate Requirements for High Dynamic Range Video. In: *SMPTE Motion Imaging Journal* 125 (2016), July, Nr. 5, S. 52–60. <http://dx.doi.org/10.5594/JMI.2016.2569920>. – DOI 10.5594/JMI.2016.2569920. – ISSN 1545–0279
- [79] LÜBBE, E. : *Colours in the Mind - Colour Systems in Reality: A formula for colour saturation*. Books on Demand GmbH, 2008
- [80] LÜBBE, E. : Experimental Evidence of the Formula of Saturation. In: *Journal of Physical Science and Application* 3 (2013), Nr. 2, S. 78–81
- [81] MANTIUK, R. ; MANTIUK, R. ; TOMASZEWSKA, A. ; HEIDRICH, W. : Color correction for tone mapping. In: *Computer Graphics Forum* 28 (2009), Nr. 2, 193–202. <http://dx.doi.org/10.1111/j.1467-8659.2009.01358.x>. – DOI 10.1111/j.1467–8659.2009.01358.x. – ISSN 1467–8659
- [82] MANTIUK, R. ; DALY, S. ; KEROFISKY, L. : Display Adaptive Tone Mapping. In: *ACM Trans. Graph.* 27 (2008), Aug., Nr. 3, 68:1–68:10. <http://dx.doi.org/10.1145/1360612.1360667>. – DOI 10.1145/1360612.1360667. – ISSN 0730–0301
- [83] MANTIUK, R. ; EFREMOV, A. ; MYSZKOWSKI, K. ; SEIDEL, H.-P. : Backward Compatible High Dynamic Range MPEG Video Compression. In: *ACM SIGGRAPH 2006 Papers*. ACM (SIGGRAPH '06). – ISBN 1–59593–364–6, 713–723
- [84] MANTIUK, R. ; MYSZKOWSKI, K. ; SEIDEL, H.-P. : Lossy compression of high dynamic range images and video. In: ROGOWITZ, B. E. (Hrsg.) ; PAPPAS, T. N. (Hrsg.) ; DALY, S. J. (Hrsg.): *Human Vision and Electronic Imaging XI* Bd. 6057. San Jose, USA : SPIE, February 2006 (SPIE), S. 60570V
- [85] MCCARTHY, S. : Quantitative Evaluation of Human Visual Perception for Multiple Screens and Multiple Codecs. In: *SMPTE Mot. Img.* (2013), S. 36–42
- [86] MEETEREN, A. van ; VOS, J. : Resolution and contrast sensitivity at low luminances. In: *Vision Research* 12 (1972), Nr. 5, 825 - IN2. [http://dx.doi.org/10.1016/0042-6989\(72\)90008-9](http://dx.doi.org/10.1016/0042-6989(72)90008-9). – DOI /10.1016/0042–6989(72)90008–9. – ISSN 0042–6989
- [87] MILLER, S. : *A Perceptual EOTF for Extended Dynamic Range Imagery*. SMPTE Monthly Webcast. <https://www.smpte.org/sites/default/files/2014-05-06-EOTF-Miller-1-2-handout.pdf>. zuletzt abgerufen am: 23.12.2019
- [88] NAKA, K. I. ; RUSHTON, W. A.: S-potentials from luminosity units in the retina of fish (Cyprinidae). In: *Journal of Physiology* 185 (1966), Aug., Nr. 3

- [89] NAYATANI, Y. ; TAKAHAMA, K. ; SOBAGAKI, H. ; HASHIMOTO, K. : Color-appearance model and chromatic-adaptation transform. In: *Color Research & Application* 15 (1990), Nr. 4, S. 210–221
- [90] NOLAND, L. H. Katy C.; Truong T. Katy C.; Truong: White Paper WHP 287 - A Survey of UK Television Viewing Conditions / BBC. 2015. – White Paper
- [91] NOVICK, S. B.: The Reproduction From Colour Telecine Systems. In: *British Kinematography Sound And Television* 51 (1969), Nr. 10, S. 342–347
- [92] PATTANAİK, S. ; YEE, H. : Adaptive Gain Control for High Dynamic Range Image Display. In: *Proceedings of the 18th Spring Conference on Computer Graphics*. ACM (SCCG '02). – ISBN 1–58113–608–0, 83–87
- [93] PATTANAİK, S. N. ; TUMBLIN, J. ; YEE, H. ; GREENBERG, D. P.: Time-dependent Visual Adaptation for Fast Realistic Image Display. In: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press/Addison-Wesley Publishing Co. (SIGGRAPH '00). – ISBN 1–58113–208–5, 47–54
- [94] PAULSEN, A. : Die Magie der besseren Pixel. In: *FKT* 69 (2015), Januar/Februar, Nr. 1-3, S. 41–44, 97–100
- [95] PERONA, P. ; MALIK, J. : Scale-space and edge detection using anisotropic diffusion. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (1990), July, Nr. 7, S. 629–639. <http://dx.doi.org/10.1109/34.56205>. – DOI 10.1109/34.56205. – ISSN 0162–8828
- [96] PETIT, J. ; MANTIUK, R. : Assessment of video tone-mapping: Are cameras' S-shaped tone-curves good enough? In: *J. Visual Communication and Image Representation* 24 (2013), Nr. 7, S. 1020–1030
- [97] POULI, T. ; ARTUSI, A. ; BANTERLE, F. ; AKYUZ, A. O. ; SEIDEL, H.-P. ; REINHARD, E. : Color Correction for Tone Reproduction. In: *CIC21: Twenty-first Color and Imaging Conference* Society for Imaging Science and Technology (IS&T), Society for Imaging Science and Technology (IS&T), 215–220
- [98] POYNTON, C. : *Digital Video and HD Algorithms and Interfaces, Second Edition*. 2. San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 2012. – 21,121,259 S.
- [99] POYNTON, C. ; FUNT, B. : Perceptual uniformity in digital image representation and display. In: *Color Research & Application* 39 (2014), Nr. 1, 6–15. <http://dx.doi.org/10.1002/col.21768>. – DOI 10.1002/col.21768. – ISSN 1520–6378

- [100] PYTLARZ, J. A. ; G., P. E.: How close is close enough? Specifying colour tolerances for HDR and WCG displays. In: *Proceedings of the IBC conference 2017*, 2017
- [101] REINHARD, E. ; DEVLIN, K. : Dynamic range reduction inspired by photoreceptor physiology. In: *Visualization and Computer Graphics, IEEE Transactions on 11* (2005), Jan, Nr. 1, S. 13–24. <http://dx.doi.org/10.1109/TVCG.2005.9>. – DOI 10.1109/TVCG.2005.9. – ISSN 1077–2626
- [102] REINHARD, E. ; POULI, T. ; KUNKEL, T. ; LONG, B. ; BALLESTAD, A. ; DAMBERG, G. : Calibrated Image Appearance Reproduction. In: *ACM Trans. Graph.* 31 (2012), Nov., Nr. 6, 201:1–201:11. <http://dx.doi.org/10.1145/2366145.2366220>. – DOI 10.1145/2366145.2366220. – ISSN 0730–0301
- [103] REINHARD, E. ; STARK, M. ; SHIRLEY, P. ; FERWERDA, J. : Photographic Tone Reproduction for Digital Images. In: *ACM Trans. Graph.* 21 (2002), Jul., Nr. 3, 267–276. <http://dx.doi.org/10.1145/566654.566575>. – DOI 10.1145/566654.566575. – ISSN 0730–0301
- [104] REINHARD, E. ; WARD, G. ; PATTANAIK, S. ; DEBEVEC, P. ; HEIDRICH, W. ; MYSZKOWSKI, K. : *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. 2. San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 2010. – 5, 243–249, 256–270, 460–468 S.
- [105] REMPEL, A. G. ; HEIDRICH, W. ; LI, H. ; MANTIUK, R. : Video Viewing Preferences for HDR Displays Under Varying Ambient Illumination. In: *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*. ACM (APGV '09). – ISBN 978–1–60558–743–1, 45–52
- [106] RICHTER, M. : *Einführung in die Farbmeterik*. Gryter, Berlin, 1981. – 149 S.
- [107] RTL: *Deutschland sucht den Superstar*. <https://www.youtube.com/user/RTLSDS2012>. zuletzt abgerufen am: 23.12.2019
- [108] SCHLICK, C. : Quantization Techniques for Visualization of High Dynamic Range Pictures. In: *Photorealistic Rendering Techniques*, Springer-Verlag, 1994, S. 7–20
- [109] SCHMIDT, U. : *Professionelle Videotechnik*. 6. Hamburg : Springer Vieweg, 2013
- [110] SCHMITT, D. : *High Dynamic Range in Videobildern - Akzeptanzermittlung von Coloristen und Konsumenten sowie die Untersuchung der HDR-Fähigkeit bestehender Kameraaufnahmen und Filmmaster*. 2015
- [111] SCHULTZ, N. : *Untersuchungen zu HDR*, Hochschule RheinMain, Masterthesis, 2015

- [112] SEETZEN, H. ; HEIDRICH, W. ; STUERZLINGER, W. ; WARD, G. ; WHITEHEAD, L. ; TRENTACOSTE, M. ; GHOSH, A. ; VOROZCOVS, A. : High Dynamic Range Display Systems. In: *ACM Trans. Graph.* 23 (2004), Aug., Nr. 3, 760–768. <http://dx.doi.org/10.1145/1015706.1015797>. – DOI 10.1145/1015706.1015797. – ISSN 0730–0301
- [113] SMPTE: *SMPTE ST 2065-1 Academy Color Encoding Specification (ACES)*. 2012
- [114] SMPTE: *SMPTE ST 2084 High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays*. 2014
- [115] SMPTE: *SMPTE ST 2086 Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images*. 2015
- [116] SMPTE: *Study Group Report High-Dynamic-Range (HDR) Imaging Ecosystem*. 2015
- [117] SMPTE: *SMPTE ST 2094-1 Dynamic Metadata for Color Volume Transform - Core Components*. 2016
- [118] SMPTE: *SMPTE ST 2113:2018 Colorimetry of P3 Color Spaces*. 2019
- [119] SONY: *Technical Summary for S-Gamut3.Cine/S-Log3 and S-Gamut3/S-Log3*. <https://pro.sony/s3/cms-static-content/uploadfile/06/1237494271406.pdf>. zuletzt abgerufen am:02.11.2020
- [120] STEVENS, J. C. ; STEVENS, S. S.: Brightness Function: Effects of Adaptation*. In: *J. Opt. Soc. Am.* 53 (1963), Mar, Nr. 3, 375–385. <http://dx.doi.org/10.1364/JOSA.53.000375>. – DOI 10.1364/JOSA.53.000375
- [121] STOKKERMANS, M. J. E. U. M.G.M.; Murdoch M. M.G.M.; Murdoch: Preference for key parameter of tone mapping operator in different viewing conditions. In: *EXPERIENCING LIGHT 2012: International Conference on the Effects of Light on Wellbeing*, Eindhoven University of Technology, 1-4
- [122] TOMASI, C. ; MANDUCHI, R. : Bilateral Filtering for Gray and Color Images. In: *Proceedings of the Sixth International Conference on Computer Vision*. IEEE Computer Society (ICCV '98). – ISBN 81–7319–221–9, 839–
- [123] TUMBLIN, J. ; HODGINS, J. K. ; GUENTER, B. K.: Two Methods for Display of High Contrast Images. In: *ACM Trans. Graph.* 18 (1999), Jan., Nr. 1, 56–94. <http://dx.doi.org/10.1145/300776.300783>. – DOI 10.1145/300776.300783. – ISSN 0730–0301

- [124] TUMBLIN, J. ; RUSHMEIER, H. : Tone Reproduction for Realistic Images. In: *IEEE Comput. Graph. Appl.* 13 (1993), Nov., Nr. 6, 42–48. <http://dx.doi.org/10.1109/38.252554>. – DOI 10.1109/38.252554. – ISSN 0272–1716
- [125] TUMBLIN, J. ; TURK, G. : LCIS: A Boundary Hierarchy for Detail-preserving Contrast Reduction. In: *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press/Addison-Wesley Publishing Co. (SIGGRAPH '99). – ISBN 0–201–48560–5, 83–90
- [126] UHD ALLIANCE: *UHD Premium*. <https://www.experienceuhd.com/uhd-premium-features>. zuletzt abgerufen am: 02.11.2020
- [127] VIDEO CODING (JCT-VC), J. C. T.: *HM Reference Software*. https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.2+SCM-3.0/. zuletzt abgerufen am: 02.11.2020
- [128] WANDELL, B. A.: *Foundations of Vision*. Stanford University, 1995
- [129] WANLASS, F. ; SAH, C.-T. : Nanowatt logic using field-effect metal-oxide semiconductor triodes. In: *Digest of Technical Papers* Bd. 6, IEEE International Solid-State Circuits Conference, 1963
- [130] WEBER, K. : *New Trends and Implementations in Imaging Technology for the Future of Live Production / GrassValley*. 2013. – White Paper
- [131] WU, H. R. ; RAO, K. R.: *Digital Video Image Quality and Perceptual Coding (Signal Processing and Communications)*. Boca Raton, FL, USA : CRC Press, Inc., 2005. – ISBN 0824727770
- [132] ZHAO, H. ; JIN, X. ; SHEN, J. : Real-Time Tone Mapping for High-Resolution HDR Images. In: *International Conference on Cyberworlds 2008, Hangzhou, China, 22-24 September 2008, Proceedings*, 256–262