



# Image quality assessment of High Dynamic Range and Wide Color Gamut images

Maxime Rousselot

## ► To cite this version:

Maxime Rousselot. Image quality assessment of High Dynamic Range and Wide Color Gamut images. Image Processing [eess.IV]. Université de Rennes, 2019. English. NNT : 2019REN1S034 . tel-02378332

**HAL Id: tel-02378332**

**<https://theses.hal.science/tel-02378332>**

Submitted on 25 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1  
COMUE UNIVERSITE BRETAGNE LOIRE

Ecole Doctorale N°601  
*Mathématique et Sciences et Technologies  
de l'Information et de la Communication*  
Spécialité: Signal, Image, Visions  
Par

« **Maxime ROUSSELOT** »

« **Image quality assessment of High Dynamic Range and Wide Color Gamut images** »

«Impact of the new TV standards on the quality perception»

**Thèse présentée et soutenue à RENNES , le 20 Septembre 2019**  
**Préparée à l'IRISA - Rennes et à Harmonic Inc.**

## **Rapporteurs avant soutenance :**

Céline LOSCOS, Professeure des universités, Université de Reims-Champagne-Ardenne

Daniel MENEVEAUX, Professeur des universités, Université de Poitiers

## **Composition du jury :**

Président : Eric MARCHAND, Professeur des universités, Université de Rennes 1

Examineurs : Céline LOSCOS, Professeure des universités, Université de Reims-Champagne-Ardenne

Daniel MENEVEAUX, Professeur des universités, Université de Poitiers

Christophe RENAUD, Professeur des universités, Université Littorale Côte d'opale

Eric MARCHAND, Professeur des universités, Université de Rennes 1

Rémi COZOT, Maître de conférences, Université de Rennes 1

Olivier LE MEUR, Maître de conférences, Université de Rennes 1

Dir. de thèse : Olivier LE MEUR Maître de conférences, Université de Rennes 1

Co-dir. de thèse : Rémi COZOT Maître de conférences, Université de Rennes 1

Invité(s)

Xavier DUCLOUX, Innovation Manager, Harmonic Inc.

# ACKNOWLEDGEMENT

---

I would like to thank the Harmonic Innovation Team for welcoming me during these three years and for giving me this opportunity.

I would like to thank Olivier Le Meur, Rémi Cozot, Xavier Ducloux and Eric Auffret for accompanying me throughout the thesis.

Special thanks to all the participants of my experiences. Without you, my thesis would not be what it is.

And finally, many thanks to my friends and family who have supported me. It would have been harder without you.

# TABLE OF CONTENTS

---

<b>Résumé en Français</b>	<b>8</b>
<b>Introduction</b>	<b>15</b>
<b>I Background and state of the art</b>	<b>21</b>
<b>Part I: Introduction</b>	<b>22</b>
<b>1 Ecosystem of HDR and WCG</b>	<b>23</b>
1.1 Introduction . . . . .	23
1.2 Definitions . . . . .	23
1.2.1 Luminance and Dynamic Range . . . . .	23
1.2.2 High Dynamic Range . . . . .	24
1.2.3 Color and Gamut . . . . .	26
1.2.4 Wide Color Gamut . . . . .	27
1.3 Image representations . . . . .	28
1.3.1 Perceptually linearized luminance . . . . .	29
1.3.2 Perceptually uniform color space . . . . .	34
1.4 HDR/WCG content compression . . . . .	39
1.4.1 Pre-encoding process . . . . .	39
1.4.2 Compression: general principle . . . . .	42
1.4.3 Adaptation of codec to HDR/WCG content . . . . .	44
1.5 Conclusion . . . . .	45
<b>2 Image quality assessment</b>	<b>46</b>
2.1 Introduction . . . . .	46
2.2 Subjective methods . . . . .	46
2.3 Available HDR image databases with subjective scores . . . . .	48
2.3.1 Image description indexes . . . . .	48
2.3.2 HDR Databases presentation . . . . .	50
2.3.3 Narwaria et al. . . . .	51
2.3.4 Korshunov et al. . . . .	52

## TABLE OF CONTENTS

---

2.3.5	Zerman et al. . . . .	52
2.3.6	Advantages and Drawbacks . . . . .	53
2.4	Image quality assessment: objective methods . . . . .	54
2.4.1	Adapting SDR image quality assessment . . . . .	55
2.4.2	Metrics designed for HDR content . . . . .	57
2.5	Assessing objective metric performances . . . . .	61
2.6	Conclusion . . . . .	62
<b>Part I: Conclusion</b>		<b>63</b>
<b>II Building HDR/WCG experimental environment and analysis</b>		<b>65</b>
<b>Part II: Introduction</b>		<b>66</b>
<b>3 New databases presentation</b>		<b>67</b>
3.1	Introduction . . . . .	67
3.2	Experimental protocol . . . . .	67
3.3	First database: HDdtb . . . . .	70
3.3.1	Images descriptions . . . . .	70
3.3.2	Distortion descriptions . . . . .	75
3.3.3	Subjective quality analysis . . . . .	78
3.4	Second database: 4Kdtb . . . . .	82
3.4.1	Images description . . . . .	82
3.4.2	Distortion descriptions . . . . .	87
3.4.3	Subjective quality analysis . . . . .	88
3.5	Conclusion . . . . .	90
<b>4 Analysis of existing metrics</b>		<b>92</b>
4.1	Introduction . . . . .	92
4.2	HDR-VDP-2 calibration . . . . .	92
4.2.1	Sensitivity to the screen spectral emission . . . . .	93
4.2.2	Sensitivity to the surround luminance . . . . .	96
4.2.3	Sensitivity to the angular resolution . . . . .	97
4.2.4	Conclusion on HDR-VDP-2 parameters . . . . .	99
4.3	Benchmark of existing HDR objective quality metrics . . . . .	99
4.3.1	The particular case of 4Kdtb images . . . . .	99
4.3.2	Existing metric performances . . . . .	100
4.3.3	Chrominance distortion sensitivity . . . . .	101

4.4 Conclusion . . . . .	102
<b>Part II: Conclusion</b>	<b>104</b>
<b>III New color metrics for HDR/WCG image quality assessment</b>	<b>105</b>
<b>Part III: Introduction</b>	<b>106</b>
<b>5 SDR metric adaptation</b>	<b>107</b>
5.1 Introduction . . . . .	107
5.2 Proposed method . . . . .	107
5.2.1 Color Space Conversion . . . . .	108
5.2.2 Remapping Function . . . . .	108
5.2.3 Evaluation of tested SDR metrics . . . . .	110
5.3 Results . . . . .	110
5.3.1 4Kdtb Database . . . . .	110
5.3.2 Zerman et al. Database . . . . .	111
5.3.3 HDdtb Database . . . . .	112
5.3.4 Korshunov et al. Database . . . . .	113
5.3.5 Narwaria et al. Database . . . . .	113
5.3.6 Results Summary . . . . .	114
5.4 Results Discussions . . . . .	114
5.4.1 Impact of the Diffuse White Luminance . . . . .	115
5.4.2 Sensibility to Chrominance Distortions . . . . .	117
5.5 Applications . . . . .	118
5.6 Conclusion . . . . .	119
<b>6 HDR-VDP-2 color extension</b>	<b>120</b>
6.1 Introduction . . . . .	120
6.2 HDR-VDP-2 color extension . . . . .	120
6.3 Training methodology . . . . .	123
6.3.1 Defining the optimisation problem . . . . .	123
6.3.2 Training databases selection . . . . .	123
6.3.3 Combining Databases . . . . .	125
6.3.4 Validation protocol . . . . .	126
6.4 Results . . . . .	127
6.4.1 Metric performances . . . . .	127
6.4.2 Analysis of trained weights . . . . .	129

## TABLE OF CONTENTS

---

6.4.3	Discussion . . . . .	131
6.5	Conclusion . . . . .	132
<b>7</b>	<b>Quality metric aggregation</b>	<b>133</b>
7.1	Introduction . . . . .	133
7.2	Proposed method . . . . .	133
7.2.1	Considered HDR quality metrics . . . . .	134
7.2.2	Chromatic-based visual features . . . . .	134
7.2.3	Image spatial information . . . . .	137
7.2.4	Mapping the features to the quality scores . . . . .	137
7.3	Training methodology . . . . .	137
7.3.1	Combining several databases . . . . .	137
7.3.2	Support Vector Regression . . . . .	138
7.4	Results . . . . .	139
7.4.1	Proposed metric performance . . . . .	139
7.4.2	Validation on an independent database . . . . .	140
7.4.3	Sensitivity to color distortions . . . . .	141
7.5	Conclusion . . . . .	142
	<b>Part III: Conclusion</b>	<b>143</b>
	<b>Conclusion</b>	<b>145</b>
	<b>Appendices</b>	<b>149</b>
<b>A</b>	<b>SDR Metrics description</b>	<b>150</b>
A.1	PSNR . . . . .	150
A.2	S-CIELab . . . . .	150
A.3	SSIM . . . . .	151
A.3.1	Color-blind SSIM: . . . . .	151
A.3.2	SSIM for color images (SSIMc): . . . . .	153
A.3.3	CSSIM . . . . .	153
A.4	Multiscale-SSIM (MS-SSIM) . . . . .	153
A.5	FSIM . . . . .	154
A.5.1	Color-blind FSIM . . . . .	154
A.5.2	FSIM for color images (FSIMc): . . . . .	155
A.6	PSNR-HVS and PSNR-HVS-M . . . . .	155
A.7	PSNR-HMA . . . . .	157
A.7.1	Color-blind PSNR-HMA . . . . .	157

A.7.2 PSNR-HMA for color images (PSNR-HMAc): . . . . .	158
<b>B Database Images</b>	<b>159</b>
B.1 Narwaria et al. . . . .	159
B.2 Zerman et al. . . . .	160
B.3 Korshunov et al. . . . .	161
B.4 HDdtb . . . . .	162
B.5 4Kdtb . . . . .	163
<b>C Performance Indexes</b>	<b>164</b>
C.1 PCC . . . . .	165
C.2 SROCC . . . . .	167
C.3 OR . . . . .	169
C.4 RMSE . . . . .	171
<b>D Metrics Sensitivity on the Chrominance Artifacts with the 4Kdtb</b>	<b>173</b>
<b>List of Abbreviations</b>	<b>175</b>
<b>List of Figures</b>	<b>178</b>
<b>List of Tables</b>	<b>183</b>
<b>Bibliography</b>	<b>185</b>

# RÉSUMÉ EN FRANÇAIS

---

## Contexte

L'industrie de la vidéo s'est toujours efforcée de fournir de meilleures images à ses utilisateurs et donc d'accroître leur qualité d'expérience. Les innovations sur les technologies de caméra et d'affichage permettent à cette industrie d'augmenter progressivement le nombre de pixels, de la télévision à la définition standard ( $720 \times 576$  ou  $720 \times 480$  pixels) à la télévision haute définition ( $1920 \times 1080$  pixels) et jusqu'à la télévision Ultra Haute Définition (UHD) ( $3840 \times 2160$  pixels).

Le nombre de pixels n'est pas le seul paramètre permettant d'améliorer l'expérience de la télévision et du cinéma. Pour améliorer la qualité d'expérience utilisateur, il est aussi possible d'augmenter la fréquence des images d'une vidéo (High Frame Rate (HFR)). Une autre approche intéressante consiste à améliorer la qualité de chaque pixel en utilisant une plage de couleurs étendue (Wide Color Gamut (WCG)) et une plage dynamique élevée (High dynamic Range (HDR)). WCG étend la plage de couleurs de la représentation d'image existante et le HDR, la plage de luminance (la quantité de lumière) qu'un pixel peut prendre. Une image HDR peut contenir des zones très claires tout en conservant les détails dans les zones sombres. Les technologies HDR et WCG permettent de restituer des images présentant presque les mêmes caractéristiques que dans le monde réel, ce qui les rend plus réalistes.

La création de contenus HDR et WCG est rendue possible en modifiant l'ensemble de la chaîne de production, de la caméra à l'affichage. Tout d'abord, la caméra doit être capable de capturer une large plage de dynamique. Une méthode populaire, introduite pour la première fois par Mann en 1993 [1], consiste à combiner plusieurs images avec différentes expositions, augmentant ainsi la plage dynamique des images et dépassant les limites des capteurs de caméra. Un exemple de caméra vidéo HDR/WCG est la caméra Arri Alexa [2]. Les écrans doivent également pouvoir afficher des images avec la même gamme de couleurs et de luminance. Plusieurs technologies ont été développées au fil des ans, telles que les écrans LCD à LED avec atténuation locale du rétroéclairage [3] ou les écrans utilisant des technologies OLED haut de gamme [4].

Une fois que les dispositifs pour capturer et afficher des vidéos HDR/WCG existent, nous devons également adapter la chaîne de transmission. HDR et WCG ont suscité un vif intérêt au cours de la dernière décennie et les organisations de normalisation ont proposé plusieurs nouvelles normes pour ajuster l'étalonnage [5], l'encodage [6] et la compression des images

[7]. C'est toute la chaîne de transmission qui doit être adaptée et évaluée, en particulier les méthodes de compression.

## Motivation

La compression est l'un des algorithmes les plus critiques pour pouvoir transmettre efficacement des images et des vidéos. En effet, son objectif est de réduire la taille des images et des vidéos et donc la bande passante nécessaire à leur transmission. Dans son livre blanc, Cisco [8] indique que le contenu vidéo représentait déjà 75% du trafic Internet mondial (77 ExaByte par mois) en 2018 et qu'il atteindrait 82% du trafic Internet d'ici 2022 (240 ExaByte par mois). Cependant, la compression est le traitement vidéo qui crée la distorsion la plus élevée et qui affecte le plus la qualité des images et des vidéos. La mesure automatique de la qualité des images et des vidéos compressées est un domaine de recherche clé pour fournir le meilleur compromis entre qualité vidéo et coûts de transmission.

Plusieurs solutions existent déjà pour évaluer la qualité perçue des images et vidéos HDR. Cependant, aucune de celles-ci ne prend en compte les images/vidéos WCG. Cela est dû au fait que le seul écran HDR professionnel abordable disponible à ce moment-là, le SIM2 HDR47ES4MB, n'a pas été en mesure d'augmenter la gamme de couleurs affichables par rapport à un écran standard et a encore diminué cette plage [9].

L'ajout du WCG dans les images et contenus vidéos nous amène à nous poser la question : ***"Est-ce que le WCG crée des artefacts chromatiques plus forts et spécifiques qui peuvent affecter la perception de la qualité?"*** Si c'est le cas et connaissant les limites des métriques existantes, la question qui en découle est : ***"Comment évaluer avec précision la qualité des images/vidéos HDR ET WCG?"***.

Répondre à ces questions peut être décomposé en deux sous-problèmes:

- L'évaluation de la qualité des images/vidéos HDR/WCG compressées peut être réalisée à l'aide d'études d'utilisateurs (ou de tests subjectifs): un panel de téléspectateurs attribue un score aux images en fonction de la qualité perçue. Si, dans la littérature, il existe déjà des bases de données d'images HDR, aucune d'entre elles n'utilise le contenu de WCG. Ainsi, les images déformées présentent moins d'artefacts chromatiques que les images codées dans un espace colorimétrique plus large. Il convient également de noter que MPEG, dans ses recommandations de compression, a identifié des artefacts chromatiques plus susceptibles d'apparaître lors de l'utilisation d'une représentation WCG [7]. Par conséquent, il est nécessaire de créer des bases de données d'images dans lesquelles les images présentent à la fois les caractéristiques HDR et WCG. En outre, ces bases de données devraient contenir des scénarios de compression réalistes,

car elles pourraient servir à évaluer les propositions MPEG. Ces bases de données peuvent également être utilisées pour évaluer la pertinence des métriques de qualité d'image existantes pour les contenus HDR/WCG et pour évaluer la robustesse de ces métriques aux artefacts chromatiques.

- Un test subjectif est la méthode la plus fiable pour évaluer la qualité des images. Cependant, effectuer des tests subjectifs prend beaucoup de temps et n'est pas toujours réalisable. L'évaluation automatique des images par des mesures objectives permettant d'émuler des tests subjectifs est souvent nécessaire pour optimiser les traitements et le sera de plus en plus avec le développement des techniques d'intelligence artificielle. Dans la littérature, plusieurs mesures ont été proposées pour évaluer les images/vidéos HDR telles que les mesures HDR-VDP-2, HDR-VQM ou PU. Cependant, ces métriques n'évaluent pas la distorsion chromatique mais uniquement la distorsion de luminance. Ceci est compréhensible car ils ont été évalués sur des images HDR uniquement. Cependant, avec les images/vidéos WCG/HDR, de nouveaux artefacts chromatiques peuvent apparaître. La création d'une métrique sensible à ces artefacts comblera une lacune existante pour l'évaluation du contenu HDR/WCG.

## Contributions

Les contributions suivantes sont présentées dans cette thèse:

- Nous proposons une première base de données d'images annotées avec des scores subjectifs, HDdtb, présentés dans le chapitre 3. Cette base de données est composée de HDR avec un gamut standard encapsulé dans un espace de couleur WCG. Nous nous concentrons sur les distorsions chromatiques pour créer cette base de données. Cette base de données est disponible à l'adresse [www-percept.irisa.fr/software/](http://www-percept.irisa.fr/software/).
- Nous utilisons cette base de données proposée pour effectuer une analyse complète de l'étalonnage de la métrique HDR-VDP-2. Cette mesure est l'une des mesures les plus précises pour évaluer la qualité du HDR et est souvent considérée comme une mesure de référence. Cette métrique modélise avec précision les premiers éléments du système visuel humain et nécessite donc beaucoup de paramètres pour être calibrée. Par exemple, cette métrique a besoin de l'émission spectrale de l'affichage, de la résolution angulaire et de la luminance ambiante. Le fait de disposer de bases de données présentant des caractéristiques différentes (par exemple différents affichages utilisant différentes technologies) augmente la pertinence de notre analyse. Cette contribution est expliquée en détail dans la publication suivante:

Maxime Rousselot, Éric Auffret, Xavier Ducloux, Olivier Le Meur, and Rémi Cozot, "Impacts of Viewing Conditions on HDR-VDP2", in: *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1442-1446, DOI:10.23919/EUSIPCO.2018.8553212

- Nous avons créé une nouvelle base de données composée d'images nativement WCG, 4Kdtb, également présentées dans le chapitre 3. Les images de cette base de données fournissent des artefacts chromatiques encore plus importants que la base de données précédente. Cette base de données est également disponible à l'adresse [www-percept.irisa.fr/software/](http://www-percept.irisa.fr/software/).
- Nous proposons également une nouvelle méthode pour adapter les métriques SDR basées sur la luminance uniquement et les métriques couleurs au contenu HDR/WCG en utilisant un espace colorimétrique HDR/WCG perceptuellement uniforme. Nous utilisons cette méthode et examinons les performances de 13 métriques SDR avec quatre espaces colorimétriques différents. Les résultats sont présentés dans le chapitre 5. L'article suivant présente cette méthode, les bases de données utilisées et notre analyse sur les performances des métriques.

Maxime Rousselot, Xavier Ducloux, Olivier Le Meur, and Rémi Cozot, "Quality Assessment of HDR/WCG Images Using HDR Uniform Color Spaces", in: *Journal of Imaging 5.1*, 2019, DOI:10.3390/jimaging5010018

Une partie de ces résultats a également été présentée à la conférence suivante:

Maxime Rousselot, Éric Auffret, Xavier Ducloux, Olivier Le Meur, and Rémi Cozot, "Adapting HDR images using uniform color space for SDR quality metrics", in: *COmpression et REprésentation des Signaux Audiovisuels (CORESA)*, 2018

- De plus, nous proposons notre propre métrique de qualité de référence complète, disponible sur [www-percept.irisa.fr/software/](http://www-percept.irisa.fr/software/). Cette métrique est adaptée au contenu HDR et WCG et est sensible aux distorsions chromatiques. La métrique proposée est basée sur deux métriques de qualité HDR existantes et des caractéristiques d'image couleur. Une régression par SVM est utilisée pour combiner les caractéristiques susmentionnées. Cette contribution est présentée en détail dans la publication suivante:

Maxime Rousselot, Xavier Ducloux, Olivier Le Meur, and Rémi Cozot,  
"Quality metric aggregation for HDR/WCG images", *in: 26th IEEE International Conference on Image Processing (ICIP)*, 2019

## Structure de la thèse

Le manuscrit de thèse est organisé en trois parties. La première partie présente en deux chapitres l'état de l'art concernant l'imagerie HDR/WCG et les bases de l'évaluation de la qualité d'image. La deuxième partie présente les bases de données d'images. Une troisième partie présente nos contributions à l'évaluation objective de la qualité d'image.

### Partie I: Background et état de l'art

- **Chapitre 1 L'environnement HDR et WCG:** Ce chapitre présente les concepts fondamentaux du HDR et du WCG et de la modification de la chaîne de transmission qu'ils impliquent. En effet, les technologies HDR et WCG étendent l'espace colorimétrique et la luminance des images/vidéos. La conséquence est que toute la chaîne de transmission d'images créée pour les formats et les représentations actuelles doit être adaptée. Par exemple, de nouveaux espaces colorimétriques perceptuellement uniforme doivent être créés. En effet, le système visuel humain ne perçoit pas la luminance proportionnellement à sa valeur physique, mais a plutôt une vision non-linéaire. Dans l'imagerie SDR (Standard Dynamic Range), un moyen classique d'obtenir un espace colorimétrique perceptuellement uniforme consiste à utiliser des fonctions appelées gamma. Cependant, la luminance des images SDR n'est pas uniformément perceptible en dehors de la plage pour laquelle la fonction gamma a été définie. Pour une plage de luminance plus élevée, de nouvelles fonctions doivent être créées. L'espace colorimétrique perceptuellement uniforme n'est pas le seul aspect nécessitant une adaptation. Le pré-traitement vidéo, y compris le sous-échantillonnage de la chrominance et la compression image/vidéo, doivent également être adaptés aux nouveaux défis liés au contenu HDR/WCG.
- **Chapitre 2 Evaluation de la qualité d'image:** Le deuxième chapitre présente tous les aspects de l'évaluation de la qualité d'image (IQA). IQA est nécessaire pour comparer différents encodages et les technologies de compression. Ce chapitre présente les deux principales catégories de méthodes d'IQA: subjective et objective. Les méthodes subjectives constituent un moyen plus robuste d'obtenir des scores de qualité d'image. Il consiste en des tests standardisés dans le cadre desquels un panel de téléspectateurs note l'image en fonction de la qualité perçue. Cependant, les tests subjectifs sont longs et

coûteux. Les métriques de qualité objective sont des algorithmes qui cherchent à émuler des tests subjectifs et évaluent automatiquement la qualité d'image sans la présence d'un humain. Cependant, dans l'état de l'art, les métriques existantes ont été adaptées pour évaluer la qualité des images SDR et non des images avec une plage de luminance plus élevée. Pour ce type de contenu, de nouvelles méthodes ont dû être inventées pour adapter les métriques SDR au HDR (PU-métriques [10]) ou pour créer de nouvelles métriques basées sur les caractéristiques HVS (HDR-VDP-2). Nous concluons ce chapitre sur les méthodes utilisées pour évaluer la performance des métriques objectives à l'aide de tests subjectifs. Il convient également de noter que, à notre connaissance, aucune métrique HDR ne prend en compte les distorsions chromatiques.

## Partie II: Création d'un environnement expérimental et d'analyse HDR/WCG

- **Chapitre 3: Présentation des nouvelles bases de données:** Dans ce chapitre, nous présentons les deux tests subjectifs que nous effectuons avec du contenu HDR/WCG. L'une contient des images avec un gamut standard, mais encapsulées dans un espace de couleur WCG, l'autre est composée d'images nativement WCG. Pour créer ces bases de données, nous nous concentrons sur les distorsions de chrominance. En effet, les distorsions de luminance sont déjà bien étudiées avec des contenus exclusivement HDR. Nous choisissons de créer des distorsions réalistes en utilisant la compression MPEG. Cela nous permet d'évaluer l'impact de recommandations MPEG sur la compression de la chrominance du contenu HDR/WCG.
- **Chapitre 4: Analyse des métriques de qualité existantes:** Dans ce chapitre, nous passons en revue les performances des métriques existantes à l'aide des bases de données nouvellement créées et ou faisant partie de l'état de l'art. Cela nous a permis de procéder à un examen complet de la calibration de l'une des mesures les plus précises: HDR-VDP-2. De même, nous examinons les performances de plusieurs métriques "PU". En particulier, nous évaluons également leur résilience aux distorsions chromatiques dues à la compression à l'aide de nos bases de données.

## Partie III: Nouvelles métriques couleurs pour l'évaluation de la qualité d'images HDR/WCG

- **Chapitre 5: Adaptation de métriques SDR:** La solution proposée par Aydın et al. [10] pour adapter toute métrique basée luminance aux images HDR consiste à transformer la luminance de HDR en une luminance perceptuellement linéaire. Ainsi, il conserve les caractéristiques des fonctions gamma SDR pour la plage de luminance correspondante

au SDR et garde une luminance perceptuellement linéaire sur la plage complémentaire. Ce chapitre présente notre proposition d'étendre ce travail aux métriques de couleur en utilisant un espace colorimétrique perceptuellement uniforme. Nous passons en revue quatre espaces colorimétriques adaptés au contenu HDR/WCG:  $ICtCp$  [6],  $J_z a_z b_z$  [11] et  $HDR-Lab$  [12] avec un blanc de référence à 100 cd/m<sup>2</sup> et à 1000 cd/m<sup>2</sup>. Nous étudions l'influence du blanc de référence sur les performances de plusieurs métriques pour chaque base de données disponible. De plus, nous analysons l'impact des distorsions chromatiques à l'aide de nos bases de données sur les métriques de qualité basées luminance et basées couleurs. Enfin, nous formulons nos recommandations pour évaluer la qualité des images à l'aide de ces métriques.

- **Chapitre 6: Extension couleur de HDR-VDP-2:** HDR-VDP-2 peut être considéré comme la mesure de référence en ce qui concerne l'évaluation de la qualité d'image HDR. Cependant, il s'agit d'une métrique basée sur la luminance uniquement. Dans ce chapitre, nous décrivons notre proposition pour étendre cette métrique aux distorsions chromatiques. Nous profitons du fait que cette mesure modélise les premières étapes de la vision et en particulier la quantité de lumière qui tombe sur les cônes L, M et S. De ces informations, nous déduisons deux canaux chromatiques pour compléter la métrique HDR-VDP-2. Nous avons réussi à obtenir des performances similaires pour la métrique HDR-VDP-2 sur nos bases de données. Cependant, nous ne sommes pas parvenus à accroître son efficacité, notamment vis-à-vis des distorsions chromatiques.
- **Chapitre 7: Fusion de métriques de qualité:** Dans ce chapitre, nous détaillons notre proposition de combiner plusieurs métriques de qualité et caractéristiques d'image, notamment des fonctionnalités chromatiques. Nous utilisons un SVM pour apprendre le meilleur moyen de combiner ces métriques. Nous entraînons cette nouvelle métrique en utilisant les bases de données à notre disposition. Avec cette méthode, nous parvenons à créer une métrique sensible aux distorsions de couleur.

# INTRODUCTION

---

## Context

The video industry has always strived to provide better images and therefore increase their quality of experience. The innovations on captures and display technologies allow the industry to gradually increase the number of pixels, from Standard Definition Television ( $720 \times 576$  or  $720 \times 480$  pixels) to High Definition Television ( $1920 \times 1080$  pixels) and then, to the Ultra High Definition (UHD) television ( $3840 \times 2160$  pixels). But the number of pixels is not the only parameter to improve the television and cinema experience. Increasing the frequency of consecutive images can be another way to improve the quality of experience (High Frame Rate (HFR)). Another interesting approach is to enhance the quality of each individual pixels using Wide Color Gamut (WCG) and High Dynamic Range (HDR). WCG extends the color range of legacy image representation and HDR extends the luminance range (the quantity of light) a pixel can take. An HDR image can contain very bright areas while retaining details in dark areas. HDR and WCG technologies allow to render images with almost the same characteristics as in the real world, making them more realistic.

HDR and WCG images are made possible by modifying the whole chain of production, from the camera to the display. First, the camera has to be able to capture a large range of dynamic. A popular method, first introduced by Mann in 1993 [1], is to combine several images with different expositions, increasing, therefore, the dynamic range of images and overcoming the limits of camera sensors. One example of HDR/WCG video camera is the Arri Alexa camera [2]. Screens also need to be able to display images with a large range of color and luminance. Several technologies were developed over the years like Led-LCD screens with local backlight dimming [3] or screens using high-end OLED technologies [4].

Once we have devices to capture and display HDR/WCG videos, we need also to adapt the transmission workflow. HDR and WCG generated a lot of interest in the last decade and standardization organizations proposed several new standards to adjust the grading [5], the encoding representation [6] and the compression of images [7]. This is the whole chain that needs to be adapted and evaluated, especially the compression methods.

## Motivations

Compression is one of the most critical algorithms to be able to transfer efficiently images and videos. Indeed, its goal is to reduce the size of images and videos and therefore the bandwidth of their transmission. In a white paper, Cisco [8] states that video content already accounted for 75% of the global internet traffic (77 ExaByte per month) in 2018 and forecasts that it will reach 82% of internet traffic by 2022 (240 ExaByte per month). However, compression is the video processing that creates the highest distortion and that affects the most the quality of images and videos. Assessing the quality of compressed images and videos is a key factor to provide the best compromise between video quality and transmission costs.

Several solutions already exist to assess the perceived quality of HDR images and videos. However, none of those takes into account WCG images/videos. This is due to the fact that the only affordable professional HDR display available at that time, the SIM2 HDR47ES4MB, is not able to increase the range of displayable colors compared to a standard screen and has a smaller range [9].

Introducing WCG in images and video contents leads us to raise the following question: ***"Does WCG create stronger and specific chromatic artifacts which can affect the perception of quality?"*** If this is the case and knowing the limitations of existing metrics, the legitimate resulting question is: ***"How accurately assessing the quality of HDR AND WCG images/videos?"***.

Answering these questions can be decomposed into two smaller problems:

- Evaluating the quality of compressed HDR/WCG images/videos can be performed by user studies (or subjective tests): a panel of viewers score images based on the perceived quality. If in the literature, there are already HDR images databases, none of them uses WCG contents. Thus, the distorted images present less chromatic artifacts than images encoded in a wider color space. It should also be noted that MPEG, in its compression recommendations, has identified chromatic artifacts that are more prone to appear when using a WCG representation [7]. Therefore, there is a need to create image databases in which images exhibit both HDR and WCG characteristics. In addition, those databases should contain realistic compression scenarios as it could serve to evaluate MPEG propositions. Those databases can also be used to evaluate the pertinence of existing image quality metrics for HDR/WCG contents and evaluate the robustness of those metrics to chromatic artifacts.
- A subjective test is the most reliable method to assess the quality of images. However, performing subjective tests is very time-consuming and not always practicable. The automatic evaluation of images quality thanks to objective metrics which emulate subjective tests is often used in algorithm research and will be more and more unavoidable with the

development of artificial intelligence techniques. In the literature, several metrics were proposed to evaluate HDR images/videos like HDR-VDP-2, HDR-VQM or PU-metrics. However, those metrics do not evaluate chromatic distortion but only luminance distortion. However, with WCG/HDR images/videos, new chromatic artifacts can appear. Creating a metric sensitive to those artifacts will correct a weakness which exists for the assessment of HDR/WCG contents.

## Contributions

The following contributions are presented in this thesis:

- We propose a first image database annotated with subjective scores, HDdtb, presented in Chapter 3. It is composed of HDR but not WCG contents encapsulated in a WCG color space. We focus on chromatic distortions to create the distorted images to evaluate. This database is available at [www-percept.irisa.fr/software/](http://www-percept.irisa.fr/software/).
- We use this proposed and already available databases to perform a complete analysis on the calibration of the metric HDR-VDP-2, a reference metric to perform image quality assessment. This metric models precisely the early-stage of the human visual system and therefore takes a lot of parameters to be calibrated. For example, this metric needs the display spectral emission, the angular resolution and the surround luminance. Having databases with different characteristics (for example different displays using different technologies) increase the pertinence of our analysis. This contribution is explained in details in the following paper:

Maxime Rousselot, Éric Auffret, Xavier Ducloux, Olivier Le Meur, and Rémi Cozot, "Impacts of Viewing Conditions on HDR-VDP2", in: *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1442-1446, DOI:10.23919/EUSIPCO.2018.8553212

- We create a new database composed of native WCG images, 4Kdtb, also presented in chapter 3. The distorted images provide even more prominent chromatic artifacts than the previous database. This database is also available at [www-percept.irisa.fr/software/](http://www-percept.irisa.fr/software/).
- We propose a new method to adapt color and color-blind SDR metrics to HDR/WCG contents using HDR/WCG perceptually uniform color space. We use this method and review the performances of 13 SDR metrics with four different color spaces. The results are

presented in chapter 5. The following article presents this method, the used databases and our analysis on the performance of the metrics.

Maxime Rousselot, Xavier Ducloux, Olivier Le Meur, and Rémi Cozot, "Quality Assessment of HDR/WCG Images Using HDR Uniform Color Spaces", in: *Journal of Imaging 5.1* , 2019, DOI:10.3390/jimaging5010018

Part of those results were also presented in the following conference:

Maxime Rousselot, Éric Auffret, Xavier Ducloux, Olivier Le Meur, and Rémi Cozot, "Adapting HDR images using uniform color space for SDR quality metrics", in: *COmpression et REprésentation des Signaux Audiovisuels (CORESA)*, 2018

- Additionally, we propose our own full-reference quality metric, available at [www-percept.irisa.fr/software/](http://www-percept.irisa.fr/software/). This metric is adapted to HDR and WCG content and is sensitive to chromatic distortions. The proposed metric is based on two existing HDR quality metrics and color image features. A support vector machine regression is used to combine the aforementioned features. This contribution is explained in detail in the following paper:

Maxime Rousselot, Xavier Ducloux, Olivier Le Meur, and Rémi Cozot, "Quality metric aggregation for HDR/WCG images", in: *26th IEEE International Conference on Image Processing (ICIP)* , 2019

## Structure of the thesis

The thesis manuscript is organized into three parts. The first part, composed of two chapters, presents the state of the art of HDR/WCG imaging and the basics of image quality assessments. The second part presents our proposed image databases annotated with subjective scores. A third part presents our contributions to objective image quality assessment.

### Part I: Background and state of the art

- **Chapter 1 Ecosystem of HDR and WCG:** This chapter gives a presentation of the fundamental concepts behind HDR and WCG and the modification of the chain of transmission

they imply. Indeed, HDR and WCG technologies extend the color space and the luminance of images/videos. The consequence is that the whole chain of image transmission which was created for legacy formats and representations, needs to be adapted. For example, new perceptually uniform color spaces have to be designed. Indeed the human visual system does not perceive luminance proportionally to its physical value but is instead not linear. In Standard Dynamic Range (SDR) imaging, a classical way to obtain a perceptually uniform color space is to use functions called gamma. However, the luminance of SDR images is not perceptually uniform outside the range for which the gamma function was defined. For a higher luminance range, new functions have to be created. The perceptually uniform color space is not the only aspects that require adaptation. The video pre-processing including the chrominance downsampling and the compression of image/video also need to be adapted to new challenges that arise with HDR/WCG contents.

- **Chapter 2 Image quality assessment:** The second chapter presents all the aspects of Image Quality Assessment (IQA). IQA is necessary to compare different encodings and compression technologies. This chapter presents the two main categories of methods to perform IQA: subjective and objective. Subjective methods are a more robust way to obtain image quality scores. It consists of standardized tests where a panel of viewers score image based on their perceived quality. However, subjective tests are time-consuming and costly. Objective quality metrics are algorithms that aim to emulate subjective tests and automatically assess image quality without the presence of a human. However, in the state of the art, metrics were tuned to assess the quality of SDR images and not images with a higher dynamic range. For this new kind of contents, new methods were designed to either adapt SDR metrics to HDR (PU-metrics [10]) or to either create new metrics based on the HVS characteristics (HDR-VDP-2). We conclude this chapter on the methods used to assess the performance of objective metrics using subjective tests. It should also be noted that, up to our knowledge, there is no HDR metric that considers chromatic distortions.

## Part II: Building HDR/WCG experimental environment and analysis

- **Chapter 3: New databases presentation:** In this chapter, we present the two subjective tests we perform on HDR/WCG contents. One contains images with a legacy gamut but encapsulated in a larger gamut, the other one is composed of native WCG images. To create those database we focus on chrominance distortions. Indeed, luminance distortions are already well studied with HDR-only contents. We choose our distortions to be realistic using MPEG compression. It allows us to evaluate the impact of the MPEG

recommendation for the compression of chrominance of HDR/WCG contents.

- **Chapter 4: Analysis of existing metrics:** In this chapter, we review the performances of existing metrics using our newly created and already created databases. This allowed us to perform a complete review of the calibration of one of the most precise metric: HDR-VDP-2. Equally, we review the performance of several PU-metrics. In particular, we also evaluate their resilience to chromatic distortions due to compression using our databases.

### Part III: New color metrics for HDR/WCG image quality assessment

- **Chapter 5: SDR metric Adaptation:** The solution proposed by Aydın et al. [10] to adapt any color-blind metrics to HDR images is to transform the luminance of HDR into a perceptually linear luminance. Thus, it retains the characteristics of the SDR gamma functions for the SDR luminance range and retains perceptual linearity otherwise. This chapter presents our proposition to extend this work to color metrics using perceptually uniform color space. We review four color spaces adapted to HDR/WCG content: *ICtCp* [6], *J<sub>z</sub>a<sub>z</sub>b<sub>z</sub>* [11] and *HDR-Lab* [12] with a diffuse white at 100 cd/m<sup>2</sup> and at 1000 cd/m<sup>2</sup>. We study the influence of the diffuse white on the performance of several metrics on each available database. In addition, we analyse the impact of chromatic distortions using our proposed database on color-blind and color metrics. Finally, we give our recommendations to assess the quality of images using legacy metrics.
- **Chapter 6: HDR-VDP-2 color extension:** HDR-VDP-2 can be considered as the reference metric concerning HDR image quality assessment. However, this is a color-blind metric. In this chapter, we describe our proposition to extend this metric to chromatic distortions. We take advantages that this metric models the early-stage of vision and especially the quantity of light falling on L,M,S cones. From this information, we deduce two chromatic channels to complete the HDR-VDP-2 metric. We succeed to obtain similar performance for the HDR-VDP-2 metric on our available database. However, we do not succeed to increase its efficiency especially towards chromatic distortions.
- **Chapter 7: Quality metric aggregation:** In this chapter, we detail our proposition to combine several image quality metrics and features including chromatic features. We use a support vector machine to learn the best way of combining those metrics. We train this new metric using the available databases. With this method, we succeed to create a metric sensitive to color distortions.

PART I

# **Background and state of the art**

---

# PART I: INTRODUCTION

---

Images and videos are efficient tools to represent the world and to communicate about a captured instant. However, this representation is always limited by technology. The capture and the displays of images impact the characteristics of the light emitted by screens as if it was directly received by our eyes. On one hand, camera sensors are not able to capture all the range of luminance and color perceived by the Human Visual System (HVS). On the other hand, screens are not able currently to reproduce such a variety of color and luminance.

Legacy image format and processing were designed with the limitation of existing technology. The luminance dynamic range, as well as the color gamut that can be represented, are limited due to the technology. For instance, the perceptually uniform color spaces are only uniform for a small portion of the color space covered by the HVS and Image Quality Assessment (IQA) metrics are tailored only for legacy image format.

In recent years, with the evolving technology, the capacity to capture and display images has grown a lot. Recent monitors can now display images with a High Dynamic Range (HDR) and with a Wide Color Gamut (WCG). The legacy format and processing adapted to the Standard Dynamic Range (SDR) are not efficient anymore for those kinds of content. New formats and new representations were created to handle those new pictures. Moreover, new IQA metrics were proposed and were tested on new databases of images in comparison with scores given by panels of viewers.

This part gives, in the first chapter, a presentation about the HDR and WCG image representation and compression. The second chapter is an overview of subjective and objective IQA metric methods for HDR/WCG content. The last section concludes this chapter.

# ECOSYSTEM OF HDR AND WCG

---

## 1.1 Introduction

The design of legacy image representations and compression have always reflected the limits of technologies (capture and display devices). High Dynamic Range and Wide Color Gamut technologies make obsolete those formats by pushing the limits of what is captured and what is displayed. Those technologies allow more freedom than ever to create images. New adapted representations are required and the existing compression codec need to be modified.

The first section of this chapter, We define HDR and WCG main concepts. In the second section, we describe the new image representations that were created to handle the new color spaces. Finally, in the last section, we describe the methods employed to adapt existing compression workflow to the new images color spaces.

## 1.2 Definitions

To fully understand the value of HDR and WCG systems, it is necessary to understand the ability of the HVS system to perceive light. The main concepts of HDR and WCG formats are presented in this section.

### 1.2.1 Luminance and Dynamic Range

The luminance is a photometric measure which describes the amount of light that is perceived by the eye or falls into a camera sensor. It is expressed in  $\text{cd/m}^2$ . For the reference, a moonless night have a luminance about  $10^{-3} \text{ cd/m}^2$ , the luminance of a sunny day is about  $10^6 \text{ cd/m}^2$  and the luminance of the sun can reach a luminance above  $10^9 \text{ cd/m}^2$

The dynamic range of an image corresponds to the range between its brightest and darkest pixels. To be closer to the human perception, it is often expressed on a logarithmic scale express in log units. The Dynamic Range (DR) of an image is then calculated as follows:

$$DR = \log_{10}(Y_{max}) - \log_{10}(Y_{min}) \quad (1.1)$$

Where  $Y_{min}$  and  $Y_{max}$  are the minimum and the maximum luminance in  $\text{cd/m}^2$ , respectively. Note that the available dynamic range of a display is sometimes expressed in f-stop. In this

case, the dynamic range is expressed as follows:

$$DR_{f-stop} = \log_2(Y_{max}) - \log_2(Y_{min}) \quad (1.2)$$

The HVS is adapted to perceive a wide range of luminance going from  $10^{-6}$  cd/m<sup>2</sup> and up to  $10^8$  cd/m<sup>2</sup> [17]. However, the HVS cannot perceive both extremes of the scale at the same time. Indeed, the HVS is adapting to an appropriate light sensitivity due to various photochemical, mechanical and neuronal processes. The luminance subset the HVS is able to operate while being in a state of full adaptation is called the steady-state or simultaneous dynamic range. The HVS has a simultaneous dynamic range around 4 log units [18] [19].

#### Photopic, mesopic and scotopic vision

The human eye is composed of four kinds of photosensors. Three are the cones which are better to perceive luminance during daylight conditions and are responsible for the color perception. The fourth photosensors are the rods, which are adapted to low light conditions. Due to the differences between photoreceptors, the HVS light sensitivity can be separated into three modes of perception. First, the scotopic vision (or night vision) corresponds to vision in which the perception of light is only performed by rods (between  $10^{-6}$  to  $10^{-3}$  cd/m<sup>2</sup>). There is no color perception under such a mode of perception. Then, photopic vision (or daylight vision) corresponds to the vision in which only the cones are receptive (between  $10$  and  $10^8$  cd/m<sup>2</sup>). Finally, the mesopic vision occurs when both cones and rods are sensitive to luminance (between  $10^{-3}$  to  $10$  cd/m<sup>2</sup>).

### 1.2.2 High Dynamic Range

A SDR display is not able to display all luminances visible to the HVS or even to cover the HVS simultaneous dynamic range. For example, the Dell 20 (E2016H) [20], a midrange screen, has a peak luminance at 250 cd/m<sup>2</sup>, a black point luminance at 0.25 cd/m<sup>2</sup> and an available dynamic range of 3 log units.

An HDR screen better takes advantage of the capabilities of the HVS by extending the dynamic range it can display. According to the Ultra HD Forum [21], a display can be considered HDR if it has a dynamic range equal to or greater than 13 f-stops (around 4 log units). The increase of dynamic range leads to luminance values that are closer to real life scenes, thus, making them more realistic. There are several screens that are already able to fulfill the

Ultra HD Forum requirement. We can mention the Dolby Pulsar monitor (not commercially available) which claims to have a peak luminance at  $4000 \text{ cd/m}^2$  and a black point luminance at  $0.005 \text{ cd/m}^2$  [22], the Sony BVM-X300 with a peak luminance at  $1000 \text{ cd/m}^2$  and a black point luminance at  $0.0005 \text{ cd/m}^2$  [23] or the Sim2 HDR47ES6MB with a peak luminance at  $4000 \text{ cd/m}^2$  and a black point luminance at  $0.001 \text{ cd/m}^2$  [24]. The dynamic range of human vision and the previously mentioned displays are illustrated in Figure 1.1.

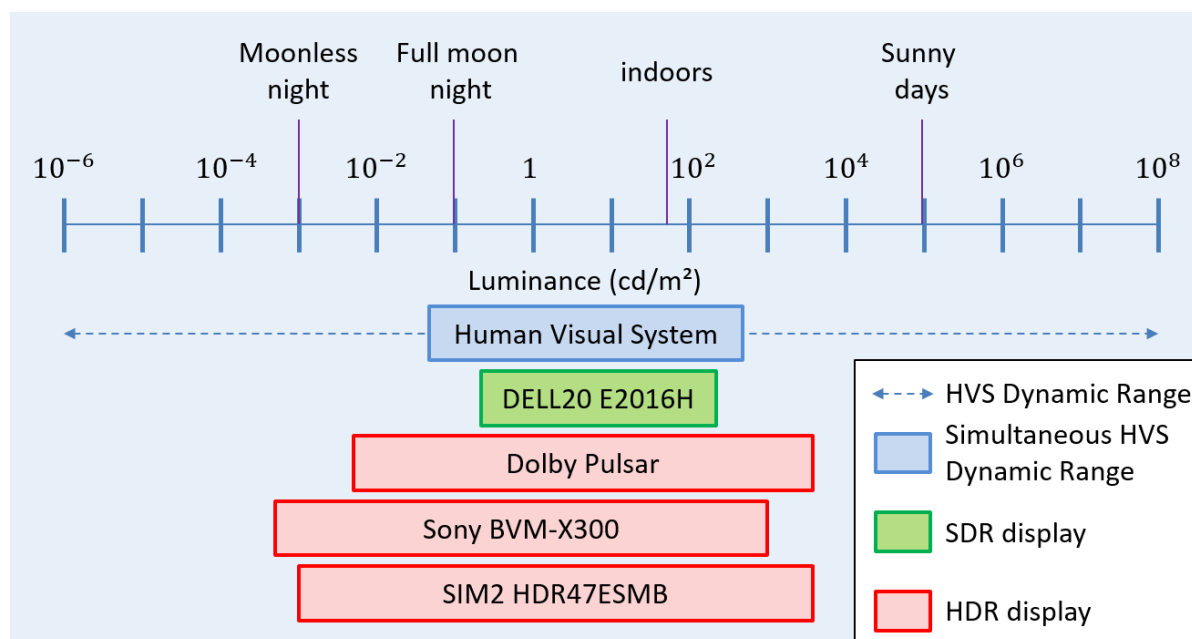


Figure 1.1: Dynamic Range of the Human Vision System and displays

### Diffuse White

The diffuse white of an image is an important characteristic to describe an HDR signal. The diffuse white is defined as "the signal level of 100% Lambertian reflector placed at the center of interest within a scene under controlled lighting". It corresponds to the light coming from a white card without any specular reflections. It should not be confused with the peak luminance as luminance coming from emissive light sources or specular reflections can be higher than the diffuse white luminance. Those luminances are called the highlights. The standard ITU-R BT.2408 recommend a diffuse white at  $203 \text{ cd/m}^2$  [5] if the screen peak luminance is  $1000 \text{ cd/m}^2$ .

The ability to accurately render the highlights is one of the key features to differentiate HDR and SDR. HDR is able to render more accurately the luminance of emissive light sources and specular reflections, thus making the image more realistic. Specular reflections, in particular, improve depth perception [25] [26].

Moreover, HDR displays are not only able to display brighter pixels but also darker pixels. Being able to better render dark regions is an important feature of HDR display. A study found that to meet the preferences of 90% of the viewers, a black point luminance of 0.005 cd/m<sup>2</sup> is needed [27].

The increase of luminance dynamic range improves the perceived quality of images. However, the luminance only describes the achromatic information of the images. Color of pixels does not depend of the luminance. More dimension are necessary to describe the light.

### 1.2.3 Color and Gamut

The luminance gives only the achromatic information of light. The HVS also perceives color. This color perception is performed by three kinds of photoreceptors called cones: *L*, *M* and *S*. Each cone is more sensitive to light with a Long, Medium and Short wavelength respectively. Similarly, most color model represent the color sensitivity of the Human Visual System with three dimensions.

To take advantage of this fact, most capture/display devices use the trichromatic color photo-receptor/transmitter: red, green and blue corresponding to the *L*, *M* and *S* cones. Those three colors are called the color primaries of the device. Thus, such RGB color spaces are a common way to define image pixels. However, as each device might use a different red, green and blue, there is a need for common color spaces to be able to exchange images.

In 1931, the CIE (Comission Internationale de l'éclairage) defined a color space that can represent any visible color: the CIE *XYZ* color space. It is defined as follows:

$$\begin{cases} X = \int_{380}^{780} J(\lambda) \times \bar{x}(\lambda) d\lambda \\ Y = \int_{380}^{780} J(\lambda) \times \bar{y}(\lambda) d\lambda \\ Z = \int_{380}^{780} J(\lambda) \times \bar{z}(\lambda) d\lambda \end{cases} \quad (1.3)$$

where  $J(\lambda)$  is the spectral power distribution of the light and  $\bar{x}$ ,  $\bar{y}$  and  $\bar{z}$  are standard color matching functions defined by the CIE. Those matching functions are akin, although different from, the cones responses. The color matching function  $\bar{y}$  was created so that the *Y* value represents the photopic luminance.

Based on this representation, another color space can be created to separate the luminance

and the chrominance:  $xyY$ .  $Y$  corresponds to the photopic luminance and  $x$  and  $y$  to the chromatic information.  $x$  and  $y$  are calculated as follow:

$$\begin{cases} x = \frac{X}{X + Y + Z} \\ y = \frac{Y}{X + Y + Z} \end{cases} \quad (1.4)$$

$x$  and  $y$  are widely used to define the chromatic information of light sources. It is also in this space that the CIE  $xy$  Chromaticity diagram is represented (cf. Figure 1.2). This diagram represents all the visible Chromaticity. One of the advantages of this representation is that taken two points of color on the Chromaticity diagram, all colors which can be created by mixing these two colors are found in a straight line between the two points. By extension, all colors created mixing three colors can be found in the triangle form by those three color chromaticity coordinates  $x$  and  $y$ .

Due to those characteristics, the  $xy$  coordinates are used to define many RGB color spaces. The triangle formed by the three color primaries chromaticity coordinates represents the RGB color space gamut, i.e. the ensemble of colors a device can handle. Only four points are needed to describe such space: the coordinates for the three primaries, red green and blue and the coordinates of the white point. The white point is the Chromaticity point obtained when each primary is used with the same intensity. This point is often defined as the standard illuminant D65 which coordinates are  $x_{D65} = 0.31271$  and  $y_{D65} = 0.32902$ .

To represent the colors displayable by all screens, a standard RGB color space was created.

#### 1.2.4 Wide Color Gamut

In 1990, for High Definition (HD) image and video, the ITU recommended the use of one particular RGB color space in the standard BT.709 [28]. This color space was created in order to cover the color displayable by a TV screen based on CRT (Cathode Ray Tube) technologies [29]. This gamut is widely used to encode images and videos. Moreover, it is recognized by most compression codec. However, screen technology has evolved and new technology emerged like Led-LCD or OLED. The gamut BT.709 cannot encompass all the color displayable by the last generations of screens. In 2012, the ITU proposed a new standard, BT.2020 [30] which describes a new color space for the development of a new generation of video, named Ultra High Definition (UHD), which could provide a higher user experience, thanks to an increased resolution, frame rate and color gamut. Note that there is, up to our knowledge, no screen that is able to fully display the BT.2020 gamut. However, some of UHD screens, launched in the market during the last years, are really close.

The  $xy$  chromaticity coordinates for the BT.709 and the BT.2020 color spaces can be found in Table 1.1. The gamuts of these color spaces are represented in the CIE  $xy$  in Figure 1.2.

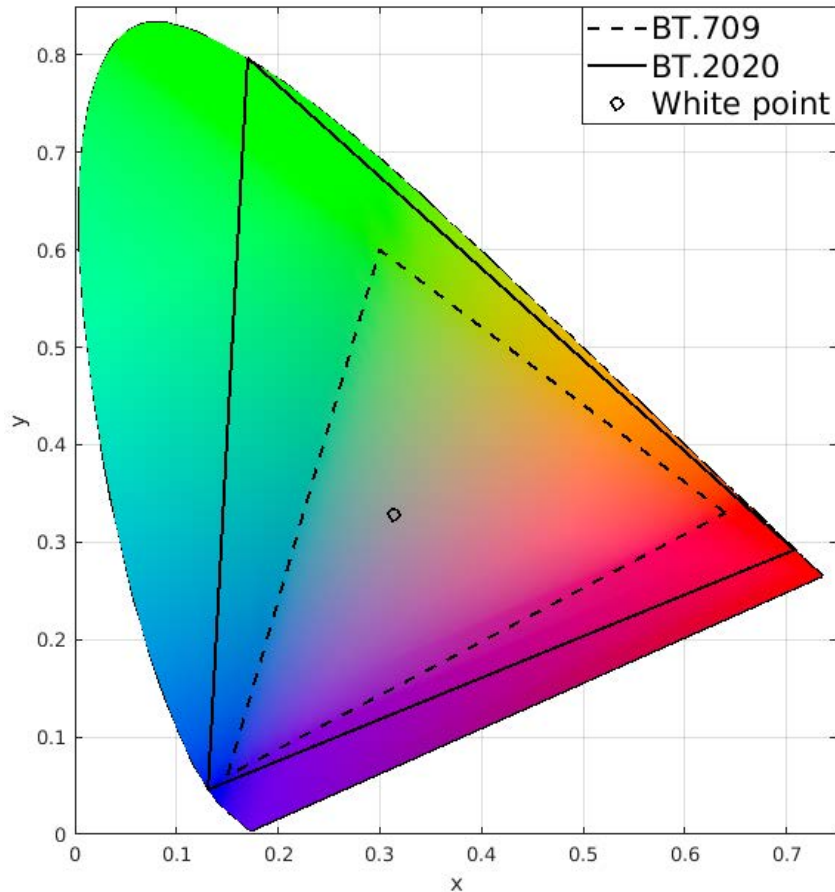


Figure 1.2: gamut of the BT.709 and BT.2020 color space and HVS gamut represented with a CIE  $xy$  chromaticity diagram

The  $XYZ$  color space is useful to circumscribe  $RGB$  color space. However, it is not representative of human perception of light.

### 1.3 Image representations

A uniform (or perceptually uniform) color space is defined so that the difference between two values always corresponds to the same amount of visually perceived change. Those color spaces are essential to optimize the quantization of pixel values or to obtain efficient IQA objective metrics. Indeed, perceptually uniform color spaces better represent the HVS perception of light.

Table 1.1: Chromaticity coordinates for the color primaries and the white point of BT.709 and the BT.2020 color space

	Red		Green		Blue		White	
	x	y	x	y	x	y	x	y
BT.709	0.640	0.330	0.300	0.600	0.150	0.060	0.3127	0.3290
BT.2020	0.708	0.292	0.170	0.797	0.131	0.046	0.3127	0.3290

### 1.3.1 Perceptually linearized luminance

The luminance express in  $\text{cd/m}^2$  is said to be linear. However, the Human visual system does not perceive the luminance proportionally to its value in  $\text{cd/m}^2$ . The human perception of luminance is not linear. This is why in most cases, the linear luminance is transformed using transfer functions. In the following of this manuscript, we differentiate three transfer function [31]:

- **OETF**: Opto-Electronic Transfer Function. This function converts the linear scene light into the image or video signal.
- **EOTF**: The Electro-Optical transfer Function. This function converts the image/video signal into the linear light of the display.
- **OOTF**: Opto-Optical Transfer Function. This is the function of a whole system that converts the linear scene light into the display light. We can differentiate the reference OOTF and the artistic OOTF. The reference OOTF is the concatenation of the OETF and the EOTF. This function is not linear. Indeed, a non-linearity is applied to compensate for differences in tonal perception between the environment of the camera and that of the display. The artistic OOTF also includes artistic adjustment, a modification of the signal to enhance the image or to apply a rendering intent.

Figure 1.3 represents the end-to-end workflow from the capture of light to the display of an image and the relation between transfer functions. In the rest of this section, for the sake of clarity, the artistic adjustments are neglected.

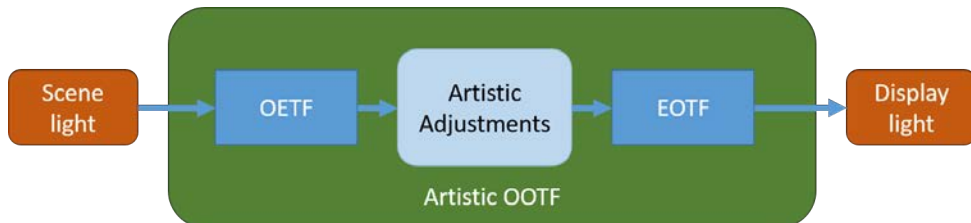


Figure 1.3: End-to-end workflow from the capture of light to the display of an image [31]

## SDR transfer functions

The SDR transfer functions are called the gamma functions. The OETF is defined in the standard BT.709 [28] and the EOTF in the standard BT.1886 [32].

The BT.709 OETF function is defined as follows:

$$\begin{cases} V = 1.099L^{0.45} - 0.099 & \text{if } 1 \geq L \geq 0.018 \\ V = 4.5L & \text{if } 0.018 > L \geq 0 \end{cases} \quad (1.5)$$

where  $L$  is the normalized luminance of the scene (between 0 and 1) and  $V$  is the corresponding image signal.

The BT.1886 EOTF function is defined as follows:

$$F_D = (L_W^{\frac{1}{\gamma}} - L_B^{\frac{1}{\gamma}})^{\gamma} \times (\max[0, \left( V + \frac{L_B^{\frac{1}{\gamma}}}{L_W^{\frac{1}{\gamma}} - L_B^{\frac{1}{\gamma}}} \right)^{\gamma}]^{\gamma}) \quad (1.6)$$

where  $F_D$  is the luminance of the screen in  $\text{cd/m}^2$  and  $V$  the image signal calculated by equation (1.5).  $\gamma$  is the exponent of power function,  $\gamma = 2.40$ .  $L_W$  is the peak luminance of the screen and  $L_B$  is the luminance for its black point.

Due to the way it is defined, it is impossible to know the EOTF and, thus, the OOTF without knowing the final display. Such kind of workflow is called scene-referred as the image signal only depends on the scene luminance characteristics and not on the screen ability to display luminance. In this kind of workflow, the OOTF is considered to be in the display. Figure 1.4 describes a generic end-to-end scene-referred workflow.

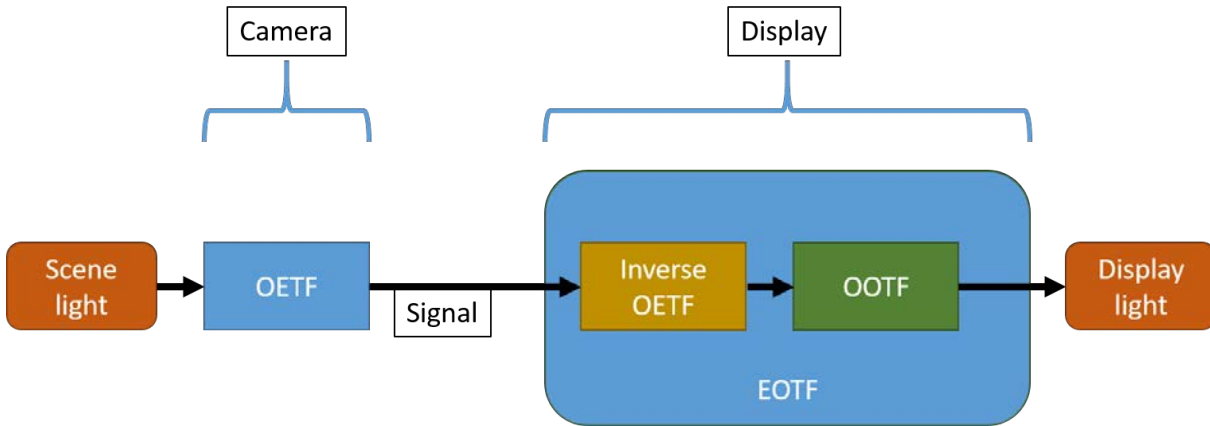


Figure 1.4: End-to-end scene-referred workflow from the capture of light to the display of an image [31]

The BT.709 EOTF and the BT.1886 OETF were created and adapted to screens based on

CRT technology. The EOTF BT.1886 were especially created based on the measured characteristics of CRT screens so that other displays based on different technologies could use images encoded with BT.709 content. Those functions held two advantages. First, the BT.1886 EOTF did not need to be implemented in CRT displays as it is directly the response of such a screen. As CRT technology is no longer the mainstream screen technology, this is no longer an advantage. Second, the EOTF is a rough but sufficient estimation of the HVS perception of luminance in the dynamic range of all HD SDR screens, being CRT, Led/LCD or OLED.

However, new UHD HDR screens are able to display a much wider range of luminance. The BT.709/BT.1886 transfer functions are a rough approximation of the HVS luminance perception in this case. Using those functions and assuming a peak brightness of the display at  $1000\text{cd/m}^2$ , even 12 bits quantization is not enough to avoid banding artifacts, especially in the dark areas [33].

For HDR imaging, two transfer functions were created and normalized in the BT.2100 standard: the Hybrid-Log-Gamma (HLG) transfer functions and the Perceptual Quantization (PQ) Transfer Function.

### Hybrid-Log-Gamma

First, BBC and NHK proposed functions called Hybrid-Log-gamma (HLG) transfer functions normalized in [34]. As the BT.709/BT.1886 system, it is a scene-referred workflow. HLG OETF is defined as follows:

$$\begin{cases} V = a \times \ln(12L - b) + c & \text{if } 1 \geq L \geq \frac{1}{12} \\ V = \sqrt{3L} & \text{if } \frac{1}{12} > L \geq 0 \end{cases} \quad (1.7)$$

where  $L$  is the normalized luminance of the scene (between 0 and 1) and  $V$  is the corresponding image signal. The constants are defined as follow:  $a = 0.17883277$ ,  $b = 1 - 4a$  and  $c = 0.5 - a \times \ln(4a)$ .

As in SDR systems, the reference OOTF of the HLG workflow also presents a  $\gamma$  non-linearity:

$$F_D = OOTF(L) = L_W L^\gamma \quad (1.8)$$

But this time, the factor  $\gamma$  is dependent on the screen peak luminance  $L_W$ :

$$\gamma = 1.2 + 0.42 \times \log(L_W/1000) \quad (1.9)$$

The HLG EOTF can be deduced from the OOTF and the OETF as follows:

$$\begin{aligned} F_D &= \text{EOTF}[\max(0, (1 - \beta)V + \beta)] \\ &= \text{OOTF} \left[ \text{OETF}^{-1}[\max(0, (1 - \beta)V + \beta)] \right] \end{aligned} \quad (1.10)$$

where,  $F_D$  is the luminance of the screen in  $\text{cd/m}^2$  and  $V$  the image signal calculated by equation (1.7).  $\beta$  is dependent on  $L_W$ , the peak luminance of the screen and  $L_B$ , the luminance for its black point:  $\beta = \sqrt[3]{3(L_B/L_W)^{1/\gamma}}$ .

The function  $\text{OETF}^{-1}$  is the exact inverse of the OETF:

$$\begin{cases} L = (\exp((V - c)/a) + b)/12 & \text{if } 1 \geq V \geq \frac{1}{2} \\ L = V^2/3 & \text{if } \frac{1}{2} > V \geq 0 \end{cases} \quad (1.11)$$

where  $L$  is the normalized luminance of the scene and  $V$  is the corresponding image signal. The constant  $a$ ,  $b$  and  $c$  are the same as for the OETF (cf. Equation 1.7).

The HLG model claims to be backward compatible. This means that the images encoded with the HLG OETF could be decoded on SDR television, using the BT.1886 EOTF (equation (1.6)), without any other form of tone-mapping [35]. The main argument of such a system is to reduce the distribution cost of HDR content for television broadcast as it would not require to broadcast two different streams, one HDR for HDR display and one SDR for SDR display. Though, in reality, SDR deployed channels use a BT.709 color space, while HDR channels are associated to the extended BT.2020 color space. Since the backward compatibility of HLG is true only if SDR and HDR channels share the same color space, it may be in practice difficult to use one single HDR channel to address all kinds of displays.

### Perceptual quantization transfer function

The perceptual quantization (PQ) transfer functions were designed by Dolby and first normalized in [36]. It optimizes the luminance quantization in order to use the most efficient bit depth to represent the signal without the apparition of banding artifact for luminance going up to  $10\,000 \text{ cd/m}^2$ . In [33], the authors show that 11 bits quantization was sufficient to avoid the banding artifacts and that a 10 bits quantization was sufficient most of the time. Unlike the precedent workflow to perceptually linearize the luminance, the PQ workflow is not a scene referred workflow. Indeed, the PQ model states that the OOTF is applied on the video signal inside the camera. It implies that the image/video signal is adapted to one reference display. Thus, each value in the image/video signal directly corresponds to a value in  $\text{cd/m}^2$ . This kind of workflow is called display-referred. Figure 1.5 describes a generic end-to-end display-referred workflow.

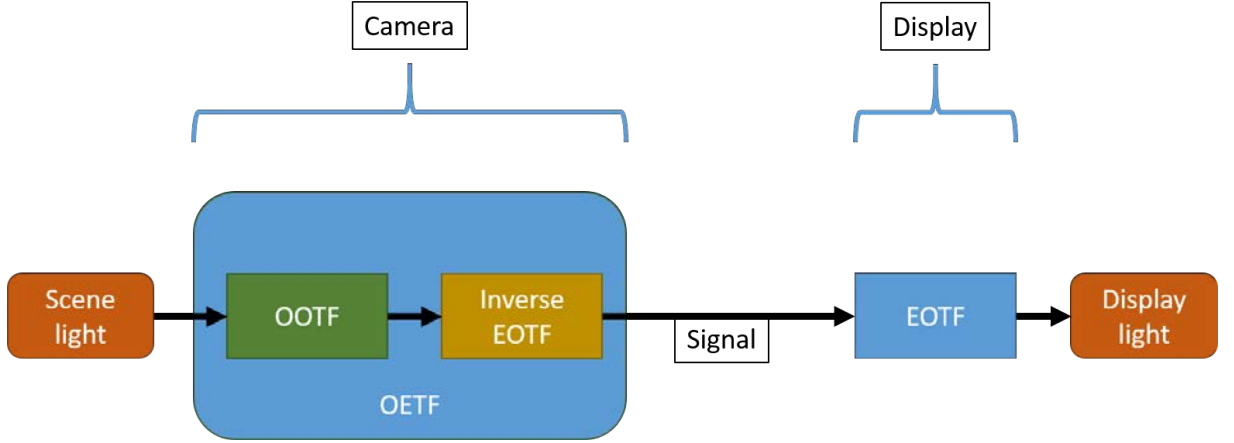


Figure 1.5: End-to-end display-referred workflow from the capture of light to the display of an image [31]

The EOTF is defined as follow:

$$F_D = L_W \times \left( \frac{\max[(V^{1/m_2} - c_1), 0]}{c_2 - c_3 \times V^{1/m_2}} \right)^{1/m_1} \quad (1.12)$$

where,  $F_D$  is the luminance of the screen in  $\text{cd/m}^2$  and  $V$  the image signal. The constant values are:  $m_1 = 2610/16384$ ,  $m_2 = 2523/4096 \times 128$ ,  $c_1 = 3424/4096$ ,  $c_2 = 2413/4096 \times 32$  and  $c_3 = 2392/4096 \times 32$ .

The reference OOTF is in the same form as the SDR OOTF:

$$F_D = OOTF(L) = G_{1886}[G_{709}[L]] \quad (1.13)$$

where

$$G_{709}[L] = \begin{cases} 1.099(L \times 59.5208)^{0.45} - 0.099 & \text{if } 1 \geq L \geq 0.0003024 \\ 267.84L & \text{if } 0.0003024 > L \geq 0 \end{cases} \quad (1.14)$$

and

$$G_{1886}[L'] = 100L'^{2.4} \quad (1.15)$$

The OETF can then be deduced from the reference OOTF and the EOTF:

$$\begin{aligned} V &= OETF[V] \\ &= EOTF^{-1}[OOTF[L]] \\ &= EOTF^{-1}[F_D] \end{aligned} \quad (1.16)$$

The inverse EOTF is defined as follows:

$$\text{EOTF}^{-1}[F_D] = \left( \frac{c_1 + c_2 Y^{m_1}}{1 + c_3 Y^{m_1}} \right)^{m_2} . \quad (1.17)$$

where,  $Y = F_D/10000$  and the constants  $c_1, c_2, c_3, m_1$  and  $m_2$  are the same constants than for the EOTF.

Figure 1.6 represents all the presented EOTF. For the HLG and the gamma EOTH, we have considered two different display, one with a black point at  $0.001\text{cd/m}^2$  and a peak brightness at  $4000\text{cd/m}^2$ , the other one with a black point at  $0.001\text{cd/m}^2$  and a peak brightness at  $10000\text{cd/m}^2$

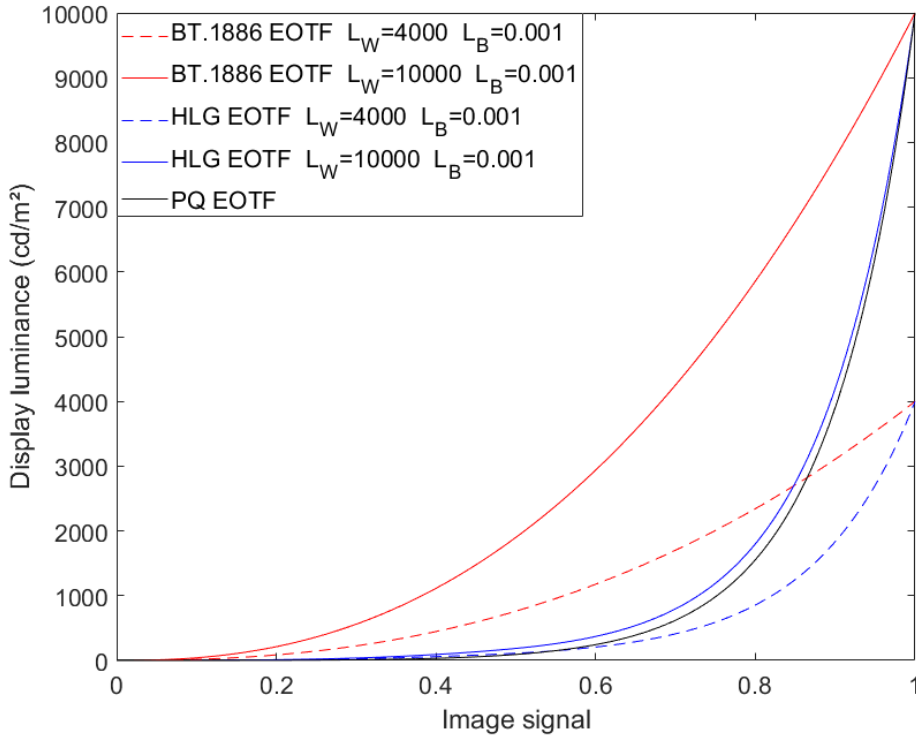


Figure 1.6: Different EOTF

### 1.3.2 Perceptually uniform color space

In the previous section, we describe how to perceptually linearize the luminance. However, as described in section 1.2, the image signal also contains chromatic information. Most of the time, to create a perceptually uniform color space, a color-opponent model is used. Those models state that the HVS uses the differences between the cones responses rather than each

cone's response individually. Color opponent models are often used as an approximation of the HVS color perception.

In the following sections, four color space are described:

- *HDR-Lab* [12], the HDR extension of the CIE 1976  $L^*a^*b^*$  [37].
- Two color spaces compatible with recent compression codecs and normalized in the ITU standard BT.2100 [6],  $Y'C'_bC'_r$  and  $ICtCp$ .
- The more recent  $J_z a_z b_z$  color space [11].

In the rest of the manuscript, the luminance component of the uniform color spaces is called uniform luminance instead of, according to the case, lightness, brightness or luma to avoid unnecessary complexity. For example, the uniform luminance of *HDR-Lab* should be called lightness while the uniform luminance of  $J_z a_z b_z$  should be called brightness. All those term always refer to the HVS luminance perception.

### $Y'C'_bC'_r$

$Y'C'_bC'_r$  normalized in the BT.2100 [6] standard is defined for HDR/WCG signal on the same model than color space  $Y'C_B C_R$  described in the BT.709 standard for SDR signal. This color space is compatible with most of the recent compression codecs. The computation of  $Y'C'_bC'_r$  is described below:

It can be used with the PQ and the HLG transfer functions.

- First, the linear  $RGB$  values (in the BT.2020 gamut) are transformed using either the PQ  $EOTF^{-1}$  or the HLG OETF into the non linear space  $R'G'B'$ :

$$\begin{cases} R' = EOTF_{PQ}^{-1}(R) & R' = OETF_{HLG}(R) \\ G' = EOTF_{PQ}^{-1}(G) & \text{OR } G' = OETF_{HLG}(G) \\ B' = EOTF_{PQ}^{-1}(B) & B' = OETF_{HLG}(B) \end{cases} \quad (1.18)$$

- Then the  $Y'$  luminance component and the  $C'_b$ ,  $C'_r$  chrominance components are calculated as follows:

$$\begin{cases} Y' = 0.227 \times R' + 0.6780 \times G' + 0.0593 \times B' \\ C'_b = \frac{B' - Y'}{1.8814} \\ C'_r = \frac{R' - Y'}{1.4746} \end{cases} \quad (1.19)$$

### *ICtCp*

*ICtCp* has better chrominance and luminance decorrelation and has better hue linearity than the  $Y'C'_bC'_r$  color space [38]. This color space is calculated in three steps:

- First, the linear *RGB* values (in the BT.2020 gamut) are converted into *LMS* values which correspond to the quantity of light absorbed by the cones:

$$\begin{cases} L = 0.41210938 \times R + 0.52392578 \times G + 0.06396484 \times B \\ M = 0.16674805 \times R + 0.72045898 \times G + 0.11279297 \times B \\ S = 0.02416992 \times R + 0.07543945 \times G + 0.90039062 \times B \end{cases} \quad (1.20)$$

- Second, the PQ EOTF<sup>-1</sup> is applied to each *LMS* component:

$$\begin{cases} L' = \text{EOTF}_{PQ}^{-1}(L) \\ M' = \text{EOTF}_{PQ}^{-1}(M) \\ S' = \text{EOTF}_{PQ}^{-1}(S) \end{cases} \quad (1.21)$$

- Finally, the luminance component *I* and the chrominance components *Ct* and *Cp* are deduced as follows:

$$\begin{cases} I = 0.5 \times L' + 0.5 \times M' \\ Ct = 1.61376953 \times L' - 3.32348632 \times M' + 1.70971679 \times S' \\ Cp = 4.37817382 \times L' - 4.37817383 \times M' - 0.13256835 \times S' \end{cases} \quad (1.22)$$

- The obtain color space is efficient to create good interchange format and to perform compression. However, each channel has to be re-scaled to obtain a perceptually uniform color space [39]:

$$\widehat{ICtCp} = [I, Ct, Cp] \times [720, 360, 720] \quad (1.23)$$

This color space can also be used with the HLG OETF instead of the PQ EOTF<sup>-1</sup>. However, the perceptual linearity is only guaranteed for a display with a peak luminance at 1000 cd/m<sup>2</sup> instead of 10000 cd/m<sup>2</sup>. The *ICtCp* color space is particularly well adapted to video compression and more importantly to the PQ EOTF as defined in BT.2100 [6].

### *HDR-Lab*

One of the most popular SDR uniform color spaces is the CIE 1976  $L^*a^*b^*$  or CIELAB which is suited for SDR content. An extension of this color space for HDR images was proposed

in [12]. The proposition is to tailor CIELAB for HDR by changing the non-linear function applied to the pixel  $XYZ$  values. This color space is calculated as follows:

$$\begin{cases} L_{HDR} = f(Y/Y_n) \\ a_{HDR} = 5[f(X/X_n) - f(Y/Y_n)] \\ b_{HDR} = 2[f(Y/Y_n) - f(Z/Z_n)] \end{cases} \quad (1.24)$$

where  $X_n$ ,  $Y_n$  and  $Z_n$  are the  $XYZ$  coordinates of the diffuse white. The non-linear function  $f$  is used to output perceptually linear values.  $f$  is defined for HDR as follows:

$$f(\omega) = 247 \frac{\omega^\epsilon}{\omega^\epsilon + 2^\epsilon} + 0.02 \quad (1.25)$$

$$\epsilon = 0.58 / (sf \times lf) \quad (1.26)$$

$$sf = 1.25 - 0.25(Y_s/0.184) \quad (1.27)$$

$$lf = \log(318) / \log(Y_n) \quad (1.28)$$

where the exponent  $\epsilon$  of the function  $f(\omega)$  is modified using a factor to take into account the surround  $sf$  and a factor to take into account the luminance level  $lf$ .  $Y_s$  is the relative luminance of the surround and  $Y_n$  is the absolute luminance of the diffuse white or reference white.

The use of *HDR-Lab* color space is somewhat difficult since it requires to know the relative luminance of the surround,  $Y_s$ , as well as the diffuse white luminance,  $Y_n$ . Unfortunately, these two parameters are most of the time unknown for HDR contents. Figure 1.7 illustrates the impact of  $Y_s$  and  $Y_n$  on  $L_{HDR}$ .

In addition to the *HDR-Lab* color space, Fairchild et al. [12] also proposed the *HDR-IPT* color space, which aims to extent the IPT color space [40] to HDR content. This color space is not studied in this manuscript due to its high similarity with *HDR-Lab*.

### $J_z a_z b_z$

$J_z a_z b_z$  [11] is a uniform color space allowing to increase the hue uniformity and to predict accurately small and large color differences, while keeping a low computational cost. It is computed from the  $XYZ$  values in five steps:

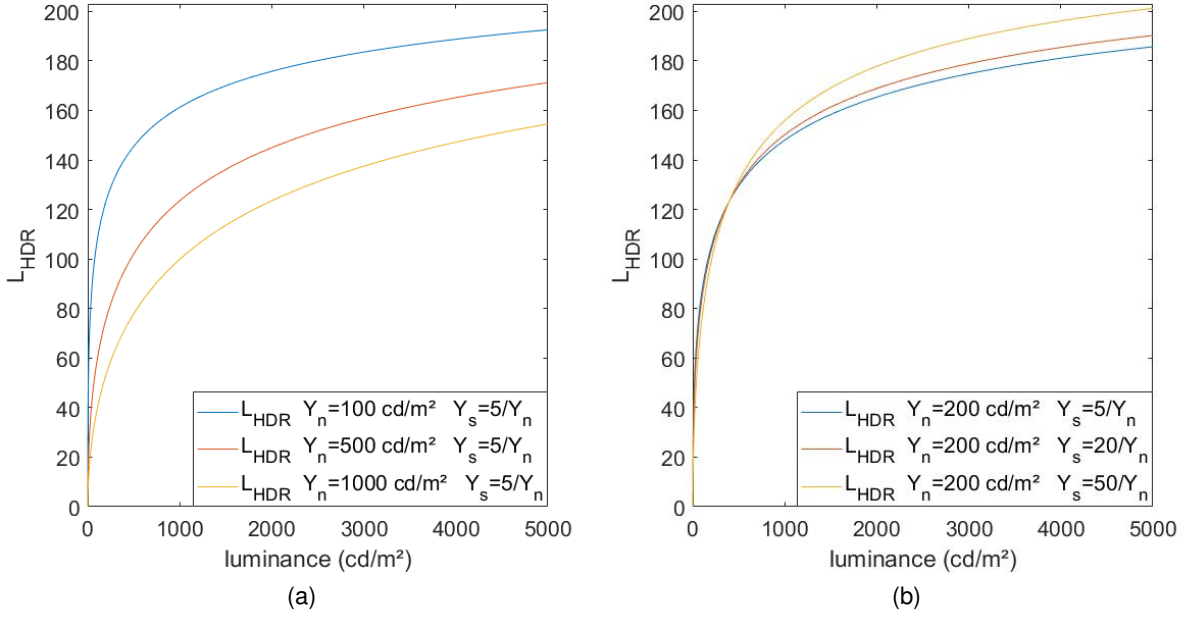


Figure 1.7:  $L_{HDR}$  in function of the luminance for **(a)** different diffuse white luminances, and for **(b)** different surround luminances.

- First, the  $XYZ$  values are adjusted to remove a deviation in the blue hue.

$$\begin{bmatrix} X' \\ Y' \end{bmatrix} = \begin{bmatrix} bX \\ gY \end{bmatrix} - \begin{bmatrix} (b-1)Z \\ (g-1)X \end{bmatrix} \quad (1.29)$$

where  $b = 1.15$  and  $g = 0.66$ .

- Second, the  $X'Y'Z$  values are converted to  $LMS$  values

$$\begin{cases} L = 0.41478972 \times X' + 0.579999 \times Y' + 0.0146480 \times Z \\ M = -0.2015100 \times X' + 1.120649 \times Y' + 0.0531008 \times Z \\ S = -0.0166008 \times X' + 0.264800 \times Y' + 0.6684799 \times Z \end{cases} \quad (1.30)$$

- Third, as for  $ICtCp$ , the PQ EOTF<sup>-1</sup> is applied on each  $LMS$  component

$$\begin{cases} L' = \text{EOTF}_{PQ}^{-1}(L) \\ M' = \text{EOTF}_{PQ}^{-1}(M) \\ S' = \text{EOTF}_{PQ}^{-1}(S) \end{cases} \quad (1.31)$$

- Fourth, the luminance  $I_z$  and the chrominance  $a_z$  and  $b_z$  are calculated

$$\begin{cases} I_z = & 0.5 \times L' + 0.5 \times M' \\ a_z = & 3.5240000 \times L' - 4.0667080 \times M' + 0.5427080 \times S' \\ b_z = & 0.1990776 \times L' + 1.0967990 \times M' - 1.2958750 \times S' \end{cases} \quad (1.32)$$

- Finally, to handle the highlight, the luminance is adjusted:

$$J_z = \frac{(1 + d) \times I_z}{1 + (d \times I_z)} - d_0 \quad (1.33)$$

where  $J_z$  is the adjusted luminance,  $d = -0.56$  and  $d_0$  is a small constant:  $d_0 \approx 1.62955 \times 10^{-11}$ .

To define interchange format and to compress video, main standardization organizations recognize two color spaces:  $Y'C'_B C'_R$  and  $ICtCp$  with the PQ or HLG transfer functions. Those organizations also proposed different methods to adapt the compression to those two color spaces.

## 1.4 HDR/WCG content compression

Several options are available to compress HDR/WCG signal. In this thesis, the focus is on video compression and more precisely the Moving Picture Experts Group (MPEG) proposition to adapt existing codecs to HDR/WCG content (namely, the Advanced Video Coding (AVC/h.264) codec and the High Efficiency Video Coding (HEVC/h.265) codec). However, solutions dedicated to still images also exist. For example, the Joint Picture Expert Group develops JPEG-Xt [41] which is a backward-compatible HDR image compression standard.

### 1.4.1 Pre-encoding process

Before the compression is performed, the image has to be in the right format. This step is called pre-processing. This conversion can be decomposed in three main steps:

- **Color space conversion:** The image has first to be converted into the right color space: HEVC recognizes two different color space:  $Y'C'_b C'_r$  and  $ICtCp$ . The two transfer functions models HLG or PQ can be chosen. However, in the case of HLG, the RGB linear values should represent the scene light (no OOTF applied) while in the case of PQ, the RGB linear values should represent the display light (the OOTF should have been applied).

- **Quantization:** The image value has to be quantized, going from a floating point representation to a fixed length representation. AVC and HEVC define multiple bit-depths from 8 to 12. A good compromise for HDR WCG images and the most chosen bit-depth for video delivery is 10.
- **Chroma sub-sampling:** This step allows reducing the size of the image by reducing the amount of chromatic information. It takes advantage that the HVS is less sensitive to high frequency variation in color than in luminance. The original image in a chroma sampling format called 4:4:4. Each pixel is composed of one value for the luminance and two values for the chrominance. The chrominances are downsampled to a chroma sampling format called 4:2:0. For a chrominance value, there are four luminance values. The difference between the two chroma sampling formats is illustrated in Figure 1.8.

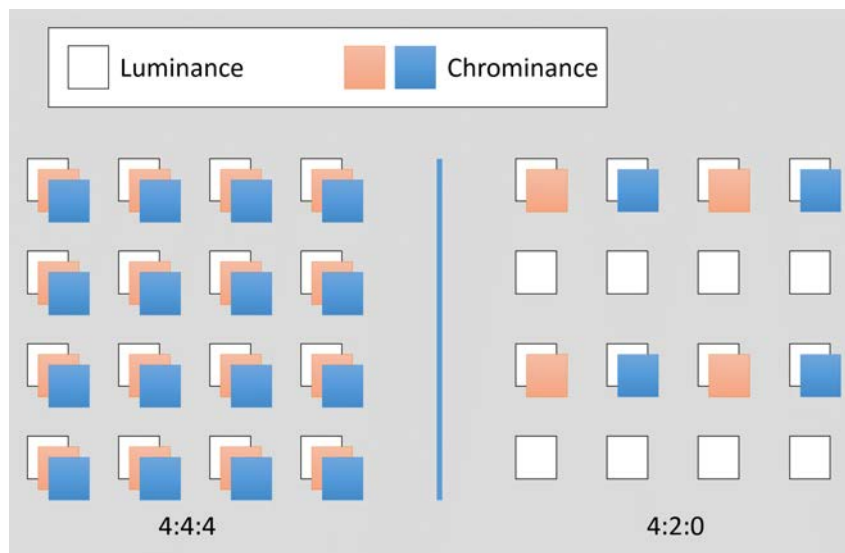


Figure 1.8: Chroma sampling format

Figure 1.9 represents the pre-encoding process using the  $Y'C'_bC'_r$  color space and Figure 1.10 represents the pre-encoding process using the  $ICtCp$  color space.

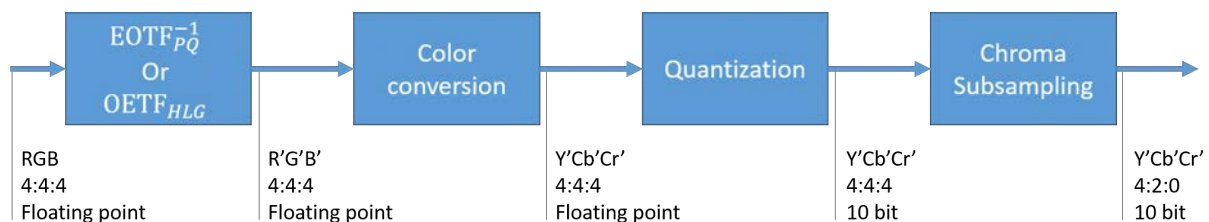
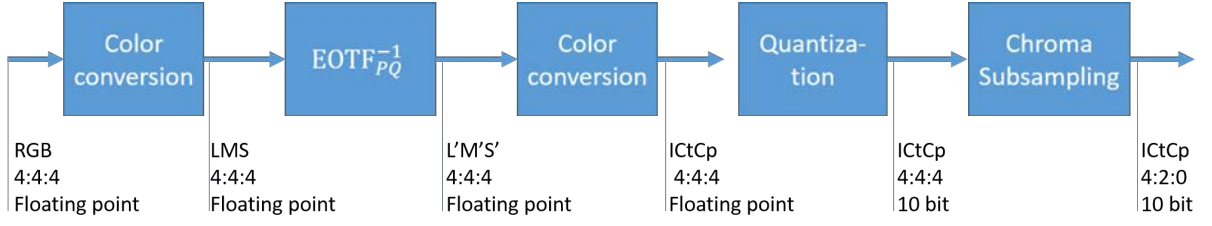


Figure 1.9: Pre-encoding process system using the  $Y'C'_bC'_r$  color space

Figure 1.10: Pre-encoding process system using the  $ICtCp$  color space

If the chosen color space is  $Y'C'_bC'_r$ , the chrominance subsampling can create visible artifacts. This is due to the not-so-good decorrelation between luminance and chrominance information, especially in the saturated area [42]. This phenomenon is particularly significant with the PQ transfer function. This is not the case with the  $ICtCp$  color space because this color space performs a better decorrelation between luminance and chrominance [43].

The ITU recommends a solution to prevent those artifacts [7] by adjusting the luminance in function of the error on the  $C'_b$  and  $C'_r$  components (cf. Figure 1.11). This recommendation proposed two methods to perform this processing: a recursive solution which consists of successive estimation of the luminance adjustment and a closed form solution which is more direct to estimate the luminance. Only the closed form [44] of this solution is used in this manuscript as it is the fastest proposed solution. The improvement bring by the luma adjustment method is illustrated by Figure 1.12.

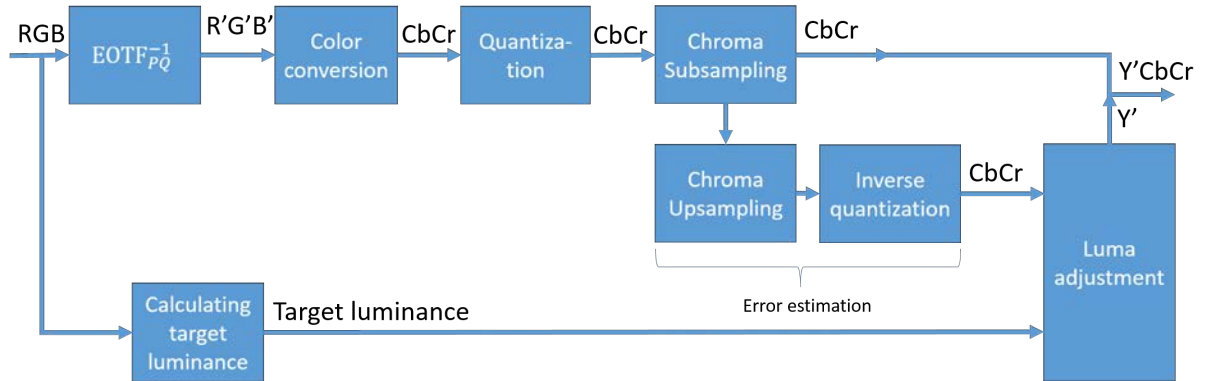


Figure 1.11: Example of a luma adjustment method (source [7])

Once the processing to adapt the picture format has been applied, the video can be compressed. In the next section we present the main mechanism involve to compress images and videos.



Figure 1.12: Tone-mapped examples showing improvements of the luma adjustment method **(a)** original image **(b)** classic chrominance subsampling **(c)** chrominance subsampling with luma adjustment method (source [45])

### 1.4.2 Compression: general principle

The main idea behind compression is to reduce the redundant information inside an image to the minimum and to only retain the information perceptible by the HVS. Figure 1.13 represents a simplified structure of MPEG (HEVC and AVC) encoder. The main principle behind video compression are:

- **Block partitioning:** The image is first segmented into several blocks. Those blocks do not have necessarily the same size.
- **Prediction:** This is a key step to reduce the redundancy of an image. It uses the fact that videos are not random signals and that pixels have often relation between each other. There is two ways to predict a block:
  - **Spatial prediction:** The block is predicted using the already processed neighbor blocks. The encoder cannot use non-processed blocks. Moreover, the encoder

should not use the source blocks but the blocks after the full encoding/decoding process as it will be the only information available to the decoder.

- **Temporal prediction:** The block is predicted using blocks of already processed images. Note that the encoder does not necessarily consider the images in temporal order.

Both, the prediction mode and the error committed during the prediction, called the residual, are transmitted to the decoder. If the prediction is accurate, its value should be close to zero.

- **Transform:** This is the second used method to reduce image redundancy. The goal of a transformation is to compact the block signal into the smaller possible number of coefficients. This is usually performed with the Discrete Cosinus Transform (DCT) although Discrete Sinus Transform (DST) is also possible in HEVC for 4X4 blocks.
- **Quantization:** The transformed residual is quantized. The quantization parameter (QP) is used in HEVC and AVC to control the quantization step size. This step reduces the amount of bits at the output of the encoder, but it also reduces the quantity of available information. Thus, the QP is often viewed as the parameter to adjust a trade-off between the amount of compression and the quality of the decoded signal.
- **Entropy encoding:** This is a classical lossless encoding method which allows optimizing the length of the bitstream. This method is particularly efficient with quantized residual blocks as most of their coefficients are null.

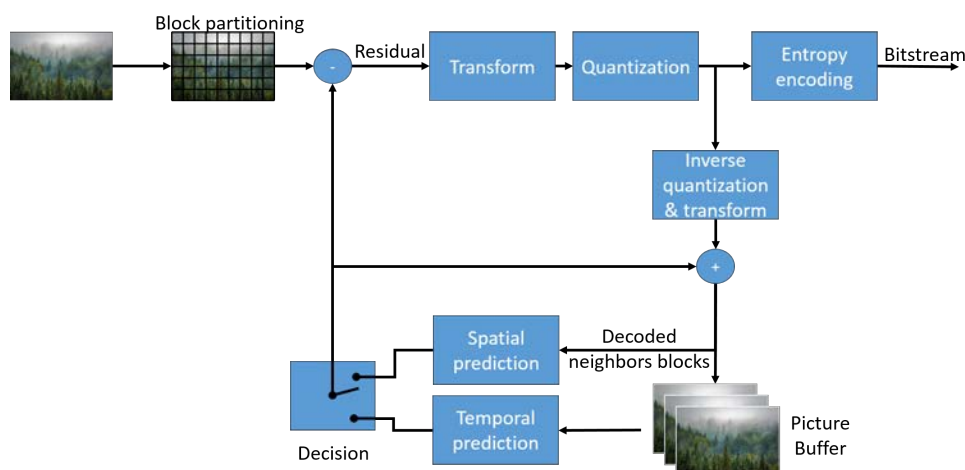


Figure 1.13: simplified structure of an MPEG video encoder

### 1.4.3 Adaptation of codec to HDR/WCG content

The recommendations of the ITU-T H.Supp1.15 [7] to adapt the HEVC and the AVC compression mainly concern the Quantization Parameter (QP) presented in the previous section. Those recommendations first concern the  $Y'C'_bC'_r$  color space with the PQ transfer function but it can be extended to the  $ICtCp$  color space and to the HLG transfer function [46]. There are two proposed algorithms to adapt the QP values to HDR/WCG content, one for luminance blocks and one for chrominance blocks:

- **Luma-dependant adaptative quantization:** The encoders achieved a good balance between dark and bright areas for SDR content but using the same setting with HDR content could ensure that too few bits are allocated to bright areas and too many to dark areas due to the characteristics of the transfer functions. The solution recommended by the ITU is to calculate the average luma value of each block and to add an offset dQP to the QP. The figure 1.14 represents the dQP value in function of the average luminance of the block in the case the chosen color space is  $Y'C'_bC'_r$  and the chosen transfer function is PQ.

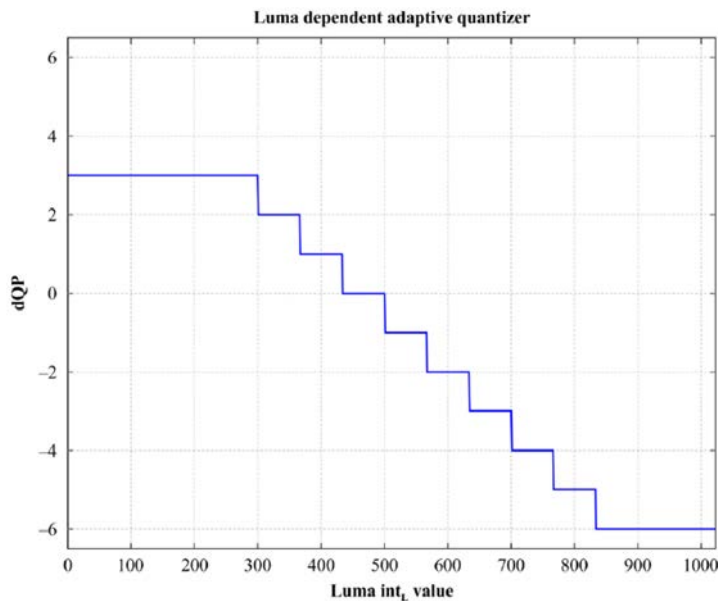


Figure 1.14: QP offset dQP in function of the average luma value in 64x64 block (source [7])

- **Chroma QP offset algorithm:** The chroma QP offset algorithm is used to overcome a compression issue with the chrominance component Cr and Cb in HDR/WCG. In WCG, most of the chrominance values tend to be near their mean value (i.e., 512, the value of grey) while the  $Y'$  component tends to use most of its range. This is even more significant for BT.709 content encapsulated in a BT.2020 gamut. This behavior creates a

shift in the bitrate allocation from the chrominance to the luminance. This can potentially create visible chrominance artifacts. The ITU proposes to add a QP offset specifically to chrominance components of picture blocks. An example can be found in [7] for a  $Y'C'_bC'_r$  color space with the PQ transfer functions:

$$\begin{aligned} QPoffsetCb &= \max(-12, \min(0, \text{Round}(c_{C'_b} \times (k \times QP + l)))) \\ QPoffsetCr &= \max(-12, \min(0, \text{Round}(c_{C'_r} \times (k \times QP + l)))) \end{aligned} \quad (1.34)$$

where QP is the quantization parameter,  $k = -0.43$ ,  $l = 9.26$ ,  $c_{C'_r} = c_{C'_b} = 1$  if the content is natively BT.2020  $c_{C'_b} = 1.14$  and  $c_{C'_r} = 1.78$  if this is a BT.709 content encapsulated in a BT.2020 content.

## 1.5 Conclusion

In this chapter we explain the main concept behind HDR and WCG. Those technologies greatly improve the viewers quality of experience. In addition we present the modifications required in the chain of transmission to handle such content. Those modifications are not negligible as the HVS have a widely different comportment in the color space areas that were not cover by the legacy SDR format. New color space had to be created to represent HDR/WCG gamut content and the compression had to be adapted.

The quality of HDR/WCG transmission have to be evaluated to select the best solution to compress images/videos. In the next chapter we present the methods to perform quality assessment on images.

# IMAGE QUALITY ASSESSMENT

---

## 2.1 Introduction

Image/video Quality Assessment (IQA) methods aim to evaluate the perceived quality of images from the humans's point of view.

IQA is an important tool to estimate the level of acceptable degradation occurring during the various steps of the image distribution: the capture, the processing, the compression, the transmission and the display of images. There are two kinds of IQA methods: Subjective and objective.

Subjective quality assessment methods are the most reliable methods to assess the quality of images. Subjective tests consist in asking to a panel of viewers to evaluate the quality of distorted images/videos with/without the reference image. Objective image quality assessment methods, on the other hand, aim to emulate subjective tests using computational models. They are less reliable but are faster and less costly to achieve as it does not require to recruit any viewer. Subjective tests are the ground truth on which are evaluated any objective methods.

In this chapter, we first introduce the different standardized methods to perform subjective tests. Then in a second section, we review three existing HDR databases annotated with subjective scores. Finally, we review the different methodology to perform objective image quality assessment on HDR content.

## 2.2 Subjective methods

In this section, we present the methodologies used to retrieve viewers' opinions. In most tests, reference images are distorted using different levels or different modes of distortion. Each kind of distortion corresponds to a hypothesis on the configuration of an image/video system. Each hypothesis is called a hypothetical reference circuits (HRC).

Several methodologies were created over the years to obtain the opinions of a panel of viewers. Most of those methods are normalised in the ITU standards BT.500 [47] and P.910 [48]. Those methodologies can be categorized into two types:

- **Single Stimulus method:** Each distorted images are rated individually independently

from the reference images. One example is the Absolute Category Rating (ACR) [48]. In this test, each images is evaluated by the viewers on a five-label scale: "bad", "poor", "fair", "good" and "excellent". Those labels are then translated to values between 1 and 5. A variant of this test is the Absolute Category Rating with Hidden Reference (ACR-HR). In this test, the reference images are also included in the test, without informing the subjects of its presence.

- **Double-stimulus or multiple stimulus:** Each distorted image is rated based on its reference image. For example, in the Double Stimulus Impairment Scale (DSIS) methodology [47] also called Degradation Category Rating (DCR) [48], the viewers observe both the reference and the distorted images (at the same time or sequentially). The scale of this kind of test go from "impairments are imperceptible" to "impairments are very annoying".

Once the scores are obtained, the mean score of each image is calculated. This score is called the Mean Opinion Score (MOS) and corresponds to the quality of an image:

$$\text{MOS}_j = \frac{1}{N} \sum_{i=1}^N s_{ij} \quad (2.1)$$

where  $\text{MOS}_j$  is the MOS of the image  $j$ ,  $N$  the number of viewers and  $s_{ij}$  the opinion of the viewer  $i$  on the image  $j$ .

As all data obtained by experimental test, this note has a confidence interval. This confidence interval is derived from the standard deviation of each sample [47]. The 95% confidence interval is given by:

$$CI_{95} = [\text{MOS}_j - \delta_j, \text{MOS}_j + \delta_j] \quad (2.2)$$

where:

$$\delta_j = 1.96 \times \frac{\text{std}_j}{\sqrt{N}} \quad (2.3)$$

where  $\text{std}_j$  is the standard deviation of the image  $j$  scores.

Both single and double stimulus methodologies have their own advantages. Single Stimulus methods are not biased by the reference and are easier to implement than Double-stimulus methods. ACR-HR also helps to reduce the scene bias (viewers liking or disliking a reference video). However, those kinds of tests are unable to evaluate some defaults that impact the "look" of an image. For example, dulled colors [48] might remain undetected. Double-stimulus methods are more adapted to also evaluate them. Moreover, those methods are particularly well suited for high quality system evaluation in the context of multimedia. For example, the DSIS and DCR methodologies allow us to obtain information on the viewer's impairment detection threshold using the two scales "imperceptible" and "perceptible but not annoying".

The obtained images annotated with quality scores are particularly useful to evaluate the accuracy of objective metrics.

## 2.3 Available HDR image databases with subjective scores

Several HDR image databases annotated with quality scores were created in the last few years to compare different compression methods or to evaluate the performance of image quality metrics. In this section, we review three available HDR image databases: Narwaria et al. [49], Zerman et al. [50] and Korshunov et al. [24]. Those databases were created to evaluate compression solutions as well as objective metrics and are composed of a variety of images with different characteristics.

Therefore, the different authors aimed to create database representative of the diversity of HDR content. First, we propose to characterize these images thanks to a set of indexes described in the following section. Result and discussion are presented in Section 2.3.2.

### 2.3.1 Image description indexes

To describe objectively the images of each database, four indicators are used:

- **The image dynamic range (DR).** It represents the luminance range of the signal. It is not a measure of the brightness of an image. For example, a night scene have a large DR due to emissive light sources (like car lights) but appears to be very dark. The image DR is similar to the dynamic range calculated for display (cf. section 1.2.1):

$$DR = \log_{10}(Y_{max}^*) - \log_{10}(Y_{min}^*) \quad (2.4)$$

where  $Y_{min}^*$  and  $Y_{max}^*$  are the minimum and the maximum luminance (in the  $XYZ$  linear domain), respectively. They are computed after excluding 1% of the brightest and darkest pixels to be more robust to outliers. For SDR images, a dynamic range is typically below 3 log units.

- **The key of the picture.** It is a measure of its overall brightness (in the range [0,1]) of a scene. This metric is complementary to the DR to characterize the luminance of an image. If the key is low (below 0.3), it means that most of the pixels are near the lowest limit of the image DR. In this case the image is perceived as dark. If the key is high (above 0.6) the pixels are nearer to the highest limit of the image DR. In this case the image is perceived as bright. The key is calculated as follows:

$$\text{key} = \frac{\overline{\log_{10}(Y)} - \log_{10}(Y_{min}^*)}{\log_{10}(Y_{max}^*) - \log_{10}(Y_{min}^*)} \quad (2.5)$$

$\overline{\log_{10}(Y)}$  is computed as follows:

$$\overline{\log_{10}(Y)} = \frac{1}{N} \sum_{i=1}^I \sum_{j=1}^J \log_{10}(Y(i, j) + \delta) \quad (2.6)$$

where  $Y(i, j)$  is the luminance of pixel  $(i, j)$ ,  $N$  the total number of pixels and  $\delta$  a small offset to avoid a singularity when the luminance is null.  $Y_{min}^*$  and  $Y_{max}^*$  are calculated in the same way as for dynamic range.

- **The spatial information (SI)** [48]. It describes the complexity of an image. The SI index corresponds to the standard deviation of the image perceptually uniform luminance plane  $Y'$  which has been filtered by a Sobel filter:

$$SI = \text{std}[\text{Sobel}(Y')] \quad (2.7)$$

For high definitions images, a SI of 45 means that the image is very simple and a SI above 100 means that the image is very complex. On SDR images, to obtain an approximately perceptually uniform luminance, this index is used after the OETF, usually a gamma function. as stated in section For HDR content, to obtain SI values comparable to the SI values applied to SDR content, we apply the PU function (cf. section 2.4.1) on the linear luminance of the HDR images.

- **The colorfulness.** The colorfulness of the perception of the chromatic component of an image. The more an image is perceived as colored, the more the image have a high colourfullness. In [51], the authors proposed an index that measures the colourfullness based on a psychophysical category scaling experiment. The authors proposed three variants to measure the colorfullness. In this thesis, we use the variant called  $\hat{M}^{(1)}$ . This index is considered to be the most natural by the authors. Moreover, this index is calculated on image converted in the CIELab space, a space that can be adapted to HDR (cf. Section 1.3.2). This metrics is calculated as follows:

$$\hat{M}^{(1)} = \sigma_{ab} + 0.37 \times \mu_{ab} \quad (2.8)$$

where

$$\sigma_{ab} = \sqrt{\sigma_a^2 + \sigma_b^2} \quad (2.9)$$

and

$$\mu_{ab} = \sqrt{\mu_a^2 + \mu_b^2} \quad (2.10)$$

where  $\sigma_a$  and  $\sigma_b$  are the standard deviations along the  $a$  and  $b$  axis, respectively.  $\mu_a$  and  $\mu_b$  are the means of the  $a$  and  $b$  component, respectively. The combination of  $\sigma_{ab}$  and  $\mu_{ab}$

was learned with the psycho-physical experiment.

The image files sometimes report luminance above or below the limits of the display. To obtain realistic values, before calculating these indicators, the image luminance is limited to the display available dynamic range used in the respective subjective tests.

### 2.3.2 HDR Databases presentation

In this subsection we review three HDR databases: Narwaria et al. [49], Zerman et al. [50] and Korshunov et al. [24]. We select those databases because they all use HDR images with various characteristics, they create realistic distortions and they are publicly available online.

To score the images, each database used a different subjective test protocol. Two used a double stimulus method and one a single stimulus method. If the protocols are different, the authors used the same HDR SIM2 display (HDR47ES4MB) which has a measured dynamic range going from 0.03 to 4250 cd/m<sup>2</sup>. This display also have one main limitations: it cannot handle the BT.2020 gamut.

All of the presented databases include compression artifacts using well known codec (JPEG, JPEG-Xt and JPEG 2000). Some of them use a backward compatible compression. This method allows the images to be displayable with SDR equipment while preserving the dynamic range for the display on HDR screens. A Tone Mapping Operator (TMO) is used to tone map the HDR images into SDR range. These tone-mapped images are then compressed using different codecs. After the decoding process, they are tone expanded to recover their HDR characteristics.

Table 2.1, summarized the characteristics of the databases including, the number of observers, the number of images, the subjective test protocol, the kind of distortions, the used display, the used gamut and the image sizes.

Table 2.1: Database characteristics.

Name	#Obs	#Img	Protocol	Distortion	Display	Gamut	Size
Narwaria et al. [49]	27	140	ACR-HR	JPEG	SIM2 HDR47ES4MB	BT.709	1920 × 1080
Korshunov et al. [24]	24	240	DSIS (side by side)	JPEG-Xt	SIM2 HDR47ES4MB	BT.709	944 × 1080
Zerman et al. [50]	15	100	DSIS	JPEG, JPEG-Xt JPEG2000	SIM2 HDR47ES4MB	BT.709	1920 × 1080

### 2.3.3 Narwaria et al.

Narwaria et al. [49]'s database (Available at [http://ivc.univ-nantes.fr/en/databases/JPEG\\_HDR\\_Images/](http://ivc.univ-nantes.fr/en/databases/JPEG_HDR_Images/)) is composed of 10 source images, which have been distorted by a backward compatible compression with the JPEG codec and the iCam TMO [52]. This database was used along with others to tune the objective metric HDR-VDP-2 [53] in its version 2.2.1 [54]. The angular resolution used during the subjective test was 60 pix/degree and the surround luminance was 130 cd/m<sup>2</sup>. The characteristics of each reference image are reported on Figure 2.1.

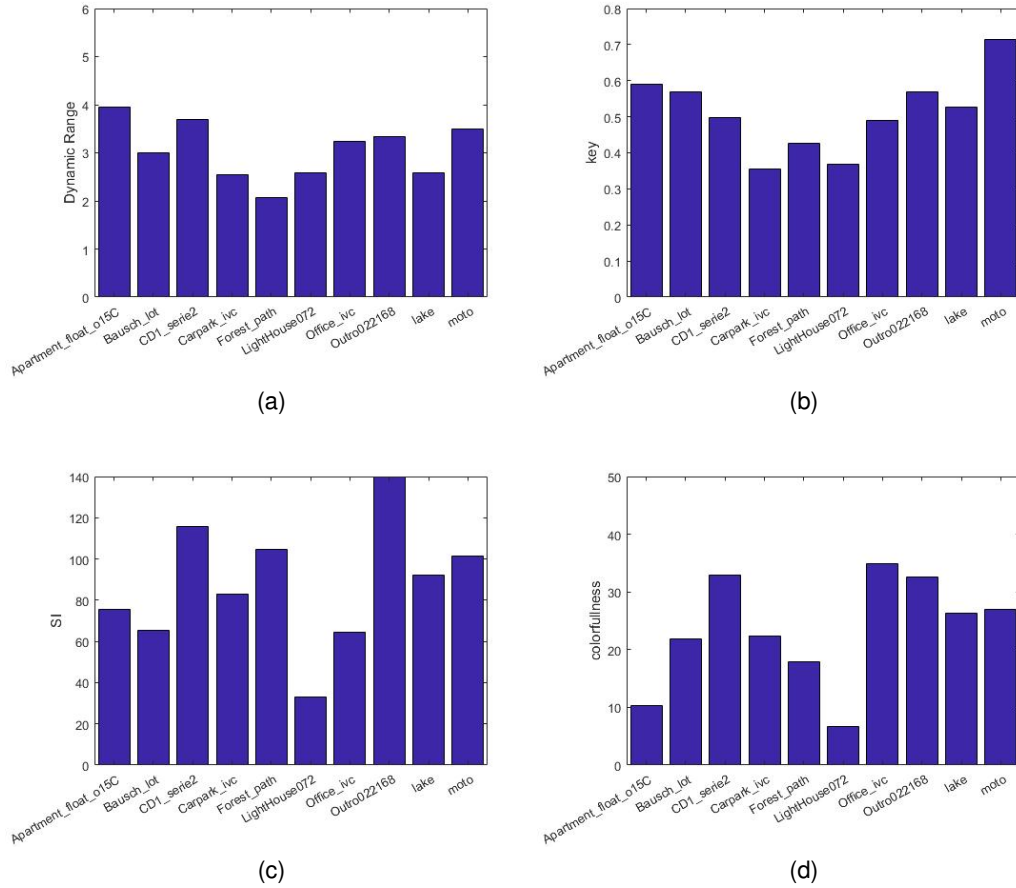


Figure 2.1: Characteristics of the Narwaria et al. [49] images: **(a)** The dynamic range, **(b)** key, **(c)** spatial Information, **(d)** Colorfulness.

### 2.3.4 Korshunov et al.

Korshunov et al. [24]’s database (Available at <http://mmspg.epfl.ch/jpegxt-hdr>) consists in images distorted with a backward-compatible compression scheme using the JPEG-XT standard and either the Mantiuk et al. [55] or the Reinhard et al. [56] TMO. The angular resolution used during the subjective test was 60 pix/degree and the surround luminance was 20 cd/m<sup>2</sup>. The characteristics of each reference images are reported on Figure 2.2.

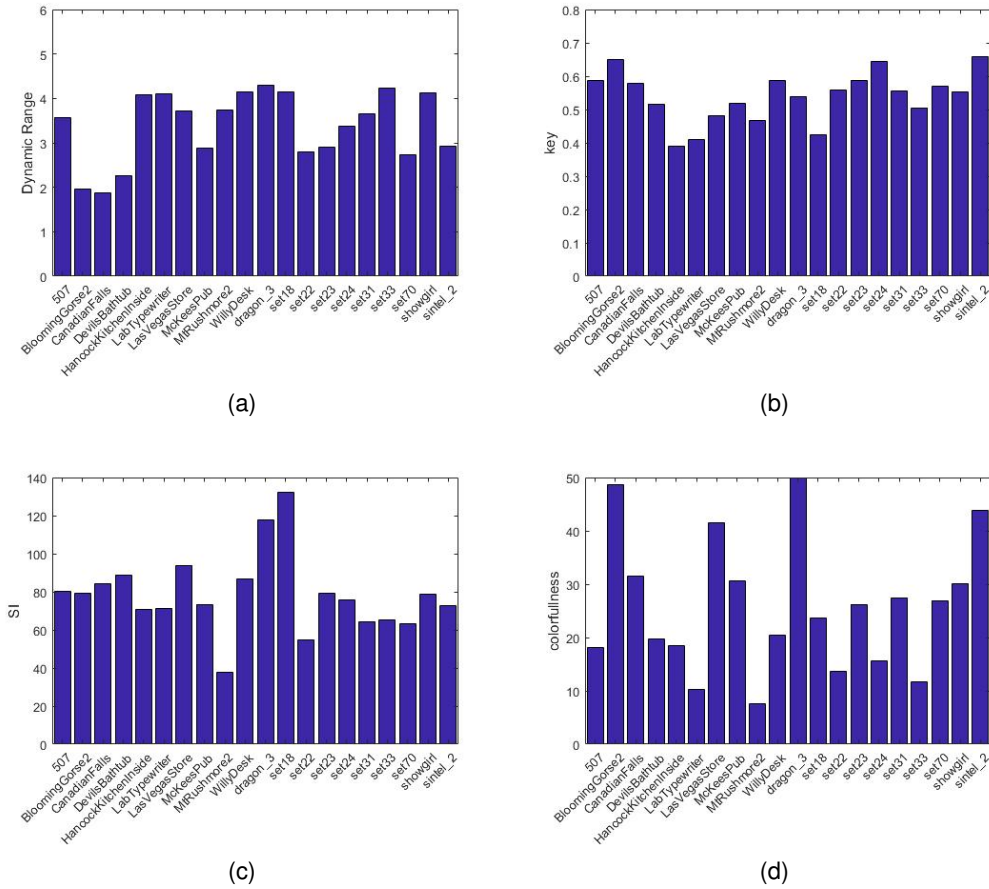


Figure 2.2: Characteristics of the Korshunov et al. [24] images: **(a)** The dynamic range, **(b)** key, **(c)** spatial Information, **(d)** Colorfulness.

### 2.3.5 Zerman et al.

Zerman et al. [50]’s database (Available at <http://webpages.12s.centralesupelec.fr/perso/giuseppe.valenzise/>) is an extension of a smaller database created by Valenzise et al. [57]. The distorted images are generated by both backward-compatible compression using the TMO proposed by Mai et al. [58] and a non backward-compatible compression with the use

of the PQ function for the EOTF. The compression was performed using the JPEG, the JPEG-XT and the JPEG2000 codec. The angular resolution used during the subjective test was 40 pix/degree and the surround luminance was 20 cd/m<sup>2</sup>. The characteristics of each reference images are reported on Figure 2.3.

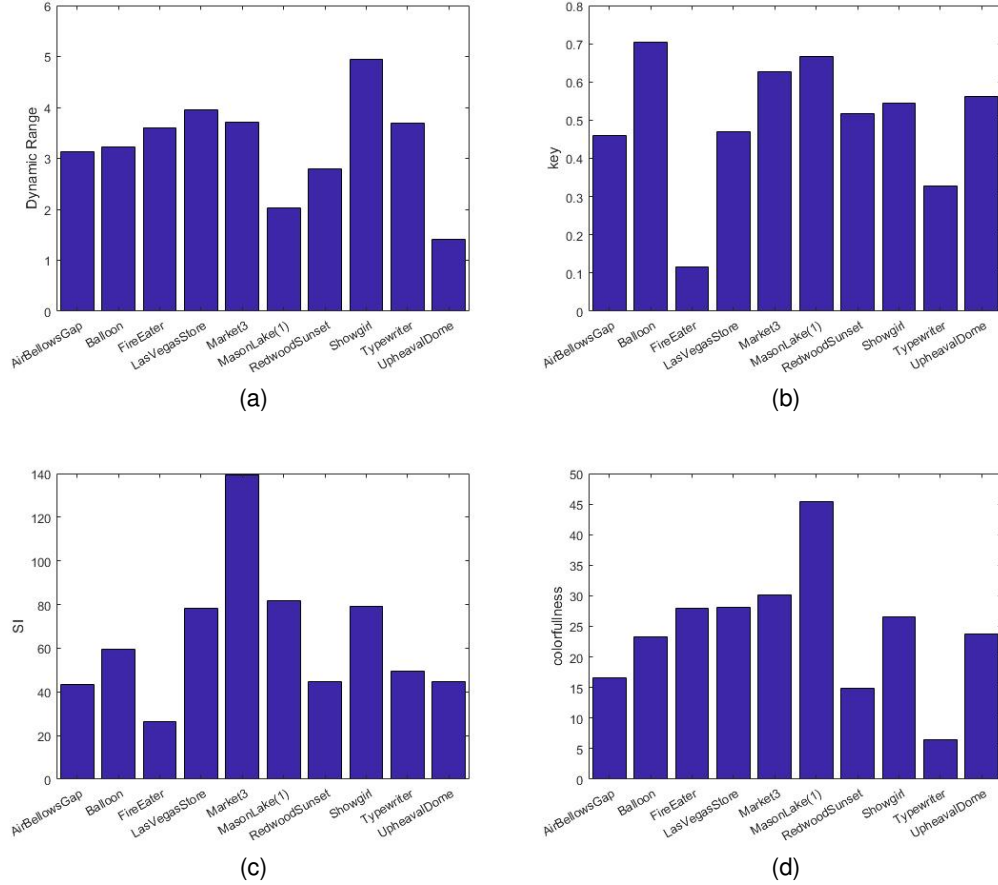


Figure 2.3: Characteristics of the Zerman et al. [50] images: **(a)** The dynamic range, **(b)** key, **(c)** spatial Information, **(d)** Colorfulness.

### 2.3.6 Advantages and Drawbacks

The databases are composed of various images with very different characteristics:

- They all have images with a dynamic range that falls below a SDR display range and images with the maximum range available for the used HDR display.
- They all have images with various key. From very bright image like "moto" of the Narwaria et al. database to "FireEater" from the Zerman et al. database. Zerman et al. is the only

database to have an images have one image wit a key below 0.2. However, it might correspond to a very unlikely extreme condition. Other databases do contain night scenes like "LightHouse072" for the Narwaria et al. database and "LabTypeWriter" for the Korshunov et al. database. Korshunov et al have no images with a key above 0.7 contrary to other databases.

- All databases have a wild range of complexity
- Korshunov et al. and Zerman et al. have a wild range of colorfulness. However, Narwaria et al. has no images with a colorfulness above 40.

In general, the three databases are intended to represent the diversity of HDR content. While some features are missing in some databases, such as images with a high key in Korshunov et al. or very colorful images in Narwaria et al., it is too early to conclude that these databases are not representative. Indeed, the HDR images are always confidential and the usages are still evolving. This type of image could correspond to a very small minority of content and would not be essential for a database to be representative of HDR content with a BT.709 gamut.

There are three main limitations on those databases:

- First, the limited BT.709 gamut is used during the databases creation. The wider BT.2020 color gamut is more and more associated with HDR images and videos. In addition, standard organizations, such as DVB, recommend the use of the BT.2020 gamut with HDR content [59] to accelerate the development of both technologies.
- Second, because they used the smaller BT.709 gamut, the databases have low color artifacts due to the compression. Therefore they are not representative to the chromatic artifacts that can appears on a larger gamut (cf. Section 1.4.3).
- All of these databases were created with the same screen (SIM2 HDR47ES4MB). This screen uses LED-LCD technology and has a specific maximum dynamic range and brightness. A different display using a different technology (such as OLED) and a different dynamic range may impact the image rendering.

## **2.4 Image quality assessment: objective methods**

Objective metrics aims to mimic the results from subjective tests. Those metrics are a key tool to evaluate the performances of image and video compression. Objective IQA metrics became an important field of research and numerous solutions exist to assess accurately the quality of images.

IQA objective metric can be separated into three categories:

- **Blind reference metrics:** Blind reference metrics or no reference metrics do not need to use the reference image to evaluate distorted images. They are particularly suited to evaluate camera performances as for this kind of system, no reference is available. Among the most use no-reference metric, we can mention DIIVINE [60], BLIINDS [61] or BRISQUE [62]. More recently Kottayil et al. [63] proposed a no-reference metric specialized for HDR contents.
- **Full-reference-metrics:** Full-reference-metrics use the reference and the distorted image. They are particularly suited for compression evaluation as the goal of such algorithms is to maintain as much of possible the integrity of a reference image. A complete review of such metrics for SDR images was perform in [64]. In this review, authors estimate that FSIMc [65] and SFF [66] provided the best performances.
- **Reduced-reference metrics:** Reduced-reference metrics use only partial information of the reference image. It can be useful when the effort to obtain the reference image is too costly. Among the most use reduced-reference metric, we can mention WNISM [67] which has been recognized as the standard method for reduced reference IQA, FEDM [68] or again RRED [69].

In this thesis, the focus is put on video compression, so only the full-reference metrics are studied.

IQA metrics used for Standard Dynamic Range (SDR) images using the legacy gamut BT.709 [28] are not able to process the information brought by the new HDR/WCG representations. To overcome this limitation, there are two distinct strategies. First, the SDR metrics can be adapted to HDR. This is the method proposed by aydin et al. [10]. The second strategy is to create specific metrics dedicated to HDR content such as HDR-VDP-2 [53] and HDR-VQM [70].

In the first part of this section, we present the method to adapt SDR metrics and in a second part, we describe metrics specifically designed for HDR content and especially HDR-VDP-2.

### 2.4.1 Adapting SDR image quality assessment

Objective image quality metric has been an active topic for many years. Many methods were created from the simplest distortion metric such as color difference metric and PSNR, to more complex metric based on phase congruence (FSIM) passing by metric based on the information theory (VIF). Thirteen SDR metrics commonly used in academic research, standardization or industry are summarized by Table 2.2. There are seven achromatic or color-blind metrics (PSNR, SSIM, MS-SSIM, FSIM, VIF, PSNR-HVS-M and PSNR-HMA) and six metrics including chrominance information ( $\overline{\Delta E}$ ,  $\overline{\Delta E^S}$ , SSIMc, CSSIM, FSIMc, PSNR-HMAc). A more detailed description of those metrics can be found in appendix A.

Table 2.2: Selected SDR quality metrics.

Name	Color	Reference	Main Principle
<b>PSNR</b>			Ratio between the range of the signal and the mean square error
$\overline{\Delta E}$	✓		Mean of the color difference metrics
$\overline{\Delta E^S}$	✓	Zhang et al. [71]	Mean of the color difference metrics and the mean the blurring effect of the HVS. Also known as S-CIELab
<b>SSIM</b>		Wang et al. [72]	Metrics based on the comparison of three characteristics of the images: the luminance, the contrast and the structure.
<b>SSIMc</b>	✓	Wang et al. [73]	Linear combination of the SSIM applied on the three components $Y'$ , Cr and Cb of the images.
<b>CSSIM</b>	✓	Hassan et al. [74]	Combination of SSIM and $\overline{\Delta E^S}$
<b>MS-SSIM</b>		Wang et al. [75]	Multi-scale SSIM
<b>FSIM</b>		Zhang et al. [65]	Comparison of the phase congruency and the gradient magnitude
<b>FSIMc</b>	✓	Zhang et al. [65]	Color extension of FSIM. Adds two comparisons corresponding to the two chrominance components
<b>VIF</b>		Sheikh et al. [76]	Metric based on the information theory and natural scene statistics
<b>PSNR-HVS-M</b>		Ponomarenko et al. [77]	PSNR on the DCT blocks of the images using CSF and visual masking
<b>PSNR-HMA</b>		Ponomarenko et al. [78]	Improvement of the PSNR-HVS-M. Takes into account the particularities of the mean shift and the contrast change distortions
<b>PSNR-HMAc</b>	✓	Ponomarenko et al. [78]	Linear combination of the PSNR-HMA on the three components $Y'$ , Cr and Cb of the images.

Those metrics were tuned for SDR content encoded with the gamma transfer functions. In [10], authors proposed to use a function called Perceptual Unit (PU) function instead of using a classical gamma transfer function. The PU function corresponds to the gamma EOTF (defined in BT.1886 [32]) for luminance value between 0.1 and 80 cd/m<sup>2</sup> while retaining per-

ceptual linearity above. This method can be used for any metrics using the luminance-corrected with the gamma function, such as PSNR, SSIM, MS-SSIM. In this manuscript, the metrics using the Perceptually Uniform (PU) function have the prefix PU- (PU-PSNR, PU-SSIM).

Figure 2.4 represents the PU function as well as the gamma function (calculated for a range between 0.01 and 100  $\text{cd/m}^2$ ).

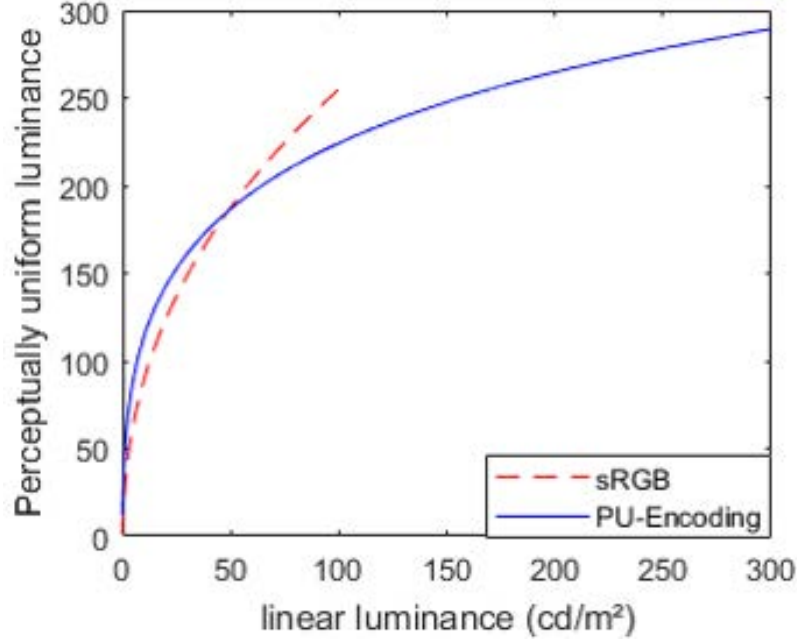


Figure 2.4: PU function and gamma function

However, this method can only be used on luminance-only (or color-blind) metric as it only adapts the color space in the luminance dimension. The authors did not use one of the perceptually uniform color space presented in the first section because those color spaces did not exist at the time.

### 2.4.2 Metrics designed for HDR content

There are two main metrics that were specifically designed with HDR in mind: HDR-VDP-2 [53] for still images and HDR-VQM [70] for video. Both metrics aim to model the HVS mechanisms.

HDR-VDP-2 which has been recognized as the standard metric to perform HDR objective IQA. As this metric was thoroughly studied during this thesis, we describe its main components in the following.

HDR-VDP-2 is a metric that models precisely the response of the early stage of the HVS. It takes into account the higher dynamic range of HDR images in the dark regions (mesopic

and scotopic vision) as well as brighter regions (photopic vision). This metric can provide a distortion map and a quality score. Figure 2.5 represents the block diagram of this metric.

The HDR-VDP-2 metric models step by step light path accross the HVS from the cornea to the brain:

- HDR-VDP-2 first models the intra-ocular scattering of the light. It corresponds to the scatter that occurs in the eye chamber and the retina. It reduces high frequencies and reduces the contrast of the light. This effect is also known as the disability glare [79]. HDR-VDP-2 models this phenomenon with a Modulation Transfer Function (MTF):

$$\mathcal{F}\{L_O[c]\} = \mathcal{F}\{L[c]\} \cdot \mathbf{MTF} \quad (2.11)$$

where  $L[c]$  the map of the  $c$  component (usually Red, Green and Blue emission from a screen), and  $L_O[c]$  the modified luminance.  $\mathcal{F}$  corresponds to the Fourier transform.

- After simulating the intra-ocular light scattering, the metric estimates the response of the L, M and S cones and the rods R responses using the photoreceptor spectral sensitivity :

$$v_{L|M|S|R}[c] = \int_{\lambda} \sigma_{L|M|S|R}(\lambda) \cdot f[c](\lambda) d\lambda \quad (2.12)$$

$$R_{L|M|S|R} = \sum_c L[c] \cdot v_{L|M|S|R}[c] \quad (2.13)$$

with  $\sigma$  the spectral sensitivity of L, M and S cones and R the rods,  $f[c]$  the spectral emission of the  $c$  component of the images and  $R$  the amount of light sensed by each photoreceptor.

This response corresponds to the quantity of light that is perceived by each photoreceptor. However, the sensibility of this photo-receptor is highly non-linear (as stated in subsection 1.3.1). This phenomenon is sometimes called luminance masking.

- To transform the response  $R_{L|M|R}$  into a perceptually linear response  $P_{L|M|R}$ , HDR-VDP-2 uses a non-linear transducer function  $t_{L|M|R}$  :

$$P_{L|M|R} = t_{L|M|R}(R_{L|M|R}) \quad (2.14)$$

In this part, the S cones are neglected because they have very little impact on the luminance perception. This function is calculated using the estimated intensity sensitivities of each photo-receptor  $s_{L|M|R}$ :

$$t_{L|M|R} = s_{peak} \int_{r_{min}}^r \frac{s_{L|M|R}(\mu)}{\mu} d\mu \quad (2.15)$$

with  $s_{peak}$  the adjustment for the peak sensitivity of the visual system and  $r_{min}$  the minimum detectable intensity. The sensitivity of the rods comes from experimental data obtain with color-blind persons. The sensitivity of the cones was deduced from the combined sensitivity of all photo-receptors  $s_A$  :

$$S_{L|M} = 0.5(s_A(2r) - s_R(2r)) \quad (2.16)$$

The global sensitivity  $s_A$  is captured in a contrast sensitivity function :

$$s_A(l) = \max_{\rho}(CSF(\rho, l)) \quad (2.17)$$

with  $l$  the luminance and  $\rho$  the spatial frequency. The chosen CSF was established from [80].

- The perceived luminance map of each image is then estimated:

$$P = P_L + P_M + P_S \quad (2.18)$$

- This response is, then, decomposed into subbands of frequency  $f$  and orientation  $o$  using the steerable pyramid [81]. The images are decomposed into four orientations bands and the maximum possible number of spatial frequency bands given the image resolution. Each subband of the reference image and the distorted images are then subtracted to each other :

$$D[f, o] = \frac{|B_T[f, o] - B_R[f, o]|^p}{\sqrt{N_{nCSF}[f, o]^{2p} + N_{mask}[f, o]^2}} \quad (2.19)$$

with  $B_R[f, o]$  the subband of frequency  $f$  and orientation  $o$  of the reference image and  $B_T[f, o]$  the subband of frequency  $f$  and orientation  $o$  of the distorted image. The exponent  $p$  is the gain that controls the shape of the masking function ( $p = 3.5$ ).  $N_{nCSF}$  is a noise corresponding to the neural CSF. The neural CSF is a CSF without the intra-ocular scattering and without the luminance masking.  $N_{mask}$  corresponds to the contrast masking. This is a signal-dependent noise component necessary because the visibility of differences depends on its background. For example, if the differences have the same orientation and spatial frequency, they are therefore easier to see. On the contrary, if the differences are superimposed on a different background, they become harder to detect.

- Once each subband of distortion is calculated, it can be used to calculate at one hand a distortion map and on the other hand, the estimated quality value of the distorted images.

The quality measure is calculated as follows:

$$Q = \frac{1}{F \cdot O} \sum_{f=1}^F \sum_{o=1}^O w_f \log \left( \frac{1}{I} \sum_{i=1}^I D^2[f, o](i) + \epsilon \right) \quad (2.20)$$

with  $F$  the number of frequencies,  $O$  the number of decomposition,  $I$  the number of pixels in the subband and  $\epsilon$  a small constant to avoid singularities. The weights  $w_f$  were trained to fit in the best way possible, the Mean Opinion Score (MOS) obtained from subjective tests. The weights  $w_f$  were optimized for HDR content in [54].

#### The case of unaligned frequencies

In practice, the frequencies associated with each image subband during the steerable pyramid decomposition are defined using the angular resolution (pixel by degree). This value depends on the display size and viewing distance. The same image may have different angular resolution for each viewing condition.

For HDR-VDP-2, weights  $w_f$  are defined for images with an angular resolution of 60 pixels / degree. The frequencies assigned to each subband image are then as follows: 30, 15, 7.5, ... cpd (cycle per degree). However, if the images have an angular resolution of 40 pix/deg, the frequency associated with each sub-band is then: 20, 10, 5, ... cpd. In this case, no weight is directly associated with the subband.

To assign a weight to this subband anyway, HDR-VDP-2 uses a linear interpolation. For example, since the weight associated with a subband of 15 cpd is 0.2832 and the weight associated with a subband of 7.5 cpd is 0.2142, the weight associated with a subband of 10 cpd calculated through linear interpolation is then 0.2372.

HDR-VDP-2 uses color information to correctly calculate photopic and scotopic luminance. This means that it models the Purkinje shift: at low luminance, the luminance peak sensitivity of the eye shifts toward the blue end of the color spectrum. However, as for PU-metrics, HDR-VDP-2 does not consider chromatic distortions [82]. This is a color blind quality metric as it only consider the luminance.

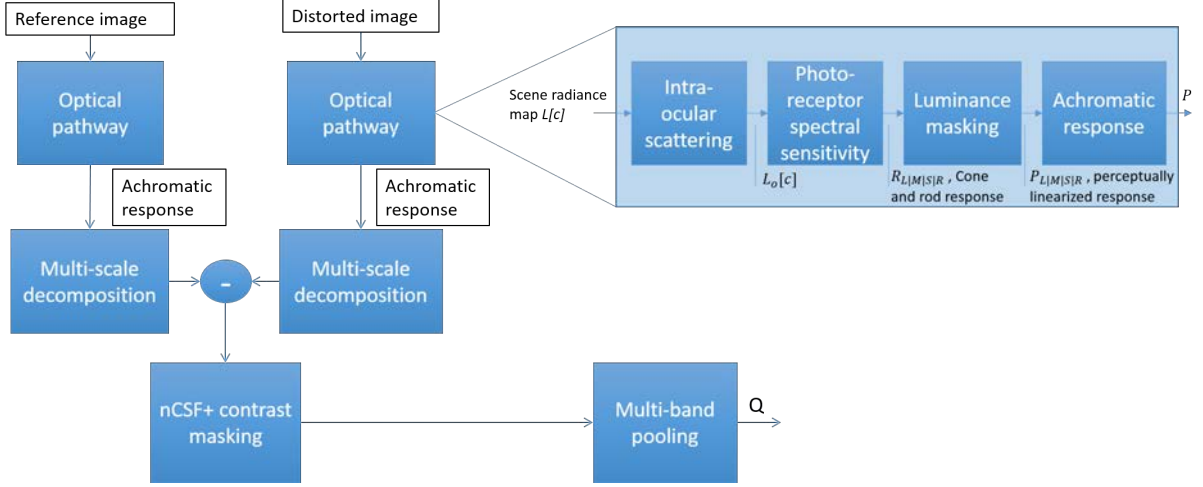


Figure 2.5: HDR-VDP-2 principle

## 2.5 Assessing objective metric performances

The performances of the different objective quality metrics can be assessed using subjective tests. Objective metrics attempt only to model the results from subjective tests. The results coming from objective metrics can be compared to the score of image assessed with subjective scores using different performance indexes that will be presented below.

However, the subjective score or Mean Opinion Score (MOS) given to an image is not necessarily the absolute quality score of this image. The same score coming from different databases can have a different meaning. To tackle this issue a mapping of the objective score has to be applied.

In [83], authors recommend, before computing any performance indexes, to apply a non-linear regression to the objective quality scores thanks to a logistic function:

$$\tilde{Q}_i = a + \frac{b}{1 + e^{-\frac{(Q_i - c)}{d}}} \quad (2.21)$$

where  $Q_i$  is the score of the quality metrics on the image  $i$  and  $\tilde{Q}_i$  the mapped quality score. Parameters  $a$ ,  $b$ ,  $c$  and  $d$  are determined by the non-linear regression. In the following of the thesis, corrected objective score  $\tilde{Q}$  are called  $Q$  to simplify the equations.

Once its objective score have been adapted, the pertinence of a metric can be estimated given several indexes.

Four widely use performance indexes are given below:

- **The Pearson correlation coefficient (PCC)** measures the linearity between two vari-

ables :

$$\text{PCC}_{S,Q} = \frac{\text{cov}(S, Q)}{\sigma_S \sigma_Q} \quad (2.22)$$

where,  $S$  corresponds to the subjective scores (MOS),  $Q$  to the corrected predicted quality score,  $\text{cov}(S, Q)$  is the covariance of  $S$  and  $Q$  and  $\sigma_S$  (resp.  $\sigma_Q$ ) is the standard deviation of  $S$  (resp.  $Q$ ).

- **The Spearman Rank Order Correlation coefficient (SROCC)** measures of the monotony between two variables. Raw scores  $S$  and  $Q$  are first converted to ranks  $rg_S$  and  $rg_Q$ . The SROCC corresponds to the PCC of these two new variables :

$$\text{SROCC}_{S,Q} = \frac{\text{cov}(rg_S, rg_Q)}{\sigma_{rg_S} \sigma_{rg_Q}} \quad (2.23)$$

- **The Outlier Ratio (OR)** [84] represents the quality metrics consistency. It represents the number of outlier point to total points  $N$ :

$$\text{OR} = \frac{\text{Total of Outlier}}{N} \quad (2.24)$$

An outlier is defined as a point for which the error exceeds the 95 percent confidence interval of the mean MOS value (cf. Section 2.2).

- **The Root Mean Square Error (RMSE)** measures the accuracy of the quality metrics :

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (S(i) - Q(i))^2} \quad (2.25)$$

where  $N$  is the number of quality score.

Those indexes allow to compare the significance of each objective metric.

## 2.6 Conclusion

In this chapter we have presented the main techniques to perform image quality assessment whether using subjective tests, whether using objective quality metrics. In addition we also have presented this techniques in the contexts of HDR images. We have presented three existing HDR databases annotated with MOS. In addition, we also have presented the existing objective metrics adapted to HDR.

However, neither the issue of WCG images and neither the issue of chromatic distortions were tackled by the existing IQA methods. There is no existing subjective tests of WCG and there is no objective metrics sensitive to chromatic distortions.

# PART I: CONCLUSION

---

In this part, we have introduced the main concepts behind HDR and WCG content and why it increases our capacity to create more natural images. We have also described the new image color spaces designed to handle this new image/video contents and the adaptation which needs to be made consequently on video codecs. We have also tackled the issue of image quality assessment.

Moreover, we have reviewed several HDR image databases annotated with quality scores. Those databases have some limitations: first, they all use exactly the same HDR display model. However, the HDR displays should be quite diverse in the future with various dynamic ranges and peak brightness available. But more importantly, those databases, due to the screen limitations, do not consider WCG images and, therefore, provide images with low distortions on chromatic components.

The same problem occurs with the objective metrics as there is no HDR metric sensitive to color distortions. They also have been only tested against standard gamut images. WCG related artifacts are not being considered yet.

As chromatic errors can occur during the compression, especially when using a wider color gamut, this thesis focus in tackling the following issues: creating image databases with realistic chromatic errors and creating metrics that can be sensitive to HDR/WCG chromatic errors.



PART II

# **Building HDR/WCG experimental environment and analysis**

---

## PART II: INTRODUCTION

---

In the field of multimedia in general and image and video in particular, subjective tests are necessary for quality assessment to reflect the human perception of quality. Indeed, the goal of many signal processing is to transmit information, in our case images or videos, to another human. Therefore, the quality of many image/video processing like compression depends on the human appreciation. Thus, such processing cannot afford to avoid any kind of user studies.

Large and numerous subjective tests adapted to the legacy SDR image and video format (LIVE [85], CSIQ [86], TID2008 [87]) were created. However, because the topic is recent, subjective tests performed on HDR/WCG contents are rather small or not representative of the variety of distortions that can occur with this kind of contents. The three limited HDR databases presented in the previous chapter, only study HDR images with the NT.709 gamut. This was due, until recently, to the lack of professional displays able to handle the BT.2020 gamut. They all used the SIM2 HDR47ES4MB display, which was one of the only HDR professional display available at the time. The recent display Sony BVM-X300 extends the gamut available to display images. Thus in this part, we propose two new databases annotated with MOS called HDdtb and 4Kdtb respectively. We focused those databases on the chrominance distortions that can occur with this larger gamut.

New databases allow us to extend the previous analysis carried out on the state of the art [88] [89] [90] [50] [91] [57]. In this part, we also evaluate the pertinence of the existing metrics against the experimental data at our disposal: the three pre-existing databases Narwaria et al., Korshunov et al. and Zerman et al. (cf. section 2.3) and the two newly created databases, HDdtb and 4Kdtb.

This part is organized as follows. In the first chapter, the two new databases are presented and their results commented. Then in a second chapter, we benchmark existing objective metrics. We analyze specially HDR-VDP-2 as it is a metric that is often considered as a reference metric for HDR content.

# NEW DATABASES PRESENTATION

---

## 3.1 Introduction

In this chapter we present experimental data we have acquired across two subjective tests to complement existing subjective studies by considering the following part:

- First, both databases deal with distortions on content encoded using the BT.2020 WCG. Previous database used the smaller gamut BT.709 instead. The first one, HDdtb, is a database composed of BT.709 contents encapsulated in a BT.2020 gamut and the second one, 4Kdtb, is composed of native BT.2020 content.
- Second, both present realistic chromatic artifacts that are meant to challenge luminance-only objective quality metrics.
- Third, these databases allow us to evaluate the gain in term of quality of the chroma Qp algorithm proposed by MPEG and presented in Section 1.4.1.

The two new databases are available at [www-percept.irisa.fr/software/](http://www-percept.irisa.fr/software/).

This chapter is separated in three sections: the first presents the test methodology, the second presents HDdtb and the third the 4Kdtb.

## 3.2 Experimental protocol

In order to deal with the limitations of existing datasets, we have designed new subjective tests. In the following we present the key ingredients of those tests.

First, we use a different display than the state of the art databases: the SONY BVM-X300 (cf. Figure 3.1). Previous database used the Sim2 HDR47ES6MB. the display SONY BVM-X300 is a professional HDR video monitor able to faithfully display the brightness of signals [92]. This display uses a different technology (OLED) than the display used by previous database (Led-LCD). One of its most interesting feature is its ability to handle most of the BT.2020 gamut. This make this display particularly suitable to analyze the impact of chromatic distortions. It has a peak brightness at  $1000 \text{ cd/m}^2$  and a luminance of a black pixel that was too low to be

measured by our equipment ( $<0.2 \text{ cd/m}^2$ ). In the rest of the thesis, we assume a luminance for the black pixel at  $0.001 \text{ cd/m}^2$ . It has a diagonal of 30 inches (750.2 mm) and is compatible with 4K images ( $4096 \times 2160$  pixels). This monitor also allows us to force the use of a chosen EOTF without having to consider the image metadata.



Figure 3.1: SONY BVM-X300 display

To display the images on the screen, we used the b<>com \*Ultra Player\* which allows distributing uncompressed YUV content with a 10 bits quantization and 4:2:0 chroma sub-sampling. For the down-sampling we applied the luma adjustment methodology presented in subsection 1.4.1. The connection to the screen was done using 3G-SDI cables.

For both subjective tests, we used the Double-Stimulus Impairment Scale (DSIS) variant I methodology [47] with a side-by-side comparison. Side-by-side comparison aims to make the viewers more sensitive to chromatic distortion. Pairs of images were presented to the viewers. One side of the screen is always the reference. To avoid a bias due to the position of the image on the screen, 50% of participants had the reference always on the right-hand side, 50% always on the left-hand side. To avoid bias with the order of presentation, the pairs of images were randomized for each participant with the condition that the same content was never shown twice consecutively. Figure 3.2 represents a image pair the viewers had to score.

To limit visual fatigue, the test is limited in times, thus limiting the number of images we can show to the viewers. Each image pair was shown 10 second. Between each image pair, the viewer has 5 second to vote. During this time a neutral grey is display on the screen to avoid that the quality perception of an image is not impacted too much by the previous one. The test sessions last 35 minute (including instructions and training time) with a 5 min pause in the middle of the test. The viewers were placed at 3.2 times the height of the displays or 1.15 meter. We measured the luminance of the room at  $40 \text{ cd/m}^2$ .

The viewers were asked to rate the quality of each distorted images on a scale from 0 to 100 on scale similar as the one presented in [57], thus to allows the viewer to note precisely the quality score of an image. This scale was associated with five labels: "Very annoying", "Annoying", "Slightly annoying", "perceptible but not annoying", "imperceptible". This scale is illustrated on Figure 3.3.



Figure 3.2: Example of a pair of images presented to the viewer (Reinhard et al. TMO [93]).

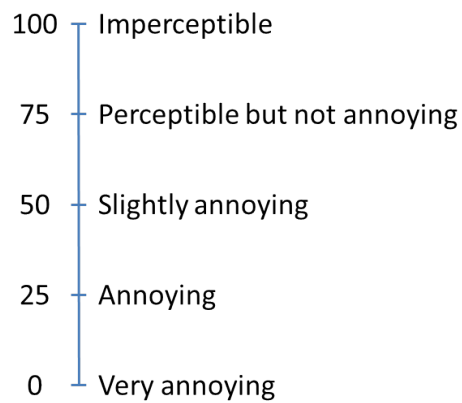


Figure 3.3: Proposed subjective tests scale

### 3.3 First database: HDdtb

The first database we propose is composed of HDR images with a BT.709 gamut. We encapsulate those images inside the BT.2020 gamut to create more chromatic distortion during the compression. This dataset is called HDdtb.

With this database, we select 8 images with a large variety of characteristics. The goal is to be representative of the diversity of HDR contents. We select four different kinds of distortions. All of them aim to create chromatic artifacts.

To create this database, we used a side-by-side methodology DSIS. However, we had only one display Sony BVM-X300. To be able to display simultaneously both the reference image and the distorted image on the same screen, we used half of HD images ( $1920 \times 1080$  pixels). Moreover, we add a little black space between the two images. The new image dimensions become  $944 \times 1080$  pixels. As the viewers were placed at 1.15 meter, it means that the images have an angular resolution 60 pixels per degree (pix/deg).

Fifteen naive subjects participated in this test (11 males, 4 females) with an average age of 25.8. All declared normal or corrected-to-normal vision. One participant was removed from the analysis using the methodology described in [47].

#### 3.3.1 Images descriptions

To create this database, eight images were selected from 3 collections: two are from the MPEG HDR sequences (FireEater and Market) [94], one is from the Stuttgart HDR Video Database (Showgirl) [95] and the remaining five images are from the HDR photographic survey [96]. Note that these images also belong to Zerman et al.'s database [50]. The characteristics of the images are not exactly the same as in Zerman et al. [50] because we used only half of the images.

Tables 3.1 to 3.3 present the detailed characteristics of the images: a thumbnail of the images using the Reinhard et al. TMO [93], the histogram of the pixels luminance and finally pixel chromaticity coordinates are plotted in the CIE xy diagram. For the last, as the  $XYZ$  color space (and the  $xy$  coordinates) makes sense mostly under scotopic conditions, we only consider the pixels that have at least a luminance of  $10\text{cd/m}^2$ . The image description indexes (presented in section 2.3.1) can be found on Figure 3.4.

In the following, we describe one by one each selected reference images:

- **FireEater**: This image represents a fire eater shot during the night (cf. Table 3.1). This is the darkest image of the database (lowest key). This is also observable on the image luminance histogram. Due to its subject, FireEater possesses a high dynamic range: the background is very dark and the fire very bright. The image possesses also the highest

colorfulness but most of its pixels  $xy$  coordinates are located in the red corner of the BT.709 gamut. It is also the simplest (lowest SI) image of the database.

- **LasVegasStore**: This image represents a street in Las Vegas shot during the night with a lot of neon lights (cf. Table 3.1). This image has a high dynamic range due to the contrast between the dark night and the neon lights but less than FireEater and Showgirl. Those two images were shots with a camera that was able to capture darker pixels. Its key is higher than FireEater as the neon lights are brighter and more numerous than the Fire. It is also a colorful image and its pixels cover almost entirely the BT.709 gamut. This image is more complex than FireEater.
- **Market3**: This image represents a market on a sunny day (cf. Table 3.1). This is the brightest image of the database with some regions as bright as the peak brightness of the display. It is also quite colorful and its pixels cover the entire BT.709 gamut. This image is also the most complex of the database.
- **MasonLake(1)**: This image represents a lake on a sunny day (cf. Table 3.2). This image has the lowest dynamic range of all images. It is a colorful image but its pixels are mostly situated in the blue area of the BT.709 gamut. This is also a simple image, large parts being almost uniform blue regions (the sky and the lake).
- **RedwoodSunset**: This image represents a beach that was shot during sunset (cf. Table 3.2). It can be decomposed in two regions. The grey and dimmed area of the beach and the brighter area of the sunset. This image possesses an average dynamic range and an average key. This image is also not very colorful and possesses a low complexity.
- **Showgirl**: This image represents a woman reflecting herself on a dressing table mirror (cf. Table 3.2). Having a portrait is an important asset of a database as faces are regions of interest for the HVS [97]. This is the image with the highest dynamic range due to the presence of black bands on the top and on the bottom of the image. Its dynamic range falls drastically when using only the core of the image from 6 log units to 2.85 log units. This image possesses a high key but as for the dynamic range, the key fall when considering only the interesting part of the image and pass from 0.68 to 0.36 so the overall brightness is not very important for this image. The complexity and the colorfulness do not change significantly using the complete or the crop image. It possesses a complexity similar to LasVegasStore. The image possesses an average colorfulness. Many of its pixels are located in the skin tone region of the BT.709 gamut.
- **Typewriter**: This image represents a typewriter and a book with text (cf. Table 3.3). Writings are generally difficult to compress. Indeed, compression artifact can make the

text hardly readable. This image possesses an average dynamic range and is quite dark (even with the presence of an emissive light source). This is the less colorful images and is mostly composed of black and white. The complexity is not very high but due to the presence of text, the SI indicators might not be in this case the best complexity measure.

- **UpheavalDome:** This image represents a rocky region during the day (cf. Table 3.3). This image possesses a low dynamic range and a key superior to the average of the database. It possesses a low spatial complexity and a medium colorfulness. It possesses a lot of rock texture.

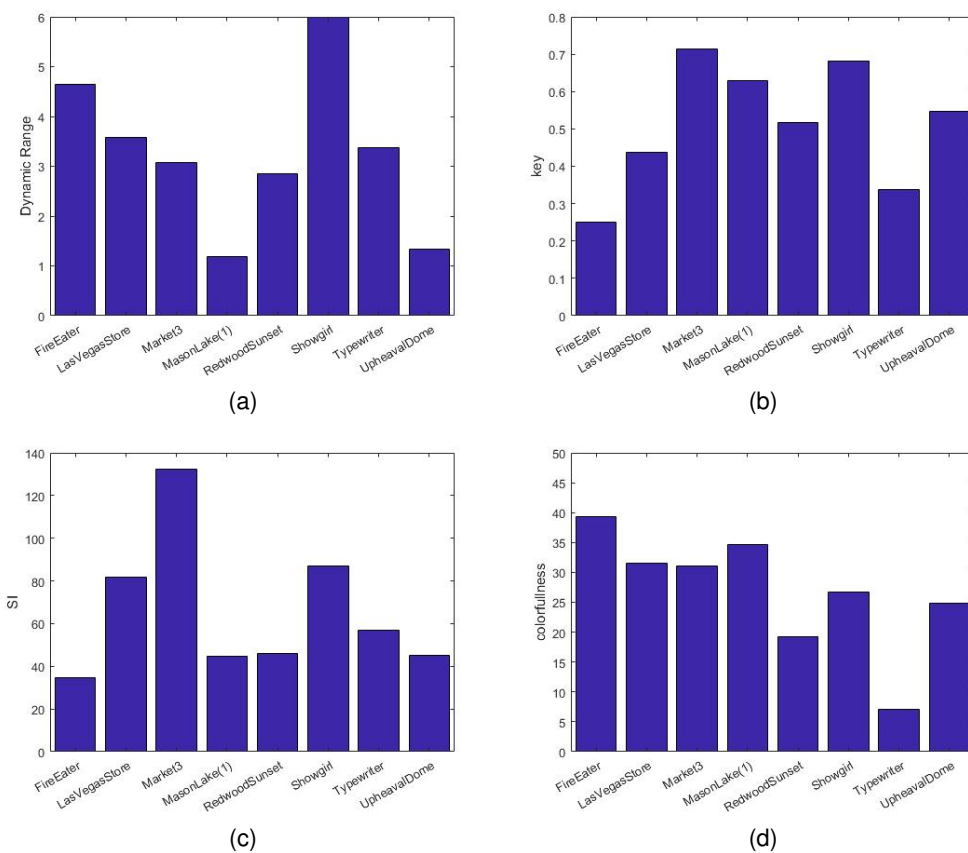


Figure 3.4: Characteristics of the HDdtb images: **(a)** The dynamic range, **(b)** key, **(c)** spatial Information, **(d)** Colorfulness.

Table 3.1: Description of HDdtb images (FireEater, LasVegasStore and Market3).


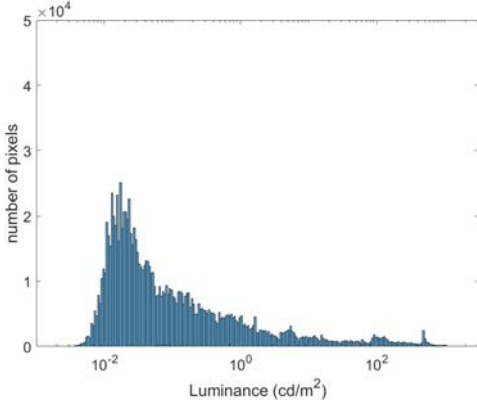
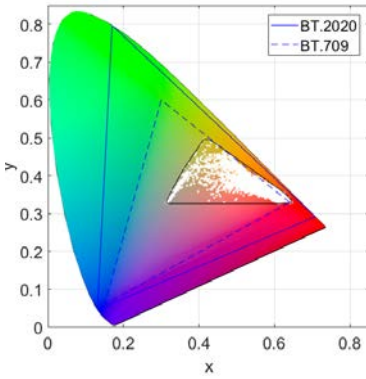

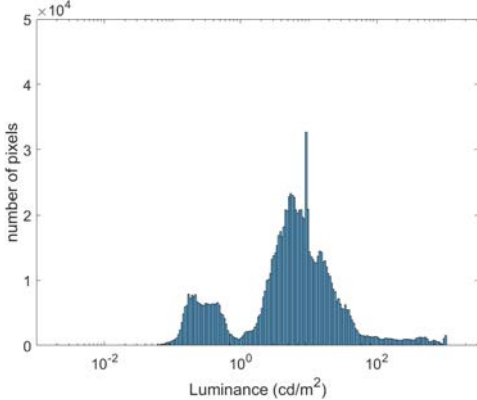
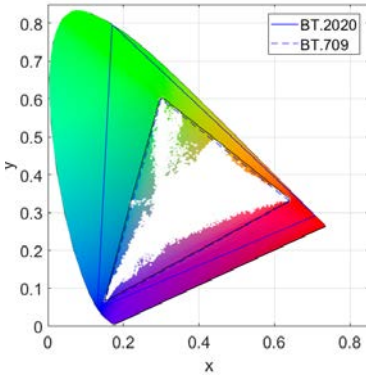

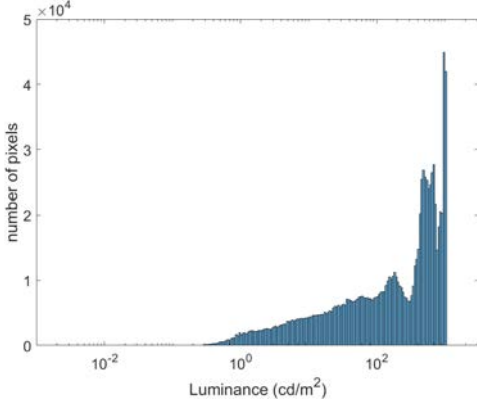
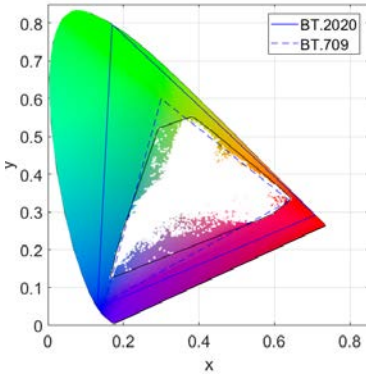
Thumbnail (Reinhard et al. TMO [93])	Luminance histogram	Pixels coordinates in the CIE xy diagram
	<b>FireEater</b>	
		
	<b>LasVegasStore</b>	
		
	<b>Market3</b>	
		

Table 3.2: Description of HDdtb images (MasonLake(1), RedwoodSunset and Showgirl).


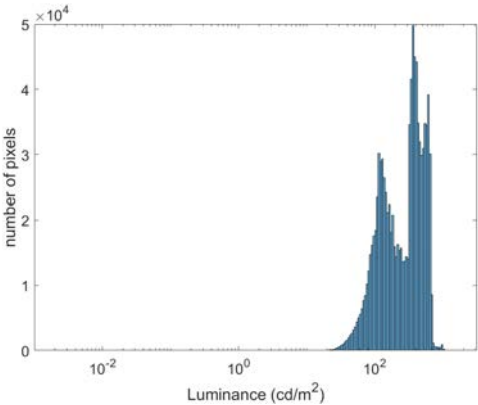
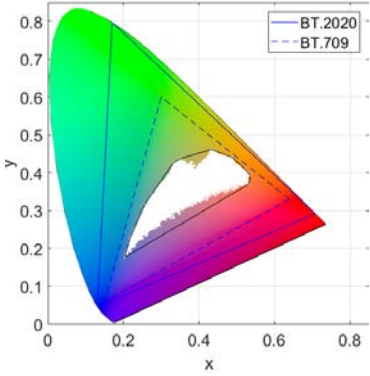

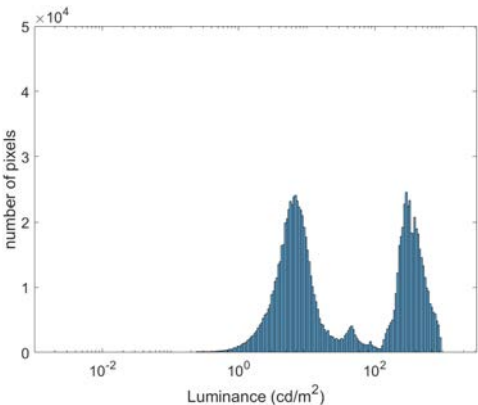
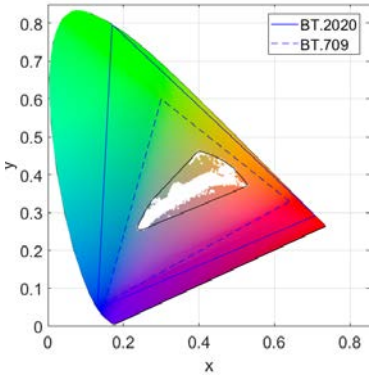

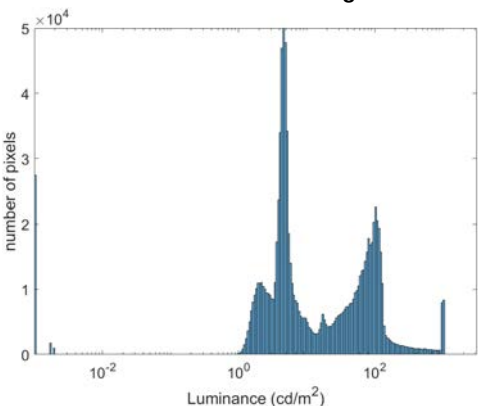
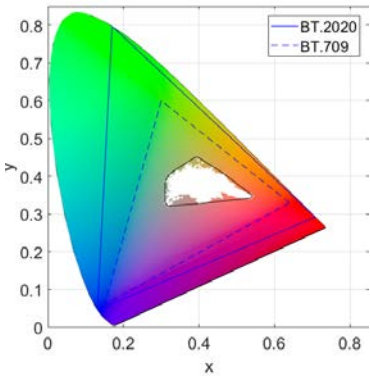

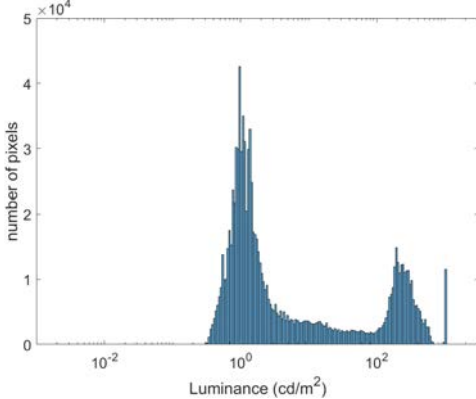
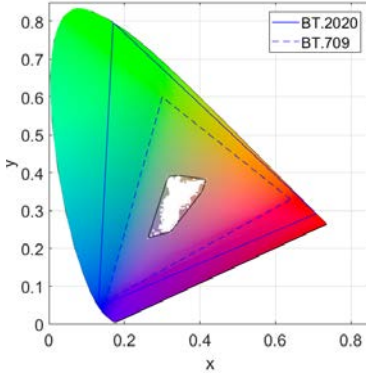

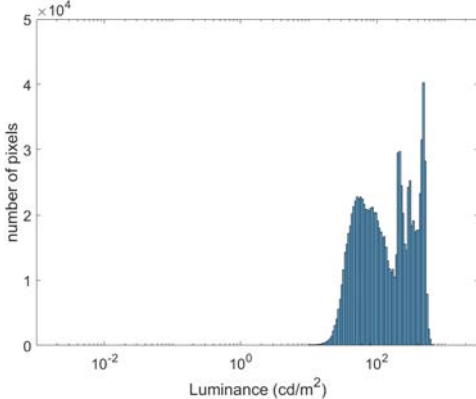
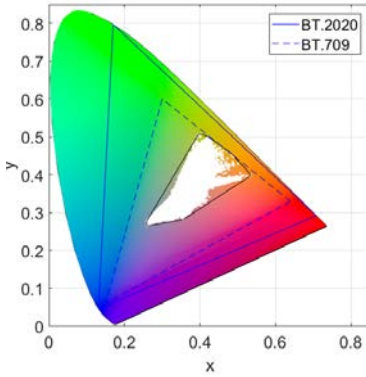
Thumbnail (Reinhard et al. TMO [93])	Luminance histogram	Pixels coordinates in the CIE xy diagram
	<b>MasonLake(1)</b>	
		
	<b>RedwoodSunset</b>	
		
	<b>Showgirl</b>	
		

Table 3.3: Description of HDdtb images (Typewriter, UpheavalDome).

Thumbnail (Reinhard et al. TMO [93])	Luminance histogram	Pixels coordinates in the CIE xy diagram
<b>Typewriter</b>		
		
<b>UpheavalDome</b>		
		

### 3.3.2 Distortion descriptions

To create the distortions, we used HDRTools (v0.16) (Available at <https://gitlab.com/standards/HDRTools/>) to apply format conversion, chrominance sub-sampling or gamut conversion. For the compression and decompression of the images, we used the reference software of HEVC, the HEVC Test Model (v16.17) (Available at <https://hevc.hhi.fraunhofer.de/>). Four kinds of distortions have been chosen:

- HEVC compression using the recommendation ITU-T H Suppl.15 [7]. Four different Qp were selected for each image for this distortion. The selected Qp are 15, 27, 31 and 39 for all images except two. Compression artifacts are harder to detect for the image "Market3" so we increase the selected Qp value: 23, 31, 39 and 43. For the image "TypeWriter", the presence of text have the tendency to highlight compression artifact: the text become

easily unreadable. For this image the selected Qp are: 15, 23, 27, 31.

- HEVC compression without the chroma Qp adaptation (cf. section 1.4.3) leading to more chrominance distortions. For each image, we use the three higher Qp selected for the previous distortion. The presence of chromatic artifact is hardly detectable for the lower Qp.

Figure 3.5 shows an example of the difference of compression between the two selected mode of compression. It represents a tone-map version of the image Showgirl compressed with a Qp 39 with and without the chroma Qp adaptation. It should be noted that the tone-mapping, the quantization to 8 bits and the resizing of the image have the tendency to lessen the effect of chromatic artifacts. This is why in this case we use a high Qp to illustrate this distortion. We can observe chromatic distortion on the image that is not using the chroma Qp adaptation.

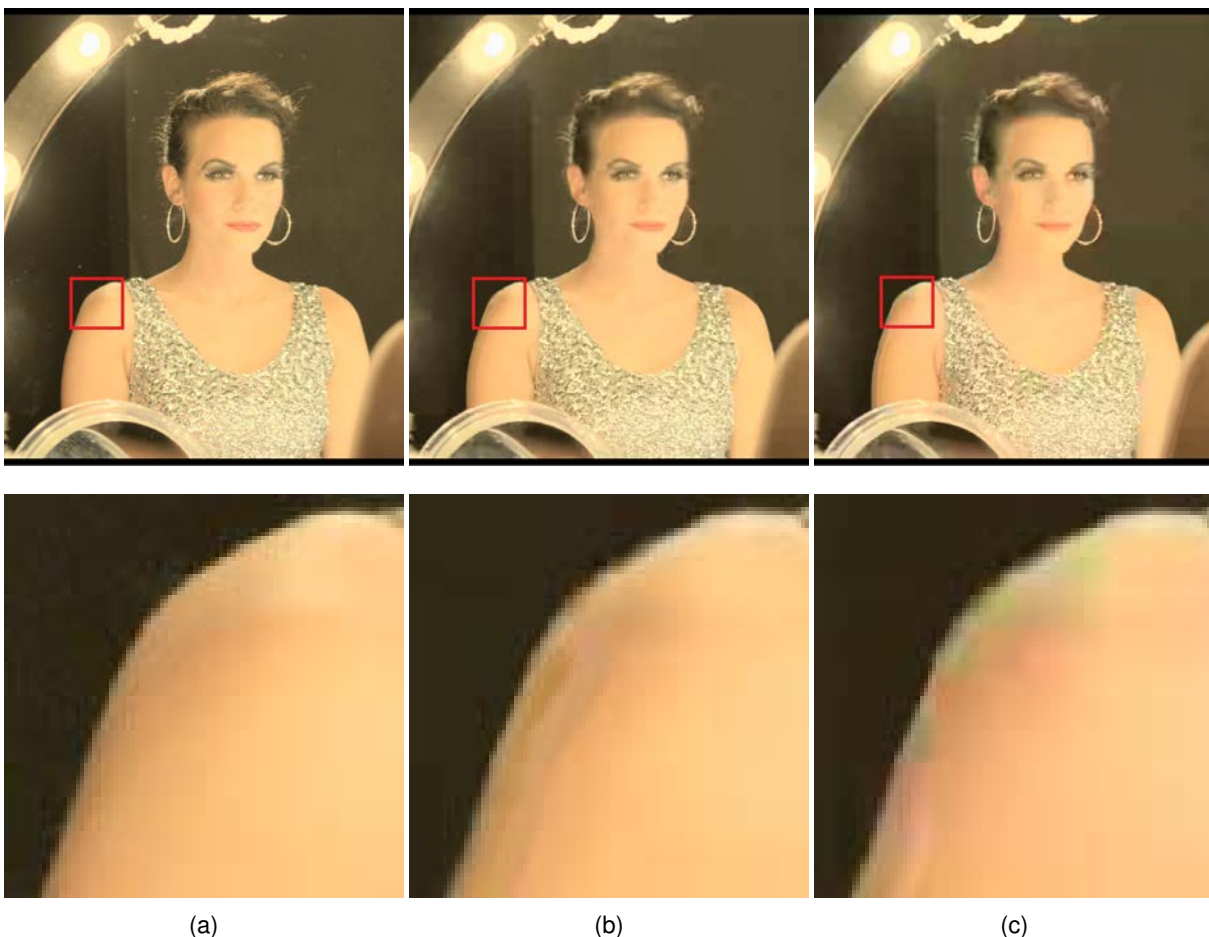


Figure 3.5: Example of compression artifacts on the image Showgirl: **(a)** The original image, **(b)** Compressed image, Qp 39, with Chroma Qp adaptation, **(c)** Compressed image, Qp 39, without Chroma Qp adaptation.

- Gaussian noise on the chroma components using the  $Y'C'_bC'_r$  color space with a PQ transfer function: 3 levels of noise measured using a signal to noise ratio (SNR) were selected: 0.5, 1 and 3. Figure 3.6 illustrates this distortion for the smallest SNR, the tone-mapping having the tendency to lessen the noise visibility.

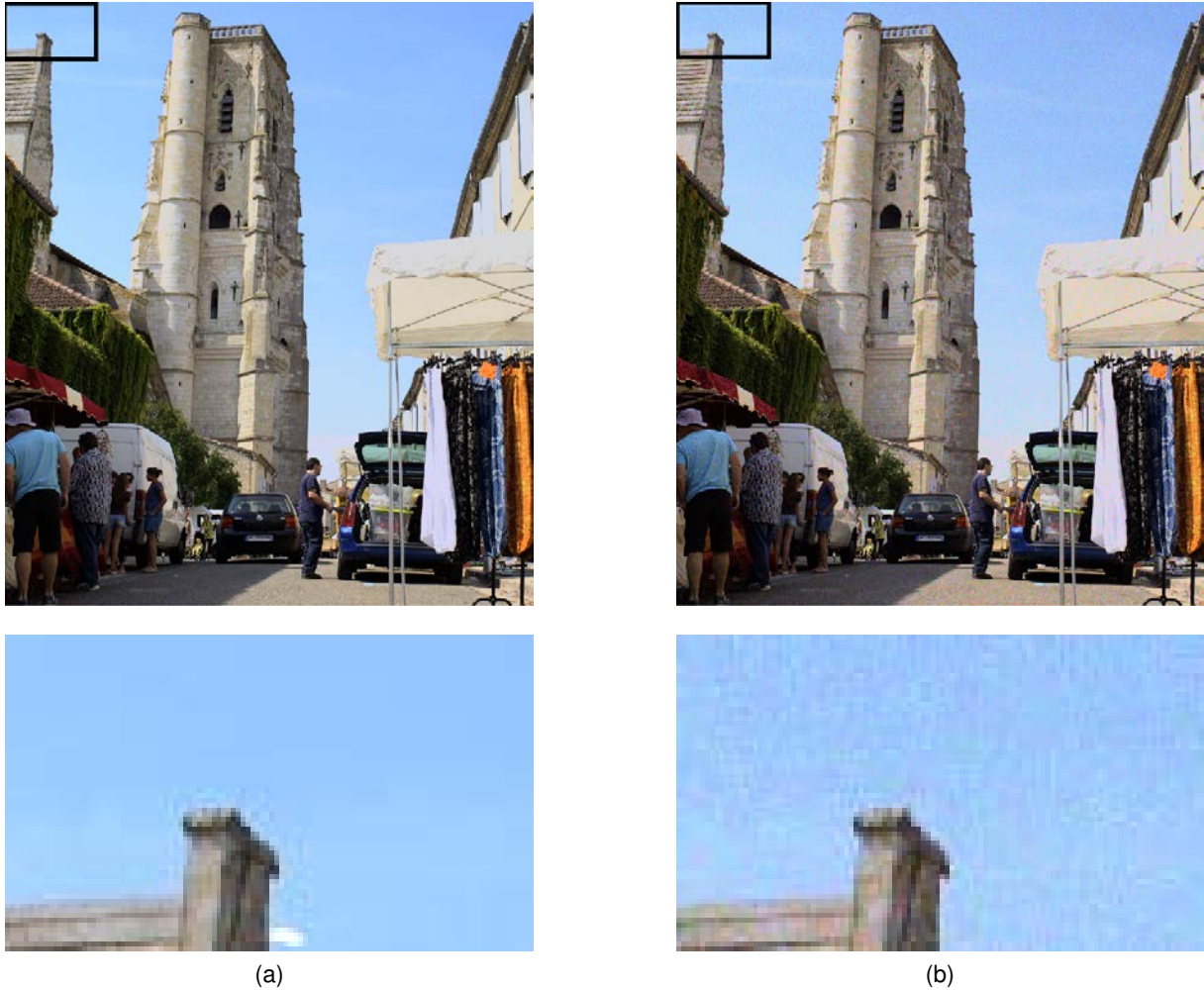


Figure 3.6: The image Market3: **(a)** The original image, **(b)** with Gaussian noise on the chroma (SNR=0.5)

- Gamut mismatch: two kinds of distortion were created: on one hand, the BT.709 images were considered as if they had been already encapsulated in a BT.2020 gamut leading to more saturated images. On the other hand, we took images already encapsulated in a BT.2020 gamut and considered them as BT.709 images and re-encapsulated them in a BT.2020 gamut (using HDRTools). This creates less saturated images. Thus the BT.2020 image is displayed as a BT.709 image. Figure 3.7 illustrates this default. We also plot the chromaticity coordinates on the CIE xy diagram. We can observe the variation in term of used gamut.

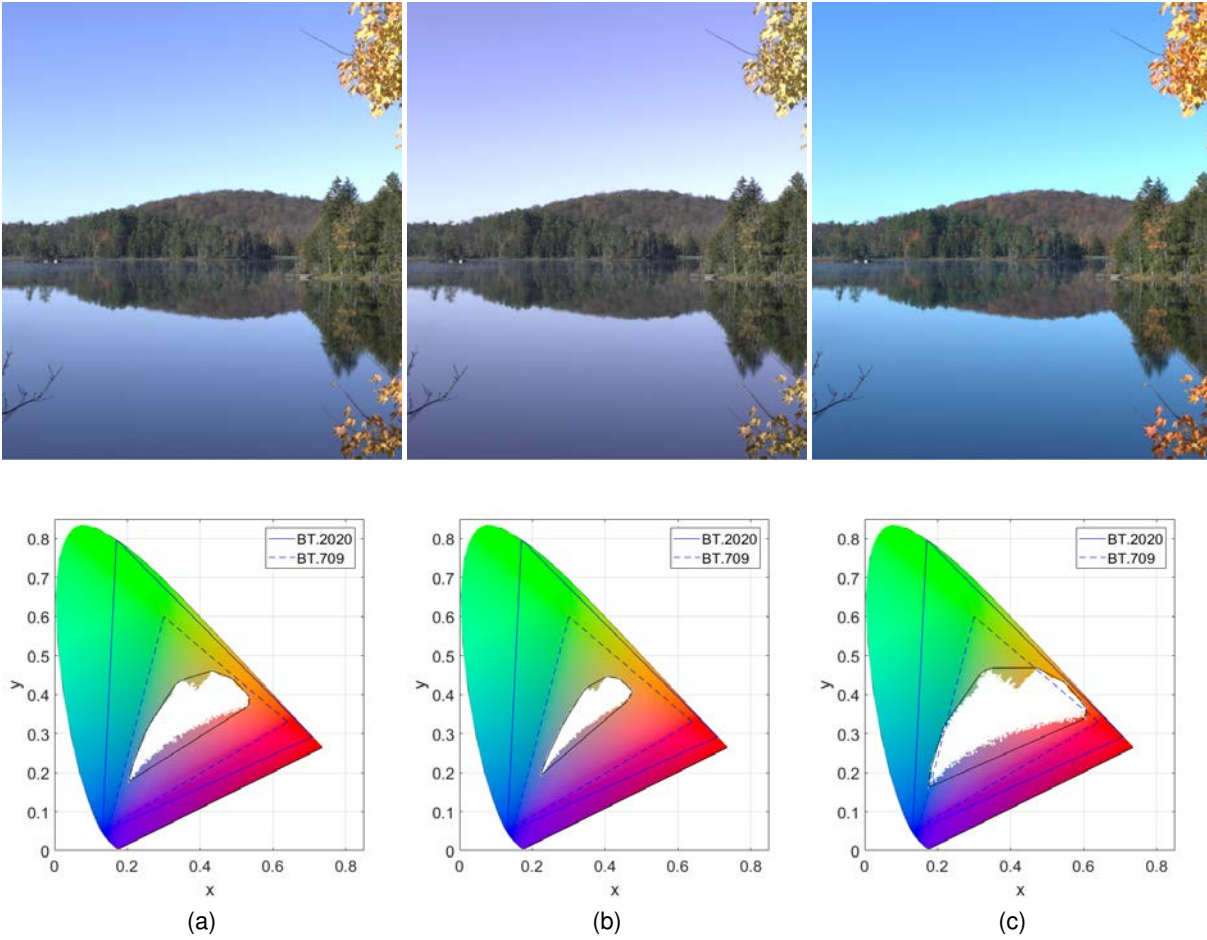


Figure 3.7: The image MasonLake(1) under three conditions: **(a)** The original image, **(b)** gamut mismatch: BT.2020 image displayed as a BT.709 image **(c)** gamut mismatch: BT.709 image displayed as a BT.2020 images.

### 3.3.3 Subjective quality analysis

Figure 3.8 represents the repartition of the MOS scores for each kind of distortion. We can observe that the MOS of compressed or Gaussian noise corrupted images are well distributed across all the quality scale although there is less image with a score above 75.

The chromatic distortion created by the compression without using the chroma Qp adaptation is not impacting the quality as much as we expected on this database. We observe that the MOS scores obtained for compressed images with and without chroma Qp adaptation are strongly correlated ( $pcc=0.95$ ). This suggests that observers are consistent in their quality scoring whether or not the chroma Qp adaptation is used. As illustrated by Figure 3.9 that represents the MOS scores of each image in function of the Qp, there is no significant difference between compression with and without chroma Qp adaptations except for one image "Red-

WoodSunset". This observation raises some questions about HDR images and their quality evaluation.

Concerning the images with the "gamut mismatch" artifact, they all have a MOS score above 50 (slightly annoying). Most of them have a score between slightly annoying and perceptible but not annoying. For most images, the distortion is important and clearly observable, but it is not necessarily associated with a loss of quality. Moreover, this distortion was reported quite hard to score according to the test participants. This default has also the highest standard error because it was depending too much on the viewer personal taste. This questions the pertinence of our subjective test with this distortion. On one hand, it is interesting to assess that some artifacts do not directly affect the quality perception of images. On the other hand, such artifact visibly modify the image and can change the artistic impact of the images (color way more saturated or more dulled ...).

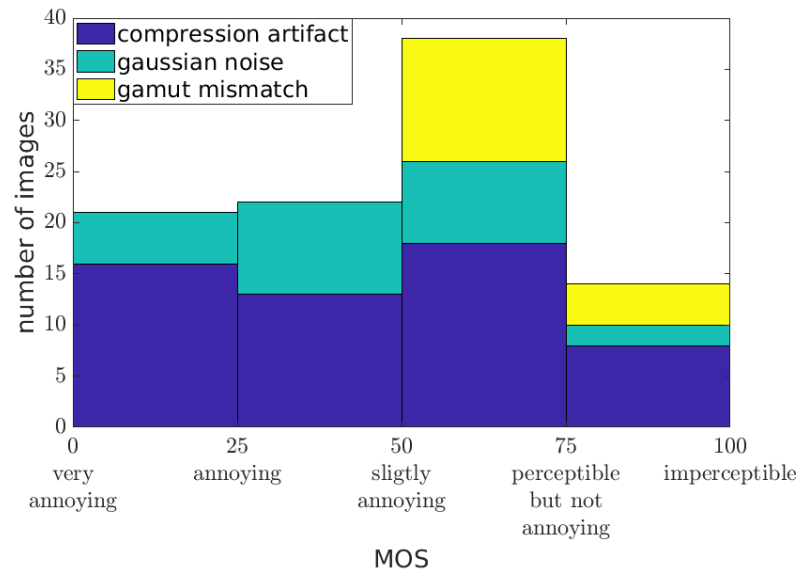


Figure 3.8: Repartition of HDdtb MOS scores.

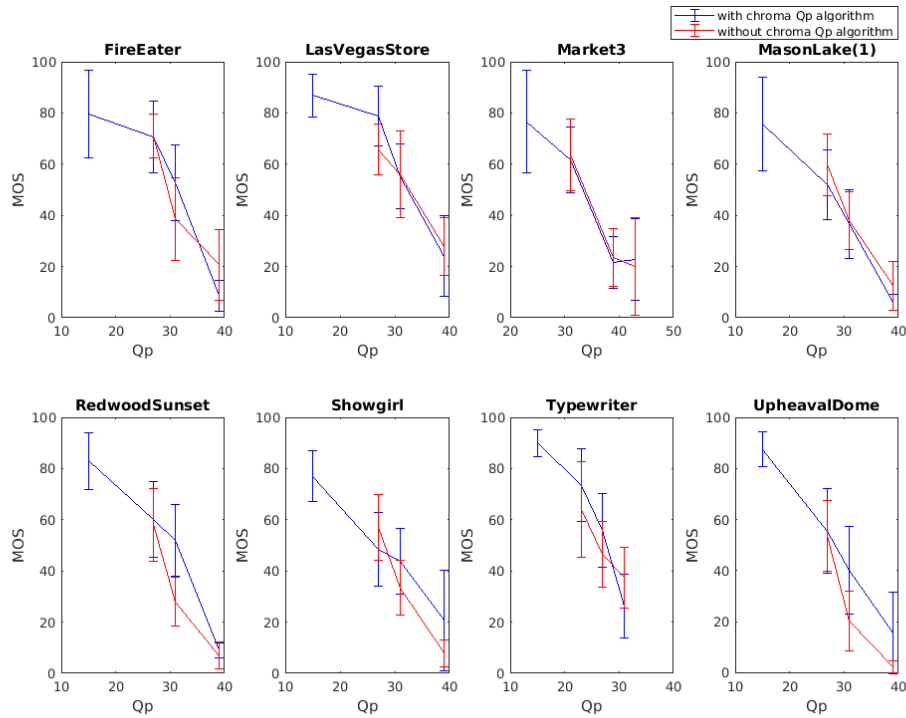


Figure 3.9: MOS score obtained from 14 naive viewers in function of the Qp.

As HDR images are still quite confidential, we may ask the following question: are naive observers able to evaluate small to medium distortions on this new format? To answer this question, 9 video quality experts rated the images using the same methodology. This is illustrated by Figure 3.10 that represents the MOS reported by naive viewers and the MOS reported by experts. The two MOS series have a huge correlation ( $pcc=0.9455$ ). Figure 3.11 represents the expert MOS scores of each image in function of the Qp. It shows that, on the image "Market3", the chroma Qp adaptation brings a small quality gain that was not seen by naive viewers. However, the difference in quality perceived by the experts was not high enough to exceed the MOS confidence intervals thus making it not statically significant.

There is no significant difference in the MOS obtained from naive viewers and the MOS obtained from expert viewers. Two factors can explain this:

- the low number of people that have rated the images (which increase the confidence interval)
- the chroma Qp adaptation gain in overall quality is too small for those images and compression Qp.

This database does not meet all of our goals. Especially, it does not provide sufficient chromatic distortion. This database might not be a huge challenge for objective metrics. As a result we decide to create another database with more prominent chromatic artifacts.

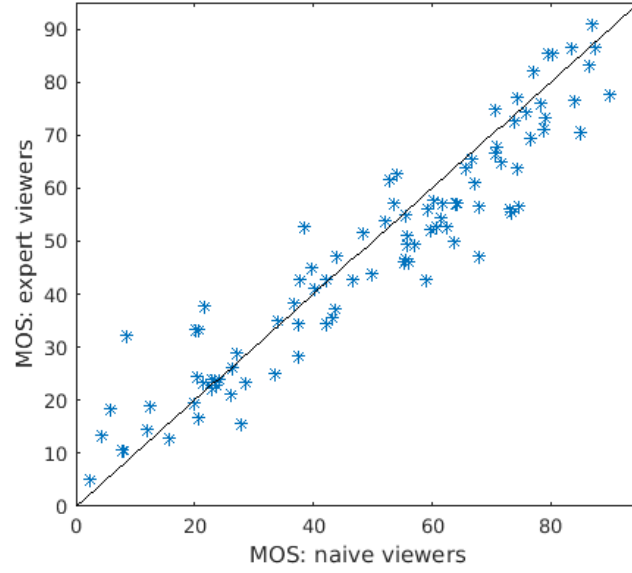


Figure 3.10: MOS obtained from naive viewer in function of MOS obtained from expert.

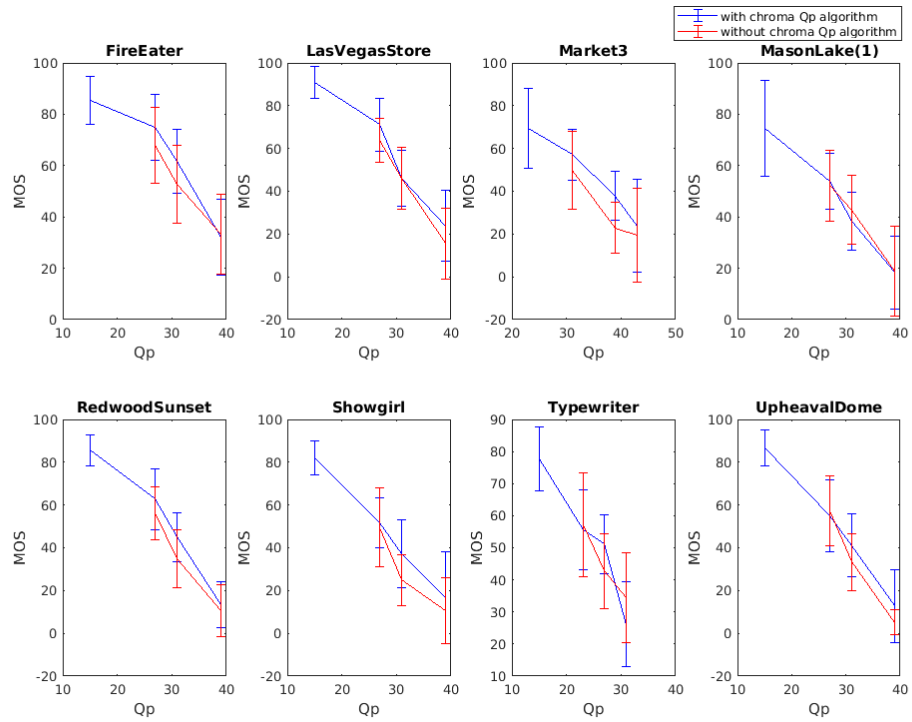


Figure 3.11: MOS score obtained from 9 experts in function of the Qp.

### 3.4 Second database: 4Kdtb

To overcome the limitation of the previous database we have created a second database more representative of the future HDR images and videos:

- The image are native BT.2020 content and not BT.709 content encapsulated in a BT.2020 gamut. Then, the images are more prone to chromatic artifacts.
- The images are cropped images with a 4K resolution ( $3840 \times 2160$  pixels). In the future, the HDR/WCG images will probably be synonym to a high resolution. We name this database 4Kdtb for this reason.

Tone-mapped versions of the selected images are available in appendix B.

As for HDdtb, we can only use half of the screen. We used only part of these 4K images so the reference and the distorted image could fit on our display ( $1890 \times 2160$  pixels) with a band of 60 black pixels. Since we used the same viewing distance as HDdtb, the angular resolution of the images increases and becomes 120 pix/deg.

Thirteen experts or sensitized subjects participated in this test (11 males, 2 females) with an average age of 40. All declared normal or corrected-to-normal vision.

#### 3.4.1 Images description

For the second database, we used eight 4K images produced by Harmonic Inc and extracted from two different clips Bike and Regatta. The characteristics of the images are given in Figure 3.12. This database possess less dark images than the previous database (no key below 0.4). However those kind of images only appears in very extreme and unlikely conditions. Moreover, the SI of this database can not be compared with the other database SI. Indeed, SI indexes values is dependent of the image resolution resolution. As we did not modify the sobel filter of the SI index, it is attended that is values will be lower. Both clips are natural outdoor scene where scene like TypeWriter of the previous database with very low colourfulness are quite unlikely.

In the following, we describe one by one each selected reference images:

- **Bike\_110s:** This image represents a woman crossing a street at the end of sunset (cf. Table 3.4). This is the darkest image of the database especially the street region. There is a lot of specular light due to the water on the street. This is amongst the most colorful images of the database due to the sunset and the cars light. Pixels chromaticity coordinates are mostly concentrated in the red and yellow region of the gamut.
- **Bike\_20s:** This image represents a cyclist in a city and was shot during the day (cf. Table 3.4). It possesses a high dynamic range and a high key. It is a very contrasted image with a dark area under the stairs of houses and bright area in the sky. It is also the more complex image of the database. The image is not very colorful, with a lot of pixel chromaticity coordinates situated in the grey and blue zone of the gamut.

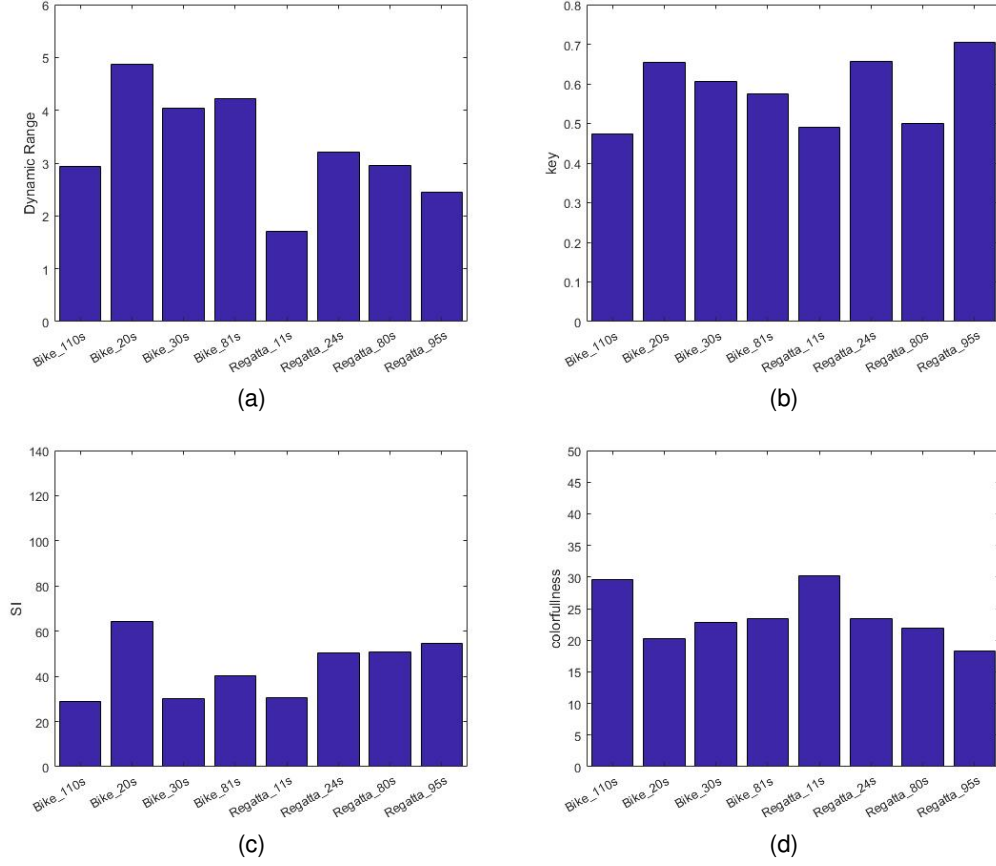


Figure 3.12: Characteristics of the 4Kdtb images: **(a)** The dynamic range, **(b)** key, **(c)** spatial Information, **(d)** Colorfulness.

- **Bike\_30s:** This image is a portrait of a man in a square at the end of the day (cf. Table 3.5). For this database, this image possesses an average dynamic range and an average key. It is an image with low complexity (the sky and the square are almost uniform). The colorfulness is not very high. Most of the saturated color comes from under an arch where the sun shine with yellow and red color. The rest of the pixels is composed mostly of grey and blue area.
- **Bike\_81s:** This image represents a cyclist in an city that was shot at the end of the day (cf. Table 3.5). It possesses a high dynamic range and an average key. It has an average complexity. Most of its colorfulness comes from the sunset with red and yellow region.
- **Regatta\_11s:** This image represents buoys on the water that was shot early in the morning (cf. Table 3.5). It has the lowest dynamic range and is amongst the image with the lowest complexity. This is the image that covers the smallest region of the gamut mostly in the yellow region.
- **Regatta\_24s:** This image represents a woman paddling on a river on a sunny day (cf. Table 3.6). It has an average dynamic range and possesses the second highest key

amongst the database. This image has an average colorfulness. One of the interesting features of this image is the huge specular highlight on the paddle.

- **Regatta\_80s:** This image represents also a woman paddling (cf. Table 3.6). This image has almost the same dynamic range but have a much lower key than Regatta\_24s
- **Regatta\_95s:** This image represents women lifting a boat (cf. Table 3.6). This last image has a rather low dynamic range compare to the other image of the database but it is the image with the highest key of our database. This is also the less colorful image of the database.

Table 3.4: characteristics of 4Kdtb images (Bike\_110s and Bike\_20s).

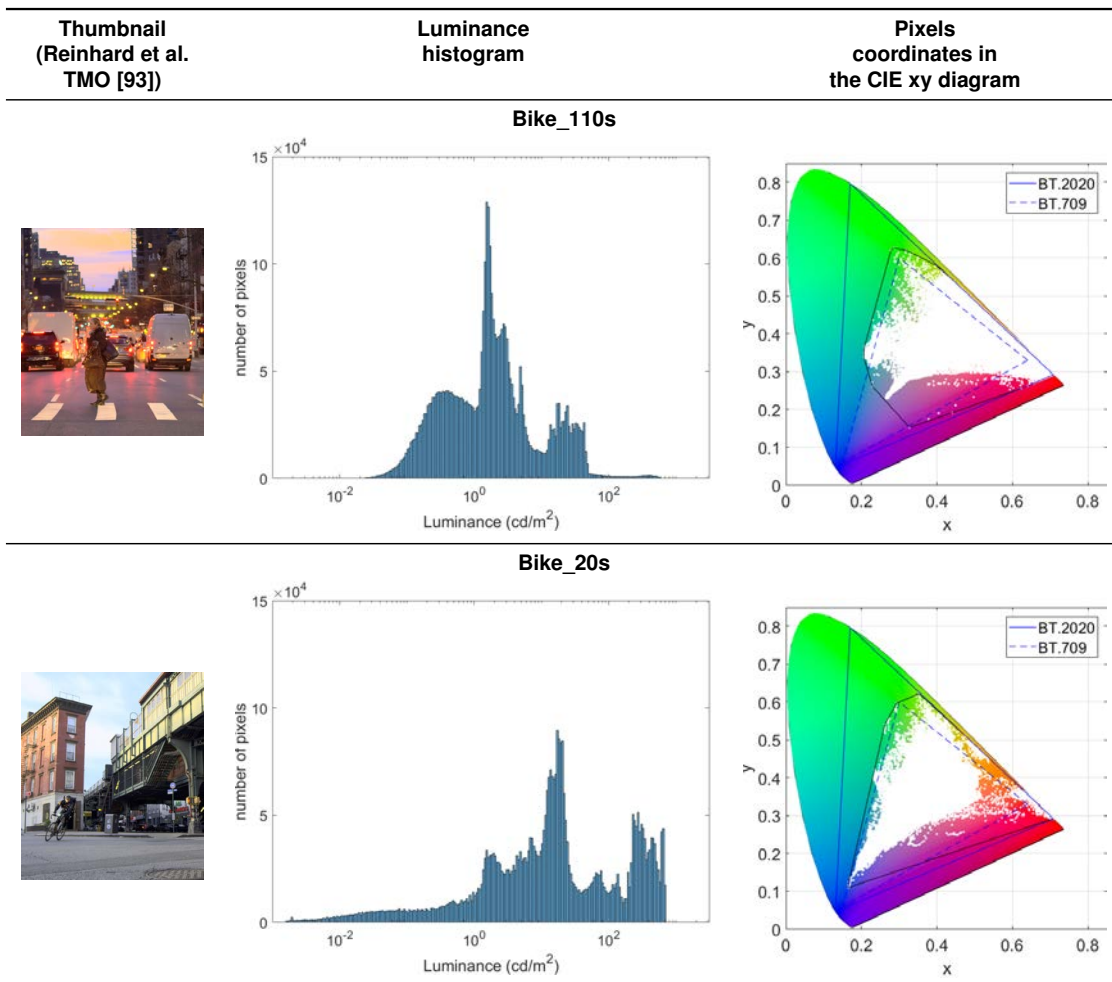


Table 3.5: characteristics of 4Kdtb images (Bike\_81s, Bike\_30s and Regatta\_11s).

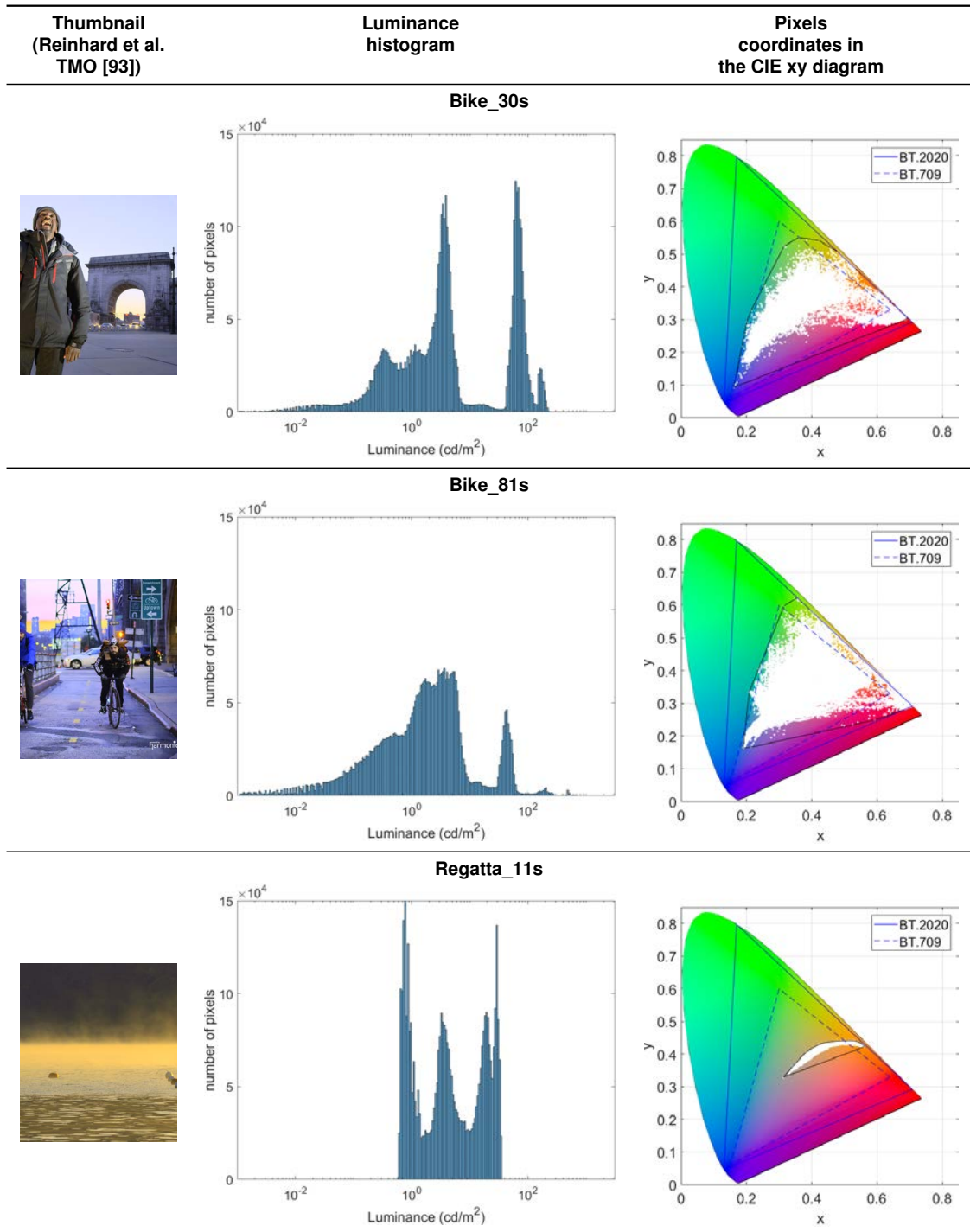

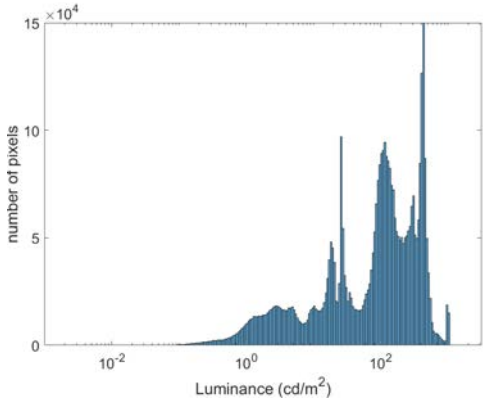
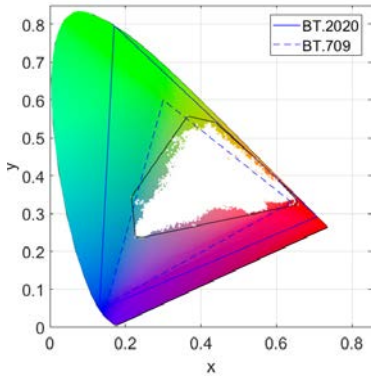

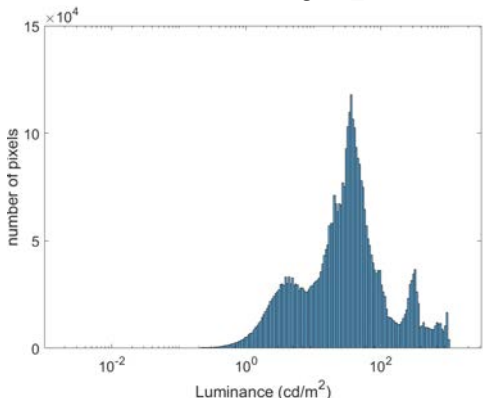
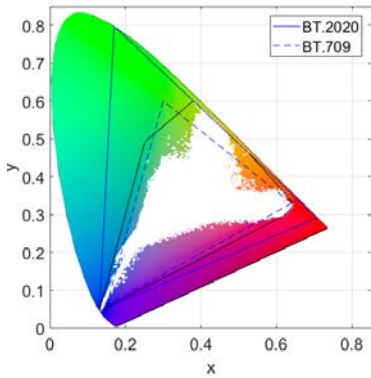

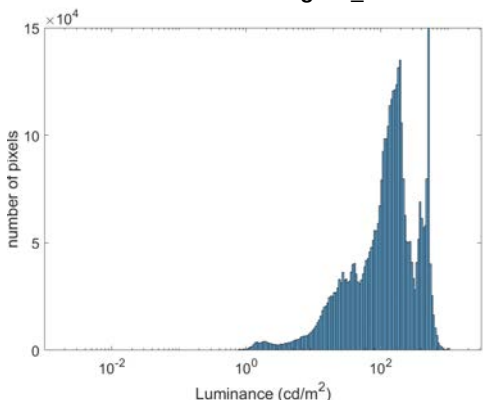
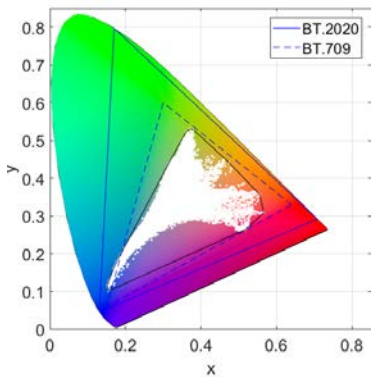


Table 3.6: characteristics of 4Kdtb images (Regatta\_24s, Regatta\_80s and Regatta\_95s).

Thumbnail (Reinhard et al. TMO [93])	Luminance histogram	Pixels coordinates in the CIE xy diagram
<b>Regatta_24s</b>		
		
<b>Regatta_80s</b>		
		
<b>Regatta_95s</b>		
		

### 3.4.2 Distortion descriptions

With this database, we aim to create more visible color artifacts than in the HDdtb database. We compressed the images with four different Qp (20,27,31,36) with three different options for the compression:

- HEVC compression using the recommendation ITU-T H Suppl.15 [7].
- HEVC compression without the chroma Qp adaptation.
- HEVC compression with 8 bits quantization for the chroma instead of 10 during the compression. The chroma was re-sampled to 10 bits before displaying images on the screen.

Figure 3.13 illustrates the different defaults on the image Regatta\_11s. On this figure, green and red artifacts appear. Those artifacts are attenuated by the tone-mapping.

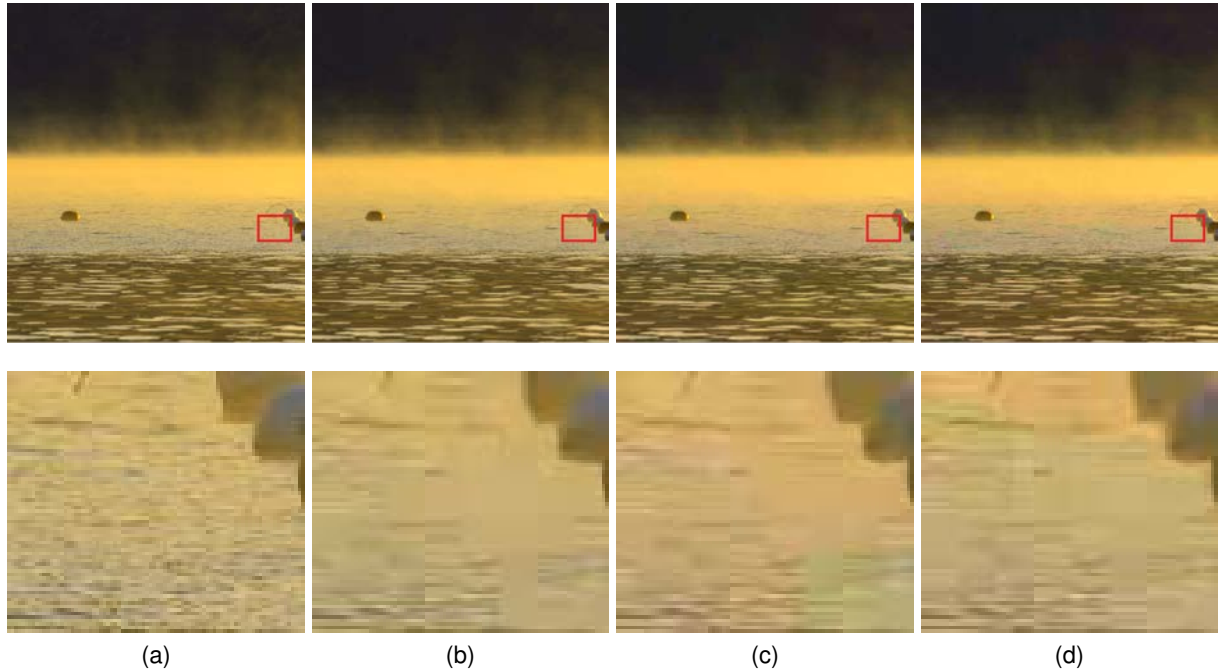


Figure 3.13: The image Regatta\_11s under three conditions: **(a)** The original image, **(b)** Compressed image, Qp 36, with Chroma Qp adaptation, **(c)** Compressed image, Qp 36, without Chroma Qp adaptation, **(d)** Compressed image, Qp 36, 8 bits quantization for the chrominance.

It should be noted that chromatic artifacts appear with a 8 bits quantization on the chroma even when the image is compressed with a low Qp. For example, on image Bike\_30s, quantization artifacts are easily visible on the sky and on the road when it is compressed with a Qp of 20 (cf. Figure 3.14).

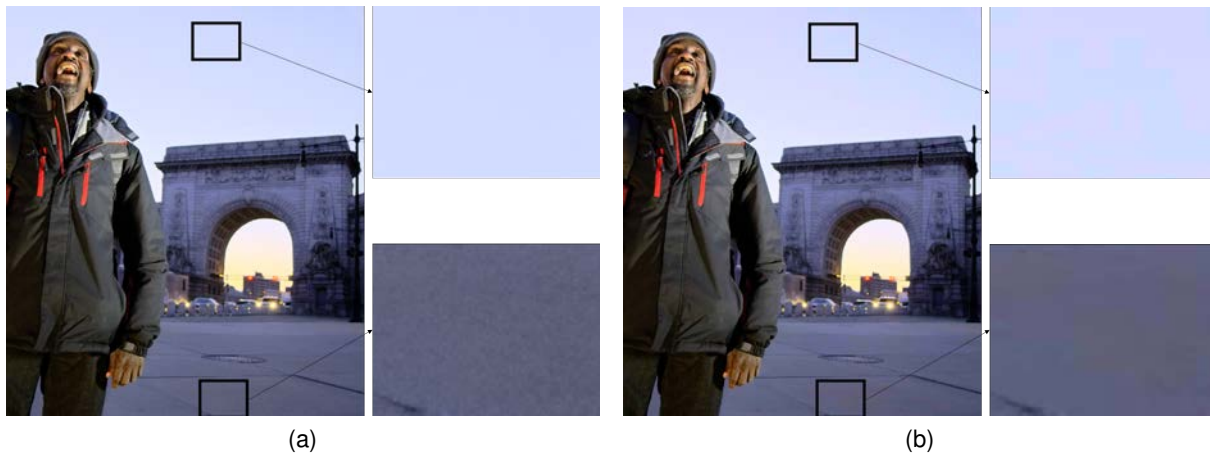


Figure 3.14: The image Bike\_30s under two conditions: **(a)** The original image, **(b)** Compressed image, Qp 20, 8 bits quantization of the chrominance

### 3.4.3 Subjective quality analysis

Figure 3.15 represents the repartition of MOS scores. We can observe that there are less images with a low MOS than images with other quality. The reason is that we wanted to focus more on the quality range where chrominance artifacts due to the compression is the most important. The previous database HDdtb shows that the impact of the chroma Qp offset algorithm was the most important for Qp around 27. If the Qp is lower, the chroma Qp adaptation is not necessary as there are enough bits allocated in the chroma. If the Qp is higher, the images are too damaged and the chroma Qp adaptation cannot improve the quality. Moreover, we feared that a too big quality gap would result for the viewer to put into perspective the smaller quality gain offers by the chroma Qp adaptation. As a result, fewer images were considered to have very poor quality.

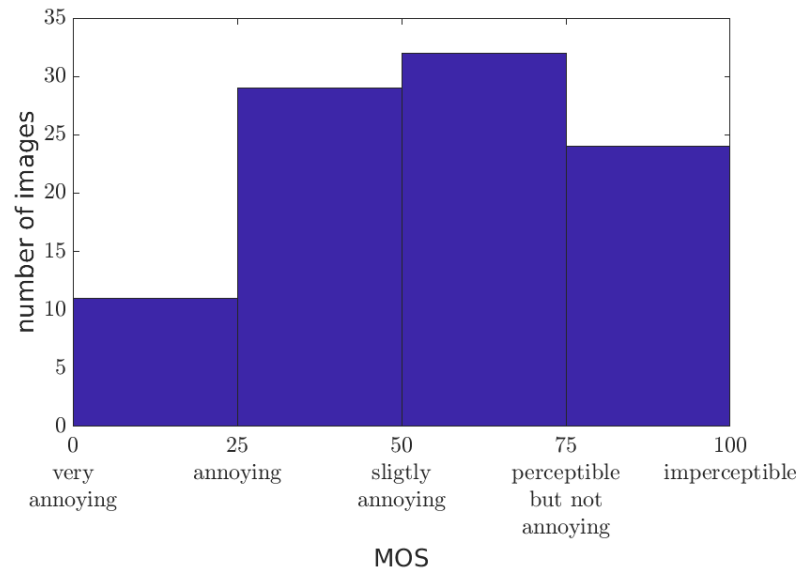


Figure 3.15: Histogram of the repartition of 4Kdtb MOS score

Figure 3.16 represent the MOS score in function of the compression  $Q_p$ . It shows that the distortions caused by an 8 bits quantization of the chroma induce a significative loss in quality for 6 on 8 images significant distortion compared to the compression with chroma  $Q_p$  adaptation even for the highest quality. Images compressed without the chroma  $Q_p$  adaptation have generally lower quality than images compressed with chroma  $Q_p$  adaptation quality but higher quality than images compressed with an 8 bits quantization of the chroma.

This database provides significant chromatic distortion as the only difference of treatment between the mode of compression is the handling of chrominance. It can be a challenging ground truth for quality metrics.

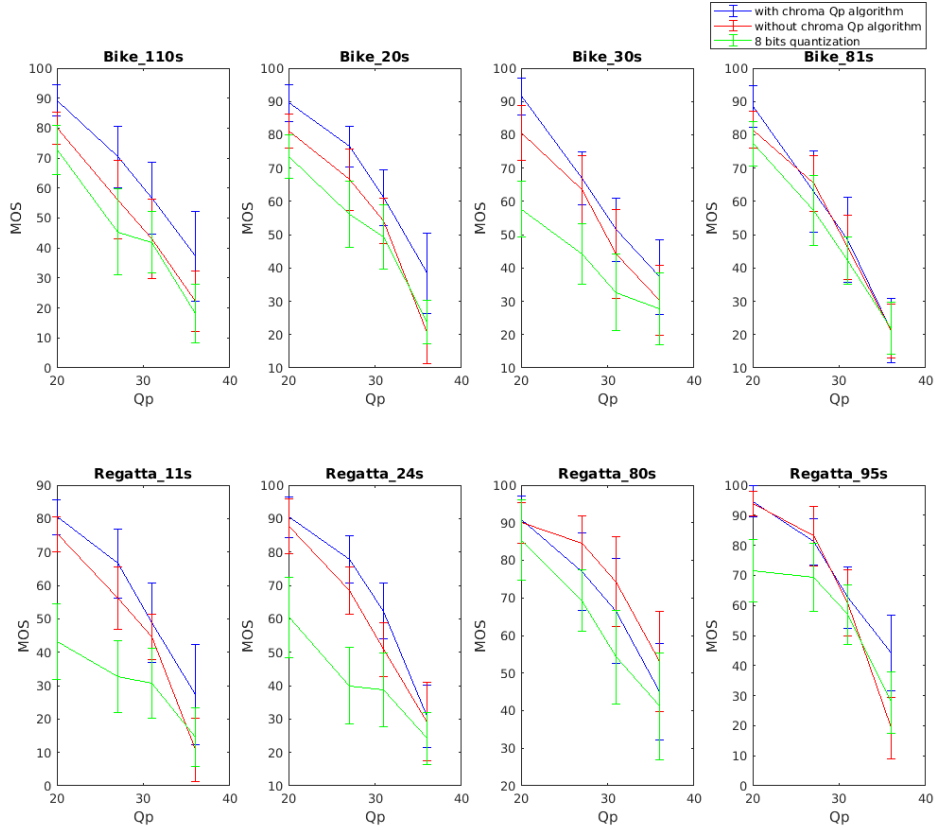


Figure 3.16: MOS score obtained for the database 4Kdtb in function of the Qp

### 3.5 Conclusion

In this chapter, we proposed two new HDR image databases annotated with subjective scores: HDdtb and 4Kdtb. Those databases are complementary to the existing databases as they proposed MOS scores obtain with a different display and with distortions specific to WCG. Another goal of those databases is to create realistic chrominance distortions, an aspect that was neglected by the previous databases.

The first database is composed of BT.709 images encapsulated in a BT.2020 gamut. The second is composed of native BT.2020 contents. If the impact of chromatic compression distortions is limited in the database HDdtb, it is easily perceptible in the database 4Kdtb.

There is still room to improve the databases. For example, we did not provide images exceeding the BT.709 gamut in the green area of the HVS gamut. We have tried to create a database that can represent the diversity of HDR/WCG content with the videos at our disposal. However, the production and the deployment of HDR/WCG contents are only at the beginning. It is difficult to know in advance what the characteristics of the real content will be. It is currently

very challenging to select the most representative images possible.

The two databases help us to understand the impact of chromatic distortion on quality perception. In the following chapter, those databases will also help us to evaluate the resilience and the sensitivity to such distortion of the objective IQA metrics.

# ANALYSIS OF EXISTING METRICS

---

## 4.1 Introduction

In this chapter, we study in details state-of-the-art HDR quality metrics. The calibration and the performances of existing objective metrics are analysed using the two new databases and the three already existing databases presented in section 2.2, namely, Narwaria et al. [49], Korshunov et al. [24] and Zerman et al. [50].

In the first part, we study in details the parameters of the metric HDR-VDP-2, which can be considered as the reference metric for HDR contents. In the second part, we compare the performances of all the existing metrics.

### Display model

Note that before applying any metric, the images have to be correctly formatted. Indeed, sometimes the image files report luminance above and below the ability of the displays. To simulate the effect of the display, we limit the luminance of images to the available dynamic range of the display used in the respective subjective tests. The luminance outside this range is cropped. In this thesis, we consider the display SIM2 HDR47ES4MB to be able to display luminance between 0.03 to 4250  $\text{cd/m}^2$  and the Sony BVM-X300 0.001  $\text{cd/m}^2$  to 1000  $\text{cd/m}^2$ . Some articles consider slightly different values. This can explain the small differences between our results and the reported results of other studies using the same database. However, those differences are not significant.

## 4.2 HDR-VDP-2 calibration

Recent studies [91] [89] have pointed out that HDR-VDP-2 is one of the best objective metrics for HDR still images. Unfortunately, HDR-VDP-2 is quite complex to use due to numer-

ous and sometimes hard-to-know parameters such as display emission spectrum, surround luminance and angular resolution. In this subsection, we evaluate the influence of those user-defined parameters on the ability of the metric to predict quality scores.

#### 4.2.1 Sensitivity to the screen spectral emission

In this subsection, we will discuss the sensitivity of HDR-VDP-2 to the spectral emission of the display. Indeed, HDR-VDP-2 allows to precisely model the quantity of light falling in each photo-receptors. That implies to define the spectral emission of each photo-transmitter of the display (Red Green and Blue).

For this purpose, we measured the spectral emission of 5 HDR displays. The Table 4.1 summarizes the characteristics of the considered displays.

Table 4.1: Considered displays

Name	Technology	Market
SONY BVM-X300	OLED	Professional
LG OLED 65 E6V	OLED	Consumer
Loewe Bild 7.55	OLED	Consumer
SONY KD-75X9405C	LED LCD	Consumer
SIM2 HDR47ES4MB	LED LCD	Professional

Three patches of pure blue, green and Red were measured on displays using the probe X-rite Eye-One Pro 2. We then measured a white patch to assess the spectral additivity of the components. The considered spectrum are plotted on Figure 4.1. We can observe that the Red and Green photo-transmitter of the Sony BVM-X300 have a narrower spectrum than the photo-transmitter of other displays. This was expected as this display can almost cover the BT.2020 gamut. Indeed, this gamut primary colors are supposed to be equivalent to monochromatic light sources or in other term to have a spectrum composed of one unique wavelength.

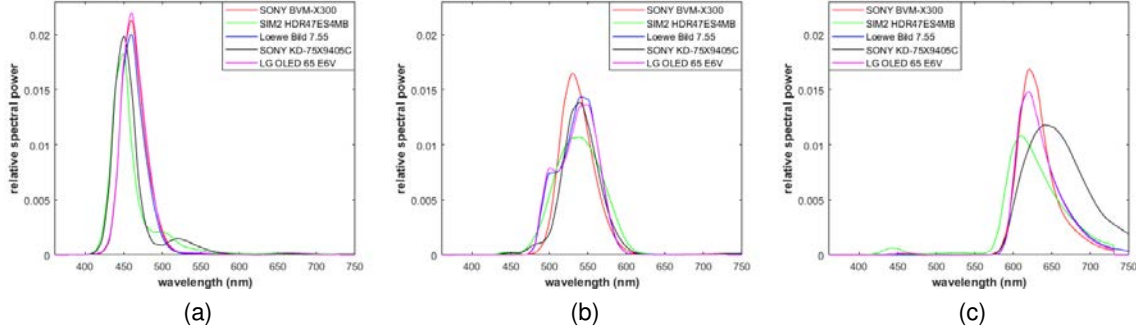


Figure 4.1: Real display spectral emissions **(a)** Blue component **(b)** Green component **(c)** Red component

To complete the study, we also consider three unrealistic spectral emission (cf. Figure 4.2):

- **a white spectrum:** the same intensity is given to each wavelength of the spectrum.
- **a Dirac spectrum:** The spectrum of each photo-transmitters is simulated by one Dirac. Each Dirac corresponding to the primaries wavelengths defined by the CIE.
- **a spectrum with only the blue component:** We consider the blue photo-transmitter to be a white spectrum. The spectrum of other photo-transmitters are void. This is the most unrealistic spectrum, created to really stressed the HDR-VDP-2 model.

These unrealistic spectral emissions have been designed with the following constraint: the luminance of a pixel using those spectrals emissions for which the value of the components R, G and B are identical must correspond roughly to the luminance of the same pixel using a real spectrum.

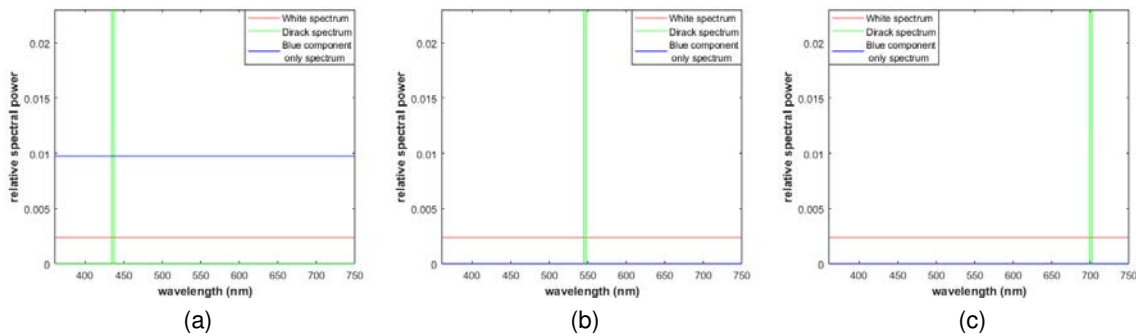


Figure 4.2: Unrealistic spectral emissions **(a)** Blue component **(b)** Green component **(c)** Red component

We then tested the performance of HDR-VDP-2 when modifying the display spectral emissions. As we can see on Tables 4.2 and 4.3, spectrums have a very low influence on the performance metrics. The results are similar with other databases. The variations of all the values are not significant. The performance metrics are slightly lower when using the white spectrum for the HDdtb database due to an overestimation of the chroma artifact (especially the chroma Gaussian noise) and only drop with a very unlikely spectrum, where only the blue component is kept. It shows that for the kind of artifact present in the databases, only a rough estimation of the display spectrum is required for evaluating the quality score. Figure 4.3 illustrates the fact that changing the spectrum have only a low impact on HDR-VDP-2. This result could be explained by the fact that the luminance of the different emission spectra is approximately preserved for the gray pixels.

Table 4.2: Performance indexes of HDR-VDP-2 for Zerman et al.

	<b>Spectrum</b>	<b>PCC</b>	<b>SROCC</b>	<b>OR</b>	<b>RMSE</b>
<b>Real spectrums</b>	SONY BVM-X300	0.94	0.93	0.45	10.4
	LG OLED 65 E6V	0.94	0.93	0.47	10.4
	Sony KD-75X9405C	0.93	0.93	0.47	10.4
	Loewe Bild 7.55	0.93	0.93	0.47	10.6
	SIM2 HDR47ES4MB	0.93	0.92	0.46	10.6
<b>Unrealistic spectrums</b>	white spectrum	0.93	0.91	0.55	11.8
	dirac spectrum	0.91	0.92	0.50	11.0
	blue-only spectrum	0.88	0.87	0.55	13.9

Table 4.3: Performance indexes of HDR-VDP-2 on HDdtb

	<b>Spectrum</b>	<b>PCC</b>	<b>SROCC</b>	<b>OR</b>	<b>RMSE</b>
<b>Real spectrums</b>	SONY BVM-X300	0.89	0.87	0.48	12.5
	LG OLED 65 E6V	0.90	0.87	0.47	12.3
	Sony KD-75X9405C	0.90	0.87	0.46	12.4
	Loewe Bild 7.55	0.90	0.87	0.48	12.3
	SIM2 HDR47ES4MB	0.90	0.87	0.48	12.2
<b>Unrealistic spectrums</b>	white spectrum	0.84	0.87	0.59	14.9
	dirac spectrum	0.90	0.87	0.46	12.4
	blue-only spectrum	0.67	0.67	0.68	20.5

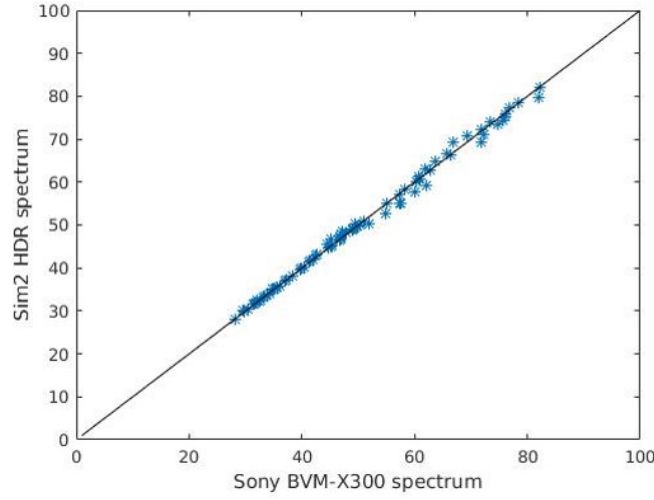


Figure 4.3: HDR-VDP2 score for two different spectrums (Sim2 HDR display and Sony BVM-X300 display) applied on the database HDdtb.

#### 4.2.2 Sensitivity to the surround luminance

In this subsection we discuss the impact of the surround luminance on HDR-VDP-2 results. The surround luminance is used by HDR-VDP-2 to estimate the luminance surrounding the image. We measure the difference in performance whih different values of suround luminance, from  $5 \text{ cd/m}^2$  (dark room) to  $200 \text{ cd/m}^2$  (lighted room). The PCC and SROCC are reported on Figure 4.4. OR and RMSE show a similar trend. The surround luminance has a low impact on the HDR-VDP-2 performances for all databases. The performance slightly decreases for low luminance for our database and Narwaria et al.

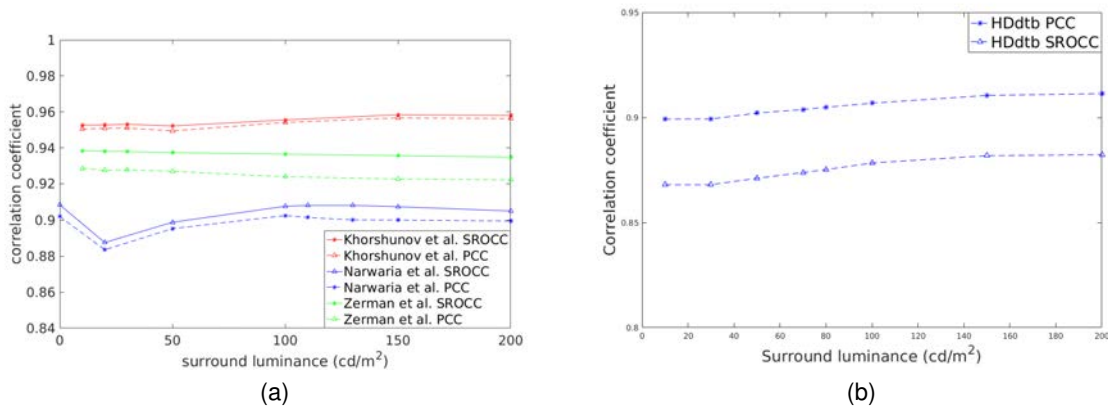


Figure 4.4: Performance of HDR-VDP-2 in function of the surround luminance for **(a)** existing database **(b)** the proposed database HDdtb

This parameter has a very low impact on the metric ability to predict quality. Moreover, the value of the HDR-VDP-2 score itself is not impacted as it is illustrated on Figure 4.5. The low impact of surrounding luminance was expected. Indeed, this parameter is used only to estimate the distortions visibility the near the extremities of the image and does not affect the estimation of the distortions visibility of the major part of the image.

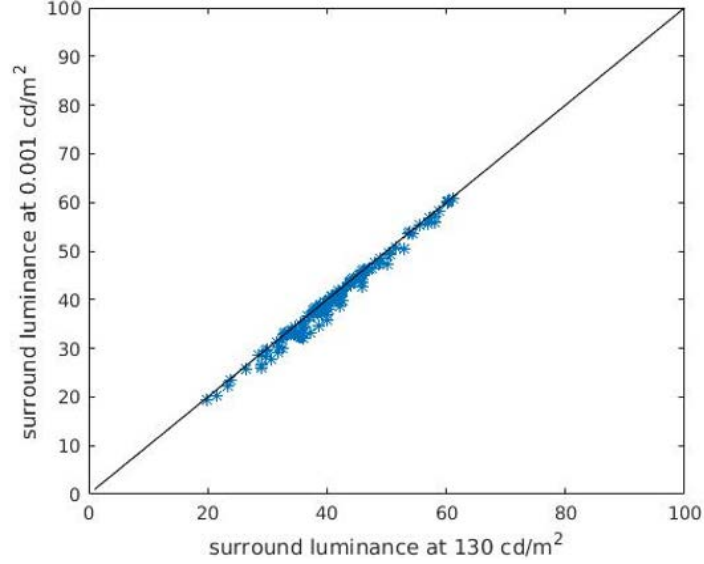


Figure 4.5: HDR-VDP2 score for two different surround luminance applied on the database Narwaria et al.

### 4.2.3 Sensitivity to the angular resolution

In this subsection we discuss the impact of the image angular resolution on HDR-VDP-2 results. The angular resolution correspond to the number of pixel by degree of the viewing angle. It depends of the size of the screen and the viewing distance. For example, for the Sony BVM-X300, 60 pix/deg (pixels per degree) correspond to a viewing distance of 1.15 meter, 80 pix/deg correspond to 1.6 meter and 20 pix/deg to 0.3 meter. This parameter is primordial for HDR-VDP-2. Indeed, this metric rely heavily on the decomposition of the image into different subbands of frequency (and orientation). The angular resolution is essential to associate a frequency to each subband. We try different values of this parameter and measured the resulting performances. The PCC and the SROCC of HDR-VDP2 are on Figure 4.6. The performance begins to significantly drop when the angular resolution is low because, during the pooling phase of HDR-VDP2, the weight given to the lowest frequency sub-bands are almost null. But, in term of correlation, only a rough estimation is needed without losing performances. However, in this case, the analysis of the HDR-VDP2 score should be really carefully made.

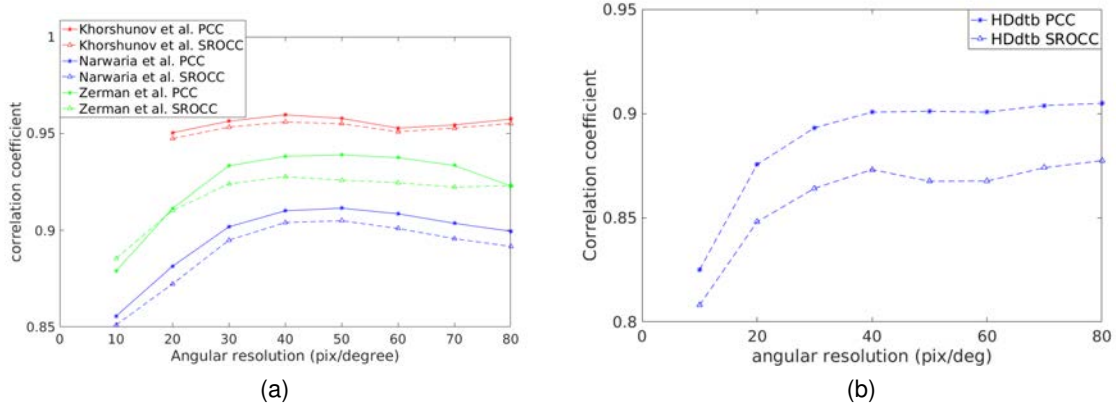


Figure 4.6: Performance of HDR-VDP-2 in function of the angular resolution for **(a)** existing database **(b)** the proposed database HDdtb

Indeed, if the values of the metric did not change significantly when changing the spectrum or the surround luminance, this is not the same with the angular resolution. The scores are shifted without losing in correlation. For example, the mean value on our database HDdtb is 60.3 when the parameter is put to 30 pixels/degree (a viewing distance of 0.55 meter) and 48.9 when put on its real value 60 pixels/degree (a viewing distance of 1.15 meter). The Figure 4.7 illustrates this phenomenon. This is a counter-intuitive result as it would mean that the nearer the viewer is of the screen, the higher HDR-VDP2 score is. The effect of the angular resolution is not completely handled for the HDR-VDP2 quality score and the same HDR-VDP-2 quality score obtain with different angular resolutions could mean something different in terms of absolute quality.

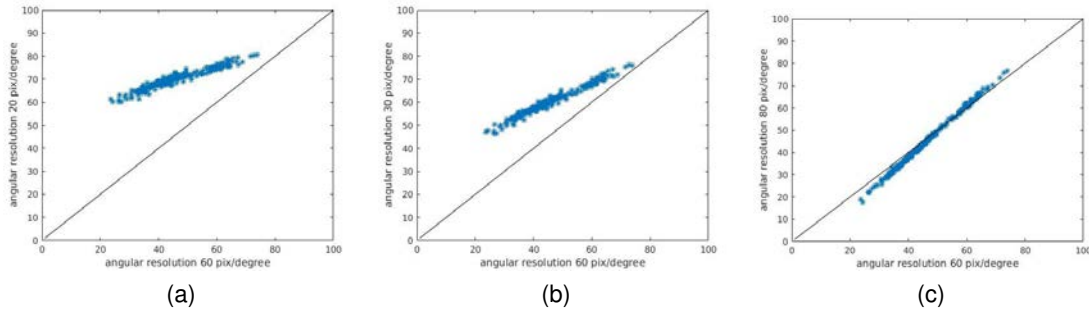


Figure 4.7: HDR-VDP2 score for different angular resolutions applied on the database Khorshunov et al. **(a)** 20 pix/deg in function of 60 pix/deg **(b)** 30 pix/deg in function of 60 pix/deg **(c)** 80 pix/deg in function of 60 pix/deg

#### **4.2.4 Conclusion on HDR-VDP-2 parameters**

We tested the sensibility of three parameters, used by HDR-VDP2 and concluded that only a rough estimation of the surround luminance and the spectral emission of the display is needed. HDR-VDP2 quality scores are not sensitive to those two environmental conditions. However, the angular resolution can impact the value of HDR-VDP-2 as it relies on the frequency to evaluate the visibility of a distortion. It should be noted that this is valid only for the quality score and not for the distortion map.

Thus, the HDR-VDP-2 quality metric turns to be simpler to use than these parameters suggest. However, the low impact of these parameters questions the pertinence to describe HVS mechanism as precisely as the HDR-VDP-2 intends to do.

### **4.3 Benchmark of existing HDR objective quality metrics**

In this section, we study the performances of existing HDR metrics with the proposed databases and the already existing ones. First, we study the impact of the images resolution on the metrics performances. Second, we compare the performances of the existing HDR metrics to determine which one is the most relevant to estimate HDR quality. Finally, we highlight a limitation of existing metrics: their lack of consideration for the chrominance distortion.

#### **4.3.1 The particular case of 4Kdtb images**

Because the database 4Kdtb has a higher resolution than others, the number of pixels per viewing angle increases. The angular resolution increases and becomes 120 pix/degree. Because some quality metrics are not adapted to this kind of resolution, we choose to study the performances of existing metrics on this database with and without downsampling the images. This downsampling allows the images to have resolution divided by 2 and an angular resolution of 60 pix/degree. The downsampling was performed on the luminance of images after the PU function was applied for PU-metrics and on the linear luminance for HDR-VQM and on the linear RGB values for HDR-VDP-2. The SROCC performances for the tested metrics are given on Table 4.4.

The considered metrics have better performances using the downsampled images (especially PSNR, SSIM, and MS-SSIM). Even if the gain is negligible for some metric (like PU-FSIM) there is no benefit for using the original images as it will only increase the complexity to calculate the metrics.

In the rest of the thesis, metrics are calculated on the downsampled images of the 4Kdtb.

Table 4.4: SROCC for the 4Kdtb database with and without downsampling the images.

Quality Metric	Without downsampling	With downsampling
PU-PSNR	0.5686	<b>0.7261</b>
PU-SSIM	0.4454	<b>0.7066</b>
PU-MS-SSIM	0.7629	<b>0.8517</b>
PU-FSIM	0.9027	<b>0.9054</b>
PU-VIF	0.8169	<b>0.8658</b>
PU-PSNR-HVS-M	0.7623	<b>0.8401</b>
PU-PSNR-HMA	0.7623	<b>0.8403</b>
HDR-VQM	0.7618	<b>0.7735</b>
HDR-VDP-2	0.8508	<b>0.8678</b>

#### 4.3.2 Existing metric performances

Several studies have already evaluated the performances of HDR metrics using different metrics and methodology. In [91], the authors assessed the performances of 35 quality metrics using the database Korshunov et al. They conclude that HDR-VDP2 (version 2.2.1 [53]), HDR-VQM and PU-MS-SSIM were the best-performing metrics. In [89], the authors came to the conclusion that HDR-VDP2 (version 2.1.1) can be successfully used for predicting the quality of video pair comparison contrary to HDR-VQM. In [88], authors found that HDR-VDP2, HDR-VQM, PU-VIF, and PU-SSIM provided similar performances for video quality estimation using MPEG Cfe (Call for Evidence) video clips. In [90], results indicate that PU-VIF and HDR-VDP2 have similar performances for video quality estimation, although PU-VIF has slightly better reliability than HDR-VDP2 for lower quality scores. Zerman et al. [50] estimate that HDR-VQM is the best full-reference HDR quality metric. They combine five different databases to compare the metrics (including Narwaria et al., Korshunov et al. and Zerman et al.). HDR-VDP-2 provides similar performances if one of the databases is excluded but have significantly lower result on this particular database.

Depending on the studied database or on the content (image or video), the different studies give slightly different, sometimes contradictory results, even if there is some constants. HDR-VDP-2 and HDR-VQM are often the best metrics and PU-based metrics provide a still good and simpler alternative.

Table 4.5 provides the SROCC for 7 PU-metrics plus HDR-VDP-2 and HDR-VQM. The best metric is in red, the second in blue and the third in green. More performance indexes are reported in appendices C.

In agreement with the state-of-the-art articles, the most performing metrics depend on the database. If the results are not completely conclusive, some perform generally better: HDR-

Table 4.5: SROCC of the existing quality metrics on the considered databases

Quality Metric	4Kdtb	HDdtb	Narwaria et al.	Korshunov et al.	Zerman et al.
PU-PSNR	0.7261	0.7802	0.5331	0.8597	0.8266
PU-SSIM	0.7066	0.8430	0.7240	0.9280	0.8315
PU-MS-SSIM	0.8517	0.8640	0.8656	0.9583	0.9165
PU-FSIM	0.9054	0.8149	0.8773	0.9553	0.8912
PU-VIF	0.8658	0.7464	0.7704	0.9322	0.8863
PU-PSNR-HVS-M	0.8401	0.7803	0.5624	0.9331	0.9035
PU-PSNR-HMA	0.8403	0.8218	0.7634	0.9369	0.9041
HDR-VQM	0.7735	0.8330	0.8995	0.9572	0.9170
HDR-VDP-2	0.8678	0.8685	0.8906	0.9516	0.9289

VDP-2 is always in the first or second best performing metric except for the database Korshunov et al. but in this case, the difference between HDR-VDP-2 SROCC and the best metric SROCC (PU-MS-SSIM) is inferior at 0.01 which is not significant.

Performances of other metrics differ more from one database to another. For example, HDR-VQM performances are amongst the top three metrics for three databases but this metric has more difficulty to handle the two new databases, especially 4Kdtb. PU-FSIM is the best performing metric on 4Kdtb and is even outperforming HDR-VDP-2 but have more difficulty to handle the database HDdtb. PU-MS-SSIM also have in general term good performances even if it has difficulty to handle the Narwaria et al. database.

Our results are similar to the one obtained from the previous studies and it is difficult to conclude with this table which metric is the best, as the answer depends on the studied database.

#### 4.3.3 Chrominance distortion sensitivity

In this subsection, we assess whether existing metrics are sensitive to chromatic distortions or not. Indeed, all of them only consider luminance distortion. We use the 4Kdtb as this database was conceived to create chrominance error. In the process, we also assess that the difference of quality for this database between the three modes of compression is due to chrominance artifacts. Figure 4.8 represents the MOS scores and the HDR-VDP-2 scores in function of the Qp. We can observe that the viewers clearly saw a difference in quality between the three modes of compression. However, HDR-VDP-2 scores are sensibly the same. Results are similar for the other images of the 4Kdtb and for the other metrics.

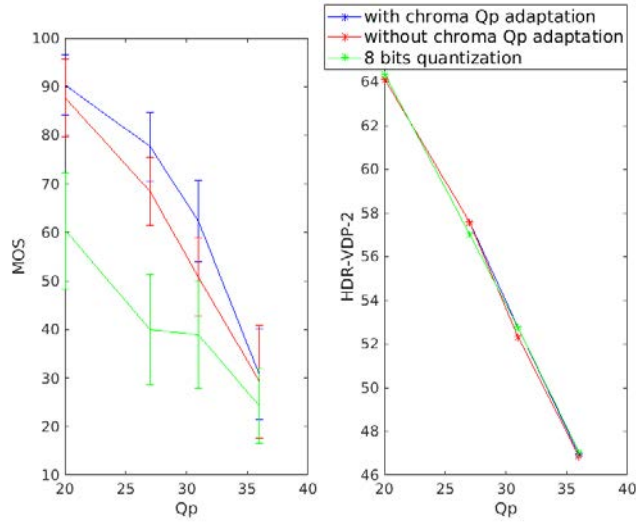


Figure 4.8: MOS scores and HDR-VDP-2 scores in function of Qp for the image Regatta\_24s of the database 4Kdtdb

As expected 4Kdtdb presents chrominance artifacts that directly impact the quality perception of images. Moreover, as there are no metrics that consider such distortion, adding chrominance consideration inside the metrics is an interesting direction to improve quality estimation. The 4Kdtdb can also be used to assess qualitatively the metrics sensitivity to meaningful chrominance distortions even in the presence of important luminance distortion. Figure 4.8 is illustrating well the difference of quality perception between the different modes of compression and a metric able to be discriminating between those modes can help to chose the best method to compress chrominance information.

## 4.4 Conclusion

In this chapter, we studied existing objective IQA metrics. We take advantage of the creation of two new databases to complement the previous metrics reviews of the state of the art.

First, we thoroughly analyzed the best existing IQA metrics, and, especially, its behavior against the parameters modeling the viewing conditions. Second, we review the performance of nine existing metrics. We get results similar to previous metric reviews: The best metric is different depending on the database and HDR-VDP-2 is, always, amongst the metrics with the best performances. However, this metric is a very complex which limits its usability for real-time application.

The tested metrics are all based on very different model. For example, FSIM is based on the phase congruency, MS-SSIM measure the structural similarity on different scale and HDR-

VDP-2 model precisely the HVS. It is quite surprising that even if those metrics are based on very different hypothesis, they are all relevant to assess the quality of images. It shows that very different image features can be used to predict an image quality.

During the analysis, we point out one limitation of existing metrics: they do not take into consideration chromatic distortions. Moreover, we have shown that the 4Kdtb could be used to assess whether a metric is sensitive or not those distortions. It is also a way of improvement. Metrics could perform better if they considered chromatic distortions

## PART II: CONCLUSION

---

In this part, we proposed two new databases HDdtb and 4Kdtb which present new kind of artifacts. If with the HDdtb, we failed to create compression artifacts that were significantly important, we succeeded with the 4Kdtb. In a broader perspective, the relevance of subjective tests can also be questioned. For example, on the proposed database HDdtb, viewers did not perceive the gamut mismatch artifact as a loss of quality. However, this kind of artifact changes completely the appearance of images. In some cases, asking the viewers to not assess only the quality of the images but also their fidelity to the image appearance can be valuable to fully evaluate image processing algorithms.

In addition, we analyzed thoroughly existing objective metrics using the newly created databases. We carefully reviewed the impact of HDR-VDP-2 parameters on its performances and found that only the angular resolution can affect the quality score. We also confirmed that HDR-VDP-2 is one of the best existing metrics but that for several databases, PU-FSIM and PU-MS-SSIM, two simpler metrics could provide similar or even better performances. Finally, we also assessed the sensitivity of existing metrics to color distortion. As existing metrics is only sensitive to luminance, the metrics are not able to differentiate the different mode of compression of the 4Kdtb. However, those different modes affect the perception of the images. New metrics, sensitive to chrominance distortions, are needed for such a case.

PART III

# **New color metrics for HDR/WCG image quality assessment**

---

## PART III: INTRODUCTION

---

Subjective tests and other user studies are the more reliable methods to evaluate the quality of images. Thus, with subjective tests, it is possible to infer the best methods of compression supposing that the selected subset of images is representative of all images. However, subjective tests are often very time-consuming, costly and not always possible in many applications such as the quality monitoring for live video applications.

This is the reason why objective metrics were created: to automatically assess the quality of images without the need of human viewers. The best performing metrics attempt to emulate the human perception of quality. In the previous chapter, we reviewed several existing metrics adapted to HDR content: PU-metrics and the bio-inspired metrics like HDR-VDP-2 for images and HDR-VQM for videos. We created databases to study the ability of these metrics to emulate human perception of quality. We highlight one limit of existing metrics. They are non-sensitive to visible chromatic distortion as they only assess luminance artifacts. However, chrominance errors can occur during the compression as it is assessed by MPEG (cf. Section 1.4), especially in the case of WCG images. This was highlighted by our database 4Kdtb (cf. Section 3.4).

In this chapter, we propose three methods to tackle this issue. First, we propose to adapt existing metrics, using perceptually linear color spaces and not only perceptually linear luminance as it is proposed for PU-metrics. Second, we propose to extend the HDR-VDP-2 metric towards color artifacts perception. Indeed, HDR-VDP-2 is a luminance only metrics that models the human perception of luminance. We attempt to complete this model using chrominance information. Third, we propose to aggregate several metrics and visual features, including features sensitive to chrominance distortions. The idea is to take advantage of each metric or feature strength while compensating individual weakness.

# SDR METRIC ADAPTATION

---

## 5.1 Introduction

If most SDR metrics are not based on the HVS perception of differences, they often assume that the representation of images is perceptually linear. However using a color space adapted to SDR metrics like  $Y'Cr'Cb'$  [28] is not a solution to obtain perceptually uniform HDR/WCG images (cf. Section 1.3.2). However, it should be possible to adapt SDR metrics if the image are encoded using an appropriate color space.

Aydin et al. proposed in [10] a solution to extend color-blind SDR metrics using a function more appropriate than the gamma function. This function is called PU and is adapted to HDR contents (cf. Section 2.4.1).

This methodology has a limitation as it only concerns luminance. In this chapter, we propose a method to adapt any SDR metric using the perceptually uniform color spaces presented in section 1.3.2. A second limitation is that this function implies a pre-process of the images to estimate the linear luminance of the screen. Most images will be encoded using color space like  $ICtCp$  already recognized in compression codecs. Using directly such space could reduce the complexity of the metrics.

The first section of this chapter presents our proposed method to adapt SDR metrics. The second section presents the results. A third section, present a detailed analysis of those results. A fourth section presents our recommendations to use the SDR metrics adaptation method and the last section concludes this chapter.

## 5.2 Proposed method

For adapting SDR metrics to HDR/WCG images, the reference and distorted images are first converted in a perceptually linear color space. A remapping function is then applied. Finally, the SDR metrics are used to determine the quality score. Figure 5.1 presents the diagram of the proposed method. These key steps are described in the following subsections.

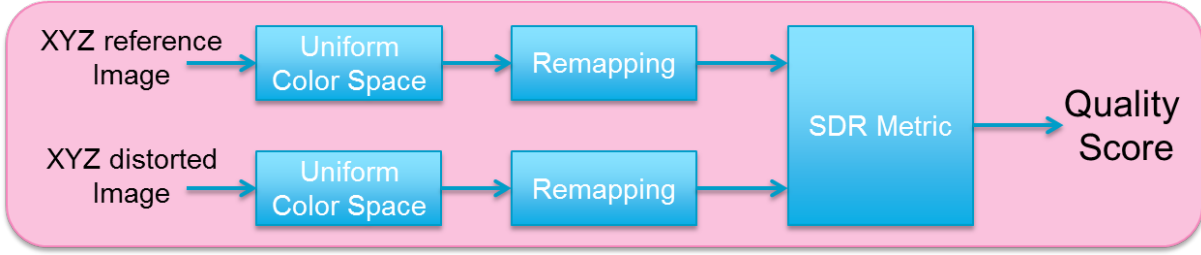


Figure 5.1: Diagram of the proposed method to adapt SDR metrics to HDR/WCG contents.

### 5.2.1 Color Space Conversion

Most SDR metrics were designed with the assumption that the images are encoded in the legacy  $Y'Cr'Cb'$  color space [28]; this color space is approximately perceptually uniform for SDR content.

To use SDR metrics with HDR images, we propose to leverage perceptually uniform color spaces adapted to HDR and WCG images such as  $ICtCp$ ,  $J_z a_z b_z$ , and  $HDR-Lab$ .

Note that we used the perceptually uniform version of  $ICtCp$ :  $\widehat{ICtCp}$  presented in Section 1.3.2. For the sake of clarity, in this chapter, we call this color space  $ICtCp$ .

The use of  $HDR-Lab$  color space requires to know the diffuse white luminance. Unfortunately this parameter is most of the time unknown for HDR contents. To cope with this issue, we consider two different diffuse whites to compute the  $HDR-Lab$  color space, i.e.  $100 \text{ cd/m}^2$ , the peak brightness of SDR screen, and  $1000 \text{ cd/m}^2$ , the peak brightness of the HDR display Sony BVM-X300. These two color spaces are named  $HDR-Lab_{100}$  and  $HDR-Lab_{1000}$ , respectively.

To illustrate the importance of using uniform color space, we also consider two non-uniform color spaces, namely  $XYZ$  and  $Y'Cr'Cb'$  color spaces as defined in the BT.2020 recommendation [30]. This last space is a WCG but SDR color space. It cannot be considered as approximately uniform for HDR content as it uses the classical gamma function. The gamma EOTF takes for parameter the display peak brightness (cf. equation 1.6). To fix this value, we choose the maximum value taken by the studied HDR images:  $4250 \text{ cd/m}^2$ .

### 5.2.2 Remapping Function

The six aforementioned color spaces, i.e.  $XYZ$ ,  $Y'Cr'Cb'$ ,  $HDR-Lab_{100}$ ,  $HDR-Lab_{1000}$ ,  $ICtCp$  and  $J_z a_z b_z$ , have a different range of values. As most of SDR metrics have constant values defined for pixel values between 0 and 255, it is required to adapt the color spaces. We remap them in a way that their respective perceptually linear luminances fit a similar range as the luminances encoded with the PU transfer function between 0 and  $100 \text{ cd/m}^2$ . We choose  $100 \text{ cd/m}^2$  as a normalization point because it roughly corresponds to the peak brightness of

an SDR screen. Moreover, the PU-encoding is used as a reference to remap the color spaces because it is already adapted to SDR metrics. The goal of this process is to obtain HDR images with the same luminance scale than SDR images in the range 0 to 100  $\text{cd/m}^2$  while preserving the perceptual uniformity of the color spaces. The remapping of the perceptual color spaces is done as follows:

$$\begin{pmatrix} \widehat{J}_z(i, j) \\ \widehat{a}_z(i, j) \\ \widehat{b}_z(i, j) \end{pmatrix} = \frac{\alpha_{PU}}{\beta_{J_z}} \times \begin{pmatrix} J_z(i, j) \\ a_z(i, j) \\ b_z(i, j) \end{pmatrix} \quad (5.1)$$

where  $J_z(i, j)$ ,  $a_z(i, j)$  and  $b_z(i, j)$  correspond to the value in the  $J_z a_z b_z$  domain of the pixel with the spatial coordinates  $i$  and  $j$ .  $\widehat{J}_z(i, j)$ ,  $\widehat{a}_z(i, j)$  and  $\widehat{b}_z(i, j)$  correspond to the same pixel value after the remapping.  $\alpha_{PU}$  is the luminance value in the PU space when linear luminance value is 100  $\text{cd/m}^2$ .  $\beta_{J_z}$  is the value of the  $J_z$  component of the  $J_z a_z b_z$  color space when linear luminance value is 100  $\text{cd/m}^2$ . A similar operation is applied to  $ICtCp$  and  $HDR-Lab$ ,  $XYZ$  and  $Y'Cr'Cb'$  color spaces. The resulting luminances for the aforementioned color-space as well as the PU-encoding luminance are plot on Figure 5.2. For these figures, we chose a surround luminance of 20  $\text{cd/m}^2$  for the two  $HDR-Lab$  color spaces, the surround luminance of two of the five studied databas.

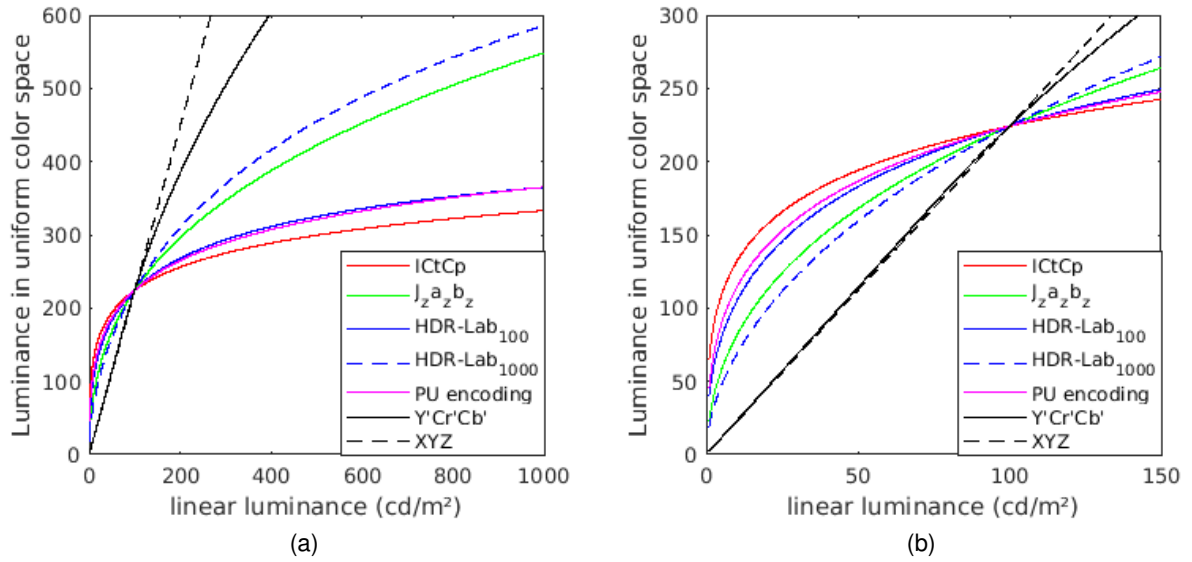


Figure 5.2: Different perceptually uniform luminances as a function of the linear luminance: **(a)** for the range 0–1000  $\text{cd/m}^2$ , **(b)** for the range 0–150  $\text{cd/m}^2$ .

### 5.2.3 Evaluation of tested SDR metrics

Once the images are remap, we can apply any SDR metrics that used images in the  $Y'Cr'Cb'$  color space. In this chapter, we study the performances of 6 SDR color-blind metrics: PSNR, SSIM [72], MS-SSIM [75], FSIM [65], PSNR-HVS-M [77] and PSNR-HMA [78] and 6 color metrics: Mean of the color difference metrics  $\overline{\Delta E}$ , the spatial extension of  $\overline{\Delta E}$ ,  $\overline{\Delta E^S}$  [71], SSIMc [73], CSSIM [74], FSIMc [65], PSNR-HMAc [78].

A short description of those metric can be found in Section 2.4.1 and a more detailed description in Appendix A.

Note that, to adapt the  $\overline{\Delta E^S}$  metrics, the blurring model used in this metrics (cf. Appendix A.2) is first applied to the  $XYZ$  color space of the images and then the different color difference metrics are calculated.

In the following sections, the naming convention used for all metrics is  $\text{Metrics}_{ColorSpace}$ . For example, the PSNR metrics used with the  $ICtCp$  color space is called  $\text{PSNR}_{ICtCp}$ .

## 5.3 Results

In this section, we present the performances of the different metrics over five databases. For the sake of completeness, we also study the performances of the following color-blind HDR metrics: HDR-VDP2 [53] (HDR-VDP-2) [54] and HDR-VQM [70]. Figures 5.3–5.7 present the SROCC performances for each database and each metric. Numerical values of the performance indexes (SROCC, PCC, OR, RMSE) can be found in Appendix C.

### 5.3.1 4Kdtb Database

For the proposed 4Kdtb database(cf. Figure 5.3), for each color-blind metrics, the best color spaces are always the  $ICtCp$ ,  $HDR-Lab_{100}$  and the PU-encoding.  $J_z a_z b_z$  and  $HDR-Lab_{1000}$  provide the lowest performances. The best performing color-blind metrics is FSIM used with the PU-encoding, closely followed by  $\text{FSIM}_{ICtCp}$  and  $\text{FSIM}_{HDR-Lab_{100}}$ . MS-SSIM used with the PU encoding,  $ICtCp$  and,  $HDR-Lab_{100}$  are almost on par with the second best performing metrics HDR-VDP2 (cf. Appendix C). The only color space that provides good performances on all color metrics is the  $ICtCp$  color space.

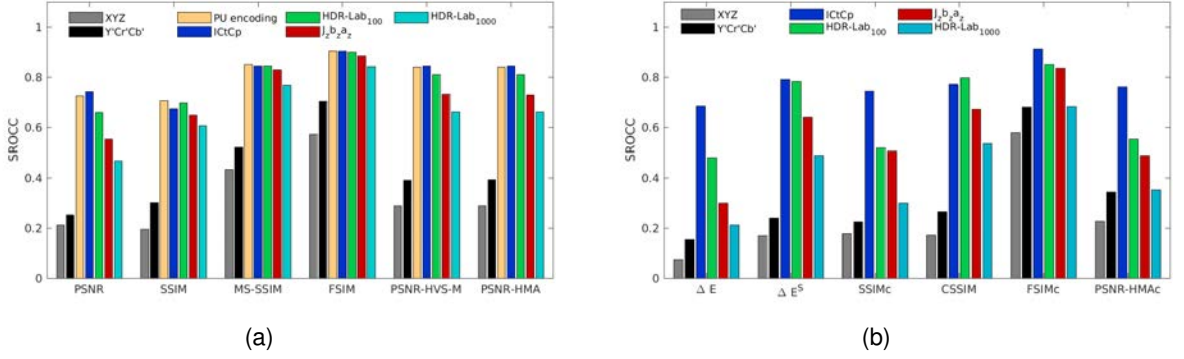


Figure 5.3: SROCC performances for the 4Kdtb database for color-blind quality metrics **(a)** and for color quality metrics **(b)**.

### 5.3.2 Zerman et al. Database

For the Zerman et al. database, as previously, the color spaces, *ICtCp*, *HDR-Lab*<sub>100</sub> and the PU-encoding provide the best performances for almost all color-blind metrics (cf. Figure 5.4). However, there is one exception with FSIM. Used with the following color spaces,  $J_z a_z b_z$ , *HDR-Lab*<sub>100</sub> and *HDR-Lab*<sub>1000</sub>, it provides slightly better performances than *ICtCp* and the PU-encoding. The best performing color-blind metrics are, with similar performances, HDR-VDP2, HDR-VQM and MS-SSIM<sub>ICtCp</sub>.

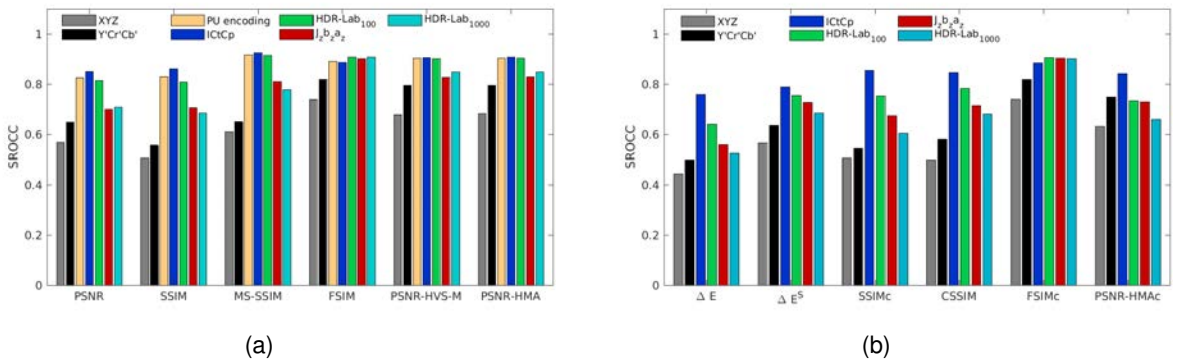


Figure 5.4: SROCC performances for the Zerman et al. database for color-blind quality metrics **(a)** and for color quality metrics **(b)**.

### 5.3.3 HDdtb Database

For our proposed HDdtb (cf. Figure 5.5), for color-blind metrics, the color space  $J_z a_z b_z$  provides slightly lower performances for all metrics, except with FSIM. For this metric, the performances with  $J_z a_z b_z$  are higher. The best performing color-blind metrics for this database are  $\text{FSIM}_{J_z a_z b_z}$ ,  $\text{FSIM}_{\text{HDR-Lab}_{1000}}$  and  $\text{MS-SSIM}_{\text{HDR-Lab}_{1000}}$ . For the color metrics, the metrics based on color difference metrics ( $\Delta E$ ,  $\Delta E^S$  and CSSIM) do not perform well. This is partially due to the presence of the gamut mismatch artifact. As noticeable on Table 5.1, discarding this artifact increases the performances of these metrics. For the participants of our subjective test, these distortions are clearly visible but are not directly associated with a loss of quality.

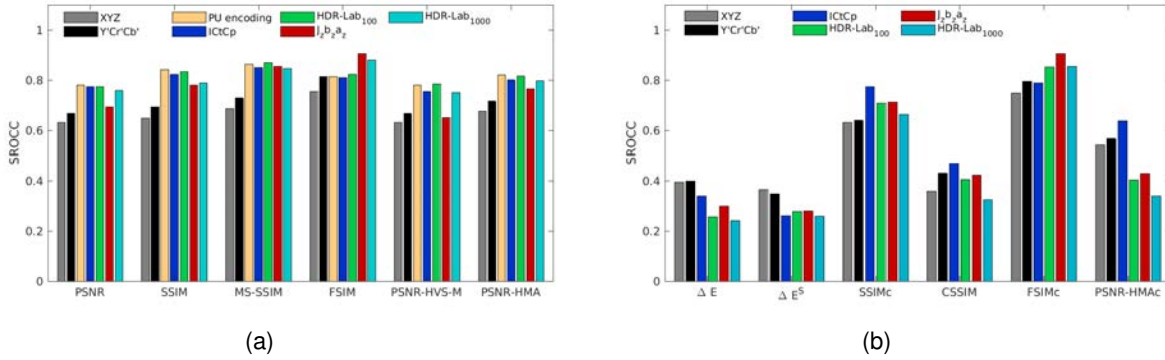


Figure 5.5: SROCC performances for the HDdtb database for color-blind quality metrics **(a)** and for color quality metrics **(b)**.

Table 5.1: SROCC for the HDdtb database with and without the gamut mismatch artifact.

Quality Metric	All Images	Without the "Gamut Mismatch" Distortion	compression Artifacts Only
$\Delta E_{\text{HDR-Lab}_{100}}$	0.2578	0.3905	0.6190
$\Delta E^S_{\text{HDR-Lab}_{100}}$	0.2784	0.5687	0.6946
$\text{CSSIM}_{\text{HDR-Lab}_{100}}$	0.4065	0.6453	0.7714

### 5.3.4 Korshunov et al. Database

The Korshunov et al. database is the less selective database (cf. Figure 5.6). Most of the metrics have high correlation coefficients and the choice of color space has close to no impact on the performances especially on color-blind metrics. Even using non-perceptually linear color space like the  $Y'Cr'Cb'$  color space impacts only moderately the performances of MS-SSIM, FSIM, PSNR-HVS-M and PSNR-HMA. For this database, the best performing color-blind metrics are  $FSIM_{J_z a_z b_z}$ ,  $FSIM_{HDR-Lab_{1000}}$  and  $MS-SSIM_{J_z a_z b_z}$ .

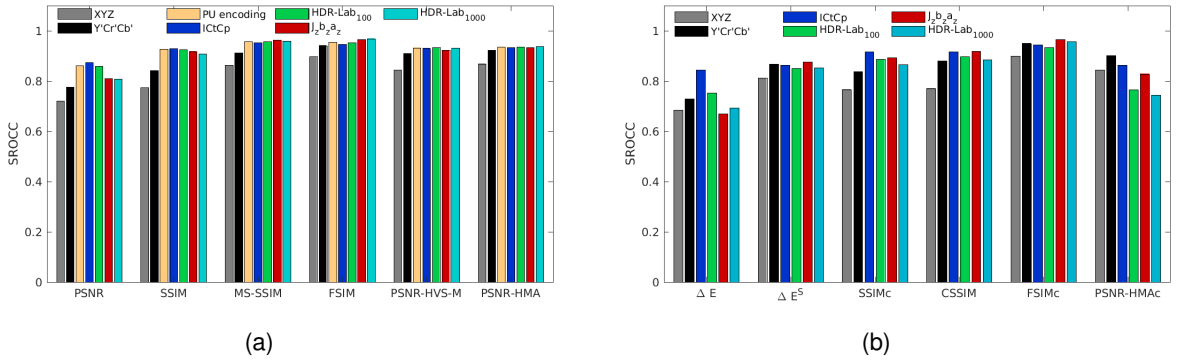


Figure 5.6: SROCC performances for the Korshunov et al. database for color-blind quality metrics (a) and for color quality metrics (b).

### 5.3.5 Narwaria et al. Database

For the Narwaria et al. database (cf. Figure 5.7),  $J_z a_z b_z$  is the best color space for SSIM and MS-SSIM while the PU-encoding and the  $HDR-Lab_{100}$  are the best color spaces for FSIM. The best metrics for this database are  $MS-SSIM_{J_z a_z b_z}$ , HDR-VDP2 and HDR-VQM. The good performances of HDR-VDP2 were expected for this database because it was part of the training set of this metric. For this database, the performances of the PSNR and the PSNR-HVS-M are relatively low compared to the other databases. The fact that PSNR-HMA with the adequate color space significantly increases the performances of PSNR-HVS-M suggests that the backward compatible compression used by Narwaria et al. (Section 2.3) creates distortions that impact the mean luminance and the contrast of the images. Indeed PSNR-HMA is an improvement of PSNR-HVS-M that takes into account these two kinds of artifacts [78].

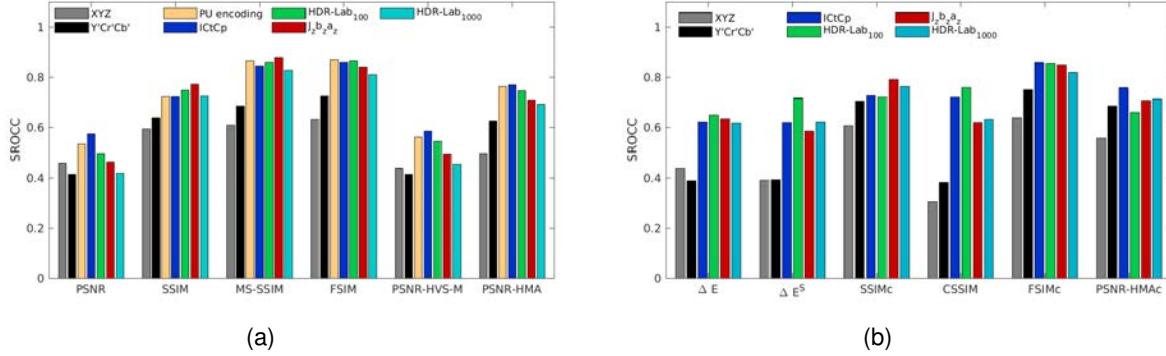


Figure 5.7: SROCC performances for the Narwaria et al. database for color-blind quality metrics (a) and for color quality metrics (b).

### 5.3.6 Results Summary

For all studied databases, HDR-VDP2 has generally high performances although it is not always on the top three metrics (cf. Appendix C). FSIM and MS-SSIM with appropriate perceptually uniform color space are often on par if not better than HDR-VDP2.

Among all metrics, FSIM is the less sensitive metrics to the choice of color space assuming that this color space is perceptually uniform.

The color extension of FSIM, namely FSIMc, does not improve the performances of FSIM even for our proposed database 4Kdtb which focuses on chromatic distortions. Worst, the metrics becomes much more sensitive to the color space choice. We observe the same behavior for the color extension of PSNR-HMA, PSNR-HMAc. The color extension of this metric also decreases its performance for all color spaces.

When using the two non-uniform color spaces  $XYZ$  and  $Y'Cr'Cb'$ , the performances of all metrics drop significantly compared to the other color spaces for all the databases and especially for our proposed database 4Kdtb, the Zerman et al. database and the Narwaria et al. database. It emphasizes the importance of perceptually uniform color space for predicting the quality of HDR images.

## 5.4 Results Discussions

We separate our analysis in two parts. First, we study the impact of the color space on the metrics performances. Moreover we emphasize the influence of the diffuse white luminance. As a reminder, the luminance of the diffuse white corresponds to the luminance of a 100% reflectance white card without any specular highlight. In HDR imaging, it is different from the peak brightness. In the second part of our analysis, we discuss the sensitivity of chrominance artifacts on color metrics using our proposed database 4Kdtb.

### 5.4.1 Impact of the Diffuse White Luminance

Our results suggest that the best color space for assessing the quality of HDR images depends on the test database. Indeed, some of the color spaces are adapted and tuned for one visualization condition.

The *HDR-Lab* color space considers two important parameters, i.e. the diffuse white and the surround luminance. Moreover, the final equation of the  $J_z a_z b_z$  color space (Equation (1.33)) was tuned using the experimental dataset called SL2 [12]. This dataset was obtained for a diffuse white at  $997 \text{ cd/m}^2$ . This explain why the  $J_z a_z b_z$  luminance has a behaviour close to the *HDR-Lab*<sub>1000</sub> luminance (cf. Figure 5.2).

The PU function and the *ICtCp* color space were not obtained through the same kind of training. They were created using Daly's Contrast Sensitivity Function model [98] and Barten's Contrast Sensitivity Function model [99], respectively. However, Figure 5.2, that represents the different color spaces luminance in function of the linear luminance, suggests that *ICtCp* and the PU encoded luminance have a behaviour closer to the *HDR-Lab*<sub>100</sub> luminance while the  $J_z a_z b_z$  luminance have a behaviour closer to the *HDR-Lab*<sub>1000</sub> luminance.

Because the color spaces are adapted for different viewing conditions, it is not easy to determine the best color space.

- With the proposed database 4Kdtb, the color spaces with a diffuse white around  $100 \text{ cd/m}^2$  (*ICtCp*, *HDR-Lab*<sub>100</sub> and the PU-mapping) give better performances than  $J_z a_z b_z$  and *HDR-Lab*<sub>1000</sub> spaces. We also observe that the performances of color metrics are more sensitive to the color space choice.
- We draw a similar conclusion on Zerman et al. database, except for FSIM and FSIMc (cf. Figure 5.4). These two metrics are less sensitive to the color space for this database.
- With the proposed database HDdtb (cf. Figure 5.6), the  $J_z a_z b_z$  color space provides the lowest performances for PSNR, SSIM, MS-SSIM and PSNR-HMA metrics but provides the highest performances with FSIM and FSIMc. However, results indicate that the PSNR, SSIM and PSNR-HMA metrics based on *HDR-Lab*<sub>1000</sub> and *HDR-Lab*<sub>100</sub> color spaces perform better than the same metrics using the  $J_z a_z b_z$  color space. This suggests that the low performances of these metrics are not due to the diffuse white characteristics of the images, but to the design of  $J_z a_z b_z$  color space which corrects a deviation in the perception of the blue hue (cf. Equation (1.29)). To test this hypothesis, we measure the SROCC of these metrics on the HDdtb database with the  $J_z a_z b_z$  color space without the blue deviation correction. We call this new space  $\widetilde{J_z a_z b_z}$ . Results, shown in Table 5.2, indicate that SROCC values of the three aforementioned metrics increase with the  $\widetilde{J_z a_z b_z}$  color space. In addition, metrics using this modified color space provide similar performances to metrics based on the *HDR-Lab*<sub>1000</sub> color space. This is consistent with the fact

that  $HDR-Lab_{1000}$  and  $J_z a_z b_z$  are adapted to almost the same diffuse white luminance. This might be due to the presence of the “gammut mismatch” artifact in this database. Indeed, the “gammut mismatch” artifact creates visible distortions that was not associated with a subjective quality loss during our test. We suspect that the blue hue deviation correction makes the  $J_z a_z b_z$  color space more sensitive to this distortion. However, this is difficult to demonstrate due to the low number of images with this kind of artifact present in this database.

Table 5.2: SROCC for the HDdtb database for three metrics based on  $J_z a_z b_z$ ,  $\widetilde{J_z a_z b_z}$  and  $HDR-Lab_{1000}$ .

Metrics	Color Spaces		
	$J_z a_z b_z$	$\widetilde{J_z a_z b_z}$	$HDR-Lab_{1000}$
PSNR	0.6933	0.7463	0.7587
SSIM	0.7831	0.7973	0.7904
PSNR-HMA	0.7664	0.7949	0.7984

- With the Narwaria et al. database, it is difficult to draw a conclusion (cf. Figure 5.7). The MS-SSIM and the SSIM metrics perform better when using the  $J_z a_z b_z$  color space. However, the FSIM and PSNR-HMA metrics perform better when using the  $ICtCp$  color space. This contrasted result might be due to the fact that the diffuse white luminance is likely not homogeneous across the entire database.

To go further into the analysis, we propose to evaluate the impact of the diffuse white on the performances of  $HDR-Lab$  metrics. The SROCC performances of three metrics (FSIM, MS-SSIM and SSIM) are evaluated for a diffuse white in the range 80 to 1000. Results are plotted in Figure 5.8. For the FSIM, the performances decrease slightly when the diffuse white luminance increases for the 4Kdtb database and the Narwaria et al. database while increasing with the diffuse white for the HDdtb. The impact of the diffuse white is more important on the MS-SSIM metric. For example, with the Zerman et al. database, the SROCC score drops from 0.9143 to 0.7791. The impact for the SSIM metrics is in the same order of magnitude as for MS-SSIM.

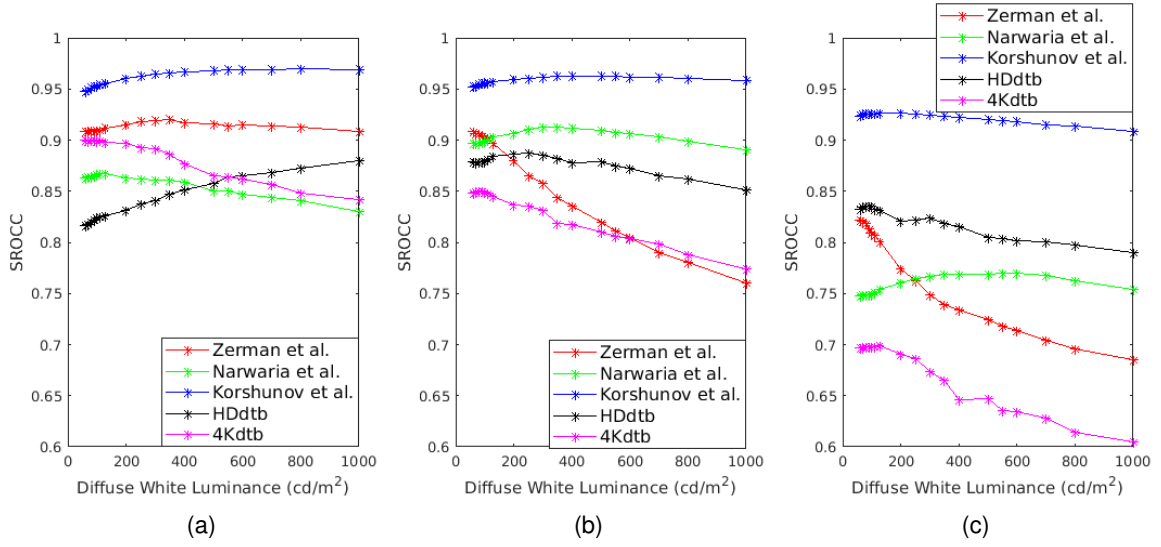


Figure 5.8: SROCC of **(a)**  $\text{FSIM}_{\text{HDR-Lab}}$ , **(b)**  $\text{MS-SSIM}_{\text{HDR-Lab}}$ , **(c)**  $\text{SSIM}_{\text{HDR-Lab}}$  in function of the diffuse white luminance.

#### 5.4.2 Sensibility to Chrominance Distortions

In this section, the ability of color metrics to take into account chrominance artifacts is discussed. The discussion is focused on the database 4Kdtb which is the only database providing significant chrominance artifacts. Also we only consider metrics using the  $\text{ICtCp}$  color space since the best performances are observed with this color space. Figure 5.9 presents the Mean Opinion Score (MOS) and objective scores for the reference image “Regatta\_24s”, for the distorted images (compressed with HEVC). The objective scores are given after applying the logistic regression presented in Section 5.3. Results for the other reference images can be found on Appendix D.

There is a clear difference of quality perception between the images compressed with the chroma  $Qp$  adaptation (cf. Section 3) (red Line) and the images compressed without the chroma  $Qp$  adaptation and a 8 bits quantization on the chrominance (blue line). The MOS of images compressed without the chroma  $Qp$  adaptation algorithm and a 10 bits quantization (green line) are in-between the two previous encodings.

As expected, the color-blind metrics, i.e. HDR-VDP2 and FSIM, are not sensitive at all to the chrominance distortions. However, more surprisingly, the color extension of FSIM, namely FSIMc, is not sensitive to the generated chrominance artifacts. The metrics was tailored for images in a BT.709 gamut with a SDR range. Its non-sensibility to the chrominance might be due to the pre-defined constant used for the color comparisons [65].

The other color metrics, i.e.,  $\overline{\Delta E^S}$ , SSIMc and CSSIM, are more sensitive to the chrominance artifacts. However, SSIMc and CSSIM have a tendency to underestimate the influence of

chrominance artifacts for images compressed with a low  $Q_p$  (so low distortion in the luminance channel) and a 8 bits quantization on the chrominance (cf. Figures D.2, D.4–D.6 and D.8).

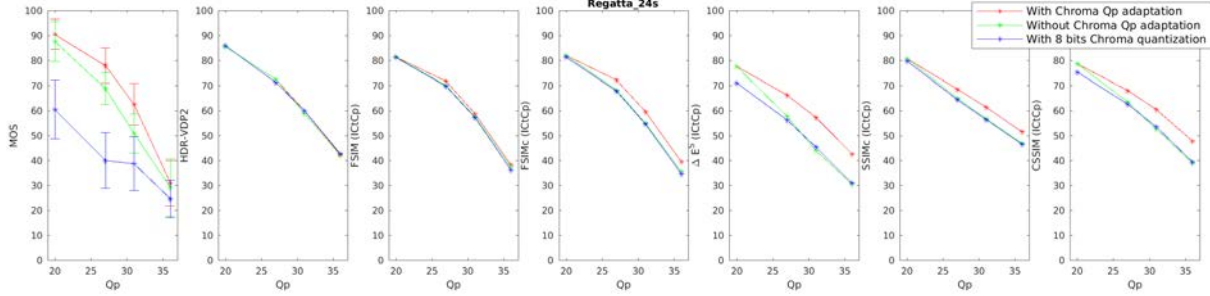


Figure 5.9: Subjective and objective scores for the image Regatta\_24s and for 6 metrics based on the  $ICtCp$  color space.

## 5.5 Applications

In this section, some recommendations are given to assess the HDR/WCG content quality in the context of image/video compression. The recommendations are listed below:

- For assessing the impact of luminance distortions, we recommend to use the FSIM metric. This is one of the best performing metrics. Moreover, it is the less sensitive to the choice of color space and to the diffuse white of the images. Using the color extension of the metrics (FSIMc) does not bring a significant added-value. In addition, it is important to underline that the FSIMc metrics is sensitive to the choice of color space (cf. Figure 5.3).
- To choose the color space, we recommend to take into account the diffuse white used during the color grading of the images. If the producer of the content follows the ITU recommendation BT.2408 [5] that defines the diffuse white luminance at  $203 \text{ cd/m}^2$ , we recommend to use the  $ICtCp$  color space. Indeed, this color space is well adapted to a low value of diffuse white. At the opposite, the  $J_z a_z b_z$  color space is well appropriate for a diffuse white luminance at  $997 \text{ cd/m}^2$ . Another benefit to use the  $ICtCp$  color space is related to its direct compatibility with popular compression codecs such as HEVC. However, in a recent HDR content analysis, in color-grading studio, there is some willingness to use higher diffuse white than recommended by the BT.2408 and that the diffuse white values is highly variable depending of the content [100].
- For application where the calculation time and the complexity are critical aspects, we recommend to be very careful with the choice of the color space. The simplest metrics, such as PSNR and SSIM, are much more sensitive than FSIM to the diffuse white luminance.

- If the chosen metrics is the PSNR, we recommend to first verify that the tested image/video processing application, such as compression codecs, does not create luminance mean shift or contrast change. These artifacts can be induced by backward compatible compression (if the image is first tone-mapped, then compressed using a legacy codec and finally tone expanded).
- For assessing the impact of both luminance and chrominance distortions, we recommend to use both the FSIM metrics and the  $\overline{\Delta E}_{ICtCp}^S$  metrics. None of the studied metrics are able to assess both at the same time luminance and chrominance distortions in a satisfactory manner.

Due to the characteristics of the tested databases, these recommendations have to be used in the context of image/video compression. Different subjective tests would be required to extend the analysis to other kinds of distortion.

## 5.6 Conclusion

In this section, we reviewed the relevance of using SDR metrics with perceptually uniform color spaces to assess the quality of HDR/WCG contents. We studied twelve different metrics along with six different color spaces. To evaluate the performances of these metrics, we used three existing HDR image databases annotated with MOS and the two databases specifically dedicated to WCG and chrominance artifacts presented in Chapter 3. We showed that the use of perceptually uniform color spaces increases, in most cases, the performances of SDR metrics for HDR/WCG contents.

In this study, we also highlights two weaknesses of state-of-art metrics. First, the relationship between the diffuse white used for grading the image and the diffuse white used for the color space is not always easy to define. In a number of cases, we do not know the value of the diffuse white used for the grading of the image. Choosing an arbitrary diffuse white for the color space may significantly alter the objective quality assessment. Further analysis of this relationship is required. A better understanding could help to evaluate compression of images using the HLG EOTF for which the diffuse white depends on the display. Second, to the best of our knowledge, the quality assessment of HDR/WCG images with chrominance distortions is still an open-issue, because of the lack of relevant objective metrics. In the following section, we proposed new metrics to tackle this issue.

# HDR-VDP-2 COLOR EXTENSION

---

## 6.1 Introduction

HDR-VDP-2 is one of the best performing metrics (cf. section 4.3 and 5.3). It aims to emulate with precision the early stage of human vision. Its principle is described in details in section 2.4. HDR-VDP-2 model precisely the human eye perception of images using the display spectral emission and the cones and rods spectral sensitivity. This metric also model precisely the contrast sensitivity function and the masking effect of the HVS. Its goal is to estimate the visibility of each difference between a reference and a distorted image. A pooling function is then used to transform a distortion map into a quality score.

This metric only takes into consideration the luminance distortion. However, it uses color information to correctly calculate photopic (daylight vision) and scotopic (night vision) luminance and thus models the Purkinje shift. To improve HDRVDP-2, we propose to extend its main mechanisms to the chrominance channels.

In this chapter, we first present our proposed extension of the HDR-VDP-2 metric toward the chrominance distortion, then in the second section, we describe the methodology to train this metric. We present the results in the third section. The last section concludes this chapter.

## 6.2 HDR-VDP-2 color extension

In this section, we present our proposition to extend the HDR-VDP-2 metric toward chromatic distortions. The architecture of this proposed HDR-VDP-2 color extension is illustrated in Figure 6.1.

We take advantage that the quantity of light perceived by each photo-receptors is precisely modeled ( $R_{L|M|S|R}$ ) for the reference and the distorted images (cf. equation 2.13).

The three maps obtained for each image by this transformation,  $R_L$ ,  $R_M$  and  $R_S$  corresponds to the images in the  $LMS$  color space after the modeling of the intra-ocular light scattering.

The obtained  $LMS$  color space can then be converted in perceptually linear color spaces like  $ICtCp$  or  $HDR-Lab$ . From this space, we can extract the two chrominance maps  $C_1C_2$  corresponding to the chrominance of the chosen color space  $CtCp$  for  $ICtCp$  or  $ab$  for  $HDR-Lab$ .

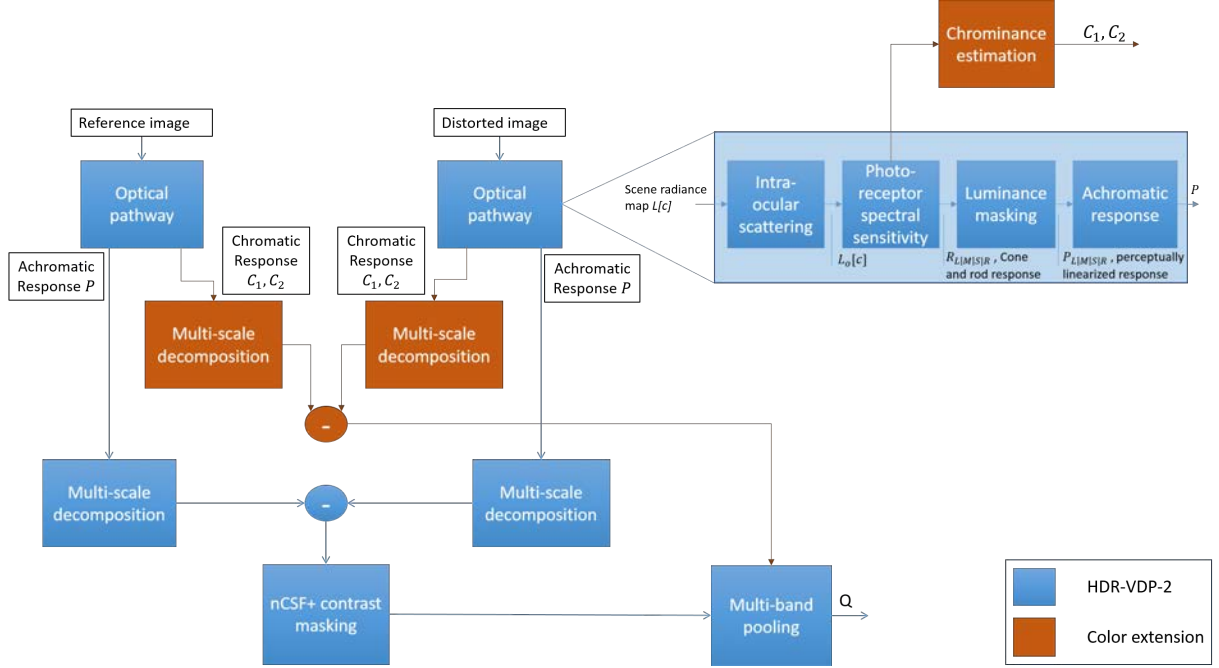


Figure 6.1: Architecture of our proposed HDR-VDP-2 color extension.

We choose to study the color spaces  $ICtCp$ ,  $HDR-Lab_{100}$  and  $HDR-Lab_{1000}$ . The resulting metrics are called  $HDR-VDP-2_{CtCp}$ ,  $HDR-VDP-2_{ab_{100}}$  and  $HDR-VDP-2_{ab_{1000}}$ .

For the sake of simplicity, we decompose those chromatic channels using the same pyramidal decomposition than for luminance, the steerable pyramid, to obtain the same number of sub-bands of frequency and orientation than for the luminance: four orientations bands and the maximum possible number of spatial frequency bands given the image resolution.

In HDR-VDP-2, the luminance subbands are subtracted to each other to obtain distortion maps as follows:

$$D[f, o](i) = \frac{|B_T[f, o](i) - B_R[f, o](i)|^p}{\sqrt{N_{nCSF}[f, o]^{2p} + N_{mask}[f, o](i)^2}} \quad (6.1)$$

with  $B_R[f, o]$  the subband of frequency  $f$  and orientation  $o$  of the reference image and  $B_T[f, o]$  the subband of frequency  $f$  and orientation  $o$  of the distorted image. The exponent  $p$  is the gain that controls the shape of the function.  $N_{nCSF}$  is a noise corresponding to the luminance neural CSF (cf. Section 2.4).  $N_{mask}$  corresponds to the contrast masking.

As for the luminance, we calculate the chrominance distortion maps:

$$D_{C_1|C_2}[f, o](i) = |B_{T_{C_1|C_2}}[f, o](i) - B_{R_{C_1|C_2}}[f, o](i)|^p \quad (6.2)$$

with  $B_{R_{C_1|C_2}}[f, o]$  the subband  $C_1$  or  $C_2$  of frequency  $f$  and orientation  $o$  of the reference image and  $B_{T_{C_1|C_2}}[f, o]$  the subband  $C_1$  or  $C_2$  of frequency  $f$  and orientation  $o$  of the distorted image. The exponent  $p$  is equal to 3.5 for the luminance distortion map ( $D_L[f, o]$ ) but have to be re-

estimated for chrominance channels. We trained the metric for different values of  $p$  for the chrominance channels: 3.5, the same value than for the luminance channel, 2 and 7.

Our proposed model neglects the contrast masking applied to the chrominance distortion maps. The contrast masking models a HVS mechanism that makes the visibility of distortions dependent on the image background. For example, those distortions can be harder to perceive if they are superimposed on a complex background. We neglect both the chrominance intra-channel contrast masking and inter-channel contrast masking (the contrast masking occurring between the two channels of chrominance and the luminance). Up to our knowledge, there is no color contrast masking model adapted to the WCG chrominance *CtCp* or *HDR-ab*. We hope that even if our model is not completely accurate, we can still improve HDR-VDP-2 performances.

Because, there is no chromatic CSF adapted to WCG content, we do not apply any corrective factor corresponding to the neural chromatic CSF noise, an equivalent for the luminance subband to  $N_{nCSF}$ . However, this factor will be learned through the training of the weights  $w_{f_{C_1}}$  and  $w_{f_{C_2}}$  (cf. Equation 6.3).

Then the pooling function (cf. equation 2.19) to obtain the quality score has to be modified:

$$\begin{aligned}
 Q = \frac{1}{F \cdot O} \sum_{f=1}^F \sum_{o=1}^O w_{f_L} \log \left( \frac{1}{I(f)} \sum_{i=1}^{I(f)} D_L^2[f, o](i) + \epsilon \right) + \dots \\
 \dots w_{f_{C_1}} \log \left( \frac{1}{I(f)} \sum_{i=1}^{I(f)} D_{C_1}^2[f, o](i) + \epsilon \right) + \dots \\
 \dots w_{f_{C_2}} \log \left( \frac{1}{I(f)} \sum_{i=1}^{I(f)} D_{C_2}^2[f, o](i) + \epsilon \right)
 \end{aligned} \tag{6.3}$$

with  $F$  the number of frequencies,  $O$  the number of decomposition,  $I(f)$  the number of pixel in the subband associated to the frequency  $f$  and  $\epsilon$  a small constant to avoid singularities. The weights  $w_{f_L}$  correspond to the weights of the luminance sub-bands with the frequency  $f$ .  $w_{f_{C_1}}$  and  $w_{f_{C_2}}$  correspond to the weights of the chrominance sub-bands. In HDR-VDP-2, there is 9 weights  $w_{f_L}$ . So in our proposed model, there is 9 weights  $w_f$  by channels. Those weights need to be trained.

The value  $Q$  does not correspond directly to a quality score. Indeed, this value is the accumulation of all the distortions of an image. The higher this value is, the lower is the quality of the image. In HDR-VDP-2(v2.2.1), the final metric is obtained as follows:

$$\hat{Q} = 100 - Q \tag{6.4}$$

where  $\hat{Q}$  is the final quality metric and  $Q$  the result of the pooling function. In the rest of the

chapter we used directly  $Q$  during the training.

## 6.3 Training methodology

### 6.3.1 Defining the optimisation problem

The weights  $w_f$  of the pooling function (cf. equation 6.3) have to be learned through training. In this case, we aim to optimize the SROCC between MOS scores of a subjective test and the objective metric scores. The SROCC have the advantage to not be affected by the non-linearity between subjective and objective scores. It is not desirable to learn this non-linearity, which is different for each image database and can be learned using a logistic regression (cf. Section 2.5).

As the pooling function gives a higher score  $Q$  if the distortion is important and the MOS score gives a higher score if the quality is better, the SROCC between MOS and  $Q$  should be as close as possible to -1.

Moreover, we consider that negative  $w_f$  is implausible as it would mean that a distortion increases the quality of an image. Thus a constraint is added to our optimization problem: all weights should be positive.

We chose to train the weights of all subbands of luminance ( $w_{f_L}$ ) and chrominance ( $w_{f_{C_1}}$  and  $w_{f_{C_2}}$ ). The chromatic distortion is often (but not always) correlated with luminance distortion. Using the original luminance weights might undermine the importance given to chromatic weights.

The optimization problem can be described as follows:

$$\begin{aligned} \underset{w_f}{\text{minimize}} \quad & 1 + SROCC(S, Q(w_f)) \\ \text{subject to} \quad & w_{f_i} \geq 0, \quad i = 1, \dots, 27 \end{aligned} \tag{6.5}$$

where,  $S$  the MOS score vector,  $Q(w_f)$  is the quality metric computed with the 27 weights  $w_f$ , 9 by channels.

### 6.3.2 Training databases selection

To obtain a database with enough images, HDR-VDP-2 (in its version 2.2.1) weights were trained using four different subjective test: Two HDR database (Narwaria et al. [70] and another database [101]) and as HDR-VDP-2 has the vocation to be a universal metric, for HDR and SDR content, two SDR databases CSIQ [86] and TID2008 [87]. Those two last were chosen due to the high number of images and the variety of their distortions.

We also chose to use multiple databases because there was no HDR database with enough images to train the metric. To illustrate this, we can train HDR-VDP-2<sub>CtCp</sub> on a unique database (using the particle swarm optimizer) and observe the result on other databases. We performed this training four times using four different databases: three HDR databases, Zerman et al., Korshunov et al. and Narwaria et al. (cf. Section 2.3) and one SDR database: TID2013. TID2013 [64] is an extension of the TID2008 database with notably more color artifacts (Image color quantization with dither, Chromatic aberration ...). For this database, we did not use the distortion called "Change of color saturation" as this default created outlier for HDR-VDP-2 (cf. Figure 6.2). The same behavior was observed for HDR-VDP-2<sub>CtCp</sub> after the training of the weights. This reminds the problem with the gamut mismatch artifacts of the HDdtb as both artifact create similar distortion(cf. Section 3.3). Those distortions are clearly visible but do not impact that much the image quality.

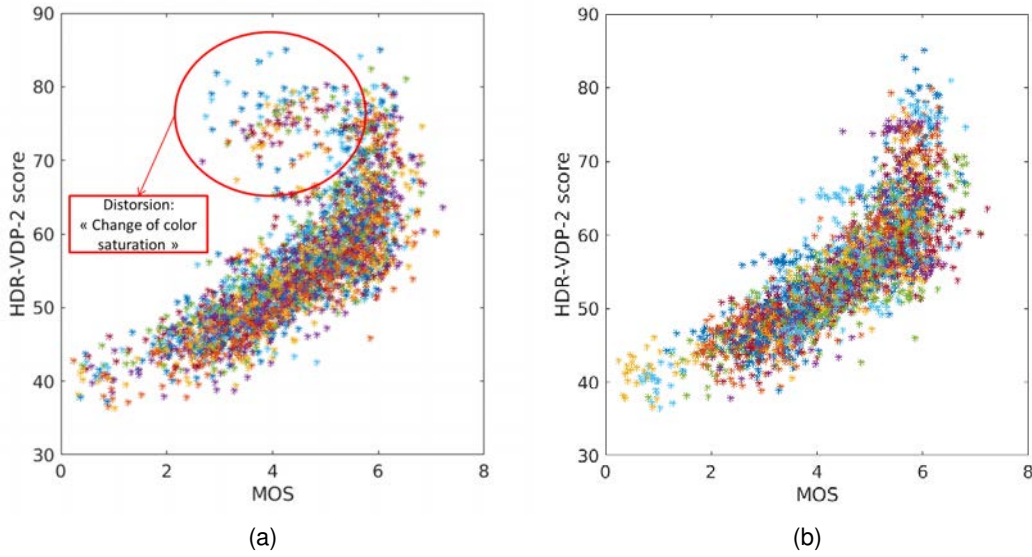


Figure 6.2: HDR-VDP-2 score in fonction of MOS for **(a)** The complete TID2013 database, **(b)** the TID2013 database without the distortion: "change in color saturation". One color represents all distorted images corresponding to one particular content.

Table 6.1 presents the performance of each trained HDR-VDP-2<sub>CtCp</sub> in term of SROCC. HDR-VDP-2 outperforms any version of HDR-VDP-2<sub>CtCp</sub> obtained from the different training. The different HDR-VDP-2 color extension trained on one database does not generalize well to other databases. To overcome the problem of databases with a small number of images, we chose to use three HDR database: the Narwaria et al and the Zerman et al., and two SDR databases selected for there high number of images: CSIQ and TID2013.

Table 6.1: Performance in term of SROCC of the proposed metric trained with different databases and HDR-VDP-2

tested on trained on	Narwaria et al.	Korshunov et al.	Zerman et al.
Narwaria et al. [70]	0.9843	0.8576	0.7765
Korshunov et al. [24]	0.7080	0.9871	0.8708
Zerman et al. [50]	0.2290	0.5199	0.9580
TID2013 [64]	0.5719	0.8607	0.7788
HDR-VDP-2	0.9012	0.9511	0.9285

### 6.3.3 Combining Databases

One problem to solve is the combination of multiple databases. Indeed, each subjective tests are different: They use a different protocol, use different images, different artifacts or different viewing conditions... Nor the subjective scores and nor the objective scores are homogeneous from one database to another. There is a solution for the subjective scores. A method like INLSA [102] allows to determine a common scale for all subjective score. However there is still the problem with the heterogeneity of objective score. Indeed, each database has its own angular resolution (cf. Table 6.2). As stated in Section 4.2.3 the impact of the angular resolution is not completely handled by the HDR-VDP-2 metric and a same HDR-VDP-2 score obtain with different angular resolutions can have different significations. The same problem arises during the training of our proposed metrics. To overcome this problem, the SROCC of each database has to be considered separately. This leads to a redefinition of the optimization problem :

$$\begin{aligned}
& \underset{w_f}{\text{minimize}} && 1 + \sum_j \alpha_j \text{SROCC}(S_j, Q_j(w_f)) \\
& \text{subject to} && w_{f_i} \geq 0, \quad i = 1, \dots, 27
\end{aligned} \tag{6.6}$$

Coefficients  $\alpha_j$  are defined in function of the number of images of the database obtain with the following constraint: SDR databases should not count for more than 50% of the score. AS TID2013 and CSIQ consist of much more images than HDR databases, using strictly the number of images as a factor to define the  $\alpha_i$  would reduce too much the impact of HDR images. A lower proportion of the loss function associated with SDR databaseq and we risk an over-fitting on the HDR databases which have less images than the SDR database. Table 6.2 summarizes the databases, their number of images, their angular resolution and the selected  $\alpha_i$  for the training.

Table 6.2: characteristics of the training databases

	database	#Image	angular resolution	$\alpha_i$
<b>SDR</b> <b>databases</b>	CSIQ	866	40	0.12
	TID2013	3000	30	0.38
<b>HDR</b> <b>databases</b>	Narwaria et al.	140	60	0.29
	Zerman et al.	100	40	0.21

#### The angular resolution problem

The use of databases with different angular resolutions raises another problem. The frequencies associated with each subband are not necessarily the same for each database. For example, for Narwaria et al., The frequencies associated with each subband are: 30, 15, 7.5, ... cpd (cycle per degree) and for the Zerman et al. database, the frequencies associated with each subband are: 20, 10, 5, ... cpd. However, the trained weights  $w_f$  must be associated with a precise frequency band. To determine the weight associated with a frequency band that is not directly defined, HDR-VDP-2 uses linear interpolation (see section 2.4.2). Thus, during the training, when calculating the quality score with a set of candidate weights  $w_f$  of an image whose subbands are not aligned with the defined frequencies, a linear interpolation is used to define the weights associated with each subband.

#### 6.3.4 Validation protocol

As the optimization problem is not convex, HDR-VDP-2 (v2.2.1) [54] used the Nelder-Mead method [103], which does not require to compute gradient descent. However, this method does not support a too high dimensionality [104]. Our proposed model triples the number of weights (passing from 9 weights for HDR-VDP-2 to 27 for our proposed extension). As a result, we choose the particle swarm optimization (PSO) [105] as this algorithm tolerates better such dimensionality without computing any gradient.

We separate each training database (cf. Section 6.3.3) in a training set (80% of images) and a test set (20% of images) with the condition that the same image (even with different distortions) can not be found in the test set and in the training set. To obtain a more robust estimation of the proposed model performance and to eliminate selection bias, we repeat this split 1000 times and perform the training of the weight 1000 times. Each time the results are evaluated on the test set.

We used the Korshunov et al. a database that presents by construction almost no chromatic distortions to verify that our proposed metric still correctly handle luminance artifact and our proposed database 4Kdtb, a database that presents such chromatic artifacts to validate our results.

## 6.4 Results

### 6.4.1 Metric performances

Table 6.3 shows the median SROCC obtain on the test sets through 1000 training. The color space choice does not have a significant impact on the metric performances, neither is the factor  $p$  (even if the trained metrics that use a factor of 7 are slightly less correlated to the MOS than other metrics).

Our proposed model does not increase the performances on the CSIQ database and is slightly better on the TID2013 database. This is harder to conclude on the other two databases as the test sets are composed of very few images: 30 for the Zerman et al. database and 42 for the Narwaria et al. database. However, we can observe that the SROCC are slightly better than HDR-VDP-2 for the Narwaria et al. database and slightly worst for the Zerman et al. database. This may be due to the higher weight  $\alpha_i$  given to the SROCC of the Narwaria et al. database during the training.

Table 6.3: Median SROCC across 1000 Train-Test Combinations database by database.

Metric	$p$	CSIQ	TID2013	Narwaria et al.	Zerman et al.
HDR-VDP-2		0.9416	0.8572	0.9213	<b>0.9432</b>
HDR-VDP-2 <sub>CtCp</sub>	2	0.9438	0.8596	0.9255	0.9212
HDR-VDP-2 <sub>CtCp</sub>	3.5	0.9411	0.8669	0.9377	0.9304
HDR-VDP-2 <sub>CtCp</sub>	7	0.9301	0.8548	0.9125	0.9265
HDR-VDP-2 <sub>ab100</sub>	2	0.9423	0.8611	0.9301	0.9225
HDR-VDP-2 <sub>ab100</sub>	3.5	0.9446	0.8651	<b>0.9318</b>	0.9405
HDR-VDP-2 <sub>ab100</sub>	7	0.9342	0.8548	0.9140	0.9195
HDR-VDP-2 <sub>ab1000</sub>	2	0.9400	0.8594	0.9245	0.9212
HDR-VDP-2 <sub>ab1000</sub>	3.5	<b>0.9456</b>	<b>0.8691</b>	0.9290	0.9255
HDR-VDP-2 <sub>ab1000</sub>	7	0.9212	0.8421	0.9100	0.9195

To validate our result, we select the median weights of 1000 training and a  $p$  factor of 3.5. Table 6.4 presents the result for the training databases and the Table 6.5 on the tested databases. On the training databases, our proposed color extension has similar performance for the CSIQ and the Zerman et al. databases and is slightly better for the TID2013 database

and the Narwaria et al. database. On the tested databases, our proposed color extension does not improve the performance of HDR-VDP-2: the difference in term of SROCC is inferior at 1%.

Table 6.4: SROCC of metrics with the median weights for the training databases.

Metric	CSIQ	TID2013	Narwaria et al.	Zerman et al.
HDR-VDP-2	0.9404	0.8528	0.8906	0.9289
HDR-VDP-2 <sub>CtCp</sub>	<b>0.9449</b>	0.8653	0.9245	0.9287
HDR-VDP-2 <sub>ab100</sub>	0.9443	0.8632	0.9243	<b>0.9297</b>
HDR-VDP-2 <sub>ab1000</sub>	0.9445	<b>0.8659</b>	<b>0.9258</b>	0.9198

Table 6.5: SROCC of metrics with the median weights on the validation databases.

Metric	Korshunov et al.	4Kdtb
HDR-VDP-2	0.9511	0.8678
HDR-VDP-2 <sub>CtCp</sub>	0.9434	0.8718
HDR-VDP-2 <sub>ab100</sub>	0.9480	0.8672
HDR-VDP-2 <sub>ab1000</sub>	0.9487	0.8717

Worse than that, the impact of the color subbands is negligible compared to the luminance subbands. This is illustrated by Figure 6.3 that represents the score given by HDR-VDP-2<sub>CtCp</sub> but without the inclusion of chrominance subbands in function of HDR-VDP-2<sub>CtCp</sub> on the 4Kdtb.

As expected, with our training, we succeed to obtain similar performances than HDR-VDP-2. However, we failed to improve its performance or to extend its sensitivity to the chrominance distortions.

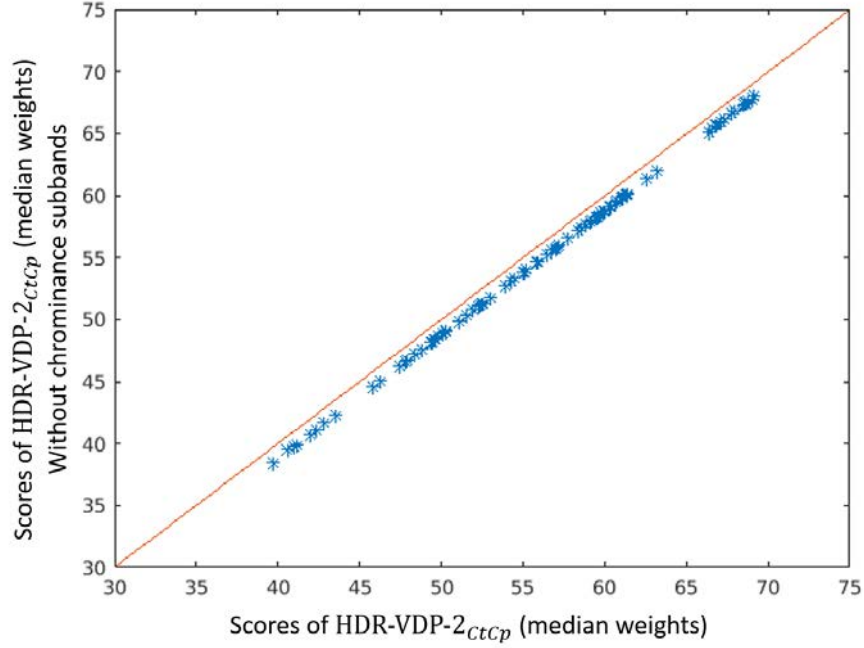


Figure 6.3: HDR-VDP-2<sub>CtCp</sub> scores without the chrominance subbands in function of HDR-VDP-2<sub>CtCp</sub> with the chrominance subbands (using median weights) (Database: 4Kdtb).

#### 6.4.2 Analysis of trained weights

To understand the results presented in the previous subsection, it is important to analyze the weights obtained with our training. In this subsection, we focus on the trained metric HDR-VDP-2<sub>CtCp</sub>, but similar results can be found with other color spaces.

During our training, in addition to the chrominance subbands, we also trained the weights of the luminance subbands. We can compare the weights of HDR-VDP-2 and the weights of HDR-VDP-2<sub>CtCp</sub>. Figure 6.4 represents the original weights of HDR-VDP-2 and the median weights obtained for the luminance channel of HDR-VDP-2<sub>CtCp</sub>. We can observe that even if the two metrics have similar performances, we obtained significantly different weights. First, we can see that the four highest frequency sub-bands 30, 15, 7.5 and 3.75 cpd contribute almost equally to the HDR-VDP-2 score, whereas for HDR-VDP-2<sub>CtCp</sub>, only two subbands (7, 5 and 3.75 cpd) are predominant for assessing quality. For HDR-VDP-2<sub>CtCp</sub>, The highest frequencies are null. Another difference is the weight associated with the frequency 0.4688 cpd. This weight is almost null for HDR-VDP-2 while it has a small but significant influence on HDR-VDP-2<sub>CtCp</sub>. If the weights are different but the performance of the metrics are similar, it shows that the problem of estimating quality metrics is complex and that it is possible to find many local minima on the basis of different training databases.

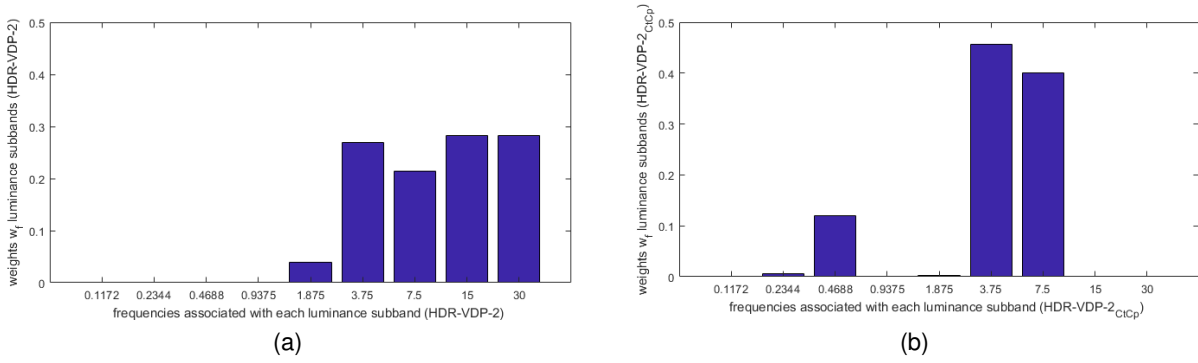


Figure 6.4: Luminance weights  $w_f$  of (a) HDR-VDP-2 and (b) HDR-VDP-2<sub>CtCp</sub> (median weights)

Figure 6.5 represents the distribution of the value taken by each chrominance and luminance weight  $w_f$  of the metric HDR-VDP-2<sub>CtCp</sub> across the 1000 formations. First, we can observe that most weights are null or almost null, even for luminance weights. The variation is not very important for weights close to zero. This confirms that, during training, the highest luminance frequencies (30 and 15 cpd) always converge toward zero, unlike the HDR-VDP-2 weights. The variation are higher for the frequency 7.5 and 3.75 cpd which can mean that the error in the two subbands are highly correlated. The fact that the chrominance weights always converge towards zero, except in a minority of cases for the 0.4688 cpd frequency, confirms that the characteristics obtained from the chrominance subbands are not efficient to evaluate the quality of the color images.

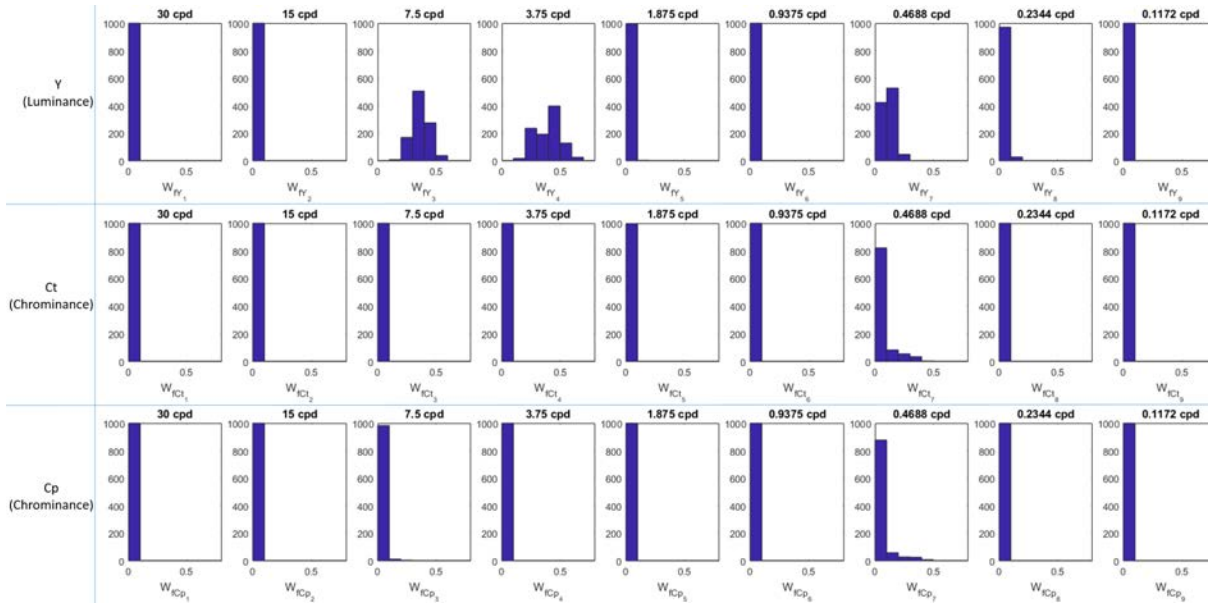


Figure 6.5: the distribution of the value taken by each  $w_f$  weight for HDR-VDP-2<sub>CtCp</sub>

### 6.4.3 Discussion

Several factors in the design of the color extension of HDR-VDP-2 and on the training protocol can explain the non sensitivity of chromatic distortion of the trained metric. We list afterwards five hypothesis that could explain those results:

- **The chromatic masking model:** As stated in the section 6.2, when designing the HDR-VDP-2 color extension, we chose to neglect intra-channel and inter-channel chrominance masking. This HVS mechanism has an impact on the visibility of the differences. Small differences are more difficult to detect on a complex background. This is an important feature of the original HDR-VDP-2 metric. The SROCC of HDR-VDP-2 goes from 0.92 to 0.81 on the Zeman et al. Database if this mechanism is not used on the luminance channel. Implementing a color masking model could improve the relevance of chromatic subbands for evaluating image quality.
- **The choice of the logarithm in the pooling function:** In the pooling function, the sum of all distortions are modified using a logarithm (cf. Equation 6.3). This choice was made for the luminance subbands in HDR-VDP-2 because it gives better result than linear values. However, it might not be the best pooling strategies for chrominance subbands.
- **The weight interpolation:** All the databases used in the training phase did not have the same angular resolution. As a result, some databases necessitate to use linear interpolation to assign weights to subbands whose frequency bands are not aligned with the candidate weights associated frequencies. This might impact the quality of the training. Obtaining more images with the same angular resolution could facilitate this training and allow us to use the INSLA methodology to combine the databases.
- **The number of weights:** Due to the lack of data, training 27 weights using algorithm like the particle swarm optimizer can create overfitting and complicate the training. Fixing the luminance weights to their original value would reduce this number. In addition, for chrominance channels, there is no need to consider the highest frequencies as they are invisible to the HVS. This would also speed up the training.
- **The database:** There are no existing HDR/WCG databases as diverse in term of distortion and with as many images as SDR databases such as TID2013. Creating such databases is expensive and time consuming. However, this database could prevent the current inclusion of less relevant SDR databases to address a lack of data. Training would be more dedicated and consequently improved.

## 6.5 Conclusion

In this section, we proposed a metric to extend the HDR-VDP-2 metric toward chrominance distortions. However, if we succeed to obtain similar performance than HDR-VDP-2 with our model, we failed to improve significantly its performances. Worse, we also failed to improve its sensitivity to chrominance distortion.

Several factors can explain the difficulties we have faced like the lack of chromatic masking model, the number of weights we choose to train or the chosen databases. The resolution of all these problems is complex and/or time consuming, without certainty that we can improve HDR-VDP-2 this way. This is why we choose to explore a different approach to create a quality metric.

In the following chapter, we present a new quality metric that consist in the aggregation of several metrics and features, including HDR-VDP-2. This approach allows us to avoid some of the difficulties involved in training the HDR-VDP-2 color extension mentioned in the previous section.

# QUALITY METRIC AGGREGATION

---

## 7.1 Introduction

To obtain a HDR metric suitable to assess color distortion, we propose to combine two full-reference quality metrics and two color image features using a Support Vector Regressor (SVR). Combining several features and quality metrics for improving the overall prediction performance was already proposed for SDR images and video. We can mention CQM [106], CF-MMF and CD-MMF [107], FVQA [108], EVQA [109] and the metric developed by Netflix, VMAF [110].. Recently, Choudhury et al. [111] proposed HDR-CQM, a combination of full-reference metrics adapted to HDR contents. The main idea is to combine, using support vector machine regression, the scores of a subset of metrics selected from a list of quality metrics. Although performing well, the proposed metric does not tackle the issue of WCG chromatic distortions.

The first section of this chapter presents our proposed metric. The second section presents the training methodology. The results are presented in a third section. The last section concludes this chapter.

## 7.2 Proposed method

The proposed metric relies on the mixture parameters of several quality metrics and image features; the combination is learned by using supervised machine learning. The main idea is to take advantage of each metric / feature strengths while compensating individual weakness. A Support Vector Regression machine with a Gaussian radial basis function kernel is used to map the features to a quality score. Figure 7.1 presents the overall architecture of the proposed method.

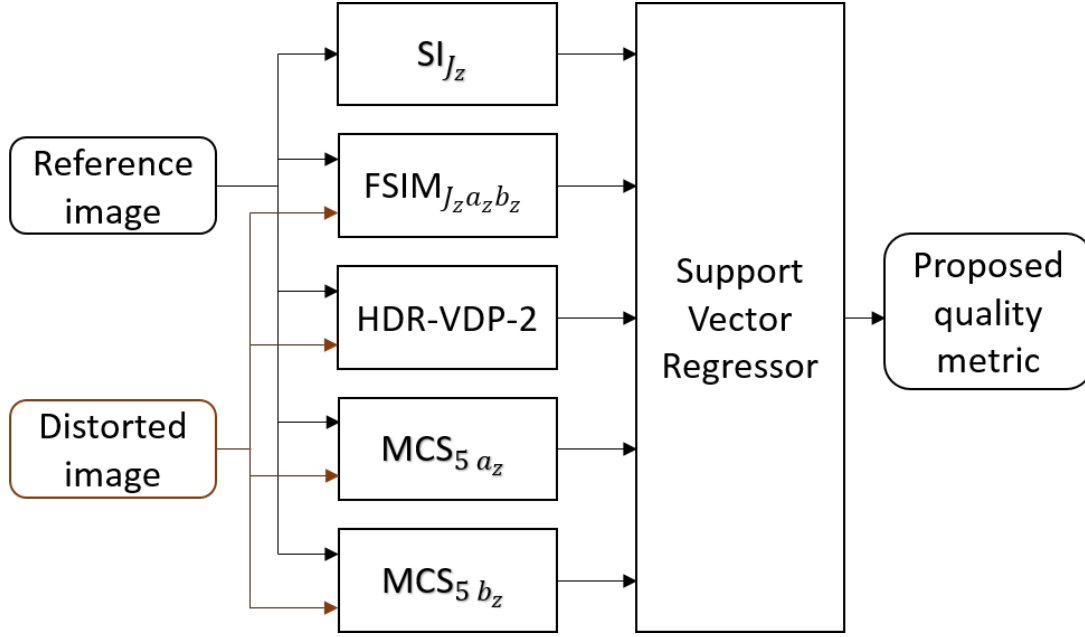


Figure 7.1: Architecture of the proposed full-reference HDR quality metric.

### 7.2.1 Considered HDR quality metrics

Two HDR metrics were selected for their good performances on several databases (cf. section 5): HDR-VDP-2 and  $\text{FSIM}_{J_z a_z b_z}$ .

- HDR-VDP-2(version 2.2.1): HDR-VDP-2 requires numerous parameters such as the angular resolution, the surround luminance and the spectral emission of the screen. To ensure that HDR-VDP-2 score is coherent, we use the same angular resolution (60 pix/deg, the angular resolution of three of the databases), for each studied database. Using different angular resolution could result in incoherence between HDR-VDP-2 scores (cf. Section 4.2).
- $\text{FSIM}_{J_z a_z b_z}$  : The FSIM metric adapted to the  $J_z a_z b_z$  color space using the methodology presented in section 5.

The two metrics considered above are only considering luminance distortions. To consider the chromatic dimension, we add two chromatic-based features.

### 7.2.2 Chromatic-based visual features

As we aim to extend the quality metric to color images, we add two features able to measure chromatic distortion. These features are inspired by the MS-SSIM features [75], the multi-scale version of the SSIM index [72]. SSIM conducts a comparison on three levels: luminance,

contrast and structure. Each factor is estimated pixel-wise through a sliding window in order to achieve a distortion map. Thus, SSIM is not comparing global but local luminance, contrast and structure. To obtain a global quality score, a mean is applied on this map. In MS-SSIM, contrast and structure are compared on different scales of the images. Our color features are created using the contrast ( $c$ ) and structure ( $s$ ) comparison function on the color components of the images. Because the HVS is less sensitive to high frequency variation in color than in luminance, we use the scale 5 of the images, i.e. the images after 5 downsampling with a ratio of 2. The goal is to capture relevant chromatic distortions. First, a distortion map,  $CS_5$  is calculated:

$$CS_5(m, n) = c(x_{mn}, y_{mn}) \times s(x_{mn}, y_{mn}) \quad (7.1)$$

where  $x_{mn}$  and  $y_{mn}$  represent pixel values with the coordinates  $m$  and  $n$  of the reference image  $X$  and the distorted image  $Y$ . Then, the feature,  $MCS_5 a_z$ , is calculated as following:

$$MCS_5(X, Y) = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N CS_5(m, n) \quad (7.2)$$

where  $M$  and  $N$  represent the image sizes. For both original and degraded images, we calculate this feature on the  $J_z a_z b_z$  color component of the images:  $MCS_5 a_z$  for the  $a_z$  component, and  $MCS_5 b_z$  for the  $b_z$  component.

Figures 7.2 and 7.3 represent, respectively, the distortions maps  $CS_5 a_z$  and  $CS_5 b_z$  of the 4Kdtb image Bike\_81s obtained when the image is compressed with a Qp of 20 with two different ways of handling color information: a 10 bits quantization of the chroma with a chroma Qp offset and a 8 bits quantization of the chroma. We can observe that  $MCS_5 a_z$  and  $MCS_5 b_z$  detect more distortions when there is an 8 bits quantization of the chroma, especially in the regions with high luminance (in this image, the sky and the street lights). Almost no distortion is detected when there is a 10 bits quantization with a chroma Qp offset.

In addition, we also add a last feature to characterize the image complexity. Indeed this characteristic affects our perception of the image distortion.

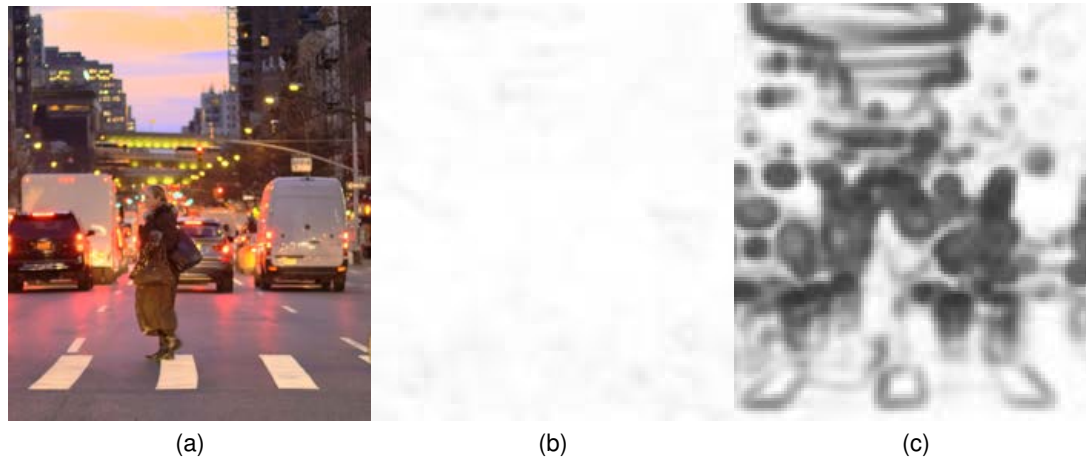


Figure 7.2: The 4Kdtb image Bike\_ 81s **(a)** and the distortion map  $CS_5_{a_z}$  obtained for the image compressed with a Qp of 20 with **(b)** a quantization of 10 bits for the chrominances **(c)** a quantization of 8 bits for the chrominances. A white pixel means that there is no differences and a black pixel that there is a huge distortion.

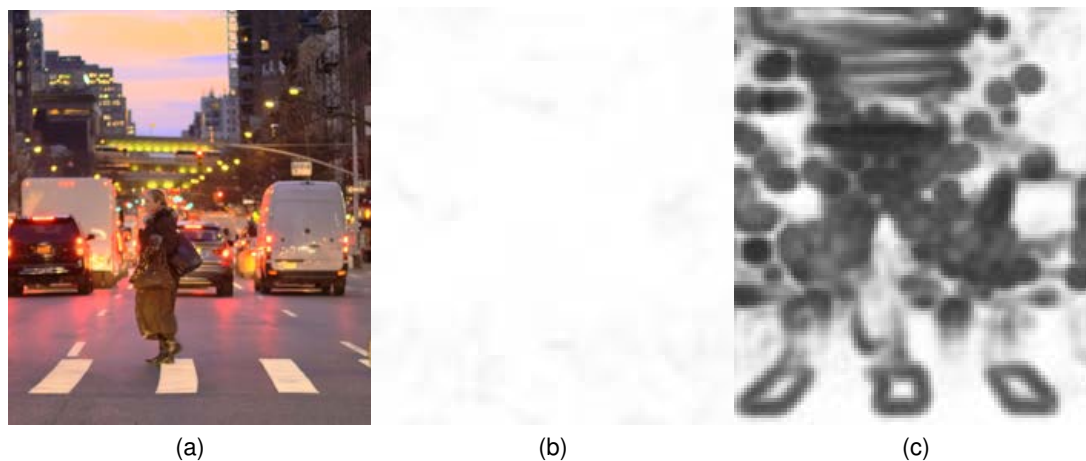


Figure 7.3: The 4Kdtb image Bike\_ 81s **(a)** and the distortion map  $CS_5_{b_z}$  obtained for the image compressed with a Qp of 20 with **(b)** a quantization of 10 bits for the chrominances **(c)** a quantization of 8 bits for the chrominances. A white pixel means that there is no differences and a black pixel that there is a huge distortion.

### 7.2.3 Image spatial information

The last feature allows us to characterize the complexity of the image. Image complexity affects the image quality perception as it is related to the HVS spatial masking effects. It is measured on the  $J_z$  component of the reference images. We selected Spatial Index SI as recommended by the International Telecommunication Union (ITU) [48]. As stated in Section 2.2, the spatial information  $SI_{J_z}$  corresponds to the standard deviation of the image luminance plane which has been filtered by a Sobel filter:  $SI = \text{std}[\text{Sobel}(X)]$ , where  $X$  is the reference image.

### 7.2.4 Mapping the features to the quality scores

To summarize, we compute from the original and distorted images a vector of 5 parameters: 2 parameters deal with the luminance component, 2 concerns chroma dimensions and one is used to characterize the image complexity. The relationship between these features and the quality scores is learned thanks to the SVR machine (cf. Figure 7.1).

## 7.3 Training methodology

### 7.3.1 Combining several databases

To train the proposed model, a large collection of annotated HDR images is required. Unfortunately, there is, to the best of our knowledge, only a small number of HDR image databases annotated with subjective quality scores. In addition those databases are composed of a rather small amount of images. (cf. Section 2.2)

To obtain a suitable database for our experiment, we considered the five databases presented in section 2.2 and 3.: Narwaria et al. [49], Korshunov et al. [24], Zerman et al. [50], 4Kdtb and HDdtb. The first four databases were used for the training phase and HDdtb was considered as an independent test database used to validate our proposed metric.

As in section 6, one problem to solve is the combination of multiple databases. For this purpose, we align the MOS of the different subjective tests into a common quality scale thanks to the Iterated Nested Least Square Algorithm (INLSA) proposed in [102]. INLSA allows to determine a common scale. As we use the same angular resolution parameter for all databases, using this algorithm does not pose the same problem as in section 6. INLSA is based on the assumption that objective quality metrics are linearly correlated with the subjective scores. Four objective metrics were selected: HDR-VDP-2,  $VIF_{J_z a_z b_z}$ ,  $MS\text{-}SSIM_{J_z a_z b_z}$  and  $FSIM_{J_z a_z b_z}$ . The Figure 7.4 illustrates the impact of the INLSA algorithm. It represents the HDR-VDP2 score in function of the MOS before the INLSA algorithm (MOS are re-scale between 0 and 1) and after the INLSA algorithm.

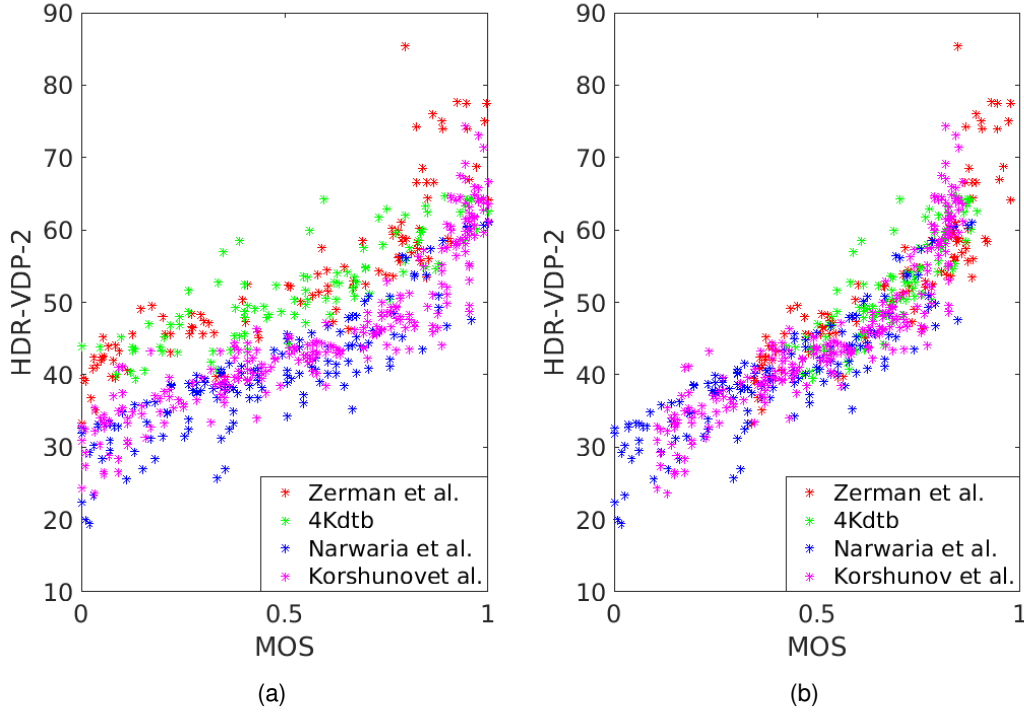


Figure 7.4: HDR-VDP2 score in function of MOS **(a)** before the INLSA algorithm **(b)** after the INLSA algorithm

### 7.3.2 Support Vector Regression

To train the SVR, we first conduct a grid-search using all of the training databases to tune the hyper-parameters to obtain optimal results. Once the best hyper-parameters are found, we split our training database into two sets of images, one for training the SVR (80% of images) and one for testing (20% of images). This split is done randomly with the constraint that the same content can not be present in the training set and testing set. We then train the SVR using the function `fitrsvm` of Matlab. To obtain a more robust estimation of the proposed model performance and to eliminate selection bias, we repeat this split 1000 times.

We evaluate the performance of our proposed metric using the median scores of four indexes: the Spearman Rank Order Correlation Coefficient (SROCC), the Kendall Rank Correlation Coefficient (KRCC), the Pearson Correlation Coefficient (PCC) and the Root Mean Square Error (RMSE).

Table 7.1: Median performances across 1000 Train-Test Combinations on the test set.

Metric	SROCC	KRCC	PCC	RMSE
MS-SSIM $_{J_z a_z b_z}$	0.8096	0.6246	0.7963	16.6
MS-SSIM $_{ICtCp}$	0.9197	0.7507	0.9237	10.74
FSIM $_{J_z a_z b_z}$	0.8955	0.7267	0.8953	12.2
FSIM $_{ICtCp}$	0.8807	0.7104	0.8757	13.52
HDR-VDP-2	0.9186	0.7496	0.9179	10.9
metric (1)	0.9234	0.7597	0.9244	10.46
metric (2)	0.9165	0.7563	0.9235	10.51
metric (3)	0.9113	0.7413	0.9101	11.36
metric (4)	0.9294	0.7692	0.9271	10.28
<b>Proposed metric</b>	<b>0.9421</b>	<b>0.7899</b>	<b>0.9376</b>	<b>9.5</b>

## 7.4 Results

### 7.4.1 Proposed metric performance

In this section, we present the performance of the proposed metric on the 1000 test sets. The median performances across 1000 Train-Test Combinations can be found in Table 7.1. We compare the score of the proposed metric to five other metrics: HDR-VDP-2 and four SDR metrics adapted to HDR using two uniform color space  $J_z a_z b_z$  and  $ICtCp$  [6] using the methodology presented in section 5.2 MS-SSIM $_{J_z a_z b_z}$ , MS-SSIM $_{ICtCp}$ , FSIM $_{J_z a_z b_z}$ , FSIM $_{ICtCp}$ . To ensure that each feature is meaningful, we also report four other trained metrics using only a subsection of all features. Metric (1) is trained with only HDR-VDP-2 and FSIM $_{J_z a_z b_z}$ , metric (2) with all features except HDR-VDP2, metric (3) with all features except FSIM $_{J_z a_z b_z}$  and metric (4) with all features except SI $_{J_z}$ . Table 7.1 shows that the proposed model provides superior performances than any other metric. To ensure that the good performances of the proposed model are homogeneous across all databases, we also report the median SROCC of the 1000 trained metrics, database by database in Table 7.2. Because the number of images in the test set coming from one database was rather small, we use the complete databases to calculate the performance indexes and not only the images used in the test set. We observe that achieving high accuracy on one database does not preclude the model performances on another database. Our proposed model reaches higher performances when compared to other metrics and especially to HDR-VDP-2 and FSIM $_{J_z a_z b_z}$ , the two metrics that are also forming the basis of our model.

Table 7.2: Median SROCC across 1000 Train-Test Combinations database by database.

Metric	4Kdtb	khorshunov et al.	Narwaria et al.	Zerman et al.
MS-SSIM <sub><math>J_z a_z b_z</math></sub>	0.8306	0.9648	0.9088	0.8109
MS-SSIM <sub>ICtCp</sub>	0.8447	0.9529	0.8714	0.9260
FSIM <sub><math>J_z a_z b_z</math></sub>	0.8849	0.9663	0.8466	0.9031
FSIM <sub>ICtCp</sub>	0.9049	0.9477	0.8645	0.8863
HDR-VDP-2	0.8678	0.9516	0.8909	0.9289
<b>Proposed metric</b>	<b>0.9095</b>	<b>0.9704</b>	<b>0.9240</b>	<b>0.9474</b>

#### 7.4.2 Validation on an independent database

In this section, we provide further evidence that the proposed metric can be generalized to other databases. More specifically, we evaluate its performance on the HDdtb database. In Table 7.3, we report results obtained on the complete HDdtb database and the results on the HDdtb images with compression artifacts only. It is worth mentioning that such artifacts are of same nature that those we trained our model for. The proposed metric outperforms five state-of-the-art metrics for all performance indexes when considering the complete database. As it was expected, the gain is higher when considering only the images with compression artifacts.

Table 7.3: Performances of several metrics for (a) the complete database HDdtb (b) Only the HDdtb images with compression artifacts.

(a)	SROCC	KRCC	PCC	RMSE
MS-SSIM <sub><math>J_z a_z b_z</math></sub>	0.8557	0.6646	0.8603	14.11
MS-SSIM <sub>ICtCp</sub>	0.8464	0.6462	0.8763	13.33
FSIM <sub><math>J_z a_z b_z</math></sub>	0.9069	0.7278	0.9187	10.94
FSIM <sub>ICtCp</sub>	0.8099	0.6080	0.8234	15.70
HDR-VDP-2	0.8685	0.6721	0.8715	12.55
<b>Proposed metric</b>	<b>0.9110</b>	<b>0.7357</b>	<b>0.9250</b>	<b>9.05</b>
(b)	SROCC	KRCC	PCC	RMSE
MS-SSIM <sub><math>J_z a_z b_z</math></sub>	0.8752	0.7165	0.8860	13.13
MS-SSIM <sub>ICtCp</sub>	0.8651	0.6749	0.8923	12.83
FSIM <sub><math>J_z a_z b_z</math></sub>	0.9173	0.7619	0.9297	10.47
FSIM <sub>ICtCp</sub>	0.8271	0.6437	0.8289	15.90
HDR-VDP-2	0.8739	0.6866	0.9018	12.28
<b>Proposed metric</b>	<b>0.9422</b>	<b>0.7931</b>	<b>0.9510</b>	<b>8.53</b>

### 7.4.3 Sensitivity to color distortions

One of the metric goals is to assess not only the luminance distortions but also chromatic distortions. To illustrate this behavior, we use the images of the 4Kdtb database. Figures 7.5 and 7.6 represent MOS score, the color-blind metrics HDR-VDP-2 and our proposed metric. The difference between the compression strategies is chrominances management. As illustrated, for example, on Figure 7.5 (c) (left), observers are sensitive to the different modes of compression. As expected, HDR-VDP-2 is not sensitive to chromatic distortions while the proposed metric is. We can still observe that color artifacts are slightly underestimated for images with a low  $Q_p$  and 8 bits quantization for the chrominance (cf. Figures 7.5 (a) and 7.6 (a), (b) and (d)).

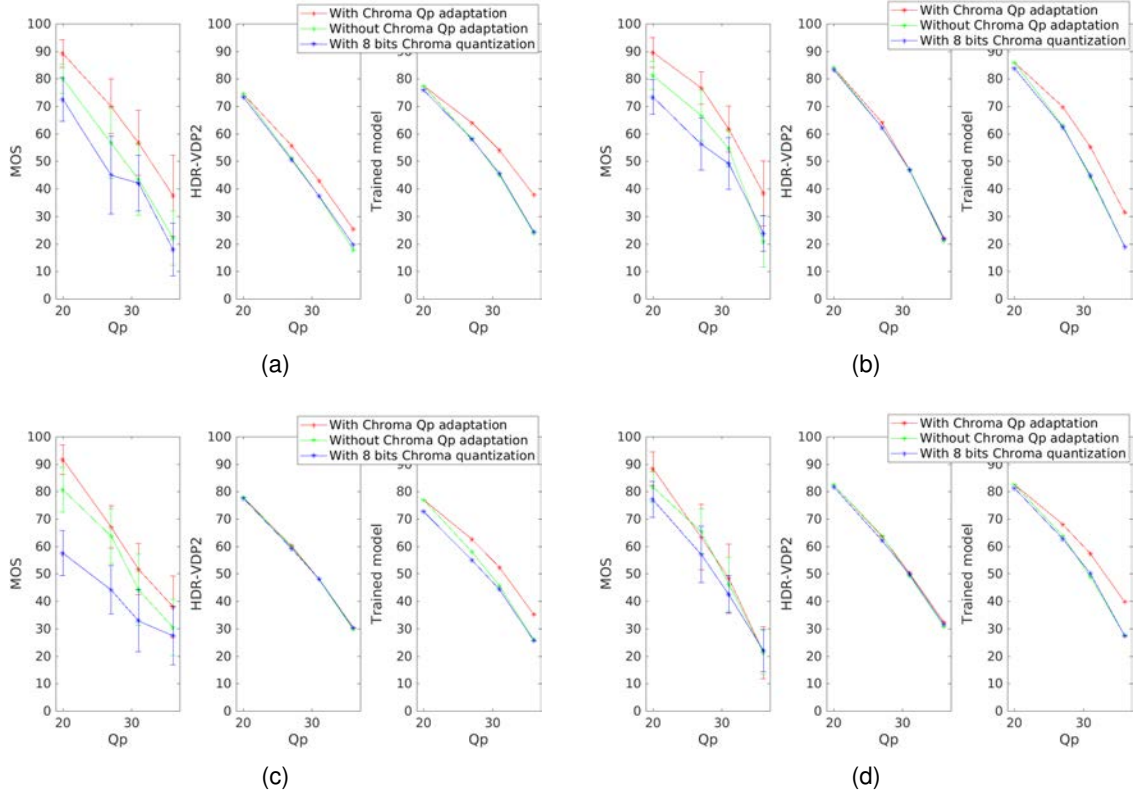


Figure 7.5: Subjective and objective scores in function of HEVC  $Q_p$  for the images (a) Bike\_110s (b) Bike\_20s (c) Bike\_30s (d) Bike\_30s.

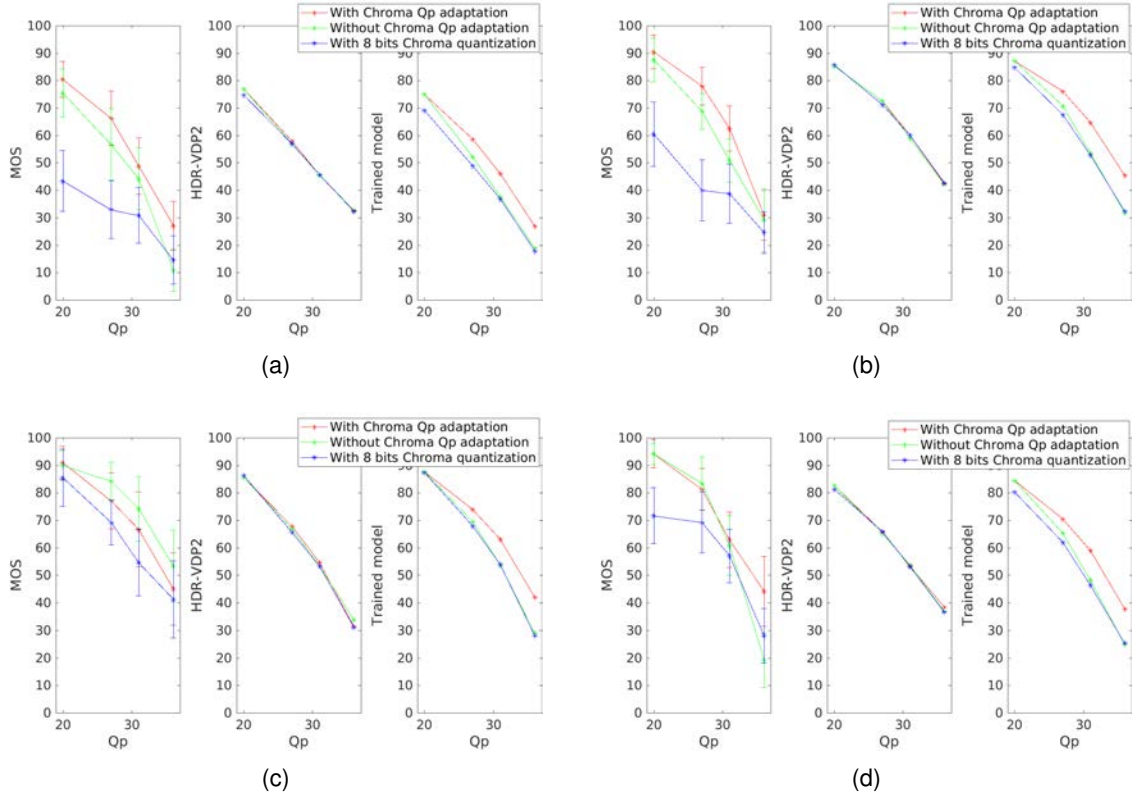


Figure 7.6: Subjective and objective scores in function of HEVC Qp for the images **(a)**Regatta\_11s **(b)** Regatta\_24s **(c)** Regatta\_80s **(d)** Regatta\_95s.

## 7.5 Conclusion

In this chapter, we proposed a new full-reference quality metric adapted to HDR/WCG content. This metric is a combination of two full-reference quality metrics as well as color image features and the image spatial information. Those features are fused into one metric with an SVR method. Experimental results show that the proposed metric can accurately assess the quality of images distorted by compression. Moreover, contrary to most HDR metrics, the metric is sensitive to chromatic distortions as well as luminance distortions. However, due to the characteristics of the database used to train and test the metric, it is not possible to assess its performance for distortion beyond image compression. Databases with more various kinds of distortions are needed to improve the proposed metric.

## PART III: CONCLUSION

---

In this part, we proposed three methods to improve the automatic quality evaluation of HDR/WCG images. Especially with chrominance artifacts.

First, we proposed to adapt SDR metric to HDR/WCG contents using perceptually uniform color space. This method is a good alternative to PU-metrics as it can be implemented with the  $ICtCp$  color space that is implemented in MPEG codec. Thus, in this case, adapting SDR metrics does not necessarily imply any color space transform. However, if the method is efficient for color-blind metric, none of the color metrics are able to evaluate the quality in a satisfactory way.

Second, we tried to extend the HDR-VDP-2 model towards the chrominance but we did not succeed to improve HDR-VDP-2 performances. This may be due to the irrelevance of the created color features. The complexity of such metric was only growing and we chose another approach to create an HDR/WCG color metric.

Finally, we proposed to aggregate several metrics and visual features to create a new metric that outperforms all other on the tested databases.

This last approach is complementary with the first approach, SDR metrics adaptation. Indeed, if the aggregation of metrics offers better performances than other metrics, it is quite complex. So it is mostly practicable to compare algorithms, for example, to drive the selection of image compression technologies. However, some applications need to obtain very fast quality estimation with very low latency. For example, a video server can evaluate the video quality of video streams to allocate and control the streaming resources. In this case,  $MS\text{-}SSIM_{ICtCp}$  and  $FSIM_{ICtCp}$  might be a better option especially coupled with  $\overline{\Delta E^S_{ICtCp}}$  to check the chrominance preservation.



# CONCLUSION

---

## Summary

In this thesis, we addressed the question of the quality assessment in the context of High Dynamic Range (HDR) and Wide Color Gamut images/videos. If the quality of HDR contents within a BT.709 gamut, was a well-studied area, it was not the case for contents that combine HDR and WCG. Indeed, screens able to display such images were only available recently at a reasonable cost. Thus, no metrics were designed to handle such contents, especially, all metrics adapted to HDR did not consider chromatic distortions. However, HDR and WCG become more and more intricate technologies. Indeed, standards development organizations strongly recommend to use both technologies at the same time to help their deployment. Therefore, there is a need to obtain adapted quality metrics.

## Experimental Data

To tackle this problematic, we created two databases of images using HDR/WCG annotated with MOS. Those databases are essential to study the performances of objective quality metrics. We created two databases, one, HDdtb, is composed of images with a BT.709 gamut encapsulated in a BT.2020 gamut. The second, 4Kdtb, is composed of images with a native BT.2020 gamut. We selected the images to obtain a large variety of characteristics, different DR, keys, colorfulness and complexity. Thus, we obtain databases that correspond to the diversity of contents allowed by the HDR/WCG formats. Moreover, we choose to create realistic artifacts that can occur during transmission of images. We focus on chromatic artifacts and how chrominance information is handled during the compression of images. Using those databases, we show that those chromatic artifacts can affect the perceived quality of images. This was especially true for the database 4Kdtb.

In chapter 4, we used three HDR databases with a standard color gamut and the two aforementioned databases to analyze in detail the state-of-the-art metrics. In particular, we studied in details the calibration of HDR-VDP-2. this metric is recognized as the standard metric to perform HDR objective IQA. We conclude that at least two parameters of the metric parameter, the display spectral emission and the surrounding luminance, have a very small impact on the metric. With the best calibration for HDR-VDP-2, we did a benchmark using the most renowned HDR metrics. We conclude that HDR-VDP-2 is always in the first or second best

---

metric. Performances of other metrics differ more from one database to another. However, for several databases, metrics with a much lower complexity than HDR-VDP-2 have similar or even outperform HDR-VDP-2 performances. As the results, it is quite complex to conclude that one metric systematically outperforms all other metrics. In addition, we have shown that the 4Kdtb can be used to illustrate the sensitivity (or the lack of sensitivity) to chromatic artifacts of quality metrics.

## Color HDR/WCG image quality assessment

The new databases allows us to study three solutions to create objective quality metrics sensitive to chromatic distortions. In addition, we wanted our metrics to retain good performances on HDR contents using standard color gamut.

The first solution, presented in chapter 5, is to adapt SDR metrics. Indeed, most SDR metrics were designed with the assumption that they are encoded using perceptually uniform color space. However, those color spaces, like  $Y'CrCb$ , are not adapted to HDR contents. Aydin et al [10] proposed to adapt the luminance using a function called PU to solve the problem. However, this approach only considers luminance distortion only and not chromatic artifacts. In this chapter, we proposed a method to adapt SDR metrics using perceptually uniform color space adapted to HDR/WCG. We have tested this method on 12 metrics including 6 metrics that include chromatic information. We have tested 4 different color spaces adapted to HDR/WCG:  $ICtCp$ ,  $J_z a_z b_z$  and two versions of  $HDR-Lab$  which differ on the diffuse white ( $100 \text{ cd/m}^2$   $1000 \text{ cd/m}^2$ ). Among all the metrics, we did not found one that was able to assess chromatic distortions while retaining the performance of the best color-blind metrics. However, we found that using these color spaces was a good complement to the PU function proposed by Aydin et al [10]. For example, using the  $ICtCp$  color space, instead of the PU function, does not involve to perform a change of color space, thus making metrics less complex. In addition, we have also shown that the grading of images might have an impact on quality perception. Especially, the image diffuse white has an impact on which color space is more suitable to perform quality assessment.

In chapter 6, we proposed another method: extending an existing metric HDR-VDP-2. This metric is often considered the reference metric concerning HDR contents. To extend this metric, we took advantage of its design. As it models precisely the perception of L, M and S cons, we can infer, from this information, two maps of chrominance distortions. However, if our training gave the same performances than the original metric, we failed to improve its performances. Worse, the trained metric was not impacted by the chromatic distortion. As we did not succeed to improve HDR-VDP-2 using HVS models, we investigated another method: metric fusion.

In chapter 7, we present a new metric which is the combination of two metrics, HDR-VDP-2 and  $FSIM_{J_z a_z b_z}$ , two color features,  $MCS_5 a_z$  and  $MCS_5 b_z$  and finally, one estimation of the

---

image complexity  $SI_{J_z}$ . We succeed to obtain a metric that has superior performances than the other studied metrics and is sensitive to chromatic distortions. This last metric is quite complex and therefore is not usable in all circumstances. The first method that adapts SDR metric can be used in those cases. Both methods are complementary.

## Future work and perspectives

### Improving the metrics

Although we provide efficient metrics to assess the quality of images, there are several aspects that can be improved.

First, we only consider still images to create our metric. Thus, we do not take into account the specific aspect of videos. Indeed, the HVS has many temporal effects which impact the perception of moving images [112]. For example, the contrast sensitivity (in luminance) of spatially low-frequency noise is significantly higher if the noise is temporally varying [113]. The temporal effect does not affect only luminance but also chromatic noise. In [114], the author shows that the sensitivity for temporal chromatic noise patterns is higher for low frequencies in comparison with static noise. Therefore, if some chromatic artifacts are benign for still images, they could become more problematic for videos.

A second strategy to improve metric is to create a bigger database. Indeed, there is no database with HDR/WCG images with as numerous images as SDR databases. For example, CSIQ [86] count 866 images and TID2013 [115] 3000. There are more images in those two databases than in the combination of all the HDR/WCG databases at our disposal. Creating a bigger, better representative database could improve the quality of the training and the universality of the resulting metrics.

The metrics performances are not the only characteristics that can be improved. Reducing the complexity of quality metrics can facilitate their use, especially the metric presented in 7. Indeed, HDR-VDP-2, one of the components of the metric is very complex and, therefore, is not usable in all circumstances. Moreover, the experience in chapter 6 shows that with an appropriate training, we could obtain similar performances than HDR-VDP-2 using only two subbands. Simplifying HDR-VDP-2 seems doable. It is not the only way of simplifying our proposed metric. In [111], authors create another combination metric HDR-CQM with similar performances than HDR-VDP-2. It is possible to replace HDR-VDP-2 by other metrics using the same methodology and adding our color features.

---

## HDR/WCG and aesthetic

HDR/WCG color space, due to their sizes, allows a lot of freedom to grade the images. The color grading of images is the process of improving the appearance of an image for specific devices. It is also often viewed as a way to generate artistic color effects to establish an aesthetic "look". However, some processes can affect this look. One example is the distortion gamut mismatch of our proposed database HDdtb. In this example, the color of the images looked more dulled or more saturated.

However, the participant of the subjective tests did not associate this distortion with a loss of quality even if the artifacts were clearly visible. It was the aesthetic of the images that has changed but viewers were focused on the compression artefacts to give a score. Most advanced metrics are not very sensitive to such gamut mismatch and really mimics the results of subjective tests.

Gamut mismatch is not the only sources of distortions that modifies the look of images. Backward-compatible compression can also create contrast change and hue shift distortions. For example, the Narwaria et al. database contains such distortions (cf. subsection 5.3.5) even if in this case, it is not clear if those distortions significantly impact the look of images. Aside from the distortions created during the transmission of images, there is also the impact of the displays that might change the aesthetic looks. Indeed, HDR displays can be very different from one another with different peak brightness. This can lead to compatibility problem that if not handled correctly might result in loss of the creative intent [116].

Those distortions are not desirable. However, they can be minimized during subjective tests. Indeed, the images/videos are presented without context and the creative intent behind the images are not known by the viewers. In such a case, it could be interesting to ask viewers to give two scores: one for the perceived quality and one for the fidelity to the source, to have a more complete estimation of the quality of an end-to-end transmission chain.

# **Appendices**

# SDR METRICS DESCRIPTION

---

In this appendix, we present in details several SDR metrics selected in section 2.4.

## A.1 PSNR

The peak Signal-to-Noise Ratio (PSNR) is a simple metric that evaluates the distortion between a reference and a distorted image.

$$\text{PSNR} = 10 \log_{10} \left( \frac{\text{DR}^2}{\text{MSE}(X, Y)} \right) \quad (\text{A.1})$$

where DR is traditionally the range of the possible luminance values (255 for an 8 bits representation). The  $\text{MSE}(X, Y)$  corresponds to the mean square error between the reference and the distorted images:

$$\text{MSE} = \frac{1}{IJ} \sum_i \sum_j [X(i, j) - Y(i, j)]^2 \quad (\text{A.2})$$

where I and J are the image resolution and i and j are the pixels spatial coordinates of the X and Y images.

## A.2 S-CIELab

The S-CIELab [71] metric is the spatial extension of the CIE-*Lab* color space. Indeed, the color difference metric  $\Delta E_{ab}$  is not correlated with the perception of differences in natural images but in large uniform patches. The goal of this extension is to take into account the blurring effect of the HVS. To achieve that, the images in the  $XYZ$  color space are transformed into an opponent color space as follows:

$$\begin{bmatrix} O_1 \\ O_2 \\ O_3 \end{bmatrix} = \begin{bmatrix} 0.279 & 0.72 & -0.107 \\ -0.449 & 0.29 & -0.077 \\ 0.086 & -0.59 & 0.501 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (\text{A.3})$$

---

Each component  $O$  is then filtered, using filters that approximate the Contrast Sensitivity Function (CSF) of the HVS. This is accomplished using filters with two-dimensional separable convolution kernels  $f$  of the form:

$$f(x, y) = k \sum_i w_i E_i(x, y) \quad (\text{A.4})$$

where

$$E_i = k_i \exp[-(x^2 + y^2)/\sigma_i^2] \quad (\text{A.5})$$

The parameters for  $w_i$  and  $\sigma_i$  are different for each component  $O_1, O_2, O_3$ , the scale factor  $k_i$  is chosen so that each  $E_i$  sums to 1. The scale factor  $k$  is chosen so that for each color plane, the kernel  $f$  sums to 1. Then, the filtered images are converted back to the  $XYZ$  color space from which the uniform color space Lab can be calculated. Once this transformation is done on the reference image and on the distorted image, a color difference map using the color difference metric,  $\Delta E$ , can be calculated. The new distortion map is called  $\Delta E^S$ . To obtain a unique value estimating the overall perceptual difference between two images, the average of this map is computed:

$$\overline{\Delta E^S} = \frac{1}{IJ} \sum_i^I \sum_j^J \Delta E^S(i, j) \quad (\text{A.6})$$

The S-CIELab metric was originally designed to measure color reproduction errors of printed digital images. It is also often used as an image quality metric because of its simple implementation.

## A.3 SSIM

### A.3.1 Color-blind SSIM:

The structural similarity index (SSIM) [72] is based on the assumption that the HVS is highly adapted to extract the structure of natural scene. SSIM is composed of the comparison of three characteristics calculated from the luminance component of two images  $X$  and  $Y$ , one being the reference image and the other one the distorted image.

- First, the luminance of each signal is compared using the  $l(X, Y)$  function:

$$l(X, Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \quad (\text{A.7})$$

- 
- Then, the  $c(X, Y)$  function which represents the contrast comparison is calculated :

$$c(X, Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \quad (\text{A.8})$$

- Finally, the structure of the images is compared using the function  $s(X, Y)$ :

$$s(X, Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3} \quad (\text{A.9})$$

where  $\mu_X$  and  $\mu_Y$  are the means of the images  $X$  and  $Y$ ,  $\sigma_X$  and  $\sigma_Y$  are their standard deviations and  $\sigma_{XY}$  is the covariance between them.  $C_1$ ,  $C_2$  and  $C_3$  are used to prevent the division by zero. These three components are combined as follows:

$$\text{SSIM}(X, Y) = [l(X, Y)^\alpha \times c(X, Y)^\beta \times s(X, Y)^\gamma] \quad (\text{A.10})$$

Usually the weights  $\alpha$ ,  $\beta$  and  $\gamma$  are set to 1. The SSIM is often calculated locally and not globally (like in previous equation) allowing to have more detailed information about the distortion. The SSIM is then calculated for each pixel through a sliding window. In [72], Wang et al. used a sliding window in which values are weighted using an  $11 \times 11$  circular-symmetric Gaussian weighting function  $W = \{w_n | n = 1, 2, \dots, N\}$ . The statistics are then calculated for each pixel  $x$  of  $X$  and  $y$  of  $Y$ :

$$\mu_x = \sum_{n=1}^N w_n x_n \quad (\text{A.11})$$

$$\sigma_x = \left( \sum_{n=1}^N w_n (x_n - \mu_x)^2 \right)^{\frac{1}{2}} \quad (\text{A.12})$$

$$\sigma_{xy} = \sum_{n=1}^N w_n (x_n - \mu_x)(y_n - \mu_y) \quad (\text{A.13})$$

A SSIM index is calculated for each pixel and its associated window using the equation (A.10). To obtain an overall quality measure of the image, the mean of the SSIM indexes is calculated:

$$\text{MSSIM}(X, Y) = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \text{SSIM}(x_{ij}, y_{ij}) \quad (\text{A.14})$$

where  $I$  and  $J$  represent the image resolution.

We keep the name SSIM instead of MSSIM to avoid confusion with the multiscale SSIM (MS-SSIM) (see section A.4).

### A.3.2 SSIM for color images (SSIMc):

The SSIM was only defined for luminance information. In [73], Wang et al. extended the SSIM to video as well as to chrominance information. They used SSIM for each component of the video sequence encoded in the Y'CrCb [28] color space. The SSIM scores for the three components are then aggregated using a weighted arithmetic mean:

$$\text{SSIMc}(i, j) = 0.8 \text{SSIM}_{Y'}(i, j) + 0.1 \text{SSIM}_{Cr}(i, j) + 0.1 \text{SSIM}_{Cb}(i, j) \quad (\text{A.15})$$

where  $i$  and  $j$  are the pixel coordinates of the SSIM map. As for the color-blind SSIM metric, an overall quality score is obtained by averaging all SSIMc scores.

### A.3.3 CSSIM

Another solution to include chrominance information in SSIM was proposed in [74]. Authors proposed to add a new comparison that use a S-CIELab  $\Delta E_{ab}^S$  distortion map (see section A.2) and is called the color comparison. As for precedent comparison, the color comparison is calculated for each pixel  $x$  of the reference image  $X$  and each pixel  $y$  of the distorted image  $Y$ :

$$c_r(x, y) = 1 - \frac{1}{k} \times \Delta E_{ab}^S(x, y) \quad (\text{A.16})$$

$k$  is a constant equal to 45. The SSIM index (cf. Equation (A.10)) is then adapted as follows:

$$\text{CSSIM} = \frac{1}{IJ} \sum_i^I \sum_j^J l(x_{ij}, y_{ij}) \times c(x_{ij}, y_{ij}) \times s(x_{ij}, y_{ij}) \times c_r(x_{ij}, y_{ij}) \quad (\text{A.17})$$

## A.4 Multiscale-SSIM (MS-SSIM)

Wang et al. [75] improved the SSIM index by using a multiscale approach (MS-SSIM). They proposed to apply the comparison functions used in SSIM at different scales of the images. The goal is to incorporate image details at different resolutions. The different scales are obtained after a low-pass filtering and a downsampling of the images. The index of the original scale is 1 and the index of the highest scale is  $M$ . The luminance comparison (equation (A.14)) is only done at the highest scale. The multiscale SSIM evaluation is obtained using the following combination:

$$\text{MS-SSIM} = [L_M(X, Y)]^{\alpha_M} \cdot \prod_{m=1}^M [C_m(X, Y)]^{\beta_m} [S_m(X, Y)]^{\gamma_m} \quad (\text{A.18})$$

Where  $X$  and  $Y$  are the reference and the distorted images,  $L$ ,  $C$  and  $S$  are the mean of the

pixelwise luminance, contrast and structure comparison functions  $l$ ,  $c$  and  $s$  defined in section A.3 for the different image scales  $m$ .  $\alpha$ ,  $\beta$  and  $\gamma$  are the parameters of the metric.

In [75], the authors proposed a 5 scales MS-SSIM where the parameters are :  $\beta_1 = \gamma_1 = 0.0448$ ,  $\beta_2 = \gamma_2 = 0.2856$ ,  $\beta_3 = \gamma_3 = 0.3001$ ,  $\beta_4 = \gamma_4 = 0.2363$  and  $\alpha_5 = \beta_5 = \gamma_5 = 0.1333$ .

## A.5 FSIM

### A.5.1 Color-blind FSIM

The feature similarity (FSIM) index [65] allows to overcome a limit of SSIM and MS-SSIM: in these algorithms, a simple average is used to pool the SSIM local distortion map (equation (A.14)), each pixel having then the same importance. The authors of FSIM made the assumption that different locations can have very different contributions to the quality perception of an image.

FSIM extracts from the image two different features: the phase congruency and the gradient magnitude.

The phase congruency model postulates that details of an image become visible where the Fourier components are maximal in phase. The phase congruency can be considered as a measure of the significance of local structure. To compute this feature, the authors used the method developed by Kovessi [117]. The comparison of the phase congruency planes of the reference image ( $PC_X(i, j)$ ) and of the distorted image ( $PC_Y(i, j)$ ),  $i$  and  $j$  being the pixels spatial coordinates, is done as follows:

$$S_{PC}(i, j) = \frac{2PC_X(i, j) \times PC_Y(i, j) + T_1}{PC_X^2(i, j) + PC_Y^2(i, j) + T_1} \quad (\text{A.19})$$

The feature called gradient magnitude is calculated to take into account contrast information (the phase congruency is contrast invariant). The comparison of the gradient magnitude planes of the reference image ( $G_X(i, j)$ ) and the distorted image ( $G_Y(i, j)$ ) is done as follows:

$$S_G(i, j) = \frac{2G_X(i, j) \times G_Y(i, j) + T_2}{G_X^2(i, j) + G_Y^2(i, j) + T_2} \quad (\text{A.20})$$

Finally, the FSIM is calculated as follows:

$$S_L(i, j) = [S_{PC}(i, j)] \times [S_G(i, j)] \quad (\text{A.21})$$

$$PC_m(i, j) = \max(PC_X(i, j), PC_Y(i, j)) \quad (\text{A.22})$$

---


$$\text{FSIM} = \frac{\sum_i \sum_j S_L(i, j) PC_m(i, j)}{\sum_i \sum_j PC_m(i, j)} \quad (\text{A.23})$$

The parameters values of FSIM are  $T_1=0.85$  and  $T_2=160$

### A.5.2 FSIM for color images (FSIMc):

To add chrominance consideration in FSIM, the authors of the metric proposed the FSIMc index [65]. Images are first converted from the *RGB* (BT.709) color space to the *YIQ* color space [118] :

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & 0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (\text{A.24})$$

and then a comparison between the chrominance components ( $I_X$  and  $I_Y$  and respectively  $Q_X$  and  $Q_Y$ ) performed for every pixel ( $i, j$ ):

$$S_I(i, j) = \frac{2I_X(i, j) \times I_Y(i, j) + T_3}{I_X^2(i, j) + I_Y^2(i, j) + T_3} \quad (\text{A.25})$$

$$S_Q(i, j) = \frac{2Q_X(i, j) \times Q_Y(i, j) + T_4}{Q_X^2(i, j) + Q_Y^2(i, j) + T_4} \quad (\text{A.26})$$

The authors incorporate these comparisons straightforwardly into the FSIM index:

$$\text{FSIMc} = \frac{\sum_i \sum_j S_L(i, j) PC_m(i, j) \times [S_I(i, j) \times S_Q(i, j)]^\lambda}{\sum_i \sum_j PC_m(i, j)} \quad (\text{A.27})$$

The values of FSIMc parameters are  $T_3=200$ ,  $T_4=200$  and  $\lambda=0.03$ .

## A.6 PSNR-HVS and PSNR-HVS-M

PSNR-HVS [119] is a modified version of the PSNR. It takes into account the contrast sensitivity function (CSF) which measures the human sensibility to spatial frequencies.

The PSNR-HVS is calculated as follows:

$$\text{PSNR-HVS} = 10 \log_{10} \left( \frac{255^2}{\text{MSE}_H} \right) \quad (\text{A.28})$$

The  $MSE_H$  is calculated as follows:

$$MSE_H = K \sum_{i=1}^{I-7} \sum_{j=1}^{J-7} \sum_{m=1}^8 \sum_{n=1}^8 ((X[m, n]_{ij} - Y[m, n]_{ij}) T_c[m, n])^2, \quad (A.29)$$

$I$  and  $J$  are the images size.  $X_{ij}$  and  $Y_{ij}$  are the  $8 \times 8$  DCT coefficients of the reference and distorted images for which the coordinates of the left upper corner are  $i$  and  $j$ .  $T_c$  is the matrix of correcting factors that takes into account the HVS spatial sensitivity.  $m$  and  $n$  are the coordinates of the DCT block coefficients. Finally,  $K$  is equal to  $1/[(I-7)(J-7)64]$ .

In [77], Ponomarenko et al. improved the performances of PSNR-HVS by adding a model of the contrast masking: the fact that any DCT coefficient  $X_{ij}$  of a block can mask any other block coefficients except the DC coefficient (the mean luminance). This metric is called PSNR-HVS-M.

To estimate the masking effect, a weighted energy  $E_w$  is calculated for each DCT block  $X$ :

$$E_w(X) = \sum_{m=1}^8 \sum_{n=1}^8 X[m, n]^2 C[m, n], \quad (A.30)$$

where  $C[m, n]$  is a matrix of correcting factors that models the contrast masking. The value of the masking effect can be too high if an image block belongs to an edge. To overcome this effect, a correction is then applied on  $E_w$ .

$$E_m(D) = E_w(D) \delta(D) / 16 \quad (A.31)$$

The correcting factor  $\delta(D)$  is calculated using the local variance  $V$  of the four  $4 \times 4$  sub-block  $D1, D2, D3$  and  $D4$  of the  $8 \times 8$  image DCT block  $D$ .

$$\delta(D) = (V(D1) + V(D2) + V(D3) + V(D4)) / 4V(D) \quad (A.32)$$

The masking model is then applied on the distorted ( $Y[m, n]$ ) and reference ( $X[m, n]$ ) images DCT blocks to obtain visible distortion blocks  $\Delta XY$  as follows:

$$\Delta XY[m, n] = \begin{cases} X[m, n] - Y[m, n] & \text{if } m = 0, n = 0 \\ 0 & \text{if } |X[m, n] - Y[m, n]| \leq E_{norm} / C[m, n] \\ X[m, n] - Y[m, n] - E_{norm} / C[m, n] & \text{if } X[m, n] - Y[m, n] > E_{norm} / C[m, n] \\ X[m, n] - Y[m, n] - E_{norm} / C[m, n] & \text{otherwise} \end{cases} \quad (A.33)$$

where

$$E_{norm} = \sqrt{\max(E_m(X), E_m(Y)) / 64} \quad (A.34)$$

---

The new MSE is then calculated as in equation (A.29):

$$\text{MSE}_{HVS\text{M}} = K \sum_{i=1}^{I-7} \sum_{j=1}^{J-7} \sum_{m=1}^8 \sum_{n=1}^8 ((\Delta XY_{ij}[m, n])T_c[m, n])^2 \quad (\text{A.35})$$

The PSNR-HVS-M is then calculated as in equation (A.28):

$$\text{PSNR-HVS-M} = 10 \log_{10} \left( \frac{255^2}{\text{MSE}_{HVS\text{M}}} \right) \quad (\text{A.36})$$

## A.7 PSNR-HMA

### A.7.1 Color-blind PSNR-HMA

PSNR-HMA [78] is an improvement of PSNR-HVS-M. PSNR-HVS-M has poor performances when used on images that contain contrast change and mean shift. PSNR-HMA was created to take into account the particularities of the HVS concerning these two distortions.

For a given reference image  $X$  and a given distorted image  $Y$  and their means  $\bar{X}$  and  $\bar{Y}$ , a new image  $C$  is calculated :

$$C = Y + Delt \quad (\text{A.37})$$

$$Delt = \bar{X} - \bar{Y} \quad (\text{A.38})$$

The correcting factor  $Popr$  is computed to assess possible contrast change :

$$Popr = \frac{\sum (X - \bar{X})(C - \bar{C})}{\sum (C - \bar{C})} \quad (\text{A.39})$$

The new image  $D$  is calculated:

$$D = (C - \bar{C}).Popr + \bar{C} \quad (\text{A.40})$$

This is the image that minimizes the Mean square error with A. Two  $\text{MSE}_{HVS\text{M}}$  are calculated (Equation (A.35)) :

$$M_1 = \text{MSE}_{HVS\text{M}}(X, C) \quad (\text{A.41})$$

$$M_2 = \text{MSE}_{HVS\text{M}}(X, D) \quad (\text{A.42})$$

---

If  $M_1 > M_2$ , the contrast change needs to be dealt with as follows:

$$M_1 = M_2 + \begin{cases} (M_1 - M_2)C_1 & \text{if } P_{opr} < 1 \\ (M_1 - M_2)C_2 & \text{if } P_{opr} \geq 1 \end{cases} \quad (\text{A.43})$$

Then the final  $\text{MSE}_{HMA}$  is calculated as follows:

$$\text{MSE}_{HMA} = M_1 + \text{Delt}^2.C_3 \quad (\text{A.44})$$

The values of the metric constants are:  $C_1=0.002$ ,  $C_2=0.25$  and  $C_3= 0.04$

### **A.7.2 PSNR-HMA for color images (PSNR-HMAc):**

To include the chrominance components into the metric, Ponomarenko et al. [78] proposed to aggregate the  $\text{MSE}_{HMA}$  for each channel using the  $Y'CrCb$  color space [28] as follows:

$$\text{MSE}_{HMA} = (\text{MSE}_{HMA}(Y) + C_4 \times \text{MSE}_{HMA}(Cr) + C_4 \times \text{MSE}_{HMA}(Cb)) / (1 + 2 \times C_4) \quad (\text{A.45})$$

The values of the metric new constant is  $C_4=0.5$ .

# DATABASE IMAGES

The following images represent tone map version of all the used reference images. The tone mapping operator (TMO) used to produce this images was the Reinhard et al. TMO [93]. We used its Matlab implementation of the HDRToolbox [120]. The reason why the same content present in several databases (like FireEater) can have a different rendering is because the Reinhard TMO was applied indifferently on BT.709 content and BT.2020 content.

## B.1 Narwaria et al.

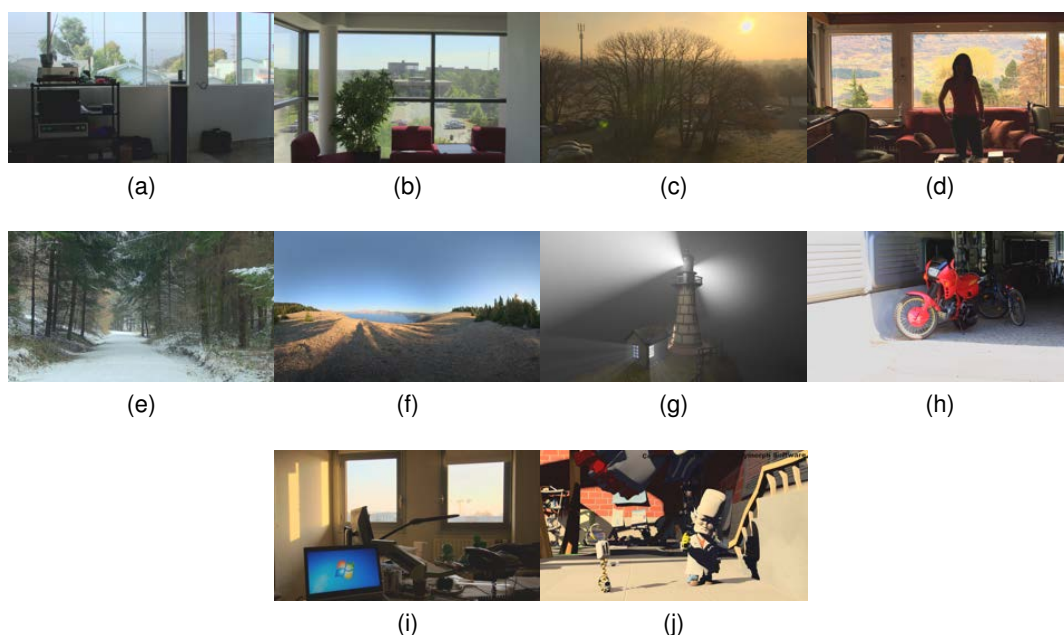


Figure B.1: Tone-mapped version (Reinhard et al. TMO [93]) of Narwaria et al. reference images. From left to right and from top to bottom: **(a)** Apartment\_float\_o15C **(b)** bausch\_lot **(c)** carpark\_ivc **(d)** CD1\_serie2 **(e)** forest\_path **(f)** lake **(g)** LightHouse072 **(h)** moto **(i)** office\_ivc **(j)** outro022168.

---

## B.2 Zerman et al.

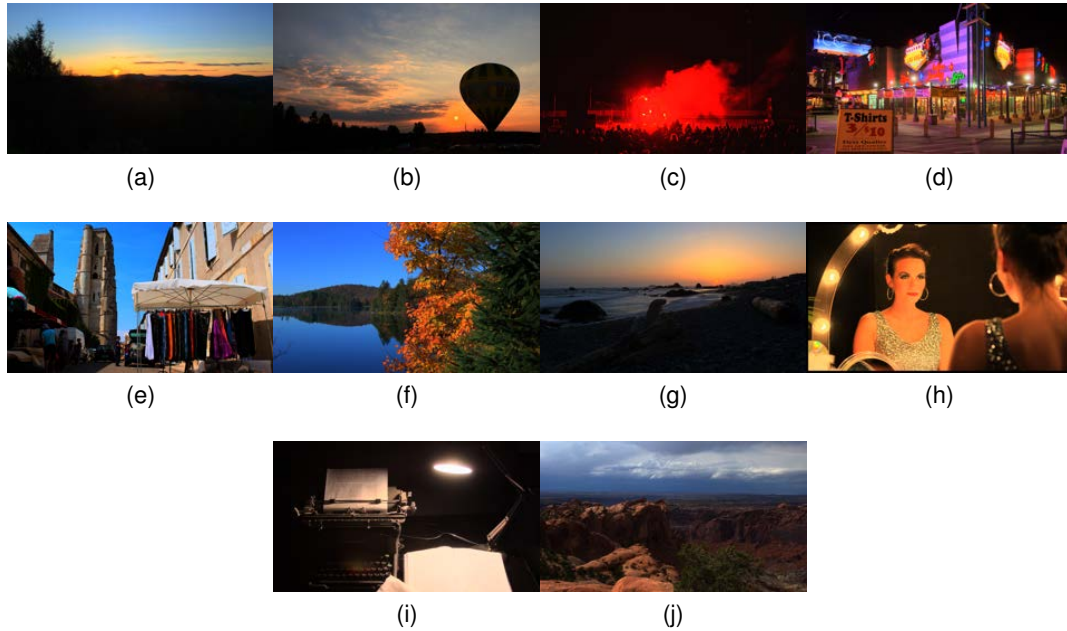


Figure B.2: Tone-mapped version (Reinhard et al. TMO [93]) of Zerman et al. reference images. From left to right and from top to bottom: **(a)** AirBellowsGap **(b)** Balloon **(c)** FireEater **(d)** LasVegasStore **(e)** Market3 **(f)** MasonLake(1) **(g)** RedwoodSunset **(h)** Showgirl **(i)** Typewriter **(j)** UpheavalDome.

### B.3 Korshunov et al.

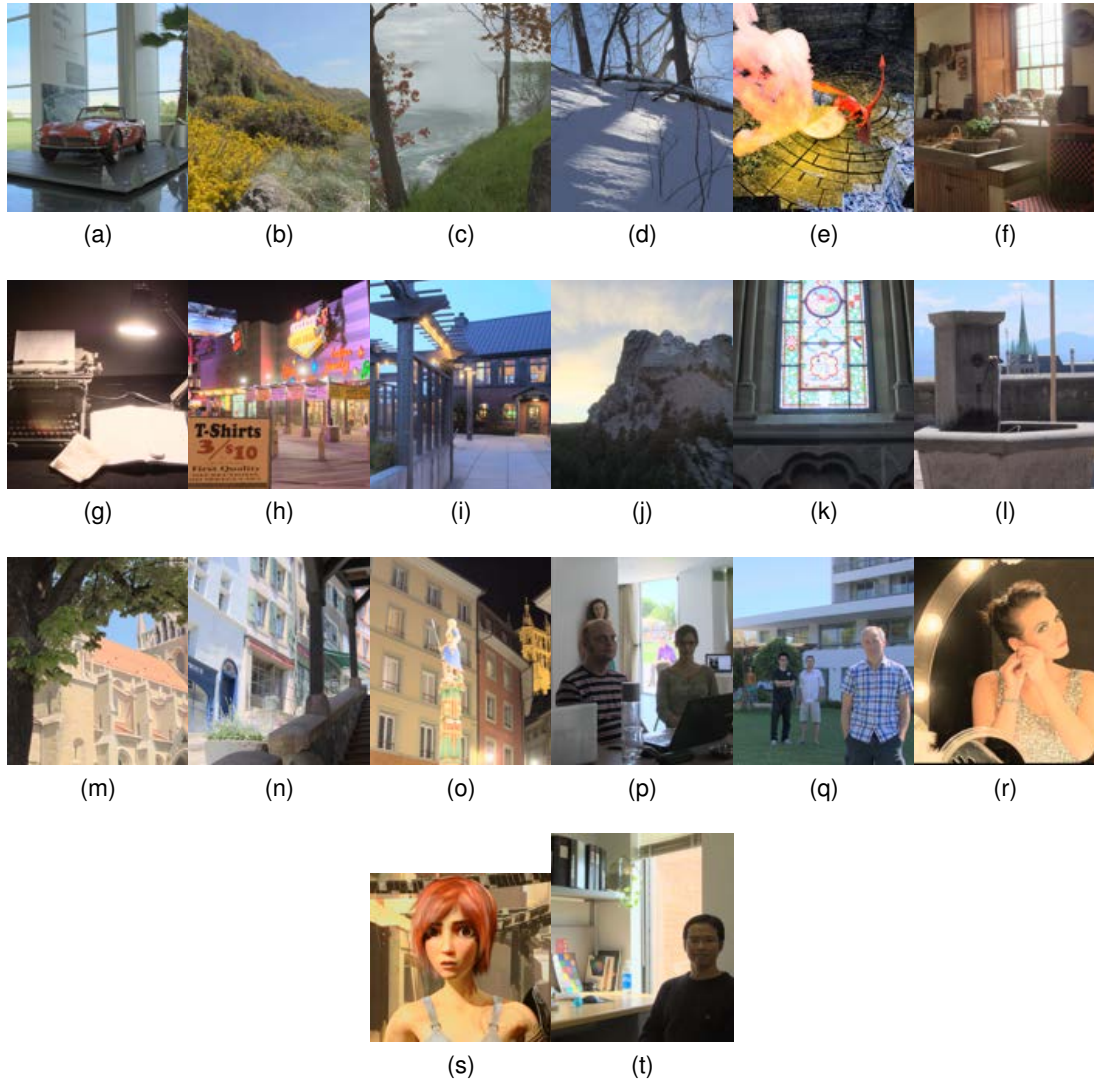


Figure B.3: Tone-mapped version (Reinhard et al. TMO [93]) of Korshunov et al. reference images. From left to right and from top to bottom: **(a)** 507 **(b)** BloomingGorse2 **(c)** CanadianFalls **(d)** DevilsBathtub **(e)** dragon\_3 **(f)** HancockKitchenInside **(g)** LabTypewriter **(h)** LasVegasStore **(i)** McKeesPub **(j)** MtRushmore2 **(k)** set18 **(l)** set22 **(m)** set23 **(n)** set24 **(o)** set31 **(p)** set33 **(q)** set70 **(r)** showgirl **(s)** sintel\_2 **(t)** WillyDesk.

## B.4 HDdtb

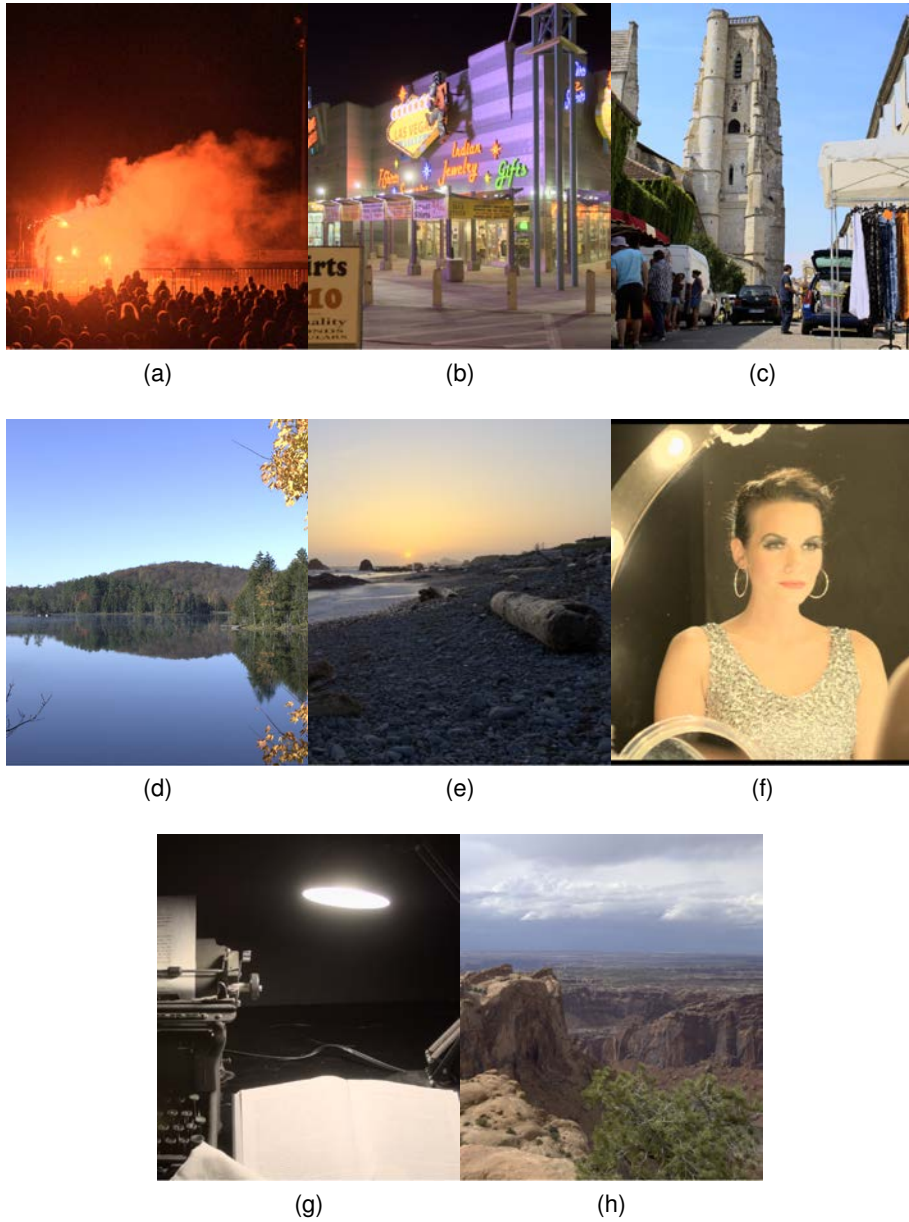


Figure B.4: Tone-mapped version (Reinhard et al. TMO [93]) of the HDdtb reference images. From left to right and from top to bottom: **(a)** FireEater **(b)** LasVegasStore **(c)** Market3 **(d)** MasonLake(1) **(e)** RedwoodSunset **(f)** Showgirl **(g)** Typewriter **(h)** UpheavalDome.

## B.5 4Kdtb



Figure B.5: Tone-mapped version (Reinhard et al. TMO [93]) of the 4Kdtb reference images. From left to right and from top to bottom: **(a)** Bike\_ 20s **(b)** Bike\_ 30s **(c)** Bike\_ 81s **(d)** Bike\_ 110s **(e)** Regatta\_ 11s **(f)** Regatta\_ 24s **(g)** Regatta\_ 80s **(h)** Regatta\_ 95s.

# PERFORMANCE INDEXES

---

In this appendix, we present the numerical value for the performance indexes of each quality metrics with each database. The metrics with the best performances in terms of SROCC and PCC is in red, the second is in blue and the third in green.

## C.1 PCC

Table C.1: PCC of the different color-blind quality metrics on the considered databases

Quality Metric	4Kdtdb	HDdtdb	Narwaria et al.	Korshunov et al.	Zerman et al.
PU-PSNR	0.6964	0.8002	0.5831	0.8597	0.8188
PSNR <sub>HDR-Lab100</sub>	0.6724	0.7950	0.5562	0.8602	0.8042
PSNR <sub>HDR-Lab1000</sub>	0.4711	0.7320	0.5344	0.8004	0.7023
PSNR <sub>ICtCp</sub>	0.7231	0.7948	0.6036	0.8697	0.8546
PSNR <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.5431	0.6899	0.5533	0.7999	0.7002
PSNR <sub>XYZ</sub>	0.1890	0.6416	0.4627	0.7017	0.5612
PSNR <sub>Y'C'r'C'b'</sub>	0.2674	0.6755	0.4996	0.7635	0.6428
PU-SSIM	0.6962	0.8520	0.7567	0.9265	0.8262
SSIM <sub>HDR-Lab100</sub>	0.6966	0.8448	0.7805	0.9243	0.8010
SSIM <sub>HDR-Lab1000</sub>	0.6025	0.7677	0.7247	0.9021	0.6596
SSIM <sub>ICtCp</sub>	0.6838	0.8366	0.7572	0.9296	0.8522
SSIM <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.6580	0.7851	0.7990	0.9174	0.6923
SSIM <sub>XYZ</sub>	0.2755	0.6174	0.6167	0.7786	0.4863
SSIM <sub>Y'C'r'C'b'</sub>	0.3435	0.6729	0.6718	0.8423	0.5343
PU-MS-SSIM	0.8479	0.8881	0.8756	0.9631	0.9324
MS-SSIM <sub>HDR-Lab100</sub>	0.8451	0.8899	0.8448	0.9606	0.9253
MS-SSIM <sub>HDR-Lab1000</sub>	0.7792	0.8395	0.8680	0.9068	0.7633
MS-SSIM <sub>ICtCp</sub>	0.8382	0.8763	0.8846	0.9575	0.9410
MS-SSIM <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.8337	0.8603	0.9157	0.9694	0.8013
MS-SSIM <sub>XYZ</sub>	0.4377	0.6422	0.6316	0.8619	0.6258
MS-SSIM <sub>Y'C'r'C'b'</sub>	0.5409	0.7062	0.7091	0.9126	0.6443
PU-FSIM	0.9000	0.8443	0.8773	0.9568	0.8988
FSIM <sub>HDR-Lab100</sub>	0.8950	0.8476	0.8726	0.9540	0.9120
FSIM <sub>HDR-Lab1000</sub>	0.8416	0.8923	0.8195	0.9733	0.9133
FSIM <sub>ICtCp</sub>	0.8992	0.8234	0.8654	0.9471	0.8883
FSIM <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.8829	0.9187	0.8466	0.9724	0.9059
FSIM <sub>XYZ</sub>	0.5817	0.7372	0.6546	0.9015	0.7402
FSIM <sub>Y'C'r'C'b'</sub>	0.7054	0.8066	0.7445	0.9215	0.8148
PU-PSNR-HVS-M	0.8169	0.7963	0.6090	0.9210	0.9120
PSNR-HVS-M <sub>HDR-Lab100</sub>	0.8200	0.8023	0.5942	0.9218	0.9110
PSNR-HVS-M <sub>HDR-Lab1000</sub>	0.6674	0.7088	0.5874	0.9252	0.8588
PSNR-HVS-M <sub>ICtCp</sub>	0.8187	0.7762	0.6269	0.9297	0.9226
PSNR-HVS-M <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.7316	0.6624	0.5807	0.9120	0.8310
PSNR-HVS-M <sub>XYZ</sub>	0.3016	0.6320	0.4517	0.8102	0.6716
PSNR-HVS-M <sub>Y'C'r'C'b'</sub>	0.4214	0.6540	0.5009	0.8972	0.7915
PU-PSNR-HMA	0.8244	0.7873	0.7625	0.9364	0.8893
PSNR-HMA <sub>HDR-Lab100</sub>	0.8129	0.7693	0.7647	0.9360	0.8850
PSNR-HMA <sub>HDR-Lab1000</sub>	0.6676	0.7545	0.7221	0.9368	0.8516
PSNR-HMA <sub>ICtCp</sub>	0.8410	0.8140	0.7904	0.9310	0.9230
PSNR-HMA <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.7298	0.7592	0.7427	0.9261	0.8322
PSNR-HMA <sub>XYZ</sub>	0.3022	0.6750	0.5069	0.8574	0.6794
PSNR-HMA <sub>Y'C'r'C'b'</sub>	0.4216	0.7203	0.6406	0.8942	0.7933
HDR-VDP2	0.8605	0.8715	0.9130	0.9518	0.9385
HDR-VQM	0.7714	0.8676	0.9061	0.9612	0.9304
PU-VIF	0.8722	0.7645	0.7571	0.9215	0.8919

Table C.2: PCC of the different color quality metrics on the considered databases.

Quality Metric	4Kdtdb	HDdtdb	Narwaria et al.	Korshunov et al.	Zerman et al.
$\overline{\Delta E}_{HDR-Lab_{100}}$	0.4582	0.2502	0.6407	0.7629	0.6012
$\overline{\Delta E}_{HDR-Lab_{1000}}$	0.2280	0.2559	0.6106	0.7024	0.5133
$\overline{\Delta E}_{ICtCp}$	0.6783	0.2548	0.6277	0.8065	0.7508
$\overline{\Delta E}_{J_z a_z b_z}$	0.4058	0.2952	0.6436	0.5536	0.5392
$\overline{\Delta E}_{XYZ}$	0.2184	0.3438	0.3375	0.5993	0.3564
$\overline{\Delta E}_{Y'Cr'Cb'}$	0.2336	0.3083	0.3220	0.6873	0.4287
$\overline{\Delta E}^S_{HDR-Lab_{100}}$	0.7513	0.1135	0.6686	0.7334	0.7464
$\overline{\Delta E}^S_{HDR-Lab_{1000}}$	0.4662	0.1027	0.5763	0.7470	0.6723
$\overline{\Delta E}^S_{ICtCp}$	0.7885	0.1355	0.5541	0.7639	0.7655
$\overline{\Delta E}^S_{J_z a_z b_z}$	0.6103	0.1331	0.5520	0.7556	0.6980
$\overline{\Delta E}^S_{XYZ}$	0.2748	0.2698	0.2761	0.7152	0.4372
$\overline{\Delta E}^S_{Y'Cr'Cb'}$	0.3093	0.2047	0.2527	0.7250	0.6325
$SSIMc_{HDR-Lab_{100}}$	0.5246	0.7050	0.7485	0.8845	0.7126
$SSIMc_{HDR-Lab_{1000}}$	0.3120	0.6618	0.7886	0.8664	0.5507
$SSIMc_{ICtCp}$	0.7376	0.7764	0.7505	0.9176	0.8275
$SSIMc_{J_z a_z b_z}$	0.5108	0.6842	0.8153	0.8914	0.6253
$SSIMc_{XYZ}$	0.2596	0.6020	0.6292	0.7713	0.4811
$SSIMc_{Y'Cr'Cb'}$	0.2851	0.6187	0.7471	0.8354	0.5185
$CSSIM_{HDR-Lab_{100}}$	0.7991	0.5193	0.7784	0.8929	0.7762
$CSSIM_{HDR-Lab_{1000}}$	0.5440	0.4249	0.6696	0.8828	0.6644
$CSSIM_{ICtCp}$	0.7699	0.5137	0.7354	0.9152	0.8372
$CSSIM_{J_z a_z b_z}$	0.6812	0.4872	0.6180	0.9197	0.7025
$CSSIM_{XYZ}$	0.2174	0.3671	0.4213	0.7590	0.4856
$CSSIM_{Y'Cr'Cb'}$	0.3065	0.4367	0.5317	0.8788	0.5649
$FSIMc_{HDR-Lab_{100}}$	0.8473	0.8598	0.8680	0.9360	0.9131
$FSIMc_{HDR-Lab_{1000}}$	0.6891	0.8654	0.8343	0.9633	0.9055
$FSIMc_{ICtCp}$	0.9080	0.8086	0.8687	0.9453	0.8894
$FSIMc_{J_z a_z b_z}$	0.8355	0.9162	0.8566	0.9717	0.9092
$FSIMc_{XYZ}$	0.5848	0.7337	0.6609	0.9029	0.7415
$FSIMc_{Y'Cr'Cb'}$	0.6783	0.7923	0.7656	0.9541	0.8162
$PSNR-HMAC_{HDR-Lab_{100}}$	0.5602	0.4489	0.6744	0.7533	0.7229
$PSNR-HMAC_{HDR-Lab_{1000}}$	0.3573	0.3986	0.6701	0.7275	0.6457
$PSNR-HMAC_{ICtCp}$	0.7540	0.6467	0.7899	0.8608	0.8195
$PSNR-HMAC_{J_z a_z b_z}$	0.4963	0.4749	0.7319	0.8196	0.7213
$PSNR-HMAC_{XYZ}$	0.2057	0.5711	0.5884	0.8328	0.6327
$PSNR-HMAC_{Y'Cr'Cb'}$	0.3580	0.6153	0.7185	0.8942	0.7432

## C.2 SROCC

Table C.3: SROCC of the different color-blind quality metrics on the considered databases.

Quality Metric	4Kdttb	HDdttb	Narwaria et al.	Korshunov et al.	Zerman et al.
PU-PSNR	0.7261	0.7802	0.5331	0.8597	0.8266
PSNR <sub>HDR-Lab100</sub>	0.6596	0.7751	0.4975	0.8602	0.8147
PSNR <sub>HDR-Lab1000</sub>	0.4673	0.7587	0.4197	0.8078	0.7086
PSNR <sub>ICtCp</sub>	0.7419	0.7745	0.5736	0.8742	0.8508
PSNR <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.5531	0.6933	0.4634	0.8102	0.7001
PSNR <sub>XYZ</sub>	0.2131	0.6311	0.4601	0.7216	0.5682
PSNR <sub>Y'C'r'C'b'</sub>	0.2519	0.6687	0.4157	0.7756	0.6493
PU-SSIM	0.7066	0.8430	0.7240	0.9280	0.8316
SSIM <sub>Z100</sub>	0.6977	0.8355	0.7494	0.9253	0.8090
SSIM <sub>HDR-Lab1000</sub>	0.6054	0.7904	0.7247	0.9085	0.6851
SSIM <sub>ICtCp</sub>	0.6752	0.8245	0.7231	0.9307	0.8618
SSIM <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.6492	0.7813	0.7721	0.9181	0.7073
SSIM <sub>XYZ</sub>	0.1965	0.6488	0.5938	0.7746	0.5065
SSIM <sub>Y'C'r'C'b'</sub>	0.3027	0.6926	0.6376	0.8421	0.5563
PU-MS-SSIM	0.8517	0.8640	0.8656	0.9583	0.9165
MS-SSIM <sub>HDR-Lab100</sub>	0.8448	0.8708	0.8200	0.9567	0.9143
MS-SSIM <sub>HDR-Lab1000</sub>	0.7684	0.8505	0.8528	0.9600	0.7791
MS-SSIM <sub>ICtCp</sub>	0.8447	0.8464	0.8714	0.9529	0.9260
MS-SSIM <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.8306	0.8557	0.9088	0.9648	0.8109
MS-SSIM <sub>XYZ</sub>	0.4334	0.6864	0.6092	0.8646	0.6104
MS-SSIM <sub>Y'C'r'C'b'</sub>	0.5202	0.7296	0.6846	0.9124	0.6499
PU-FSIM	0.9054	0.8149	0.8773	0.9553	0.8912
FSIM <sub>HDR-Lab100</sub>	0.8994	0.8239	0.8726	0.9553	0.9091
FSIM <sub>HDR-Lab1000</sub>	0.8420	0.8799	0.8195	0.9692	0.9087
FSIM <sub>ICtCp</sub>	0.9049	0.8099	0.8654	0.9477	0.8863
FSIM <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.8849	0.9069	0.8466	0.9663	0.9031
FSIM <sub>XYZ</sub>	0.5732	0.7546	0.6316	0.8986	0.7393
FSIM <sub>Y'C'r'C'b'</sub>	0.7052	0.8153	0.7264	0.9415	0.8190
PU-PSNR-HVS-M	0.8401	0.7803	0.5624	0.9331	0.9035
PSNR-HVS-M <sub>HDR-Lab100</sub>	0.8110	0.7856	0.5455	0.9333	0.9028
PSNR-HVS-M <sub>HDR-Lab1000</sub>	0.6607	0.7508	0.4557	0.9311	0.8486
PSNR-HVS-M <sub>ICtCp</sub>	0.8452	0.7554	0.5846	0.9308	0.9066
PSNR-HVS-M <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.7315	0.6501	0.4963	0.9230	0.8286
PSNR-HVS-M <sub>XYZ</sub>	0.2891	0.6314	0.4392	0.8449	0.6793
PSNR-HVS-M <sub>Y'C'r'C'b'</sub>	0.3922	0.6670	0.4157	0.9102	0.7954
PU-PSNR-HMA	0.8403	0.8218	0.7634	0.9369	0.9041
PSNR-HMA <sub>HDR-Lab100</sub>	0.8114	0.8167	0.7458	0.9365	0.9034
PSNR-HMA <sub>HDR-Lab1000</sub>	0.6607	0.7984	0.6907	0.9384	0.8493
PSNR-HMA <sub>ICtCp</sub>	0.8455	0.8011	0.7696	0.9343	0.9076
PSNR-HMA <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.7287	0.7664	0.7094	0.9339	0.8294
PSNR-HMA <sub>XYZ</sub>	0.2895	0.6773	0.4979	0.8692	0.6831
PSNR-HMA <sub>Y'C'r'C'b'</sub>	0.3926	0.7160	0.6246	0.9236	0.7954
HDR-VDP2	0.8678	0.8685	0.8906	0.9516	0.9289
HDR-VQM	0.7735	0.8330	0.8995	0.9572	0.9170
PU-VIF	0.8658	0.7464	0.7704	0.9322	0.8863

Table C.4: SROCC of the different color quality metrics on the considered databases.

Quality Metric	4Kdttb	HDdttb	Narwaria et al.	Korshunov et al.	Zerman et al.
$\overline{\Delta E}_{HDR-Lab_{100}}$	0.4807	0.2578	0.6490	0.7523	0.6394
$\overline{\Delta E}_{HDR-Lab_{1000}}$	0.2123	0.2418	0.6179	0.6945	0.5259
$\overline{\Delta E}_{ICtCp}$	0.6846	0.3401	0.6218	0.8448	0.7599
$\overline{\Delta E}_{J_z a_z b_z}$	0.3008	0.2994	0.6339	0.6694	0.5602
$\overline{\Delta E}_{XYZ}$	0.0739	0.3949	0.3896	0.6850	0.4436
$\overline{\Delta E}_{Y'Cr'Cb'}$	0.1549	0.3992	0.4377	0.7297	0.4999
$\overline{\Delta E}_{HDR-Lab_{100}}^S$	0.7827	0.2784	0.7181	0.8508	0.7559
$\overline{\Delta E}_{HDR-Lab_{1000}}^S$	0.4898	0.2585	0.6207	0.8524	0.6851
$\overline{\Delta E}_{ICtCp}^S$	0.7911	0.2606	0.6195	0.8635	0.7892
$\overline{\Delta E}_{J_z a_z b_z}^S$	0.6396	0.2804	0.5855	0.8771	0.7283
$\overline{\Delta E}_{XYZ}^S$	0.1706	0.3651	0.3911	0.8130	0.5665
$\overline{\Delta E}_{Y'Cr'Cb'}^S$	0.2392	0.3485	0.3927	0.8674	0.6365
$SSIMc_{HDR-Lab_{100}}$	0.5184	0.7085	0.7212	0.8873	0.7535
$SSIMc_{HDR-Lab_{1000}}$	0.2991	0.6641	0.7643	0.8943	0.6047
$SSIMc_{ICtCp}$	0.7437	0.7748	0.7273	0.9174	0.8545
$SSIMc_{J_z a_z b_z}$	0.5059	0.7134	0.7926	0.8943	0.6740
$SSIMc_{XYZ}$	0.1785	0.6325	0.6064	0.7670	0.5065
$SSIMc_{Y'Cr'Cb'}$	0.2259	0.6393	0.7044	0.8392	0.5443
$CSSIM_{HDR-Lab_{100}}$	0.7972	0.4065	0.7605	0.8981	0.7834
$CSSIM_{HDR-Lab_{1000}}$	0.5369	0.3257	0.6322	0.8857	0.6813
$CSSIM_{ICtCp}$	0.7712	0.4696	0.7212	0.9173	0.8464
$CSSIM_{J_z a_z b_z}$	0.6730	0.4242	0.6181	0.9197	0.7037
$CSSIM_{XYZ}$	0.1717	0.3592	0.3054	0.7713	0.4995
$CSSIM_{Y'Cr'Cb'}$	0.2657	0.4307	0.3830	0.8805	0.5805
$FSIMc_{HDR-Lab_{100}}$	0.8510	0.8531	0.8548	0.9332	0.9068
$FSIMc_{HDR-Lab_{1000}}$	0.6835	0.8560	0.8196	0.9575	0.9025
$FSIMc_{ICtCp}$	0.9127	0.7892	0.8585	0.9449	0.8852
$FSIMc_{J_z a_z b_z}$	0.8371	0.9065	0.8485	0.9657	0.9046
$FSIMc_{XYZ}$	0.5784	0.7483	0.6376	0.8999	0.7413
$FSIMc_{Y'Cr'Cb'}$	0.6799	0.7966	0.7512	0.9500	0.8196
$PSNR-HMAC_{HDR-Lab_{100}}$	0.5533	0.4042	0.6592	0.7664	0.7337
$PSNR-HMAC_{HDR-Lab_{1000}}$	0.3534	0.3394	0.7138	0.7446	0.6589
$PSNR-HMAC_{ICtCp}$	0.7618	0.6373	0.7585	0.8638	0.8418
$PSNR-HMAC_{J_z a_z b_z}$	0.4893	0.4293	0.7065	0.8287	0.7301
$PSNR-HMAC_{XYZ}$	0.2282	0.5431	0.5565	0.8455	0.6315
$PSNR-HMAC_{Y'Cr'Cb'}$	0.3443	0.5669	0.6851	0.9025	0.7486

## C.3 OR

Table C.5: OR of the different color-blind quality metrics on the considered databases.

Quality Metric	4Kdtdb	HDdtdb	Narwaria et al.	Korshunov et al.	Zerman et al.
PU-PSNR	0.6354	0.5729	0.7714	0.5833	0.6400
PSNR <sub>HDR-Lab100</sub>	0.6458	0.5833	0.8857	0.6042	0.6800
PSNR <sub>HDR-Lab1000</sub>	0.6563	0.5625	0.7714	0.6667	0.7000
PSNR <sub>ICtCp</sub>	0.5938	0.5833	0.7786	0.5958	0.6100
PSNR <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.6563	0.6042	0.8143	0.6500	0.7400
PSNR <sub>XYZ</sub>	0.6875	0.6354	0.8429	0.8125	0.7700
PSNR <sub>Y'C'r'C'b'</sub>	0.6979	0.6042	0.7714	0.7583	0.7300
PU-SSIM	0.6354	0.4792	0.7786	0.4792	0.6500
SSIM <sub>HDR-Lab100</sub>	0.5938	0.4792	0.7857	0.4792	0.6500
SSIM <sub>HDR-Lab1000</sub>	0.6458	0.4896	0.7571	0.5875	0.7100
SSIM <sub>ICtCp</sub>	0.5833	0.5417	0.7929	0.4875	0.6700
SSIM <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.6771	0.5000	0.7500	0.5542	0.6900
SSIM <sub>XYZ</sub>	0.7083	0.6562	0.8071	0.7333	0.8700
SSIM <sub>Y'C'r'C'b'</sub>	0.6979	0.5625	0.8286	0.7000	0.8000
PU-MS-SSIM	0.4063	0.5104	0.6786	0.3667	0.5000
MS-SSIM <sub>HDR-Lab100</sub>	0.4375	0.4792	0.7357	0.3915	0.5400
MS-SSIM <sub>HDR-Lab1000</sub>	0.5000	0.4792	0.7143	0.3708	0.7000
MS-SSIM <sub>ICtCp</sub>	0.4479	0.5521	0.6500	0.3958	0.4600
MS-SSIM <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.4271	0.5729	0.6857	0.3500	0.6900
MS-SSIM <sub>XYZ</sub>	0.6875	0.6250	0.8143	0.6333	0.8100
MS-SSIM <sub>Y'C'r'C'b'</sub>	0.6667	0.5625	0.7929	0.5667	0.7900
PU-FSIM	0.3545	0.5000	0.6143	0.4167	0.5000
FSIM <sub>HDR-Lab100</sub>	0.3750	0.5208	0.6714	0.4500	0.4400
FSIM <sub>HDR-Lab1000</sub>	0.4479	0.5313	0.6357	0.3333	0.5300
FSIM <sub>ICtCp</sub>	0.3229	0.5000	0.6714	0.4667	0.5200
FSIM <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.4167	0.5104	0.6500	0.3250	0.5900
FSIM <sub>XYZ</sub>	0.6562	0.5938	0.7714	0.5625	0.7800
FSIM <sub>Y'C'r'C'b'</sub>	0.5833	0.5625	0.6786	0.4667	0.6900
PU-PSNR-HVS-M	0.4896	0.6563	0.7500	0.5875	0.6600
PSNR-HVS-M <sub>HDR-Lab100</sub>	0.4583	0.6250	0.7643	0.5875	0.6400
PSNR-HVS-M <sub>HDR-Lab1000</sub>	0.5938	0.6458	0.8071	0.5333	0.6600
PSNR-HVS-M <sub>I</sub>	0.4583	0.7083	0.7857	0.5125	0.5900
PSNR-HVS-M <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.5521	0.6875	0.7571	0.5750	0.6700
PSNR-HVS-M <sub>XYZ</sub>	0.6979	0.6354	0.8500	0.7042	0.7600
PSNR-HVS-M <sub>Y'C'r'C'b'</sub>	0.6667	0.6771	0.8214	0.5875	0.7200
PU-PSNR-HMA	0.4167	0.7188	0.8071	0.5125	0.6200
PSNR-HMA <sub>HDR-Lab100</sub>	0.4792	0.7083	0.7786	0.5292	0.6100
PSNR-HMA <sub>HDR-Lab1000</sub>	0.5833	0.6667	0.7571	0.4625	0.6500
PSNR-HMA <sub>ICtCp</sub>	0.4271	0.6667	0.7643	0.5375	0.5700
PSNR-HMA <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	0.6032	0.5938	0.7643	0.5292	0.6800
PSNR-HMA <sub>XYZ</sub>	0.6979	0.6146	0.8429	0.6833	0.7500
PSNR-HMA <sub>Y'C'r'C'b'</sub>	0.6667	0.5833	0.7857	0.6333	0.7100
HDR-VDP2	0.3545	0.4576	0.6250	0.3708	0.4400
HDR-VQM	0.5313	0.5616	0.9061	0.392	0.5300
PU-VIF	0.4063	0.5938	0.7571	0.5833	0.5500

Table C.6: OR of the different color quality metrics on the considered databases.

Quality Metric	4Kdtdb	HDdtdb	Narwaria et al.	Korshunov et al.	Zerman et al.
$\overline{\Delta E}_{HDR-Lab_{100}}$	0.7083	0.7604	0.7714	0.7167	0.8100
$\overline{\Delta E}_{HDR-Lab_{1000}}$	0.7083	0.7604	0.7929	0.7500	0.8000
$\overline{\Delta E}_{ICtCp}$	0.6354	0.7708	0.8357	0.7833	0.6900
$\overline{\Delta E}_{J_z a_z b_z}$	0.6458	0.7604	0.8429	0.8542	0.8000
$\overline{\Delta E}_{XYZ}$	0.6562	0.7500	0.8643	0.8333	0.8500
$\overline{\Delta E}_{Y'Cr'Cb'}$	0.6875	0.8714	0.8429	0.7756	0.8700
$\overline{\Delta E}^S_{HDR-Lab_{100}}$	0.5417	0.7813	0.8214	0.8083	0.7100
$\overline{\Delta E}^S_{HDR-Lab_{1000}}$	0.6667	0.7708	0.8143	0.8000	0.7600
$\overline{\Delta E}^S_{ICtCp}$	0.4792	0.7813	0.8214	0.8583	0.6800
$\overline{\Delta E}^S_{J_z a_z b_z}$	0.6563	0.7813	0.8429	0.7792	0.7400
$\overline{\Delta E}^S_{XYZ}$	0.6875	0.7500	0.8714	0.8000	0.8600
$\overline{\Delta E}^S_{Y'Cr'Cb'}$	0.6771	0.7396	0.8500	0.7792	0.7500
$SSIMc_{HDR-Lab_{100}}$	0.6979	0.5625	0.7929	0.6208	0.6900
$SSIMc_{HDR-Lab_{1000}}$	0.7083	0.6458	0.7429	0.8458	0.8000
$SSIMc_{ICtCp}$	0.5729	0.5729	0.8429	0.5292	0.6400
$SSIMc_{J_z a_z b_z}$	0.6979	0.5625	0.6929	0.6167	0.7300
$SSIMc_{XYZ}$	0.7188	0.6667	0.8000	0.7292	0.8700
$SSIMc_{Y'Cr'Cb'}$	0.7188	0.6667	0.7643	0.6542	0.8000
$CSSIM_{HDR-Lab_{100}}$	0.5417	0.6979	0.8071	0.6458	0.6700
$CSSIM_{HDR-Lab_{1000}}$	0.6771	0.7604	0.7500	0.6667	0.7700
$CSSIM_{ICtCp}$	0.5000	0.8021	0.8143	0.6458	0.6900
$CSSIM_{J_z a_z b_z}$	0.6667	0.7396	0.8286	0.6125	0.7600
$CSSIM_{XYZ}$	0.6979	0.7708	0.7929	0.7167	0.6900
$CSSIM_{Y'Cr'Cb'}$	0.7083	0.7708	0.8000	0.6167	0.7600
$FSIMc_{HDR-Lab_{100}}$	0.4375	0.5625	0.5928	0.5083	0.5600
$FSIMc_{HDR-Lab_{1000}}$	0.6563	0.5000	0.6357	0.3625	0.6200
$FSIMc_{ICtCp}$	0.2917	0.5938	0.5857	0.4833	0.5500
$FSIMc_{J_z a_z b_z}$	0.4375	0.4271	0.6714	0.3208	0.5500
$FSIMc_{XYZ}$	0.6458	0.5729	0.7786	0.5667	0.7700
$FSIMc_{Y'Cr'Cb'}$	0.6562	0.5729	0.6786	0.3625	0.6900
$PSNR-HMAC_{HDR-Lab_{100}}$	0.6354	0.7500	0.7571	0.7500	0.7300
$PSNR-HMAC_{HDR-Lab_{1000}}$	0.6458	0.7813	0.7643	0.7417	0.7400
$PSNR-HMAC_{ICtCp}$	0.5104	0.7188	0.7500	0.6375	0.7600
$PSNR-HMAC_{J_z a_z b_z}$	0.6458	0.7396	0.7357	0.6542	0.7300
$PSNR-HMAC_{XYZ}$	0.6667	0.6771	0.7571	0.6875	0.7500
$PSNR-HMAC_{Y'Cr'Cb'}$	0.6771	0.6979	0.7000	0.5958	0.7500

## C.4 RMSE

Table C.7: RMSE of the different color-blind quality metrics on the considered databases.

Quality Metric	4Kdtb	HDdtb	Narwaria et al.	Korshunov et al.	Zerman et al.
PU-PSNR	15.87	17.09	20.43	16.00	17.08
PSNR <sub>HDR-Lab100</sub>	16.44	17.45	24.95	15.97	17.66
PSNR <sub>HDR-Lab1000</sub>	19.52	19.79	21.27	18.77	21.14
PSNR <sub>ICtCp</sub>	15.32	16.81	20.05	15.46	15.49
PSNR <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	18.66	20.02	20.92	18.79	21.21
PSNR <sub>XYZ</sub>	21.80	21.24	22.36	22.31	24.58
PSNR <sub>Y'C<sub>r</sub>'C<sub>b</sub>'</sub>	21.38	20.42	21.79	20.23	22.75
PU-SSIM	15.92	14.48	16.44	11.75	16.69
SSIM <sub>HDR-Lab100</sub>	15.91	14.81	15.73	11.86	17.78
SSIM <sub>HDR-Lab1000</sub>	17.70	17.08	16.12	13.52	22.41
SSIM <sub>ICtCp</sub>	15.91	15.16	16.43	11.54	15.28
SSIM <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	16.70	17.14	15.13	12.46	21.20
SSIM <sub>XYZ</sub>	21.32	21.79	19.80	19.65	25.95
SSIM <sub>Y'C<sub>r</sub>'C<sub>b</sub>'</sub>	20.89	20.49	18.63	16.88	25.11
PU-MS-SSIM	11.76	12.73	12.15	8.43	10.73
MS-SSIM <sub>HDR-Lab100</sub>	11.86	12.62	13.46	8.71	11.26
MS-SSIM <sub>HDR-Lab1000</sub>	13.62	15.03	12.49	8.12	19.18
MS-SSIM <sub>ICtCp</sub>	12.10	13.33	11.73	9.04	10.53
MS-SSIM <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	12.25	14.11	10.11	7.69	17.88
MS-SSIM <sub>XYZ</sub>	19.95	21.23	19.65	15.88	23.17
MS-SSIM <sub>Y'C<sub>r</sub>'C<sub>b</sub>'</sub>	18.66	19.61	17.74	12.81	22.71
PU-FSIM	9.67	14.84	12.07	9.11	13.02
FSIM <sub>HDR-Lab100</sub>	9.89	14.70	12.29	9.39	12.18
FSIM <sub>HDR-Lab1000</sub>	11.98	12.50	14.41	7.19	12.09
FSIM <sub>ICtCp</sub>	9.70	15.70	12.60	10.05	13.64
FSIM <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	10.42	10.94	13.39	7.31	12.58
FSIM <sub>XYZ</sub>	18.04	18.71	19.01	13.55	19.97
FSIM <sub>Y'C<sub>r</sub>'C<sub>b</sub>'</sub>	15.72	16.37	16.79	11.60	17.22
PU-PSNR-HVS-M	12.80	16.75	19.95	15.39	12.19
PSNR-HVS-M <sub>HDR-Lab100</sub>	12.70	16.53	20.23	12.14	14.20
PSNR-HVS-M <sub>HDR-Lab1000</sub>	16.52	19.54	20.36	11.88	15.21
PSNR-HVS-M <sub>ICtCp</sub>	12.74	17.45	19.60	11.54	11.45
PSNR-HVS-M <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	15.12	20.75	20.48	12.85	16.52
PSNR-HVS-M <sub>XYZ</sub>	21.15	21.46	22.44	18.36	22.01
PSNR-HVS-M <sub>Y'C<sub>r</sub>'C<sub>b</sub>'</sub>	20.12	20.95	21.77	13.83	18.15
PU-PSNR-HMA	12.55	17.07	16.28	10.99	13.60
PSNR-HMA <sub>HDR-Lab100</sub>	12.92	17.70	16.21	11.10	14.10
PSNR-HMA <sub>HDR-Lab1000</sub>	16.51	18.20	17.40	10.96	15.62
PSNR-HMA <sub>ICtCp</sub>	12.00	16.09	15.41	11.60	11.43
PSNR-HMA <sub>J<sub>z</sub>a<sub>z</sub>b<sub>z</sub></sub>	15.16	18.11	16.84	11.82	16.47
PSNR-HMA <sub>XYZ</sub>	21.14	20.43	21.68	16.12	21.79
PSNR-HMA <sub>Y'C<sub>r</sub>'C<sub>b</sub>'</sub>	20.11	19.21	19.31	14.04	18.08
HDR-VDP2	11.3	12.55	11.27	9.60	9.50
HDR-VQM	14.11	14.72	10.64	8.57	10.88
PU-VIF	10.85	17.85	15.74	12.17	13.43

Table C.8: RMSE of the different color quality metrics on the considered databases

Quality Metric	4Kdtb	HDdtb	Narwaria et al.	Korshunov et al.	Zerman et al.
$\overline{\Delta E}_{HDR-Lab_{100}}$	19.72	26.81	19.31	20.25	23.73
$\overline{\Delta E}_{HDR-Lab_{1000}}$	21.60	26.77	19.92	22.29	25.49
$\overline{\Delta E}_{ICtCp}$	16.30	27.11	19.58	19.99	19.62
$\overline{\Delta E}_{J_z a_z b_z}$	20.28	26.46	19.25	26.08	25.01
$\overline{\Delta E}_{XYZ}$	21.64	26.01	23.68	27.51	28.31
$\overline{\Delta E}_{Y' C_{r'} C_{b'}}$	21.57	26.36	23.82	22.77	26.83
$\overline{\Delta E}^S_{HDR-Lab_{100}}$	14.64	27.51	18.70	21.30	19.77
$\overline{\Delta E}^S_{HDR-Lab_{1000}}$	19.67	27.68	20.56	20.82	21.99
$\overline{\Delta E}^S_{ICtCp}$	13.64	27.61	20.94	28.91	18.10
$\overline{\Delta E}^S_{J_z a_z b_z}$	17.58	27.44	21.45	20.53	20.35
$\overline{\Delta E}^S_{XYZ}$	21.33	26.67	24.18	21.90	26.71
$\overline{\Delta E}^S_{Y' C_{r'} C_{b'}}$	21.09	27.11	24.34	21.59	23.00
$SSIM_{HDR-Lab_{100}}$	18.88	19.62	16.68	14.61	19.95
$SSIM_{HDR-Lab_{1000}}$	21.07	20.74	15.46	15.92	24.30
$SSIM_{ICtCp}$	14.98	17.44	16.62	13.50	15.77
$SSIM_{J_z a_z b_z}$	19.07	20.18	14.56	14.20	22.63
$SSIM_{XYZ}$	21.42	22.11	19.55	19.93	26.04
$SSIM_{Y' C_{r'} C_{b'}}$	21.26	21.76	16.72	17.22	25.40
$CSSIM_{HDR-Lab_{100}}$	13.34	23.59	15.79	14.10	18.72
$CSSIM_{HDR-Lab_{1000}}$	18.61	25.05	18.68	14.71	22.20
$CSSIM_{ICtCp}$	14.15	23.74	17.04	12.62	16.24
$CSSIM_{J_z a_z b_z}$	16.24	24.16	20.41	16.47	21.14
$CSSIM_{XYZ}$	21.65	25.76	22.81	20.39	26.21
$CSSIM_{Y' C_{r'} C_{b'}}$	21.11	24.91	21.30	14.94	24.51
$FSIM_{HDR-Lab_{100}}$	11.78	14.14	12.80	11.02	12.11
$FSIM_{HDR-Lab_{1000}}$	16.07	13.88	13.87	8.41	12.0
$FSIM_{ICtCp}$	9.29	16.29	12.46	10.22	13.57
$FSIM_{J_z a_z b_z}$	12.19	11.09	12.98	7.40	12.36
$FSIM_{XYZ}$	17.99	18.82	18.88	13.47	19.93
$FSIM_{Y' C_{r'} C_{b'}}$	16.30	16.90	16.18	9.38	17.17
$PSNR-HMAC_{HDR-Lab_{100}}$	18.37	24.75	18.57	20.60	20.52
$PSNR-HMAC_{HDR-Lab_{1000}}$	20.72	25.40	18.70	21.49	22.68
$PSNR-HMAC_{ICtCp}$	14.57	21.12	19.02	15.94	17.03
$PSNR-HMAC_{J_z a_z b_z}$	19.26	24.37	17.14	17.94	20.57
$PSNR-HMAC_{XYZ}$	21.71	22.75	20.34	17.34	23.00
$PSNR-HMAC_{Y' C_{r'} C_{b'}}$	20.71	21.83	17.49	14.02	19.87

# METRICS SENSITIVITY ON THE CHROMINANCE ARTIFACTS WITH THE 4KDTB

This appendix is an extension of the Section 5.4.2 where the impacts of the chrominance artifacts on quality metrics performances is studied. For each reference image of the database 4Kdtb, the subjective and objective scores for each distorted image are shown in function of the HEVC Qp. The objective scores are displayed after applying the logistic regression presented in Section 2.5.

The images compressed with the chroma Qp offset adaptation (cf. Section 1.4.3) are represented with a red line.

The images compressed without the chroma Qp offset and 10 bits quantization on the chrominance with a green line. The images compressed without the chroma Qp offset algorithm and a 10 bits quantization are represented with a blue line.

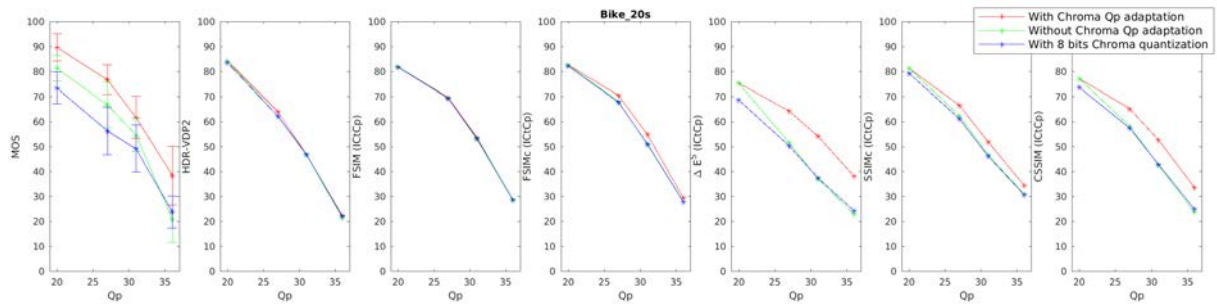


Figure D.1: Subjective and objective scores for the image Bike\_20s.

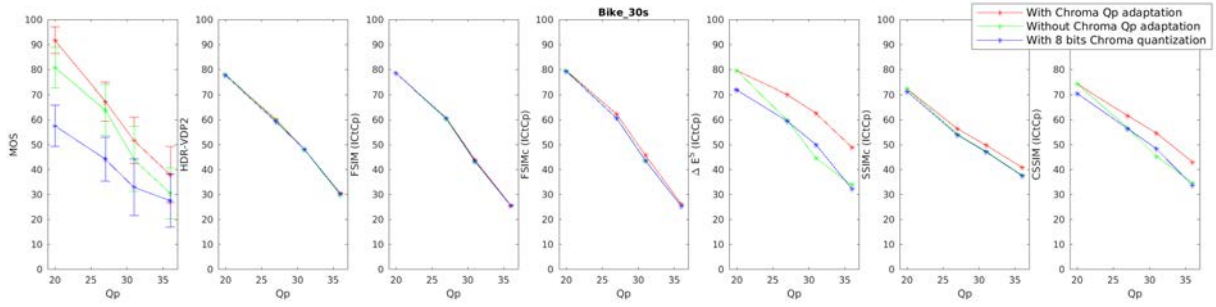


Figure D.2: Subjective and objective scores for the image **Bike\_30s**.

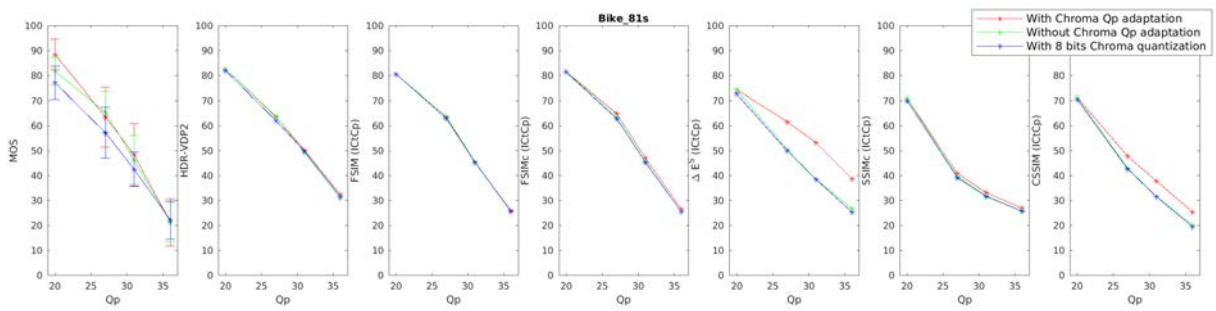


Figure D.3: Subjective and objective scores for the image **Bike\_81s**.

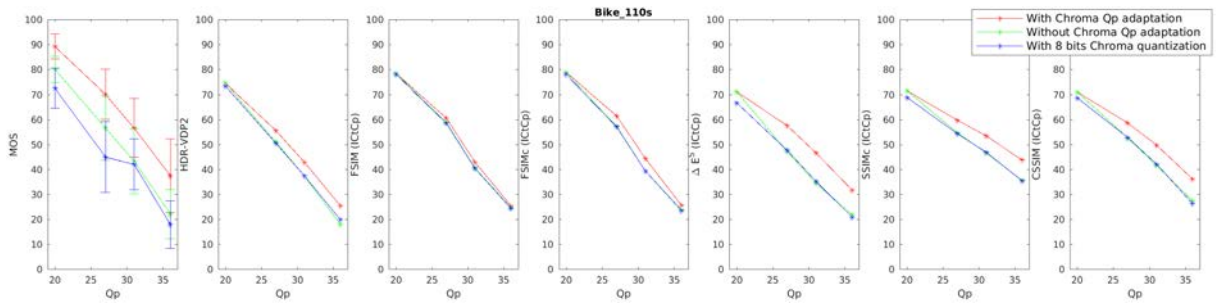


Figure D.4: Subjective and objective scores for the image **Bike\_110s**.

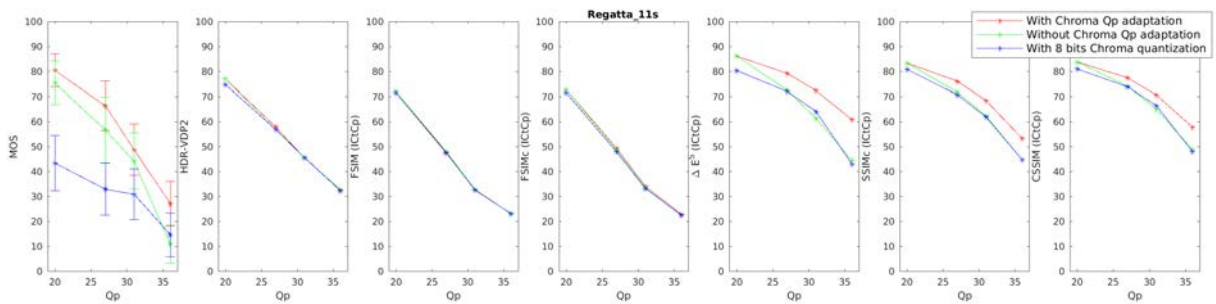


Figure D.5: Subjective and objective scores for the image **Regatta\_11s**

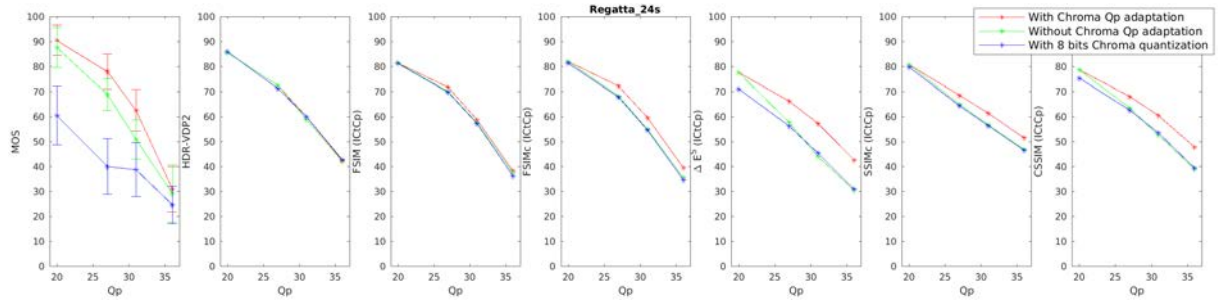


Figure D.6: Subjective and objective scores for the image Regatta\_24s.

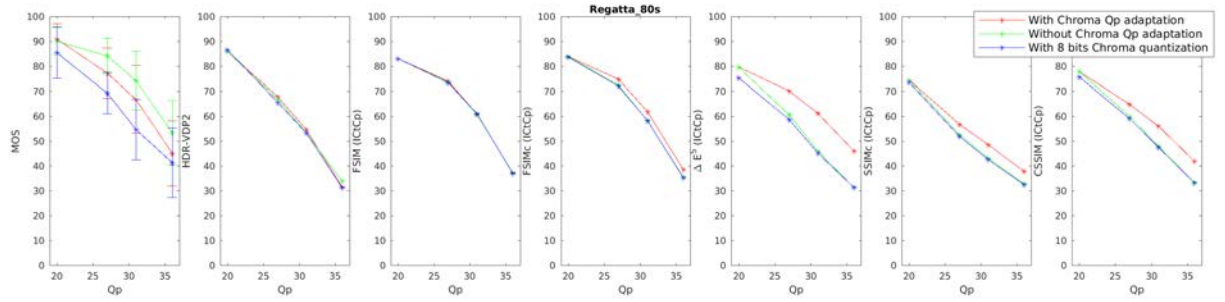


Figure D.7: Subjective and objective scores for the image Regatta\_80s.

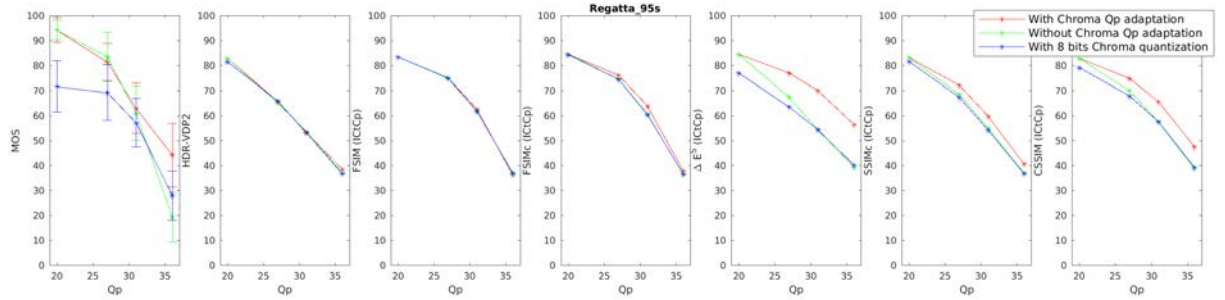


Figure D.8: Subjective and objective scores for the image Regatta\_95s.

# LIST OF ABBREVIATIONS

---

ACR	Absolute Category Rating
ACR-HR	Absolute Category Rating with Hidden Reference
AVC	Advanced Video Coding
CD-MMF	Context Dependent Multi-Method Fusion
CF-MMF	Context Free Multi-Method Fusion
CIE	Comission International de l'éclairage
cpd	cycle per degree
CQM	Combined Quality Metric
CSF	Contrast Sensitivity Function
CSSIM	Color and Structure SIMilarity
DCR	Degradation Category Rating
DR	Dynamic Range
DSIS	Double Stimulus Impairment Scale
EOTF	Electro-Optical transfer Function
EVQA	Ensemble-learning-based Video Quality Assessment
FSIM	Feature SIMilarity
FSIMc	Feature SIMilarity Color
FVQA	Fusion Video Quality Assessment
HD	High Definition
HDR	High Dynamic Range
HDR-VDP	High Dynamic Range Visual Difference Predictor

---

HDR-VQM	High Dynamic Range Video Quality Metric
HEVC	High Efficiency Video Coding
HFR	High Frame Rate
HLG	Hybrid-Log-Gamma
HRC	Hypothetical Reference Circuits
HVS	Human Visual System
INLSA	Iterated Nested Least Square Algorithm
IQA	Image Quality Assessment
ITU	International Telecommunication Union
JPEG	Joint Photographic Experts Group
KRCC	Kendall Rank Correlation Coefficient
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MS-SSIM	MultiScale Structure SIMilarity
MSE	Mean Square Error
MTF	Modulation Transfer Function
OETF	Opto-Electronic Transfer Function
OOTF	Opto-Optical Transfer Function
OR	Outlier Ratio
PCC	Pearson Correlation Coefficient
PQ	Perceptual Quantization
PSNR	Peak Signal on Noise Ratio
PSNR-HMA	PSNR-HVS-M Accounting for contrast change and hue shift
PSNR-HMAc	PSNR-HMA Color
PSNR-HVS	Peak Signal on Noise Ratio based on Human Visual System

---

PSNR-HVS-M	PSNR-HVS modelling contrast Masking
PSO	particle swarm optimization
QP	Quantization Parameter
RMSE	Root Mean Square Error
SDI	Serial Digital Interface
SDR	Standard Dynamic Range
SROCC	Spearman Rank Order Correlation
SSIM	Structure SIMilarity
SSIMc	Structure SIMilarity Color
SVR	Support Vector Regressor
UHD	Ultra High Definition
VIF	Visual Information Fidelity
VMAF	Video Multimethod Assessment Fusion
WCG	Wide Color Gamut

# LIST OF FIGURES

---

1.1	Dynamic Range of the Human Vision System and displays . . . . .	25
1.2	gamut of the BT.709 and BT.2020 color space and HVS gamut represented with a CIE $xy$ chromaticity diagram . . . . .	28
1.3	End-to-end workflow from the capture of light to the display of an image [31] . . .	29
1.4	End-to-end scene-referred workflow from the capture of light to the display of an image [31] . . . . .	30
1.5	End-to-end display-referred workflow from the capture of light to the display of an image [31] . . . . .	33
1.6	Different EOTF . . . . .	34
1.7	$L_{HDR}$ in function of the luminance for <b>(a)</b> different diffuse white luminances, and for <b>(b)</b> different surround luminances. . . . .	38
1.8	Chroma sampling format . . . . .	40
1.9	Pre-encoding process system using the $Y'C'_bC'_r$ color space . . . . .	40
1.10	Pre-encoding process system using the $ICtCp$ color space . . . . .	41
1.11	Example of a luma adjustment method (source [7]) . . . . .	41
1.12	Tone-mapped examples showing improvements of the luma adjustment method <b>(a)</b> original image <b>(b)</b> classic chrominance subsampling <b>(c)</b> chrominance subsampling with luma adjustment method (source [45]) . . . . .	42
1.13	simplified structure of an MPEG video encoder . . . . .	43
1.14	QP offset dQP in function of the average luma value in 64x64 block (source [7]) .	44
2.1	Characteristics of the Narwaria et al. [49] images: <b>(a)</b> The dynamic range, <b>(b)</b> key, <b>(b)</b> spatial Information, <b>(d)</b> Colorfulness. . . . .	51
2.2	Characteristics of the Korshunov et al. [24] images: <b>(a)</b> The dynamic range, <b>(b)</b> key, <b>(c)</b> spatial Information, <b>(d)</b> Colorfulness. . . . .	52
2.3	Characteristics of the Zerman et al. [50] images: <b>(a)</b> The dynamic range, <b>(b)</b> key, <b>(c)</b> spatial Information, <b>(d)</b> Colorfulness. . . . .	53
2.4	PU function and gamma function . . . . .	57
2.5	HDR-VDP-2 principle . . . . .	61
3.1	SONY BVM-X300 display . . . . .	68
3.2	Example of a pair of images presented to the viewer (Reinhard et al. TMO [93]).	69

---

3.3	Proposed subjective tests scale . . . . .	69
3.4	Characteristics of the HDdtb images: <b>(a)</b> The dynamic range, <b>(b)</b> key, <b>(c)</b> spatial Information, <b>(d)</b> Colorfulness. . . . .	72
3.5	Example of compression artifacts on the image Showgirl: <b>(a)</b> The original image, <b>(b)</b> Compressed image, Qp 39, with Chroma Qp adaptation, <b>(c)</b> Compressed image, Qp 39, without Chroma Qp adaptation. . . . .	76
3.6	The image Market3: <b>(a)</b> The original image, <b>(b)</b> with Gaussian noise on the chroma (SNR=0.5) . . . . .	77
3.7	The image MasonLake(1) under three conditions: <b>(a)</b> The original image, <b>(b)</b> gamut mismatch: BT.2020 image displayed as a BT.709 image <b>(c)</b> gamut mismatch: BT.709 image displayed as a BT.2020 images. . . . .	78
3.8	Repartition of HDdtb MOS scores. . . . .	79
3.9	MOS score obtained from 14 naive viewers in function of the Qp. . . . .	80
3.10	MOS obtained from naive viewer in function of MOS obtained from expert. . . . .	81
3.11	MOS score obtained from 9 experts in function of the Qp. . . . .	81
3.12	Characteristics of the 4Kdtb images: <b>(a)</b> The dynamic range, <b>(b)</b> key, <b>(c)</b> spatial Information, <b>(d)</b> Colorfulness. . . . .	83
3.13	The image Regatta_11s under three conditions: <b>(a)</b> The original image, <b>(b)</b> Compressed image, Qp 36, with Chroma Qp adaptation, <b>(c)</b> Compressed image, Qp 36, without Chroma Qp adaptation, <b>(d)</b> Compressed image, Qp 36, 8 bits quantization for the chrominance. . . . .	87
3.14	The image Bike_30s under two conditions: <b>(a)</b> The original image, <b>(b)</b> Compressed image, Qp 20, 8 bits quantization of the chrominance . . . . .	88
3.15	Histogram of the repartition of 4Kdtb MOS score . . . . .	89
3.16	MOS score obtained for the database 4Kdtb in function of the Qp . . . . .	90
4.1	Real display spectral emissions <b>(a)</b> Blue component <b>(b)</b> Green component <b>(c)</b> Red component . . . . .	94
4.2	Unrealistic spectral emissions <b>(a)</b> Blue component <b>(b)</b> Green component <b>(c)</b> Red component . . . . .	94
4.3	HDR-VDP2 score for two different spectrums (Sim2 HDR display and Sony BVM-X300 display) applied on the database HDdtb. . . . .	96
4.4	Performance of HDR-VDP-2 in function of the surround luminance for <b>(a)</b> existing database <b>(b)</b> the proposed database HDdtb . . . . .	96
4.5	HDR-VDP2 score for two different surround luminance applied on the database Narwaria et al. . . . .	97
4.6	Performance of HDR-VDP-2 in function of the angular resolution for <b>(a)</b> existing database <b>(b)</b> the proposed database HDdtb . . . . .	98

---

4.7	HDR-VDP2 score for different angular resolutions applied on the database Korshunov et al. <b>(a)</b> 20 pix/deg in function of 60 pix/deg <b>(b)</b> 30 pix/deg in function of 60 pix/deg <b>(c)</b> 80 pix/deg in function of 60 pix/deg . . . . .	98
4.8	MOS scores and HDR-VDP-2 scores in function of Qp for the image Regatta_24s of the database 4Kdtb . . . . .	102
5.1	Diagram of the proposed method to adapt SDR metrics to HDR/WCG contents. .	108
5.2	Different perceptually uniform luminances as a function of the linear luminance: <b>(a)</b> for the range 0–1000 cd/m <sup>2</sup> , <b>(b)</b> for the range 0–150 cd/m <sup>2</sup> . . . . .	109
5.3	SROCC performances for the 4Kdtb database for color-blind quality metrics <b>(a)</b> and for color quality metrics <b>(b)</b> . . . . .	111
5.4	SROCC performances for the Zerman et al. database for color-blind quality metrics <b>(a)</b> and for color quality metrics <b>(b)</b> . . . . .	111
5.5	SROCC performances for the HDdtb database for color-blind quality metrics <b>(a)</b> and for color quality metrics <b>(b)</b> . . . . .	112
5.6	SROCC performances for the Korshunov et al. database for color-blind quality metrics <b>(a)</b> and for color quality metrics <b>(b)</b> . . . . .	113
5.7	SROCC performances for the Narwaria et al. database for color-blind quality metrics <b>(a)</b> and for color quality metrics <b>(b)</b> . . . . .	114
5.8	SROCC of <b>(a)</b> FSIM <sub>HDR-Lab</sub> , <b>(b)</b> MS-SSIM <sub>HDR-Lab</sub> , <b>(c)</b> SSIM <sub>HDR-Lab</sub> in function of the diffuse white luminance. . . . .	117
5.9	Subjective and objective scores for the image Regatta_24s and for 6 metrics based on the <i>ICtCp</i> color space. . . . .	118
6.1	Architecture of our proposed HDR-VDP-2 color extension. . . . .	121
6.2	HDR-VDP-2 score in fonction of MOS for <b>(a)</b> The complete TID2013 database, <b>(b)</b> the TID2013 database without the distortion: "change in color saturation". One color represents all distorted images corresponding to one particular content.	124
6.3	HDR-VDP-2 <sub>CtCp</sub> scores without the chrominance subbands in function of HDR-VDP-2 <sub>CtCp</sub> with the chrominance subbands (using median weights) (Database: 4Kdtb). . .	129
6.4	Luminance weights $w_f$ of <b>(a)</b> HDR-VDP-2 and <b>(b)</b> HDR-VDP-2 <sub>CtCp</sub> (median weights) . . . . .	130
6.5	the distribution of the value taken by each $w_f$ weight for HDR-VDP-2 <sub>CtCp</sub> . . . .	130
7.1	Architecture of the proposed full-reference HDR quality metric. . . . .	134

---

7.2	The 4Kdtb image Bike_ 81s <b>(a)</b> and the distortion map $CS_5 a_z$ obtained for the image compressed with a Qp of 20 with <b>(b)</b> a quantization of 10 bits for the chrominances <b>(c)</b> a quantization of 8 bits for the chrominances. A white pixel means that there is no differences and a black pixel that there is a huge distortion.	136
7.3	The 4Kdtb image Bike_ 81s <b>(a)</b> and the distortion map $CS_5 b_z$ obtained for the image compressed with a Qp of 20 with <b>(b)</b> a quantization of 10 bits for the chrominances <b>(c)</b> a quantization of 8 bits for the chrominances. A white pixel means that there is no differences and a black pixel that there is a huge distortion.	136
7.4	HDR-VDP2 score in function of MOS <b>(a)</b> before the INLSA algorithm <b>(b)</b> after the INLSA algorithm . . . . .	138
7.5	Subjective and objective scores in function of HEVC Qp for the images <b>(a)</b> Bike_110s <b>(b)</b> Bike_20s <b>(c)</b> Bike_30s <b>(d)</b> Bike_30s. . . . .	141
7.6	Subjective and objective scores in function of HEVC Qp for the images <b>(a)</b> Regatta_11s <b>(b)</b> Regatta_24s <b>(c)</b> Regatta_80s <b>(d)</b> Regatta_95s. . . . .	142
B.1	Tone-mapped version (Reinhard et al. TMO [93]) of Narwaria et al. reference images. From left to right and from top to bottom: <b>(a)</b> Apartment_float_o15C <b>(b)</b> bausch_ lot <b>(c)</b> carpark_ ivc <b>(d)</b> CD1_serie2 <b>(e)</b> forest_path <b>(f)</b> lake <b>(g)</b> Light-House072 <b>(h)</b> moto <b>(i)</b> office_ivc <b>(j)</b> outro022168. . . . .	159
B.2	Tone-mapped version (Reinhard et al. TMO [93]) of Zerman et al. reference images. From left to right and from top to bottom: <b>(a)</b> AirBellowsGap <b>(b)</b> Balloon <b>(c)</b> FireEater <b>(d)</b> LasVegasStore <b>(e)</b> Market3 <b>(f)</b> MasonLake(1) <b>(g)</b> RedwoodSunset <b>(h)</b> Showgirl <b>(i)</b> Typewriter <b>(j)</b> UpheavalDome. . . . .	160
B.3	Tone-mapped version (Reinhard et al. TMO [93]) of Korshunov et al. reference images. From left to right and from top to bottom: <b>(a)</b> 507 <b>(b)</b> BloomingGorse2 <b>(c)</b> CanadianFalls <b>(d)</b> DevilsBathtub <b>(e)</b> dragon_ 3 <b>(f)</b> HancockKitchenInside <b>(g)</b> LabTypewriter <b>(h)</b> LasVegasStore <b>(i)</b> McKeesPub <b>(j)</b> MtRushmore2 <b>(k)</b> set18 <b>(l)</b> set22 <b>(m)</b> set23 <b>(n)</b> set24 <b>(o)</b> set31 <b>(p)</b> set33 <b>(q)</b> set70 <b>(r)</b> showgirl <b>(s)</b> sintel_ 2 <b>(t)</b> WillyDesk. . . . .	161
B.4	Tone-mapped version (Reinhard et al. TMO [93]) of the HDdtb reference images. From left to right and from top to bottom: <b>(a)</b> FireEater <b>(b)</b> LasVegasStore <b>(c)</b> Market3 <b>(d)</b> MasonLake(1) <b>(e)</b> RedwoodSunset <b>(f)</b> Showgirl <b>(g)</b> Typewriter <b>(h)</b> UpheavalDome. . . . .	162
B.5	Tone-mapped version (Reinhard et al. TMO [93]) of the 4Kdtb reference images. From left to right and from top to bottom: <b>(a)</b> Bike_ 20s <b>(b)</b> Bike_ 30s <b>(c)</b> Bike_ 81s <b>(d)</b> Bike_ 110s <b>(e)</b> Regatta_ 11s <b>(f)</b> Regatta_ 24s <b>(g)</b> Regatta_ 80s <b>(h)</b> Regatta_ 95s. . . . .	163

---

D.1	Subjective and objective scores for the image Bike_20s. . . . .	173
D.2	Subjective and objective scores for the image Bike_30s. . . . .	174
D.3	Subjective and objective scores for the image Bike_81s. . . . .	174
D.4	Subjective and objective scores for the image Bike_110s. . . . .	174
D.5	Subjective and objective scores for the image Regatta_11s . . . . .	174
D.6	Subjective and objective scores for the image Regatta_24s. . . . .	175
D.7	Subjective and objective scores for the image Regatta_80s. . . . .	175
D.8	Subjective and objective scores for the image Regatta_95s. . . . .	175

# LIST OF TABLES

1.1	Chromaticity coordinates for the color primaries and the white point of BT.709 and the BT.2020 color space . . . . .	29
2.1	Database characteristics. . . . .	50
2.2	Selected SDR quality metrics. . . . .	56
3.1	Description of HDdtb images (FireEater, LasVegasStore and Market3). . . . .	73
3.2	Description of HDdtb images (MasonLake(1), RedwoodSunset and Showgirl). . .	74
3.3	Description of HDdtb images (Typewriter, UpheavalDome). . . . .	75
3.4	characteristics of 4Kdtb images (Bike_110s and Bike_20s). . . . .	84
3.5	characteristics of 4Kdtb images (Bike_81s, Bike_30s and Regatta_11s). . . . .	85
3.6	characteristics of 4Kdtb images (Regatta_24s, Regatta_80s and Regatta_95s). .	86
4.1	Considered diplays . . . . .	93
4.2	Performance indexes of HDR-VDP-2 for Zerman et al. . . . .	95
4.3	Performance indexes of HDR-VDP-2 on HDdtb . . . . .	95
4.4	SROCC for the 4Kdtb database with and without downsampling the images. . . .	100
4.5	SROCC of the existing quality metrics on the considered databases . . . . .	101
5.1	SROCC for the HDdtb database with and without the gamut mismatch artifact. .	112
5.2	SROCC for the HDdtb database for three metrics based on $J_z a_z b_z$ , $\widetilde{J_z a_z b_z}$ and $HDR-Lab_{1000}$ . . . . .	116
6.1	Performance in term of SROCC of the proposed metric trained with different databases and HDR-VDP-2 . . . . .	125
6.2	characteristics of the training databases . . . . .	126
6.3	Median SROCC across 1000 Train-Test Combinations database by database. . .	127
6.4	SROCC of metrics with the median weights for the training databases. . . . .	128
6.5	SROCC of metrics with the median weights on the validation databases. . . . .	128
7.1	Median performances across 1000 Train-Test Combinations on the test set. . . .	139
7.2	Median SROCC across 1000 Train-Test Combinations database by database. . .	140
7.3	Performances of several metrics for <b>(a)</b> the complete database HDdtb <b>(b)</b> Only the HDdtb images with compression artifacts. . . . .	140

---

C.1	PCC of the different color-blind quality metrics on the considered databases . . .	165
C.2	PCC of the different color quality metrics on the considered databases. . . . .	166
C.3	SROCC of the different color-blind quality metrics on the considered databases.	167
C.4	SROCC of the different color quality metrics on the considered databases. . . . .	168
C.5	OR of the different color-blind quality metrics on the considered databases. . . .	169
C.6	OR of the different color quality metrics on the considered databases. . . . .	170
C.7	RMSE of the different color-blind quality metrics on the considered databases. .	171
C.8	RMSE of the different color quality metrics on the considered databases . . . . .	172

# BIBLIOGRAPHY

---

- [1] Steve Mann, “Compositing multiple pictures of the same scene”, in: *Proc. IS&T Annual Meeting, 1993*, 1993, pp. 50–52.
- [2] ARRI, *HDR FAQ - Common question regarding High Dynamic Range*, 2019, URL: <https://www.arri.com/en/learn-help/learn-help-camera-system/workflow-knowledge/image-science/frequently-asked-questions-on-hdr#accordion-44126> (visited on 04/23/2019).
- [3] Helge Seetzen, Wolfgang Heidrich, Wolfgang Stuerzlinger, Greg Ward, Lorne Whitehead, Matthew Trentacoste, Abhijeet Ghosh, and Andrejs Vorozcovs, “High dynamic range display systems”, in: *ACM transactions on graphics (TOG)* 23.3 (2004), pp. 760–768, DOI: 10.1145/1186562.1015797.
- [4] Stephen R Forrest, “The road to high efficiency organic light emitting devices”, in: *Organic Electronics* 4.2-3 (2003), pp. 45–48, DOI: 10.1016/j.orgel.2003.08.014.
- [5] *Operational practices in HDR television production*, Rep BT.2408-0, ITU-R, 2017.
- [6] *Image parameter values for high dynamic range television for use in production and international programme exchange*, Rec BT.2100-2, ITU-R, 2018.
- [7] *Conversion and coding practices for HDR/WCG Y’CbCr 4:2:0 video with PQ transfer characteristics*, Rec H-Suppl.15, ITU-T, 2017.
- [8] Cisco, *Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper*, 2019, URL: [https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html#\\_Toc532256803](https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html#_Toc532256803) (visited on 04/23/2019).
- [9] SIM2, *HDR47ES4MB*, 2017, URL: <http://hdr.sim2.it/hdrproducts/hdr47es4mb> (visited on 04/23/2019).
- [10] Tunç O. Aydın, Rafal Mantiuk, and Hans-Peter Seidel, “Extending quality metrics to full luminance range images”, in: *Proc.SPIE* 6806 (2008), DOI: 10.1117/12.765095.
- [11] Muhammad Safdar, Guihua Cui, Youn Jin Kim, and Ming Ronnier Luo, “Perceptually uniform color space for image signals including high dynamic range and wide gamut”, in: *Opt. Express* 25.13 (June 2017), pp. 15131–15151, DOI: 10.1364/OE.25.015131.

- 
- [12] Mark D. Fairchild and Ping-Hsu Chen, “Brightness, lightness, and specifying color in high-dynamic-range scenes and images”, in: *Proc.SPIE 7867* (2011), DOI: 10.1117/12.872075.
- [13] Maxime Rousselot, Éric Auffret, Xavier Ducloux, Olivier Le Meur, and Rémi Cozot, “Impacts of Viewing Conditions on HDR-VDP2”, in: *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1442–1446, DOI: 10.23919/EUSIPCO.2018.8553212.
- [14] Maxime Rousselot, Olivier Le Meur, Rémi Cozot, and Xavier Ducloux, “Quality Assessment of HDR/WCG Images Using HDR Uniform Color Spaces”, in: *Journal of Imaging* 5.1 (2019), DOI: 10.3390/jimaging5010018.
- [15] Maxime Rousselot, Éric Auffret, Xavier Ducloux, Olivier Le Meur, and Rémi Cozot, “Adapting HDR images using uniform color space for SDR quality metrics”, in: *COMpression et REprésentation des Signaux Audiovisuels (CORESA 2018)*, 2018.
- [16] Maxime Rousselot, Olivier Le Meur, Rémi Cozot, and Xavier Ducloux, “Quality metric aggregation for HDR/WCG images”, in: *26th IEEE International Conference on Image Processing (ICIP)*, IEEE, 2019.
- [17] Donald C Hood and Marcia A Finkelstein, “Sensitivity to light”, in: *Handbook of Perception and Human Performance: Vol. 1: Sensory Processes and Perception*, ed. by Kenneth Boff, Lloyd Kaufman, and James Thomas, New York, NY, USA: John Wiley and Sons, 1986, chap. 5.
- [18] James A Ferwerda, “Elements of early vision for computer graphics”, in: *IEEE Computer Graphics and Applications* 21.5 (July 2001), pp. 22–33, DOI: 10.1109/38.946628.
- [19] Timo Kunkel and Erik Reinhard, “A Reassessment of the Simultaneous Dynamic Range of the Human Visual System”, in: *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, APGV ’10, Los Angeles, Ca: ACM, 2010, pp. 17–24, DOI: 10.1145/1836248.1836251.
- [20] Dell, *Dell 20 Monitor: E2016H*, Feb. 2019, URL: <https://www.dell.com/en-uk/shop/accessories/apd/210-afpd> (visited on 04/23/2019).
- [21] *Ultra HD Forum:Phase A Guidelines*, tech. rep., version 1.5, Fremont, CA, USA: Ultra HD Forum, Sept. 2018.
- [22] Philippe Hanhart, Marco V Bernardo, Pavel Korshunov, Manuela Pereira, António MG Pinheiro, and Touradj Ebrahimi, “HDR image compression: A new challenge for objective quality metrics”, in: *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, Sept. 2014, pp. 159–164, DOI: 10.1109/QoMEX.2014.6982313.

- 
- [23] Xiaofei Pan, Jiaqi Zhang, Shanshe Wang, Shiqi Wang, Yun Zhou, Wenhua Ding, and Yahui Yang, "HDR video quality assessment: Perceptual evaluation of compressed HDR video", in: *Journal of Visual Communication and Image Representation* 57 (2018), pp. 76–83, DOI: doi:10.1016/j.jvcir.2018.10.016.
- [24] Pavel Korshunov, Philippe Hanhart, Thomas Richter, Alessandro Artusi, Rafał Mantiuk, and Touradj Ebrahimi, "Subjective quality assessment database of HDR images compressed with JPEG XT", in: *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, May 2015, pp. 1–6, DOI: 10.1109/QoMEX.2015.7148119.
- [25] Andrew Blake and Heinrich Bulthoff, "Shape from specularities: computation and psychophysics", in: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 331.1260 (1991), pp. 237–252, DOI: 10.1098/rstb.1991.0012.
- [26] Ron O Dror, Alan S Willsky, and Edward H Adelson, "Statistical characterization of real-world illumination", in: *Journal of Vision* 4.9 (2004), pp. 11–11, DOI: 10.1167/4.9.11.
- [27] Scott Daly, Timo Kunkel, Xing Sun, Suzanne Farrell, and Poppy Crum, "41.1: Distinguished Paper: Viewer Preferences for Shadow, Diffuse, Specular, and Emissive Luminance Limits of High Dynamic Range Displays", in: *SID Symposium Digest of Technical Papers* 44.1 (2013), pp. 563–566, DOI: 10.1002/j.2168-0159.2013.tb06271.x.
- [28] *Parameter values for the HDTV standards for production and international programme exchange*, Rec BT.709-6, ITU-R, 2015.
- [29] Maciej Pedzisz, "Beyond bt. 709", in: *SMPTE 2013 Annual Technical Conference & Exhibition*, SMPTE, Oct. 2013, pp. 1–13, DOI: 10.5594/M001507.
- [30] *Parameter values for ultra-high definition television systems for production and international programme exchange*, Rec BT.2020-2, ITU-R, 2015.
- [31] *High dynamic range television for production and international programme exchange*, Rep BT.2390-5, ITU-R, 2018.
- [32] *Reference electro-optical transfer function for flat panel displays used in HDTV studio production*, Rec BT.1886-0, ITU-R, 2011.
- [33] Scott Miller, Mahdi Nezamabadi, and Scott Daly, "Perceptual signal coding for more efficient usage of bit codes", in: *SMPTE Motion Imaging Journal* 122.4 (2013), pp. 52–59, DOI: 10.5594/j18290.
- [34] *Essential parameter values for the extended image dynamic range television system for programme production*, standard STD-B67, Association of Radio Industries and Businesses (ARIB), 2015.

- 
- [35] Tim Borer, Andrew Cotton, and Peter Wilson, “Perceptual Uniformity for High-Dynamic-Range Television Systems”, in: *SMPTE Motion Imaging Journal* 125.8 (2016), pp. 75–84, DOI: 10.5594/JMI.2016.2602118.
- [36] *High dynamic range electro-optical transfer function of mastering reference displays*, standard ST.2084, Society of Motion Picture & Television Engineers (SMPTE), 2014.
- [37] *Colorimetry — Part 4: CIE 1976 L\*A\*B\* colour space*, standard CIE S014-4/E:2007, Commision Internationale de l’Eclairage, 1976.
- [38] *What is ICTcP ?*, White paper Version 7.2, Dolby, 2017.
- [39] *Perceptual Color Volume: Measuring the Distinguishable colors of HDR and WCG displays*, White paper Version 7.1, Dolby, 2018.
- [40] Fritz Ebner and Mark D Fairchild, “Development and testing of a color space (IPT) with improved hue uniformity”, in: *color and Imaging Conference*, vol. 1998, 1, Society for Imaging Science and Technology, 1998, pp. 8–13.
- [41] Thomas Richter, “On the standardization of the JPEG XT image compression”, in: *2013 Picture Coding Symposium (PCS)*, Dec. 2013, pp. 37–40, DOI: 10.1109/PCS.2013.6737677.
- [42] Olie Baumann, Alex Okell, and Jacob Ström, “Characterization of Processing Artifacts in High Dynamic Range, Wide Color Gamut Video”, in: *SMPTE Motion Imaging Journal* 127.3 (Apr. 2018), pp. 1–7, DOI: 10.5594/JMI.2018.2792761.
- [43] Taoran Lu, Fangjun Pu, Peng Yin, Tao Chen, and Walt Husak, “Conversion and Coding Practices for HDR/WCG ICTCP 4:2:0 Video”, in: *2017 Data Compression Conference (DCC)*, Apr. 2017, pp. 13–22, DOI: 10.1109/DCC.2017.45.
- [44] Andrey Norkin, “Fast algorithm for HDR video pre-processing”, in: *2016 Picture Coding Symposium (PCS)*, Dec. 2016, pp. 1–5, DOI: 10.1109/PCS.2016.7906374.
- [45] Jacob Ström, Jonatan Samuelsson, and Kristofer Dovstam, “Luma adjustment for high dynamic range video”, in: *2016 Data Compression Conference (DCC)*, IEEE, 2016, pp. 319–328, DOI: 10.1109/DCC.2016.65.
- [46] *Signalling, backward compatibility and display adaptation for HDR/WCG video coding*, Rec H-Suppl.18, ITU-T, 2017.
- [47] *Methodology for the subjective assessment of the quality of television pictures*, Rec BT.500-13, ITU-R, 2012.
- [48] *Subjective video quality assessment methods for multimedia applications*, Rec P.910, ITU-T, 2008.

- 
- [49] Manish Narwaria, Matthieu P. Da Silva, Patrick Le Callet, and Romuald Pepion, "Tone mapping-based high-dynamic-range image compression: study of optimization criterion and perceptual quality", in: *Optical Engineering* 52 (2013), DOI: 10.1117/1.0E.52.10.102008.
- [50] Emin Zerman, Giuseppe Valenzise, and Frederic Dufaux, "An extensive performance evaluation of full-reference HDR image quality metrics", in: *Quality and User Experience* 2.1 (2017), p. 5, DOI: 10.1007/s41233-017-0007-4.
- [51] David Hasler and Sabine E. Suesstrunk, "Measuring colorfulness in natural images", in: vol. 5007, 2003, DOI: 10.1117/12.477378.
- [52] Jiangtao Kuang, Garrett M. Johnson, and Mark D. Fairchild, "iCAM06: A refined image appearance model for HDR image rendering", in: *Journal of Visual Communication and Image Representation* 18.5 (2007), Special issue on High Dynamic Range Imaging, pp. 406–414, ISSN: 1047-3203, DOI: 10.1016/j.jvcir.2007.06.003.
- [53] Rafat Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich, "HDR-VDP-2: A Calibrated Visual Metric for Visibility and Quality Predictions in All Luminance Conditions", in: *ACM Trans. Graph.* 30.4 (July 2011), 40:1–40:14, DOI: 10.1145/2010324.1964935.
- [54] Manish Narwaria, Rafal Mantiuk, Matthieu P. Da Silva, and Patrick Le Callet, "HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images", in: *Journal of Electronic Imaging* 24.1 (2015), DOI: 10.1117/1.JEI.24.1.010501.
- [55] Rafal Mantiuk, Karol Myszkowski, and Hans-Peter Seidel, "A Perceptual Framework for Contrast Processing of High Dynamic Range Images", in: *ACM Trans. Appl. Percept.* 3.3 (July 2006), pp. 286–308, ISSN: 1544-3558, DOI: 10.1145/1166087.1166095.
- [56] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda, "Photographic Tone Reproduction for Digital Images", in: *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '02, San Antonio, Texas: ACM, 2002, pp. 267–276, ISBN: 1-58113-521-1, DOI: 10.1145/566570.566575.
- [57] Giuseppe Valenzise, Francesca De Simone, Paul Lauga, and Frederic Dufaux, "Performance evaluation of objective quality metrics for HDR image compression", in: *Proc.SPIE* 9217 (2014), DOI: 10.1117/12.2063032.
- [58] Zicong Mai, Hassan Mansour, Rafal Mantiuk, Panos Nasiopoulos, Rabab Ward, and Wolfgang Heidrich, "Optimizing a Tone Curve for Backward-Compatible High Dynamic Range Image and Video Compression", in: *IEEE Transactions on Image Processing* 20.6 (June 2011), pp. 1558–1571, ISSN: 1057-7149, DOI: 10.1109/TIP.2010.2095866.

- 
- [59] *Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications*, Technical Specification ETSI TS 101 154 v2.4.1, DVB, 2018, pp. 130–132.
- [60] Anush Krishna Moorthy and Alan Conrad Bovik, “Blind image quality assessment: From natural scene statistics to perceptual quality”, in: *IEEE transactions on Image Processing* 20.12 (2011), pp. 3350–3364, DOI: 10.1109/TIP.2011.2147325.
- [61] Michele A Saad, Alan C Bovik, and Christophe Charrier, “Blind image quality assessment: A natural scene statistics approach in the DCT domain”, in: *IEEE transactions on Image Processing* 21.8 (2012), pp. 3339–3352, DOI: 10.1109/TIP.2012.2191563.
- [62] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik, “No-reference image quality assessment in the spatial domain”, in: *IEEE Transactions on Image Processing* 21.12 (2012), pp. 4695–4708, DOI: 10.1109/TIP.2012.2214050.
- [63] Navaneeth Kamballur Kottayil, Giuseppe Valenzise, Frederic Dufaux, and Irene Cheng, “Blind Quality Estimation by Disentangling Perceptual and Noisy Features in High Dynamic Range Images”, in: *IEEE Transactions on Image Processing* 27.3 (2017), pp. 1512–1525, DOI: 10.1109/TIP.2017.2778570.
- [64] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al., “Image database TID2013: Peculiarities, results and perspectives”, in: *Signal Processing: Image Communication* 30 (2015), pp. 57–77, DOI: 10.1016/j.image.2014.10.009.
- [65] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang, “FSIM: A Feature Similarity Index for Image Quality Assessment”, in: *IEEE Transactions on Image Processing* 20.8 (Aug. 2011), pp. 2378–2386, DOI: 10.1109/TIP.2011.2109730.
- [66] Hua-Wen Chang, Hua Yang, Yong Gan, and Ming-Hui Wang, “Sparse feature fidelity for perceptual image quality assessment”, in: *IEEE Transactions on Image Processing* 22.10 (2013), pp. 4007–4018, DOI: 10.1109/TIP.2013.2266579.
- [67] Zhou Wang and Eero P Simoncelli, “Reduced-reference image quality assessment using a wavelet-domain natural image statistic model”, in: *Human Vision and Electronic Imaging X*, vol. 5666, International Society for Optics and Photonics, 2005, pp. 149–160, DOI: 10.1117/12.597306.
- [68] Guangtao Zhai, Xiaolin Wu, Xiaokang Yang, Weisi Lin, and Wenjun Zhang, “A psychovisual quality metric in free-energy principle”, in: *IEEE Transactions on Image Processing* 21.1 (2011), pp. 41–52, DOI: 10.1109/TIP.2011.2161092.
- [69] Rajiv Soundararajan and Alan C Bovik, “RRED indices: Reduced reference entropic differencing for image quality assessment”, in: *IEEE Transactions on Image Processing* 21.2 (2011), pp. 517–526, DOI: 10.1109/TIP.2011.2166082.

- 
- [70] Manish Narwaria, Matthieu Perreira Da Silva, and Patrick Le Callet, "HDR-VQM: An objective quality measure for high dynamic range video", in: *Signal Processing: Image Communication* 35 (2015), pp. 46–60, ISSN: 0923-5965, DOI: doi:10.1016/j.image.2015.04.009.
- [71] Xuemei Zhang and Brian A. Wandell, "A spatial extension of CIELAB for digital color-image reproduction", in: *Journal of the Society for Information Display* 5.1 (1997), pp. 61–63, DOI: 10.1889/1.1985127.
- [72] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", in: *IEEE Transactions on Image Processing* 13.4 (Apr. 2004), pp. 600–612, DOI: 10.1109/TIP.2003.819861.
- [73] Zhou Wang, Ligang Lu, and Alan C. Bovik, "Video quality assessment based on structural distortion measurement", in: *Signal Processing: Image Communication* 19.2 (2004), pp. 121–132, ISSN: 0923-5965, DOI: 10.1016/S0923-5965(03)00076-6.
- [74] Mohammed A. Hassan and Mazen S. Bashraheel, "Color-based structural similarity image quality assessment", in: *2017 8th International Conference on Information Technology (ICIT)*, May 2017, pp. 691–696, DOI: 10.1109/ICITECH.2017.8079929.
- [75] Zhou Wang, Eero P. Simoncelli, and Alan C. Bovik, "Multiscale structural similarity for image quality assessment", in: *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers*, vol. 2, Nov. 2003, pp. 1398–1402, DOI: 10.1109/ACSSC.2003.1292216.
- [76] Hamid R. Sheikh, Alan C. Bovik, and Gustavo. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics", in: *IEEE Transactions on image processing* (2005).
- [77] Nikolay Ponomarenko, Flavia Silvestri, Karen Egiazarian, Marco Carli, Jaakko Astola, and Vladimir Lukin, "On between-coefficient contrast masking of DCT basis functions", in: *Proceedings of the third international workshop on video processing and quality metrics*, vol. 4, 2007.
- [78] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, and Marco Carli, "Modified image visual quality metrics for contrast change and mean shift accounting", in: *CAD Systems in Microelectronics (CADSM), 2011 11th International Conference The Experience of Designing and Application of*, IEEE, 2011.
- [79] JJ Vos, TJTP Van den Berg, et al., "Report on disability glare", in: *CIE collection* 135.1 (1999), pp. 1–9.

- 
- [80] Kil Joong Kim, Rafal Mantiuk, and Kyoung Ho Lee, "Measurements of achromatic and chromatic contrast sensitivity functions for an extended range of adaptation luminance", in: *Human Vision and Electronic Imaging XVIII*, vol. 8651, International Society for Optics and Photonics, 2013, 86511A, DOI: 10.1117/12.2002178.
- [81] Eero P Simoncelli and William T Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation", in: *Proceedings., International Conference on Image Processing*, vol. 3, Oct. 1995, 444–447 vol.3, DOI: 10.1109/ICIP.1995.537667.
- [82] Rafal Mantiuk, *HDR-VDP: Frequently Asked Questions*, Aug. 2017, URL: [http://hdrvdp.sourceforge.net/wiki/index.php/Frequently\\_Asked\\_Questions](http://hdrvdp.sourceforge.net/wiki/index.php/Frequently_Asked_Questions) (visited on 06/04/2019).
- [83] Ann Marie Rohaly, Philip J Corriveau, John M Libert, Arthur A Webster, Vittorio Baroncini, John Beerends, Jean-Louis Blin, Laura Contin, Takahiro Hamada, David Harrison, et al., "Video quality experts group: Current results and future directions", in: *Visual Communications and Image Processing 2000*, vol. 4067, International Society for Optics and Photonics, 2000, pp. 742–754, DOI: 10.1117/12.386632.
- [84] *Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models*, Rec P.1401, ITU-T, 2012.
- [85] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms", in: *IEEE Transactions on Image Processing* 15.11 (Nov. 2006), pp. 3440–3451, DOI: 10.1109/TIP.2006.881959.
- [86] Eric Cooper Larson and Michael Damon Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy", in: *Journal of Electronic Imaging* 19.1 (2010), DOI: 10.1117/1.3267105.
- [87] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, Marco Carli, and Federica Battisti, "TID2008-a database for evaluation of full-reference visual quality assessment metrics", in: *Advances of Modern Radioelectronics* 10.4 (2009), pp. 30–45.
- [88] Philippe Hanhart, Martin Řeřábek, and Touradj Ebrahimi, "Subjective and objective evaluation of HDR video coding technologies", in: *International Conference on Quality of Multimedia Experience (QoMEX)*, 2016, pp. 1–6, DOI: 10.1109/QoMEX.2016.7498943.
- [89] Philippe Hanhart, Martin Řeřábek, and Touradj Ebrahimi, "Towards high dynamic range extensions of HEVC: subjective evaluation of potential coding technologies", in: *Proc.SPIE* 9599 (2015), DOI: 10.1117/12.2193832.

- 
- [90] Toinon Vigier, Lukáš Krasula, Aurélien Milliat, Matthieu Perreira Da Silva, and Patrick Le Callet, “Performance and robustness of HDR objective quality metrics in the context of recent compression scenarios”, in: *Digital Media Industry and Academic Forum*, July 2016, pp. 59–64, DOI: 10.1109/DMIAF.2016.7574903.
- [91] Philippe Hanhart, Marco V Bernardo, Manuela Pereira, António MG Pinheiro, and Touradj Ebrahimi, “Benchmarking of objective quality metrics for HDR image quality assessment”, in: *EURASIP Journal on Image and Video Processing* 2015.1 (Dec. 2, 2015), p. 39, DOI: 10.1186/s13640-015-0091-4.
- [92] Sony, *Operation Manuals: BVM-X300*, 2017, URL: [https://pro.sony/fr\\_CH/product-resources/manuals/1237495186206](https://pro.sony/fr_CH/product-resources/manuals/1237495186206) (visited on 04/23/2019).
- [93] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda, “Photographic tone reproduction for digital images”, in: *ACM transactions on graphics (TOG)* 21.3 (2002), pp. 267–276, DOI: 10.1145/566654.566575.
- [94] Sébastien Lasserre, Fabrice LeLéanec, and Edouard Francois, “Description of HDR sequences proposed by Technicolor”, in: *ISO/IEC JTC1/SC29/WG11 JCTVC-P0228*, IEEE, San Jose, USA (2013).
- [95] Jan Froehlich, Stefan Grandinetti, Bernd Eberhardt, Simon Walter, Andreas Schilling, and Harald Brendel, “Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays”, in: *Proc.SPIE* 9023 (2014), DOI: 10.1117/12.2040003.
- [96] Mark D Fairchild, “The HDR photographic survey”, in: *Color and Imaging Conference*, vol. 2007, 1, Society for Imaging Science and Technology, 2007, pp. 233–238.
- [97] Antoine Coutrot and Nathalie Guyader, “How saliency, faces, and sound influence gaze in dynamic social scenes”, in: *Journal of Vision* 14.8 (July 2014), pp. 5–5, DOI: 10.1167/14.8.5.
- [98] Scott J. Daly, “Visible differences predictor: an algorithm for the assessment of image fidelity”, in: *Proc.SPIE* 1666 (1992), DOI: 10.1117/12.135952.
- [99] Peter G. J. Barten, “Formula for the contrast sensitivity of the human eye”, in: *Proc.SPIE* 5294 (2003), DOI: 10.1117/12.537476.
- [100] Erik Reinhard, Jürgen Stauder, and Michel Kerdranvat, “An Assessment of Reference Levels in HDR Content”, in: *SMPTE Motion Imaging Journal* 128.3 (Apr. 2019), pp. 20–27, DOI: 10.5594/JMI.2019.2895795.
- [101] Manish Narwaria, Matthieu Perreira Da Silva, Patrick Le Callet, and Romuald Pépion, “Impact of tone mapping in High dynamic range image compression”, in: *VPQM*, Chandler, United States, Jan. 2014, pp. 1–6.

- 
- [102] Margaret H Pinson and Stephen Wolf, “An objective method for combining multiple subjective data sets”, in: *Proc.SPIE* 5150 (2003), DOI: 10.1117/12.509909.
- [103] John A Nelder and Roger Mead, “A simplex method for function minimization”, in: *The computer journal* 7.4 (1965), pp. 308–313.
- [104] Lixing Han and Michael Neumann, “Effect of dimensionality on the Nelder Mead simplex method”, in: *Optimization Methods and Software* 21 (Feb. 2006), pp. 1–16, DOI: 10.1080/10556780512331318290.
- [105] Russell Eberhart and James Kennedy, “A new optimizer using particle swarm theory”, in: *MHS’95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, IEEE, Oct. 1995, pp. 39–43, DOI: 10.1109/MHS.1995.494215.
- [106] Krzysztof Okarma, “Combined Full-Reference Image Quality Metric Linearly Correlated with Subjective Assessment”, in: *Artificial Intelligence and Soft Computing*, Berlin: Springer Berlin Heidelberg, 2010, pp. 539–546, DOI: 10.1007/978-3-642-13208-7\_67.
- [107] Tsung-Jung Liu, Weisi Lin, and C.-C Jay Kuo, “Image Quality Assessment Using Multi-Method Fusion”, in: *IEEE Transactions on Image Processing* 22.5 (May 2013), pp. 1793–1807, DOI: 10.1109/TIP.2012.2236343.
- [108] Joe Yuchieh Lin, Tsung-Jung Liu, Eddy Chi-Hao Wu, and C.-C Jay Kuo, “A fusion-based video quality assessment (FVQA) index”, in: *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, Dec. 2014, pp. 1–5, DOI: 10.1109/APSIPA.2014.7041705.
- [109] Joe Y. Lin, Chi-Hao Wu, Ioannis Katsavounidis, Zhi Li, Anne Aaron, and C.-C Jay Kuo, “EVQA: An ensemble-learning-based video quality assessment index”, in: *2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, June 2015, pp. 1–6, DOI: 10.1109/ICMEW.2015.7169760.
- [110] Zhi Li, Anne Aaron, Ioannis Katsavounidis, Anush Moorthy, and Megha Manohara, *Toward a practical perceptual video quality metric*, ed. by The Netflix Tech Blog, 2016, URL: <https://medium.com/netflix-techblog/toward-a-practical-perceptual-video-quality-metric-653f208b9652> (visited on 04/23/2019).
- [111] Anustup Choudhury and Scott Daly, “Combining Quality Metrics for Improved HDR Image Quality Assessment”, in: *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, Mar. 2019, pp. 179–184, DOI: 10.1109/MIPR.2019.00039.
- [112] R Eric Fredericksen and Robert F Hess, “Estimating multiple temporal mechanisms in human vision”, in: *Vision Research* 38.7 (1998), pp. 1023–1040.

- 
- [113] Tamara Seybold, Betina L. Koelln, Aynur Pasha, and Harald Brendel, “Visibility of spatiotemporal noise in digital video”, in: *Color and Imaging Conference 2016.1* (2016), pp. 20–26, DOI: doi:10.2352/ISSN.2169-2629.2017.32.20.
- [114] Catherine Meininger, “Determining Visibility Thresholds for Spatial and Spatiotemporal Chromatic Noise”, in: *SMPTE Motion Imaging Journal* 128.2 (Mar. 2019), pp. 31–40, DOI: 10.5594/JMI.2018.2887289.
- [115] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lina Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al., “Color image database TID2013: Peculiarities and preliminary results”, in: *Visual Information Processing (EUVIP), 2013 4th European Workshop on*, IEEE, 2013.
- [116] Cambodge Bist, Rémi Cozot, Gérard Madec, and Xavier Ducloux, “Tone compatibility between HDR displays”, in: *Applications of Digital Image Processing XXXIX*, vol. 9971, International Society for Optics and Photonics, 2016, p. 99710D, DOI: 10.1117/12.2236965.
- [117] Peter Kovesi, “Image features from phase congruency”, in: *Videre: Journal of computer vision research* 1.3 (1999), pp. 1–26.
- [118] Christopher C Yang and Sai Ho Kwok, “Efficient gamut clipping for color image processing using LHS and YIQ”, in: *Optical Engineering* (2003), DOI: 10.1117/1.1544479.
- [119] Karen Egiazarian, Jaakko Astola, Nikolay Ponomarenko, Vladimir Lukin, Federica Battisti, and Marco Carli, “New full-reference quality metrics based on HVS”, in: *Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, vol. 4, 2006.
- [120] Francesco Banterle, Alessandro Artusi, Kurt Debattista, and Alan Chalmers, *Advanced High Dynamic Range Imaging: Theory and Practice (2nd Edition)*, Natick, MA, USA: AK Peters (CRC Press), July 2017.

## **Titre:** Estimation de la qualité d'image High Dynamic Range et Wide Color Gamut

**Mot clés :** Wide Color Gamut, High Dynamic Range, Évaluation de la qualité d'images, Compression, Traitement de l'image

**Resumé :** Ces dernières années, les technologies d'écran se sont considérablement améliorées. Par exemple, le contraste des écrans à plage dynamique élevée (HDR) dépasse de loin la capacité d'un écran conventionnel. De plus, un écran à gamut de couleur étendu (WCG) peut couvrir un espace colorimétrique plus grand que jamais. L'évaluation de la qualité de ces nouveaux contenus est devenue un domaine de recherche actif, les métriques de qualité SDR classiques n'étant pas adaptées. Cependant, les études les plus récentes négligent souvent une caractéristique importante: les chrominances. En effet, les bases de données existantes contiennent des images HDR avec un gamut de couleur standard, négligeant ainsi l'augmentation de l'espace colorimétrique due au WCG et les artefacts

chromatiques. La plupart des mesures de qualité HDR objectives non plus ne prennent pas en compte ces artefacts. Pour surmonter cette problématique, dans cette thèse, nous proposons deux nouvelles bases de données HDR/WCG annotés avec des scores subjectifs présentant des artefacts chromatique réaliste. En utilisant ces bases de données, nous explorons trois solutions pour créer des métriques HDR/WCG: l'adaptation des métriques de qualité SDR, l'extension colorimétrique d'une métrique HDR connue appelée HDR-VDP-2 et, enfin, la fusion de diverses métriques de qualité et de features colorimétriques. Cette dernière métrique présente de très bonnes performances pour prédire la qualité tout en étant sensible aux distorsions chromatiques.

## **Title:** Image quality assessment of High Dynamic Range and Wide Color Gamut images

**Keywords :** Wide Color Gamut, High Dynamic Range, Image Quality Assessment, Compression, Image processing

**Abstract :** To improve their ability to display astonishing images, screen technologies have been greatly evolving. For example, the contrast of high dynamic range rendering systems far exceed the capacity of a conventional display. Moreover, a Wide Color gamut display can cover a bigger color space than ever. Assessing the quality of these new content has become an active field of research as classical SDR quality metrics are not adapted. However, state-of-the-art studies often neglect one important image characteristics: chrominances. Indeed, previous databases contain HDR images with a standard gamut thus neglecting the increase of color space due to WCG. Due to their gamut, these databases are less prone to contain chromatic artifacts than WCG content. Moreover, most existing HDR objective quality

metrics only consider luminance and are not considering chromatic artifacts. To overcome this problematic, in this thesis, we have created two HDR / WCG databases with annotated subjective scores. We focus on the creation of a realistic chromatic artifacts that can arise during compression. In addition, using these databases, we explore three solutions to create HDR / WCG metrics. First, we propose a method to adapt SDR metrics to HDR / WCG content. Then, we proposed an extension of a well-known HDR metric called HDR-VDP-2. Finally, we create a new metric based by aggregating two quality metrics and color features. This last metric presents very good performance to predict quality while being sensitive to chromatic distortion.