# Methods for improving the backward compatible High Dynamic Range compression

David Gommelet

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

UNIVERSITÉ DE RENNES 1

# THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 601
*Mathématiques et Sciences et Technologies de l'Information et de la Communication*
Spécialité : *Informatique*

Par

**David GOMMELET**

## Methods for Improving the Backward Compatible High Dynamic Range Compression

**Thèse présentée et soutenue à Rennes, le 25 septembre 2018**
**Unité de recherche : préparée au centre de recherche INRIA Rennes Bretagne Atlantique et à l'entreprise ERICSSON / ENVIVIO de Rennes**

**Rapporteurs avant soutenance :**

| | |
|---|---|
| Marco CAGNAZZO | Maitre de conférences |
| Frédéric DUFAUX | Directeur de recherche |

**Composition du Jury :**

| | |
|---|---|
| Marc ANTONINI | Directeur de recherche, Laboratoire I3S |
| Marco CAGNAZZO | Maitre de conférences, TELECOM ParisTech |
| Frédéric DUFAUX | Directeur de recherche, L2S Centrale Supelec |
| Patrick LE CALLET | Professeur, Université de Nantes |
| Adrian MUNTEANU | Professeur, Vrije Universiteit Brussel |

**Président**

**Directeur de thèse**

| | |
|---|---|
| Christine GUILLEMOT | Directrice de recherche, INRIA |

**Co-directeurs de thèse**

| | |
|---|---|
| Aline ROUMY | Chargée de recherche, INRIA |
| Michael ROPERT | Principal Ingénieur, ERICSSON |

**Invité(s)**

| | |
|---|---|
| Julien LE TANOU | Ingénieur Senior, ERICSSON |

# Remerciements

Tout d'abord, je souhaite vivement remercier mes encadrants au cours de ces quatre dernières années: Christine GUILLEMOT, Julien LE TANOU, Michael ROPERT et Aline ROUMY. Merci à vous pour votre disponibilité, vos conseils, vos encouragements mais aussi votre gentillesse. J'ai énormément appris à vos côtés et ce tout au long de cette thèse. Votre suivi assidu a largement contribué à la réalisation de cette thèse. Je tiens aussi à remercier Thomas GUIONNET qui devait être mon encadrant industriel de thèse mais qui malheureusement a changer d'entreprise après le début de ma thèse. Merci de m'avoir choisi et fait confiance pour cette thèse.

Je souhaite ensuite remercier tout particulièrement les membres de mon jury de thèse pour le temps que vous avez consacré à mon manuscrit et à ma soutenance, ainsi que pour vos commentaires pertinents sur mes travaux. Merci à Marc ANTONINI d'avoir présidé ce jury, à Marco CAGNAZZO et Frédéric DUFAUX d'avoir accepté d'être rapporteurs, et à Patrick LE CALLET et Adrian MUNTEANU pour avoir accepté d'être examinateurs.

Par ailleurs, je remercie ERICSSON, anciennement ENVIVIO, ainsi que l'INRIA pour les conditions idéales de travail dont j'ai pu bénéficier pendant ma thèse. Merci à tous mes collègues pour leur accueil, les nombreuses discussions et les bons moments passés ensemble. Je tiens à remercier plus particulièrement : Médéric BLESTEL pour son aide et tous ses conseils et idées, Maxime BICHON qui était, avec moi, l'un des deux premiers doctorants d'ERICSSON/ENVIVIO et qui est donc passé par les mêmes problèmes et interrogations, Thierry DUMAS pour son aide et sa présence en fin de thèse, Amandine BIARD pour avoir simplifié mes démarches administratives et enfin, Huguette BECHU pour son aide, notamment durant le jour stressant de soutenance.

Je remercie également tous mes amis pour leurs encouragements et pour m'avoir permis de me changer les idées. Dans la peur d'oublier des noms, je me contenterai de remercier Guilhem, Jérémy et Jude pour être venu assister à ma soutenance.

Je termine par un grand merci à ma famille, mes parents, ma soeur et mon frère pour m'avoir soutenu tout au long de cette thèse. C'est aussi grâce à vous que je suis arrivé jusqu'ici. Finalement, un immense merci à Tiphaine pour m'avoir supporté au quotidien tout au long de cette thèse en particulier dans les moments les plus stressants. Merci pour ta patience et tes encouragements.

# Contents

# Résumé en Français

## Préambule et Contexte

La distribution de contenu vidéo tend à devenir universelle. La diffusion vidéo, initialement limitée aux télévisions, est maintenant disponible pour les ordinateurs, tablettes et smartphones. Les réseaux de diffusion comme la Transmission Numérique Terrestre (TNT), la transmission par câble, satellite, téléphone mobile et Internet sont tous utilisés pour transmettre des contenus images ou vidéos. Toutes ces évolutions bénéficient à l'utilisateur final en terme d'accessibilité. Cependant, la quantité de données transmises a énormément augmenté ces dernières années, avec l'avènement de plateformes comme YouTube ou Netflix, et elle continue d'augmenter. Malgré des capacités de diffusion améliorées, l'accroissement continu du nombre de contenus à transmettre nécessite une augmentation des performances de compression.

Parallèlement à l'évolution des dispositifs de visionnage et à l'évolution des réseaux, les exigences de qualité et d'immersion ne cessent d'augmenter. En effet, ces dernières années, les résolutions vidéos ont rapidement évolué de définition standard (SD : Standard Definition) vers la haute définition (HD : High Definition), et plus récemment vers l'ultra haute définition (UHD : Ultra High Definition). Ces améliorations permettent de reproduire des vidéos plus nettes grâce à un nombre accru de pixels par image. Simultanément, cette augmentation croissante de résolution a été accompagnée par l'introduction d'autres modalités comme l'augmentation du nombre d'images par seconde (HFR : High Frame Rate), ou la stéréoscopie (3D) pour laquelle des images différentes sont adressées à chaque oeil du spectateur. Une autre tendance récente est l'imagerie à haute dynamique (HDR : High Dynamic Range) et à gamme de couleurs étendues (WCG : Wide Color Gamut), qui permet respectivement de reproduire des images avec des niveaux de luminosité à la fois plus sombres et plus clairs, et des couleurs plus intenses qu'avec les écrans classiques. Toutes ces évolutions bénéficient à l'utilisateur final en terme d'immersion et de qualité d'expérience. Cependant, elles augmentent drastiquement la taille des contenus et nécessitent donc des capacités de traitement et de compression accrues.

Actuellement, la plus efficace des normes de compression et récemment standardisée, en 2013, est HEVC/H.265. Elle a été développée pour remplacer l'ancienne norme de compression AVC/H.264 en améliorant ses performances de compression par un facteur 2. Cependant, comme expliqué précédemment et en raison de l'exigence croissante des performances de compression, l'effort de normalisation est toujours très actif et un nouveau standard est en cours de développement : VVC/H.266. Toutes ces normes sont conçues avec les mêmes principes permettant de réduire les redondances spatiales et temporelles, et elles peuvent facilement être utilisées sur des contenus de différentes résolutions ou

fréquences d'images. Par contre, pour d'autres modalités comme la 3D ou le HDR, ces standards ne sont pas directement adaptables et des outils dédiés sont nécessaires. Ainsi, les contenus 3D ou HDR sont, par défaut, non compatibles avec les décodeurs traditionnels et donc avec les télévisions traditionnelles.

En revanche, les méthodes de compression rétro compatibles sont spécifiquement conçues pour répondre à ce problème. Cette méthode a été utilisée pour la première fois lors de la transition du noir et blanc vers la couleur, de telle façon que les contenus couleurs compressés étaient décodables par les télévisions noir et blanc traditionnelles. Une partie du signal est extraite pour reconstruire le contenu noir et blanc et l'afficher. De la même manière, des schémas de compression spécifiques peuvent être utilisés pour la compression de contenus HDR ou autres. Ces contenus HDR sont alors décodables par les nouveaux décodeurs HDR mais aussi par les décodeurs traditionnels dit SDR (Standard Dynamic Range).

# Motivations

Comme introduit dans la section précédente, l'imagerie à haute dynamique permet d'afficher des niveaux de lumière à la fois plus sombres et plus intenses que les écrans actuels. Cette quantité de lumière est généralement mesurée en unité de Luminance exprimée en candela par mètre carré ($cd.m^{-2}$ ou nits). En revanche, la dynamique de luminance est mesurée avec différentes unités. Dans tous les cas, cette dynamique est caractérisée par les valeurs minimale $L_{min}$ et maximale $L_{max}$ de luminance perceptibles. Elle peut être décrite en unités de contraste $(L_{max}/L_{min})$, en f-stops $(log_2(L_{max}) - log_2(L_{min}))$ ou en ordre de magnitude $(log_{10}(L_{max}) - log_{10}(L_{min}))$. Dans la suite de cette thèse, on utilise principalement les ordres de magnitude par souci de cohérence et clarté.

L'ultime but de l'imagerie HDR est de capturer et reproduire toute la dynamique de lumière visible et perceptible par l'oeil humain. Comme illustré en Figure 1, cette dynamique va de $10^{-6}$ à $10^{8}$ nits, pour un total de 14 ordres de magnitude de luminance. Cependant, l'oeil humain est capable de percevoir toute cette gamme de luminance uniquement grâce à plusieurs mécanismes d'adaptation. Sans ces adaptations, il peut instantanément traiter 5 ordres de magnitude de luminance dans la plage $[10^{-6}\,;\,10^{8}]$ nits.

En revanche, comme décrit en Figure 1, les appareils de capture et écrans SDR peuvent traiter uniquement 3 ordres de magnitude de luminance. Il est possible de créer des images HDR en utilisant la synthèse d'images par ordinateur ou en combinant plusieurs appareils de capture SDR avec différents niveaux d'exposition. Par contre, pour être affiché sur des écrans SDR, ces contenus HDR doivent être convertis avec des algorithmes spécifiques appelés TMO (Tone Mapping Operator). De nombreux TMO ont été développés ces dernières années avec différentes intentions.

Plus récemment, plusieurs écrans HDR ont été proposés, avec un minimum de 5 ordres de magnitude de luminance, pour correspondre au minimum à la vision instantanée de l'oeil humain. Ainsi, ils offrent une expérience de visionnage améliorée en comparaison des écrans SDR. Ces écrans HDR ouvrent aussi la possibilité de diffuser des contenus HDR au grand public, à condition de résoudre le problème de la compression HDR. En effet, les données HDR brutes sont généralement représentées avec des données flottantes et ils utilisent ainsi jusqu'à 4 fois plus d'espace de stockage que les données SDR non

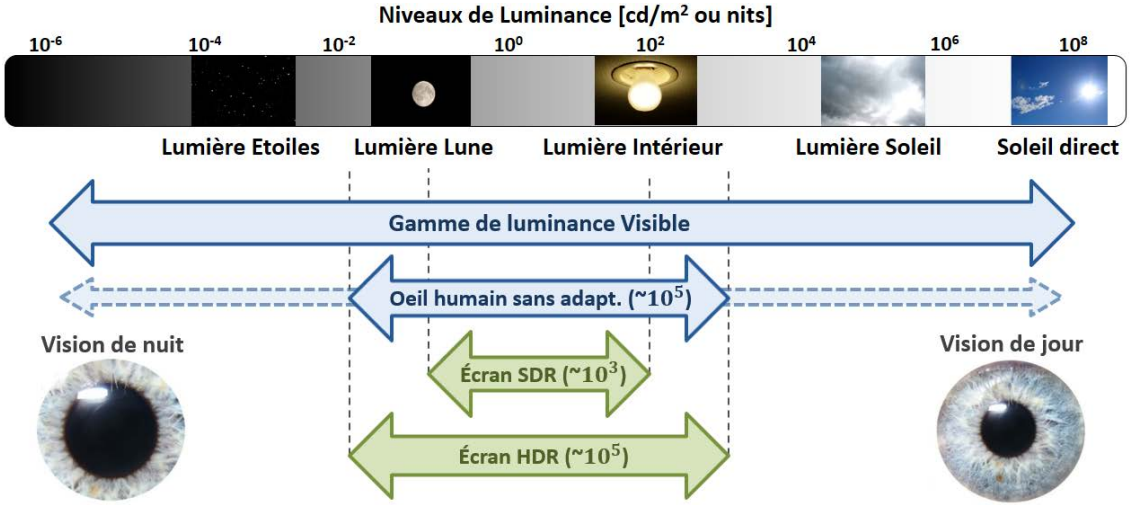



FIGURE 1 – Sources de lumières et niveaux de luminances associés - Dynamique de l'oeil humain, des écrans SDR et HDR

compressées.

Pour gérer le problème de compression des contenus HDR, un groupe de normalisation MPEG a été créé. Plusieurs solutions ont été proposées et celle qui a été sélectionnée consiste à utiliser des étapes de pré/post traitement associées au standard de compression traditionnels comme HEVC. Cette solution offre de bonnes performances de compression et est compatible avec les nouveaux écrans HDR. Cependant, la majorité des utilisateurs est actuellement équipé de télévisions classiques SDR qui ne peuvent pas interpréter ces contenus compressés HDR. Les utilisateurs renouvellent lentement leurs télévisions pour une HDR et il est donc très important de pouvoir délivrer un signal HDR qui puisse être décodé par les télévisions SDR et HDR.

Pour ce faire, la plus simple manière serait de créer un contenu SDR avec un TMO puis d'envoyer chaque contenu dans des bitstreams différents. Cette solution, appelée simulcast, est clairement sous optimale en terme de performance de compression, car beaucoup d'information redondante est envoyée dans les deux bitstreams. Comme vu dans la section précédente, une meilleure alternative est l'utilisation d'un schéma de compression rétro compatible ou scalable qui encode le contenu HDR en un seul bitstream compatible avec les télévisions HDR et SDR. Pour cela, le contenu HDR est premièrement tone mappé, avec un TMO en une version SDR, elle est encodée avec un encodeur traditionnel comme HEVC, puis envoyée avec des métadonnées. Ce signal peut être décodé par les télévisions classiques ou il peut être inverse tone mappé, à l'aide des métadonnées, pour obtenir une reconstruction du contenu HDR. Il est aussi possible d'envoyer un signal de raffinement contenant le résidu HDR pour améliorer la reconstruction du contenu HDR.

# Objectifs

L'objectif de cette thèse est de proposer de nouvelles solutions pour améliorer la compression HDR rétro compatible. En effet, le TMO est un élément essentiel pour permettre la rétro compatibilité, mais il peut être conçu avec différentes intentions. Dans un pre-

mier temps, les TMO étaient conçus pour préserver l'intention artistique du contenu HDR dans le contenu SDR obtenu. De nombreuses solutions ont été développées dans ce but mais dans le contexte de la compression, les TMO peuvent être aussi conçu pour optimiser les performances débit distorsion. Certains pourraient opposer qu'optimiser les TMO uniquement dans un but de compression résulte en un contenu SDR qui ne préserve pas l'intention artistique du contenu HDR. Ainsi, il serait intéressant de faire un compromis entre performances de compression pour le contenu HDR et qualité perceptuelle du contenu SDR.

Les contributions de cette thèse traitent principalement les deux derniers types de TMO. Elles utilisent des schémas de compression avec différentes configurations et proposent de nouvelles solutions pour répondre à ces deux enjeux. Premièrement, nous considérons un schéma de compression avec deux couches où l'on envoie une couche SDR puis une couche de raffinement contenant le résidu HDR. Dans une seconde contribution, on étudie un schéma de compression utilisant uniquement la couche de base SDR. Finalement, dans un troisième temps, on se concentre sur les aspects spécifiques à la compression vidéo. En effet, des problèmes inexistants dans la compression d'images fixes apparaissent quand on tone mappe des vidéos.

## Structure de la thèse

Ce manuscrit peut être divisé en deux grandes parties. La première traite de l'état de l'art en imagerie HDR dans le Chapitre 1 et de la compression vidéo dans le Chapitre 2, tandis que la seconde partie regroupe les contributions proposées dans les Chapitres 3, 4 and 5. Un bref résumé de chaque chapitre est donné ci-dessous :

**Chapitre 1 :** Ce chapitre fournit les informations générales et décrit les concepts de base pour comprendre l'imagerie HDR. Premièrement, on explique le système de mesure de la lumière avec la radiométrie et la photométrie, puis le système de mesure de la couleur avec la colorimétrie et les espaces de couleur. Grâce à ces fondamentaux, nous présentons les différentes méthodes pour créer du contenu HDR. Elles peuvent être classifiées en deux groupes principaux : génération assistée par ordinateur ou acquisition à l'aide de différentes techniques. Ces contenus HDR sont alors généralement représentés avec des valeurs flottantes, puis stockés avec des formats d'image dédiés pour un stockage plus efficace. Plusieurs formats d'image HDR ont été proposés dans la littérature et sont décrits dans ce chapitre. Finalement, nous présentons les différentes façons d'afficher ces contenus HDR. Ils peuvent être affichés sur les écrans classiques SDR à l'aide de TMO, ou bien sur les nouveaux écrans HDR récemment introduits. Ces derniers utilisent différentes nouveautés technologiques ayant chacunes leurs avantages et inconvénients.

**Chapitre 2 :** Bien que les formats d'images HDR présentés dans le chapitre 1 permettent de réduire significativement, en taille, la représentation des contenus HDR, ils ne sont pas suffisants pour des applications de distribution à grande échelle. La compression de ces contenus est nécessaire pour réduire drastiquement leur taille et s'adapter aux capacités des réseaux de distribution. Dans ce chapitre, on commence par présenter les principes essentiels de la compression d'image et de vidéo SDR. Ils consistent à supprimer les

redondances et l'information imperceptible du contenu, à travers des étapes de prédiction, transformation, quantification et de codage entropique. Ensuite, nous détaillons la norme de compression la plus récente HEVC/H.265 avec un résumé de chacun des principaux outils. Tous les standards de compression SDR peuvent être utilisés pour la compression SDR avec un pré/post traitement adapté. Les méthodes de compression HDR classiques et rétro compatibles sont plus précisément décrites dans ce chapitre.

A cause de l'étape de quantification additionnelle, la compression peut générer des artefacts visibles sur le contenu décodé. Moins un standard ou un algorithme de compression introduit des artefacts, plus il est considéré comme efficace en terme de compression. Ainsi, pour pouvoir comparer les algorithmes/standards entre eux, une étape d'évaluation du contenu décodé est nécessaire. Ce chapitre présente plusieurs métriques de qualité HDR et SDR couramment utilisées pour cette étape d'évaluation.

**Chapitre 3 :** Dans ce premier chapitre de contribution, on étudie un schéma de compression rétro compatible où deux couches sont envoyées pour reconstruire le contenu HDR. Une première couche de base avec le contenu SDR et les métadonnées et une couche de raffinement contenant le résidu. Ce schéma de compression est conçu de manière à améliorer les performances de compression de la couche de raffinement. En effet, ce problème est assez peu traité dans la littérature, surtout pour les TMO classiques conçus pour le rendu artistique du contenu SDR. Dans un second temps, nous proposons un TMO minimisant la distorsion du signal HDR sous une contrainte de débit total pour les deux couches, tout en préservant une bonne qualité perceptuelle/artistique pour le contenu SDR. Les modèles proposés pour la distorsion HDR et la qualité SDR sont conçus de façon à prendre en compte le débit des deux couches et sont ainsi plus optimaux que les méthodes de l'état de l'art utilisant des simplifications ou hypothèses grossières. Cette modélisation précise nous permet de déterminer que minimiser la distorsion HDR dans un schéma de compression à deux couches est équivalent à, premièrement, une optimisation du TMO de façon à minimiser la distorsion HDR dans un schéma de compression simple couche et, deuxièmement, une étape d'allocation de débit pour trouver la répartition optimale entre la couche de base et celle de raffinement. La première étape, l'optimisation du TMO, est résolue grâce à une paramétrisation de la courbe de TMO par des courbes linéaires par morceaux. Cette paramétrisation permet aussi de définir le TMO avec très peu de paramètres et ainsi de réduire la taille des métadonnées envoyées avec la couche de base. Les résultats expérimentaux sont d'abord analysés après la première étape d'optimisation dans un contexte simple couche, puis, après la seconde étape d'optimisation dans un contexte double couche. Dans les deux cas, la solution proposée fournit le meilleur compromis entre performances de compression et préservation de la qualité du signal SDR.

**Chapitre 4 :** Après avoir démontré dans le chapitre 3 qu'optimiser un schéma de compression double couche est équivalent à optimiser un schéma de compression simple couche, puis à réaliser une optimisation de l'allocation de débit entre les deux couches, le chapitre 4 se concentre sur les schémas de compression simple couche. Comme pour le chapitre précédent, nous essayons de minimiser la distorsion HDR sous une contrainte de débit tout en préservant la qualité perceptuelle du contenu SDR. Cependant, nous concevons aussi un TMO qui optimise uniquement les performances de compression. Dans ce cas, le problème de minimisation est très similaire. Il suffit simplement d'enlever la

contrainte sur la qualité perceptuelle SDR. La principale nouveauté de ce chapitre est la proposition de nouveaux modèles améliorés pour la distorsion HDR, le débit SDR et la qualité perceptuelle SDR. Plus précisement, les modèles de débit SDR et de distorsion HDR sont basés sur le gradient de l'image HDR. La précision supérieure de ces modèles, par rapport à ceux de l'état de l'art, est démontrée sur plusieurs expériences. Les deux mêmes problèmes d'optimisation que ceux traités dans le chapitre 3 sont traités grâce à une paramétrisation efficace en fonctions linéaires par morceaux. Les résultats expérimentaux sont présentés pour les deux problèmes de minimisation. Dans le premier cas, où l'on optimise uniquement les performances de compression, le TMO proposé fournit de meilleures performances débit-distorsion que les méthodes de l'état de l'art. Pour la seconde minimisation, avec la contrainte de qualité SDR, le TMO proposé fournit toujours le meilleur compromis entre performances débit-distorsion et préservation de la qualité du contenu SDR. Comme on pouvait le supposer, les performances débit-distorsion du premier TMO sont toujours meilleures que celles du second TMO.

**Chapitre 5 :** Les problèmes d'optimisation proposés dans les chapitres 3 et 4 sont conçus avec des statistiques d'image fixe et ainsi, les TMO obtenus sont testés sur des images fixes. Il est facile d'étendre ces solutions à la vidéo en calculant un TMO pour chaque image de la séquence vidéo. Cependant, tone mapper indépendamment chaque frame d'une vidéo HDR peut entraîner d'importants changements temporels lumineux dans la vidéo SDR. Cette particularité a certes un effet en terme de qualité perceptuelle du contenu SDR, mais elle impacte également fortement les performances de compression, car elle pénalise l'étape de prédiction de l'encodeur SDR. Dans ce chapitre, on se concentre sur cet aspect temporel. Grâce à un état de l'art, nous concluons que la solution optimale est de proposer un nouvel algorithme de Weighted Prediction (WP) et de nouvelles fonctions de WP au sein de l'encodeur HEVC, pour prédire plus précisément les variations lumineuses de la vidéo SDR. En effet, certaines solutions proposent d'adapter directement le TMO pour réduire les variations lumineuses du contenu SDR mais cela dégrade fortement sa cohérence temporelle. Le nouvel algorithme de WP est basé sur une compensation de mouvement par flow optique pour faire correspondre les images consécutives et ainsi identifier facilement les variations lumineuses. Pour compenser ces variations, le standard HEVC est actuellement limité à une WP globale linéaire. Dans ce chapitre, on introduit des fonctions non linéaires globales et locales paramétrisées avec des courbes linéaires par morceaux. Ces nouvelles fonctions de WP permettent de réaliser une meilleure prédiction des variations lumineuses et grâce à la paramétrisation, elles ajoutent un faible excès de débit. La contribution de chaque nouveauté proposée est analysée indépendamment dans les résultats expérimentaux, mais nous étudions aussi une mise en compétition entre les différentes fonctions de WP. Dans tous les cas, l'algorithme proposé et les fonctions de WP améliorent les performances de compression.

# Introduction

## Preamble and Context

Video content distribution tends to become universal. Video diffusion, initially limited to home television set, is now available for computer, tablets and smartphones. Broadcast networks such as terrestrial transmission, cable, satellite, mobile phone and internet are all exploited to transmit image and video content. All of these evolutions benefit the end-user in terms of accessibility. However, the amount of transmitted data has grown enormously over these last years, with the advent of platform such as YouTube or Netflix, and continues to grow. Despite improved broadcast capacities, this increasing amount of content requires increasingly efficient compression methods.

In parallel with networks and devices evolutions, the quality and immersion requirements keep on increasing. Indeed, these last years, video resolutions went from Standard Definition to High Definition (HD) and more recently to Ultra High Definition (UHD). Each of these new resolutions increases the number of pixels per image and thus produces sharper videos. The increasing resolution quality was simultaneously accompanied by the evolution of other modalities such as Higher Frame Rate (HFR), where videos are displayed with an increased number of images per second, or the stereoscopy (3D), where different images are addressed for each eye. The recent trend is now towards High Dynamic range (HDR) and Wide Color Gamut (WCG) imaging which respectively allow the reproduction of much brighter and darker images with more saturated colors. All of these evolutions benefit the end-user in terms of quality of experience and realism however, they also increase the content size and therefore require increased processing and compression capabilities.

The currently most efficient and latest released compression Standard, in 2013, is HEVC/H.265. It was designed to replace the older compression standard AVC/H.264 by improving the compression performances by two. However, as explained above and due to the increasing requirement for more efficient compression standards, the standardization effort is still pursued and a new one is currently under development: VVC/H.266. All these standards are designed with the same principles to reduce the temporal and spatial redundancies and they can easily be adapted to different resolutions or frame rates. However, for other modalities such as 3D or HDR, these standards are not directly adaptable and dedicated tools are required. In this way, such content is, by default, not compatible with standard legacy decoders and therefore with legacy displays.

On the other hand, backward compatible compression methods are especially designed to address such specific problem. This compression type was first used for the transition from black and white to color, such that colored compressed contents were decodable by

legacy black and white televisions. A part of the signal is extracted to reconstruct a black and white content. In the same way, specific compression schemes can also be used for HDR content compression, or others modalities. HDR contents can then be decoded by the new HDR decoders but also by legacy SDR (Standard Dynamic Range) ones.

# Motivations

As introduced in the previous section, High Dynamic Range imaging allows the reproduction of much brighter and darker lights than Standard Dynamic Range imaging. The amount of light is measured with Luminance unit in candela per square meter ($cd.m^{-2}$ or nits), however, the dynamic range is measured with different units. In all cases, it is characterized by the minimum $L_{min}$ and maximum $L_{max}$ perceivable luminance values. It can be described with contrast units ($L_{max}/L_{min}$), in f-stops ($log_2(L_{max}) - log_2(L_{min})$) or in orders of magnitude ($log_{10}(L_{max}) - log_{10}(L_{min})$). In the following, we mainly use the orders of magnitude unit for sake of clarity.

The ultimate aim of HDR imaging is to capture and reproduce all the dynamic range of the visible light perceivable by the human eye. As illustrated in Figure 2, this dynamic range goes from $10^{-6}$ to $10^{8}$ nits, thus providing 14 orders of luminance magnitude. However, the human is only able to perceive all this dynamic through several eye adaptation mechanisms. Without these adaptations, the human eye can instantaneously perceive 5 orders of luminance magnitude, if it belongs to the luminance range $[10^{-6}; 10^{8}]$ nits.

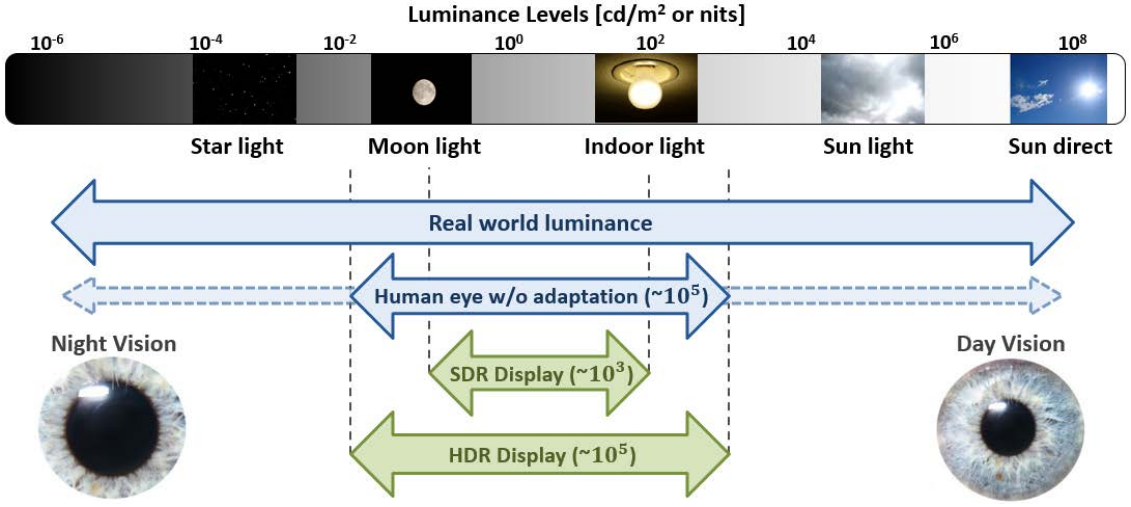



Figure 2 – Light sources and associated luminance levels - Dynamic ranges of the human eye, SDR displays and HDR displays

In contrast, as depicted in Figure 2, SDR capture devices and SDR displays can only process 3 orders of luminance magnitude. It is possible to create HDR images using computer graphics or SDR capture devices by combining multiple SDR images with different camera exposures. However, to be displayed on SDR devices, these HDR images need to be converted with specific algorithms called Tone Mapping Operator (TMO). Numerous TMO have been developed these last years, with different purposes.

More recently, several HDR displays were proposed, with a minimum of 5 orders of luminance magnitude, to at least match the instantaneous human eye vision. Therefore, they offer an enhanced viewing experience compared to SDR displays. HDR displays also open the possibility to broadcast HDR content to the general public, if the problem of HDR content compression is solved. Indeed, HDR raw data are represented with floating point numbers and therefore they use 4 times more space than uncompressed SDR data.

To address the compression problem of HDR content, an MPEG normalization group was created. Several solutions were proposed and the selected one consists in using pre or post processing associated with legacy SDR compression standard such as HEVC. This solution proposes good compression performances and can address the new HDR displays. However, the majority of consumers are currently equipped with SDR displays that cannot handle HDR compressed content. Consumers will slowly renew their display to an HDR one and it is therefore of great importance to deliver an HDR signal that can be decoded by both SDR and HDR displays.

The most simple way to address both HDR and SDR displays with the same HDR content would be to create the SDR content with a TMO then send each content with separate bitstreams. This solution called simulcast is clearly not optimal in terms of compression performances since redundant information is sent in both bitstreams. As introduced in the previous section, a better alternative is to use backward compatible or scalable compression, which encodes an HDR content into a single bitstream that can be decoded by both SDR and HDR displays. To do so, the HDR input is first tone mapped into an SDR version which is encoded with a legacy codec such as HEVC and sent along with metadata. This signal can be decoded with legacy SDR display or it can be inverse tone mapped to yield an HDR reconstructed content. One can also send an enhancement layer containing the HDR residue to improve the reconstruction of the HDR content.

# Objectives

The aim of this thesis is to propose new solutions for improving the backward compatible HDR compression. As explained in the previous section, the TMO is an essential component for backward compatibility, however it can be designed for different purposes. At first, TMO were mostly designed to preserve the HDR artistic intent in the resulting SDR content. Many TMOs have been developed for this purpose but in the context of compression, TMOs can also be designed to optimize the compression performances. Some might argue that optimizing the TMO only for compression purposes may yield SDR content which does not preserve the artistic intent of the HDR content. Therefore, it might be interesting to trade off between compression performance of the HDR content and perceptual quality of the SDR one.

The contributions in this thesis mainly address these two last use cases. They use different compression scheme configurations and propose new solutions to answer these challenges. First, we considered dual-layer compression schemes where both a base SDR layer and a residual layer are sent. In a second contribution, we studied compression schemes using only the base layer. And finally, we focused on the specific aspects of video compression. Indeed, problems not present in still image compression arise when we tone mapped videos.

## Thesis structure

This manuscript can be divided in two main parts. The first one deals with the state of the art in HDR imaging in Chapter 1 and video compression in Chapter 2, while the second part gathers our contributions in Chapters 3, 4 and 5. A brief overview of each chapter is outlined below.

**Chapter 1 :** This chapter provides background information and describes the basic concepts to understand HDR imaging. First, we review the science of measuring light with radiometry and photometry, then the science of measuring color with colorimetry and color spaces. Thanks to these fundamentals, we present the different methods to create HDR content. They can be classified in two main groups: generation with computer graphics or acquisition using several techniques. These HDR contents are usually represented with floating point values and then stored with dedicated image formats for more efficient storage. Several formats were proposed in the literature and are described in this chapter. Finally, we present the different ways to display these HDR contents. They can be displayed on SDR devices using TMO or on newly introduced HDR devices based on different technological innovations.

**Chapter 2 :** While the HDR image formats presented in the previous chapter successfully reduce the representation size of HDR content, they are not sufficient for broadcast applications. Content compression is required to drastically reduce its size and fit the capacity of distribution networks. In this chapter, we start by reviewing the main concepts of SDR image/video compression. It consists in removing the redundant and imperceptible information of the content through prediction, transform, quantization and entropy coding. Then, we further detail the most recent compression standard HEVC/H.265 with an overview of its main tools. All these SDR compression standards can be used for HDR compression with adapted pre and post processing. HDR only and backward compatible compression methods are further explained in this chapter.

Due to the additional quantization step, compression can generate visible artifacts on the decoded content. The less a compression standard or algorithm introduces artifacts, the more it is efficient. Therefore, to compare algorithms/standards between each other, a quality evaluation is required. This chapter presents several SDR and HDR metrics used to efficiently perform the quality evaluation.

**Chapter 3 :** In this first chapter of contribution, we study a backward compatible compression scheme where both a base and an enhancement layer are sent to reconstruct the HDR content. This compression scheme is designed in a specific way to improve the compression of the enhancement layer. Indeed, this subject is not often discussed in the literature, especially with classical TMOs intended for artistic purposes. In a second time, we propose a TMO minimizing the HDR distortion under a total rate constraint for both layers while preserving a good artistic/perceptual quality for the SDR content. The proposed models for the HDR distortion and the SDR quality are designed to account for both layer rates and therefore are more optimal than the state of the art ones using additional assumptions or simplifications. This precise modelization allows us to determine that minimizing the HDR distortion in a dual layer compression scheme is equivalent

to, first, a TMO optimization minimizing the HDR distortion in a single layer compression scheme and, secondly, a rate allocation step to find the optimal repartition between the base and enhancement layers. The first step, the TMO optimization is successfully solved thanks to a parameterization of the TMO curve by a piecewise linear function. This parameterization also allows to defines the TMO with few parameters and therefore allows to reduce the size of the metadata sent with the base layer. Experimental results are analyzed after the first optimization step in a single layer context and after the second optimization in a dual layer context. In both cases, the proposed solution provides the best trade-off between compression performances and quality preservation of the SDR content.

**Chapter 4 :** Since we demonstrated in Chapter 3 that optimizing a dual layer compression scheme is equivalent to optimize a single layer compression scheme with a rate allocation optimization step, Chapter 4 focus on single layer compression schemes. As for the previous chapter, we attempt to minimize the HDR distortion under a rate constraint while preserving the SDR perceptual quality. However, we also design a TMO only for compression performances. In this case, the minimization problem is very similar. It simply consists to remove the constraint on the SDR perceptual quality. The main novelty of this chapter is the proposal of improved models for the HDR distortion, the SDR rate, and the SDR perceptual quality. Especially for the SDR rate and the HDR distortion since these models are based on the gradient of the HDR image. Their superior reliability, than state of the art methods, to the real data is well demonstrated with several experiments. As for Chapter 3, the two optimization problems are solved thanks to the efficient parameterization in piecewise linear functions. Experimental results are presented for the two minimization problems. In the first case, where we only optimize the compression performances, the proposed TMO provides the optimal rate distortion performances in comparison with state of the art methods. For the second minimization, with the SDR quality constraint, the proposed TMO always provides the best tradeoff between rate distortion performances and quality preservation of the SDR content. As easily deductible, the rate distortion performances of the first TMO always outperforms the rate distortion performances of the second TMO.

**Chapter 5 :** The optimization problems proposed in Chapter 3 and 4 are computed with still image statistics and therefore the obtained TMOs are tested on still images. These solutions are easily extendable to video by computing a TMO for each frame of the video sequence. However, tone mapped each frame of an HDR video independently can result in important temporal brightness variations in the SDR video. This particularity can be discussed in terms of perceptual quality of the SDR content, however it is not optimal for the compression performances since it penalize the prediction step of the encoder. In chapter 5, we focus on this temporal aspect. Based on the literature, we conclude that the optimal solution is to propose a new Weighted Prediction (WP) algorithm and new WP functions inside the HEVC encoder to more accurately predict the SDR brightness variations. Indeed some solutions proposed to adapt the TMO to reduce the brightness variations in the SDR content, however they strongly degrade its temporal coherency. The new Weighted Prediction algorithm is based on an optical flow motion compensation to accurately match the consecutive frames and so easily identify the brightness variations. To compensate these variations, the HEVC standard is currently limited to a global linear

WP. In this chapter, we introduce global and local non linear functions parameterized with piecewise linear curves. These new WP functions allows for a finer prediction of the brightness variations and thanks to the parameterisation they results in a small rate excess. The contribution of each proposed novelty is analyzed independently in the experimental results but we also study a competition between the different WP functions. In all cases, the proposed WP algorithm and functions improves the compression performances.

# Chapter 1

# High Dynamic Range Imaging Overview

This chapter provides background information on High Dynamic Range (HDR) and forms the basis of following chapters. It describes the fundamentals concepts to understand HDR imaging.

The main aspect of HDR is to process a much larger scale of luminance than Standard Dynamic Range (SDR). Improving the dynamic range consists to increase the range between the minimum and maximum luminance values. Indeed, SDR displays only reproduce 3 orders of luminance magnitude ($10^3$) although the average human eye can instantaneously see around 5 orders of luminance magnitude ($10^5$) and even more using the pupil adaptation, around ($10^{14}$) [BADC17]. Thus, HDR imaging allows to better represent the wide luminance range of real-world scenes.

The HDR pipeline is described Fig 1.1 and exhibits the many challenges of HDR imaging [Man06]. The aim is to acquire, store and reproduce all physical values of light [RWP+10, MR11]. Indeed, the field of HDR imaging encompasses various fields related to light such as radiometry, photometry or colorimetry [RWP+10]. Each of them deals with a specific aspect of light or its perception by the average human eye [RKAJ08]. In Section 1.1, we present the fundamental aspects of light and color and explain how the physical values are related to the digital ones.

Concerning the acquisition, there are two main ways to produce HDR contents. It is either created by computer-generated imagery (CGI) or captured with HDR cameras or complex set of legacy SDR cameras. In practice, capturing the perceivable wide luminance range of real-world scenes (14 orders of luminance magnitude) faces some technical limitations. These aspects are further detailed in Section 1.2.

Most legacy SDR image representations generally use 3 components with 8 bits integer value to represent a pixel. For HDR imaging, this 8 bits-per-component representation is deprecated and the physical light values can be stored using a floating-point representation with up to 32 bits per component [RWP+10]. For efficient storage and manipulation, several HDR image formats have been proposed in the literature. They are described in Section 1.3.

These HDR image formats can reduce the representation of HDR contents to 10 bits-per-component, however this reduction is not enough and incompatible with efficient distribution or storage. Contents still need to be more efficiently compressed. To do so, you

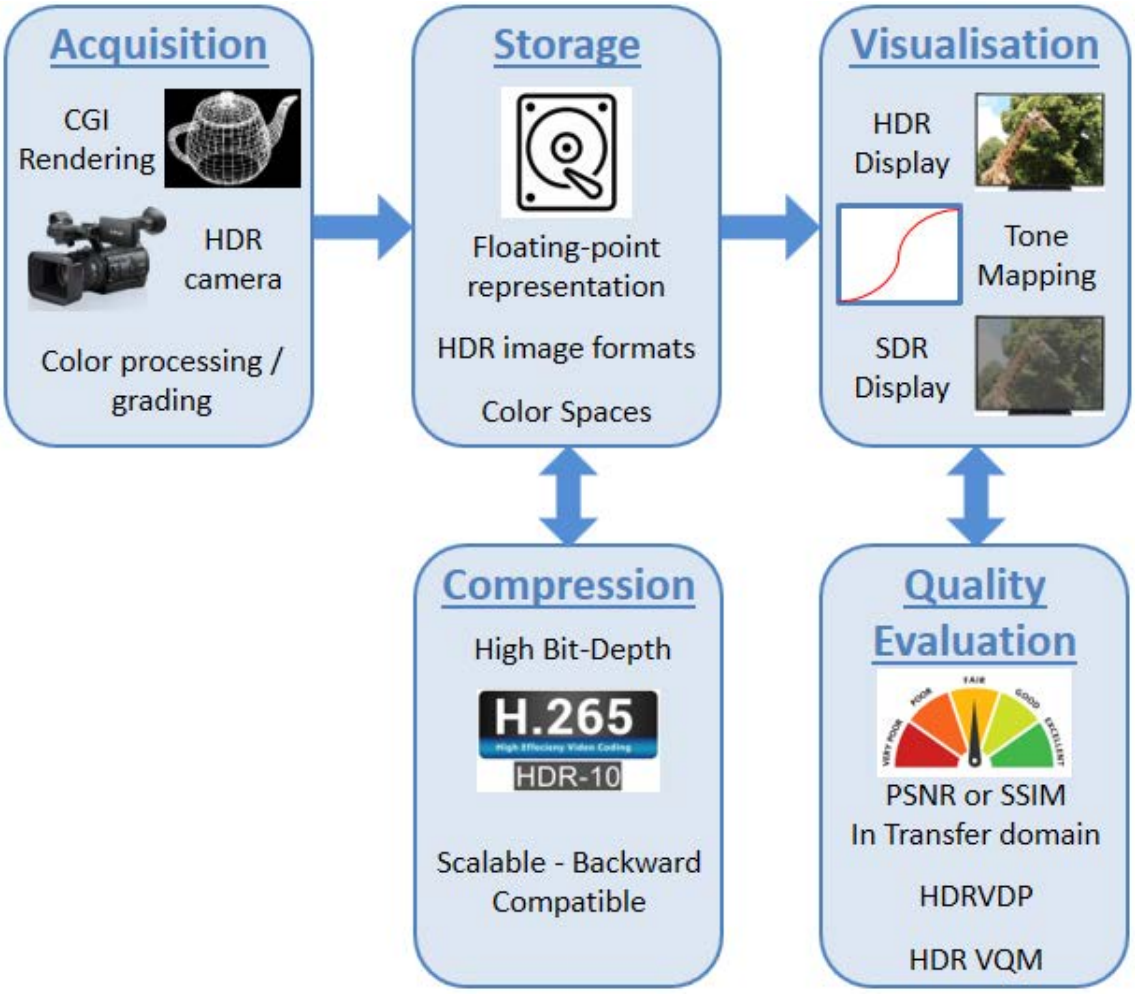



Figure 1.1 – High Dynamic Range (HDR) pipeline

can either use legacy SDR codecs with adequate pre and post processing or you can use scalable approaches. Indeed, scalable solutions allow backward compatibility with legacy SDR displays. This approach is the main topic of this thesis and for this reason the different methods to compress HDR contents are thoroughly described in Chapter 2.

HDR displays aim at representing all the perceivable luminance range (14 orders of luminance magnitude) although current technical limitations prevent it. Instead HDR displays aim at providing a dynamic at least as good as the instantaneous human eye vision (5 orders of luminance magnitude). The different displays technologies achieving this luminance range are described in Section 1.4. However, HDR displays are relatively recent compared to the field of HDR imaging. For a long time, the display of HDR contents was made on legacy SDR displays through algorithms called Tone Mapping Operator (TMO) [RWP+10]. These algorithms allow a compatibility between HDR contents and SDR displays. For this reason, Section 1.4 also describes these TMOs.

Finally, the quality of the displayed content can be assessed using several new HDR error metrics. However, these metrics mostly measure the error introduced by the compression algorithms. Therefore, the HDR metrics are described with the compression methods

in Chapter 2.

## 1.1 Light and Color - Fundamentals

### 1.1.1 Light - Radiometry

The field of Radiometry consists in measuring electromagnetic radiations. Electromagnetic radiations consist of different electromagnetic waves propagating through space and time carrying some energy, called radiant energy. They are released in the space by stars such as the Sun. Electromagnetic radiations takes many forms and are characterized by their wavelength as described in Figure 1.2. Examples of these radiations, ordered by increasing wavelength are: Gamma-rays, X-rays, UltraViolet (UV) rays, Light or Visible Light, InfraRed (IR), Microwaves and Radio waves. We refer as light, the electromag-

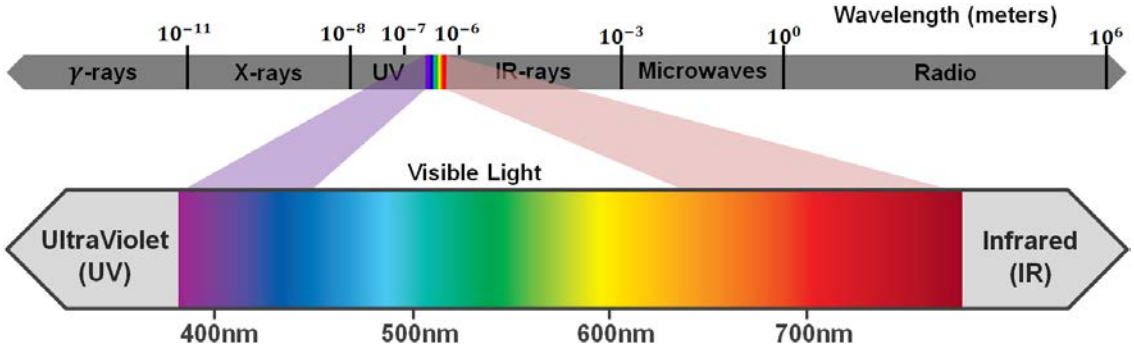



Figure 1.2 – Electromagnetic Spectrum and Visible Light Spectrum (wavelengths going from 380nm to 780nm)

netic radiations capable of producing a visual sensation for the human eye. Indeed, IR, UV rays or others are not visible. The wavelength of visible light approximately ranges from 380nm (nanometer) to 780nm [RKAJ08].

Electromagnetic radiations or radiant energy, and therefore light, is measured in Joule (J) in the International System of Units (SI). In real-world scenes, light propagates through different environments (space, air, water, ...) and interacts with different materials that either absorb, refract, reflect or transmit the light [BADC17]. To measure how light or radiant energy propagates, the "Commission Internationale de l'Éclairage" (CIE - International Commission of Lighting) standardized different quantities measuring radiant energy integrated over time, space, or angle. There are many of these quantities but the main ones are the Radiant Intensity and the Radiance. The Radiant Intensity measures the amount of light coming from different directions and converging on the same infinitesimal small point. It measures light per amount of time and per unit of direction, and is expressed in Watts per steradian ($W.sr^{-1}$). The Radiance measures the amount of light emitted/received on a point in a specific direction. It measures light per amount of time, per unit of area and per unit of direction, and is expressed in Watts per square meter per steradian ($W.m^{-2}.sr^{-1}$).

### 1.1.2 Human Vision - Photometry

While Radiometry measures physical absolute values, Photometry measures how light interact with the human visual system (HVS), or basically, the human eye. Thus, Photometry only focuses on electromagnetic radiations with wavelengths going from 380nm to 780nm (Figure 1.2).

The retina of the average human eye contains approximately 100 million light sensitive cells, named photoreceptors and divided in two types: the rods (around 90 million) and the cones (around 6 million) [BADC17]. The rods are the most light sensitive cells and therefore responsible for vision under low light conditions. They do not differentiate colors, which is why we are unable to distinguish colors at night. The cones require more light to be activated. They are responsible for color vision under normal to bright light conditions [BPNS15]. Human eye vision under low light conditions, using only the rods,

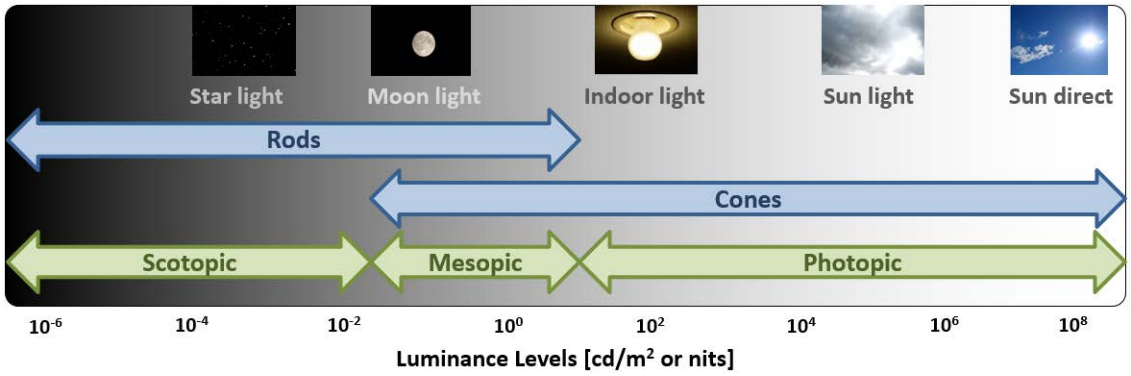



Figure 1.3 – Vision regimes and photoreceptor activation levels

is called the scotopic vision regime, while under normal-bright conditions, using only the cones, is called photopic vision regime. Between low light and normal light conditions, both cones and rods are active. The vision under these conditions is called the mesopic vision regime. Figure 1.3 illustrates these different vision regimes.

Within the visible light spectrum, these photoreceptors are not equally sensitive to all wavelengths. Stimuli with different spectral distributions but with the same radiance level provide different perception levels of brightness. This relative response of photoreceptors differs depending on their type (rod or cone). Figure 1.4 shows these relative responses called luminous efficiency functions or luminosity functions. $V(\lambda)$, the photopic luminosity function, was established in 1924 by the CIE and $V'(\lambda)$, the scotopic luminosity function, adopted by the CIE in 1951 [Poy12].

These functions are used to weight the previously defined quantities Radiant Intensity and Radiance into Luminous Intensity and, the widely used, Luminance. We do not measure a radiant power in Watts anymore but a photometrically weighted radiant power, or luminous power, in Lumens. Therefore, the Luminous Intensity is expressed in Lumen per steradian ($lm.sr^{-1}$), also named candela ($cd$), and the Luminance is expressed in candela per square meter ($cd.m^{-2}$), also named nits. Luminance is the most used photometric unit in HDR imaging since it is related to the human eye perception of brightness. Figure 1.3 illustrates the luminance levels of photoreceptors activation.

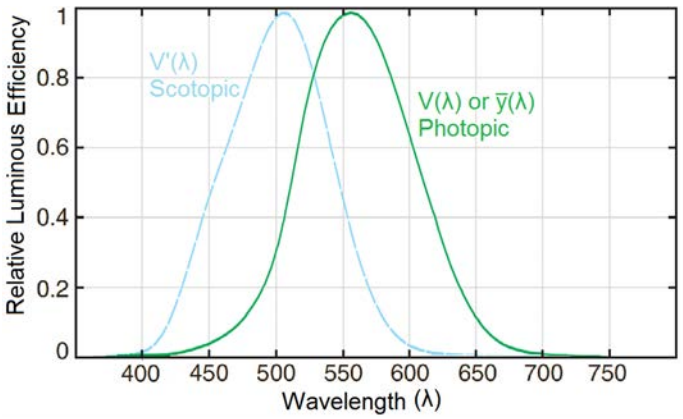



Figure 1.4 – Luminous efficiency functions

### 1.1.3 Luminance Perception Models

As stated in the introduction, the HVS is able to handle a huge range of luminance values, around 14 orders of magnitude from $10^{-6}$ to $10^{8}$ nits [BADC17]. Instantaneously, this range is reduced but still contains 4 to 5 orders of luminance magnitude [RWP+10, BADC17]. Section 1.1.2 described the response of the HVS to stimuli with different spectral distributions. In this section, we will now describe the HVS adaptation depending on the luminance level. While spectral sensitivity is only due to photoreceptor mechanisms, the HVS luminance adaptation is multifactorial (pupils dilation and contraction, rods and cones system, photoreceptor mechanisms and photo-chemical reactions). This thesis does not focus on every factor but rather on their common adaptation.

#### 1.1.3.1 Intensity-based models

The human senses (vision, hearing, smell, ...) do not perceive linearly the intensity variations of a stimulus. Adding a 1 kg mass to a 1 kg package or to a 10 kg package does not give the same sensation of mass variation. This observation was made by Weber in 1834 and generalized by Fechner in 1860 [Fec60]. Indeed, the same observation applies to brightness sensation, the perception of luminance is not linearly linked to the luminance intensity (in nits). If we define the intensity $I$ of a luminous stimuli and $P(I)$ the HVS perceived intensity, the Weber-Fechner law states that the differential change in perceived intensity $\Delta_{P(I)}$ between two luminous stimuli depends on their intensity $I$ [Hec24], more precisely:

$$\Delta_{P(I)} = k.\frac{\Delta_I}{I} \tag{1.1}$$

where $\Delta_I$ is the difference in intensity that generates a Just Noticeable Difference (JND) between the two stimuli and $k$ a constant. The ratio $\Delta_I/I$ is called the contrast [1] threshold or Weber ratio/fraction and is assumed constant for any value of $I$ (approximately 1% [RWP+10]). Integrating equation (1.1):

$$P(I) \;=\; k.\ln(I) + C \;=\; k.\ln(\frac{I}{I_0}) \tag{1.2}$$

1. The term contrast is ambiguous, it can define the difference in intensity $\Delta_I$ or the ratio $\Delta_I/I$

where $C$ is the constant of integration and $I_0$ the threshold intensity such that there is no perception, $P(I_0) = 0$. Equation (1.2) shows that the perceived luminous intensity is proportional to the logarithm of luminance intensity. This claim is still popular in HDR image/video processing. For example, many early HDR image formats are based on a logarithmic transformation (See Sec.1.3.2). However, Threshold vs. Intensity (TVI) experiments, detailed below, show that this law is only representative of the HVS for high luminance intensities (Photopic regime).

The TVI experiments consist in measuring the contrast threshold $\Delta_I/I$ for different luminance intensities using multiple observers. To do so, two adjacent patches with intensity $I$ and $I + \Delta_I$ are projected on a uniform background. The JND in intensity corresponds to the smallest $\Delta_I$ value providing a visible difference between the two patches. The contrast threshold $\Delta_I/I$ is then plotted as a function of the intensity $I$ thus providing a TVI function. Multiple TVI functions have been proposed [BB71, Bod73, FPSPS96], but the most popular one for SDR imaging is the power law.

Back in 1963, [SS63] shows that the Weber-Fechner law does not fit the experimental data and that a power law function (with exponent ranging from 1/3 to 1/2) is a better approximation. However, the power law is only representative of the HVS for normal luminance intensities (mesopic regime - from 1 to 400 nits [Man06]). Actually, it covers the luminance range of many SDR displays. This is one of the reasons why power law functions are preferred in SDR video compression. This aspect is further detailed in Section 1.3.

All these TVI functions do not take into account a feature of the HVS that is spatial frequency sensitivity, as explained in the following section.

#### 1.1.3.2 Intensity and Spatial Frequency based models

The HVS sensitivity to contrast not only depends on intensity but also changes with the spatial frequency. The JND will differ depending on the pattern used for the test. In TVI experiments, the pattern only consists in two adjacent patches, but in Constrast Sensivity Function (CSF) experiments the pattern used is sinusoidal, like in [VV72, Dal93, Bar99]. This pattern is illustrated in Figure 1.5(a). The contrast is constant along the horizontal axis, only the spatial frequency increases. If the HVS was not sensible to spatial frequency, the sensitivity curve should be a straight horizontal line. Figure 1.5(b) plots the sensitivity curves for different background luminance levels. The higher the luminance level, the higher the sensitivity to the frequency, even though a maximum seems to be reached at 1000 nits. A TVI function can be derived from the CSF by selecting one spatial frequency or, to have the highest sensitivities, by selecting the peaks of each CSF.

The luminance without TVI correction is usually refereed to as linear-light luminance. For TVI corrected luminance, several terms are used like uniform luminance, perceptually uniform luminance or brightness.

### 1.1.4 Colorimetry - Chromaticity

The field of colorimetry measures color and color appearances. It consists in providing physical values to colors perceived by the HVS. Under normal to bright light conditions, only cones are activated. The human eye has three types of cone cells, this is referred to

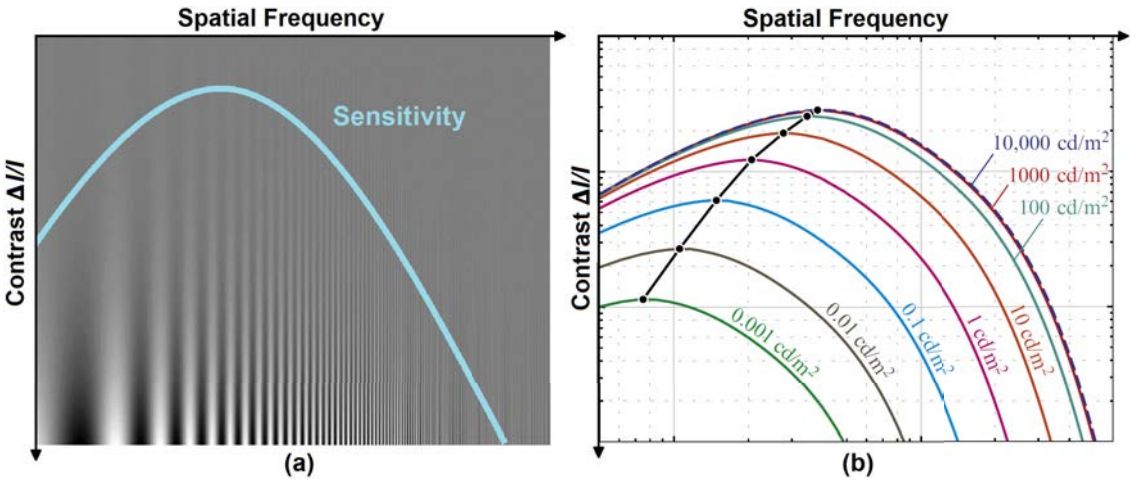



Figure 1.5 – Contrast Sensitivity Function - (a) Sinusoidal pattern - (b) Contrast sensitivity curves for different luminance adaptation levels

as trichromatic vision. Each cone type has different wavelength sensibilities, as illustrated in Figure 1.6, and their weighted combination allows the brain to interpret colors. The

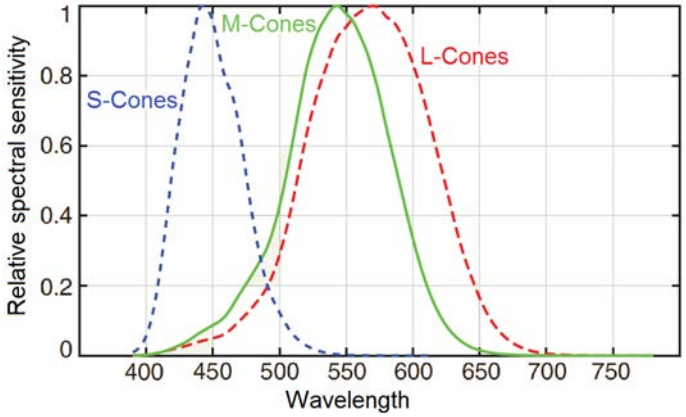



Figure 1.6 – Relative response of L, M and S cone cells to the different wavelengths

different types of cones are named as follows: S-Cones (Short wavelengths) with a peak sensitivity around 440nm, M-Cones (Medium wavelengths) with a peak sensivity around 545nm, and L-Cones (Long wavelengths) with a peak sensivity around 580nm [Hun05]. As an example, blue color is perceived by the human eye when the S cones are significantly more stimulated than the M or L cones. In the same way, yellow is perceived when L cones are stimulated slightly more than M cones and much more than the S cones.

Under low light conditions, rod cells are activated. The fact that we have only one type of rod cells prevents us to distinguish colors in the scotopic vision regime and thus explains our monochrome vision in low light conditions. Note that the function $V(\lambda)$ in Figure 1.4 represents the common response of the three L, M and S cone cells and that the response of the rod cells is in fact $V'(\lambda)$ since we only have one type of rod [Fai05]. It was very difficult to obtain the individual response of each type of cone since their wavelength sensibilites overlap a lot, as illustrated in Figure 1.6. Therefore it is impossible to stimulate

only one type of cone cell [RKAJ08]. For these reasons, the sensitivity of each cone type was measured rather recently.

#### 1.1.4.1 The CIE 1931 Color Spaces

The best way to describe the color perceived by the HVS is to use a color space. A color space is a defined range of colors were each color is described using, generally [2], three code values called components (or channels). Each of these values represents an amount of a specific color, called primary color. The use of three components is consistent with the physiology of the human eye (three types of cones) but it is based on experimental facts [Hun05]. Indeed, since the cones sensitivity was unknown, color matching experiments were made and they showed that almost each color $C$ may be reproduced using an additive weighted combination of three monochromatic primaries $\mathcal{R}$, $\mathcal{G}$ and $\mathcal{B}$ [Fai05]:

$$C = \mathrm{R}.\mathcal{R} + \mathrm{G}.\mathcal{G} + \mathrm{B}.\mathcal{B} \tag{1.3}$$

The values R, G and B represent the components of the color $C$ in the color space defined with the primaries $\mathcal{R}$, $\mathcal{G}$ and $\mathcal{B}$. Thus, each color is defined in a color space with a tristimulus value.

In 1931, the CIE standardized a set of primaries $\mathcal{R}$ (Red 700 nm), $\mathcal{G}$ (Green 546.1 nm) and $\mathcal{B}$ (Blue 435.6 nm) [SG31]. Using these primaries, it is possible to represent almost all colors of the visible light spectrum (See Fig.1.2). Figure 1.7(a) illustrates the set of tristimulus values needed to reproduce the monochromatic color at each wavelength $\lambda$. The contribution of each primary for the entire spectrum is represented by the functions

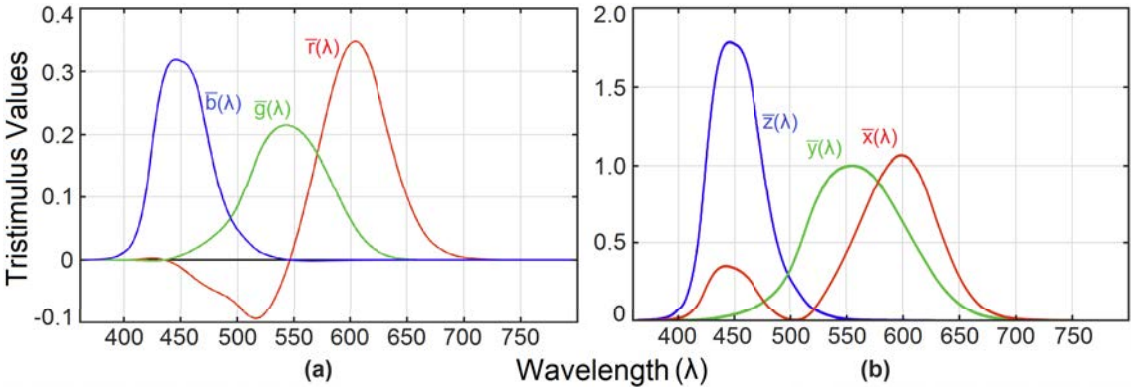



Figure 1.7 – CIE 1931 Standard - (a) RGB color matching functions - (b) XYZ color matching functions

$\bar{r}$, $\bar{g}$ and $\bar{b}$, also known as color matching functions. Using these functions, the tristimulus value of a spectral stimulus $C_\lambda$ can be computed with the following equations:

$$\mathrm{R} = \int_\lambda C_\lambda.\bar{r}(\lambda)d\lambda \qquad \mathrm{G} = \int_\lambda C_\lambda.\bar{g}(\lambda)d\lambda \qquad \mathrm{B} = \int_\lambda C_\lambda.\bar{b}(\lambda)d\lambda \tag{1.4}$$

2. Some color spaces use 4 and more components. For example, the CMYK color space, popular for printers, uses 4 components.

However, Figure 1.7(a) shows that, sometimes, a negative amount of primary is required to represent some colors. Motivated by the impracticality of negative amount of primary, the CIE standardized, also in 1931, the $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ primaries. The CIE 1931 XYZ alternative color matching functions $\bar{x}$, $\bar{y}$ and $\bar{z}$ are represented in Figure 1.7(b). These functions show that all colors of the visible light spectrum can be represented using positive amounts of primary. Another property of this color space is that the function $\bar{y}(\lambda)$ was designed to match the photopic luminosity function $V(\lambda)$ presented in Figure 1.4. In this way, the Y component of any color provides its luminance value. The only way to do so was to use imaginary primaries [RWP+10].

#### 1.1.4.2 The CIE 1931 Chromacity Diagram

The CIE 1931 XYZ color space uses a 3-dimensional (3D) representation of colors. For convenient representation, the CIE also standardized a practical 2D representation of colors called the CIE Chromaticity Diagram. This is achieved by removing the luminance information in colors, thus providing the chromaticity coordinates $x$, $y$ and $z$ defined by:

$$x = \frac{\mathrm{X}}{\mathrm{X}+\mathrm{Y}+\mathrm{Z}} \qquad y = \frac{\mathrm{Y}}{\mathrm{X}+\mathrm{Y}+\mathrm{Z}} \qquad z = \frac{\mathrm{Z}}{\mathrm{X}+\mathrm{Y}+\mathrm{Z}} \tag{1.5}$$

Mathematically, this is a projection of the point $M(X, Y, Z)$ on $M(x, y, z)$ belonging to the plane of equation x+y+z=1. Note that $z$ is usually ignored since it can be easily calculated with the two ohter coordinates: $z = 1 - x - y$. The luminance Y and the two chromaticity coordinates $x$ and $y$ form the CIE $xy$Y colorspace. These two coordinates provide the chromaticity diagram plotted in Fig 1.8, it represents all colors perceivable by the human eye. The points outside the colored shape are imaginary due to the use of imaginary primaries (See Sec.1.1.4.1). All of the monochromatic colors from the visible light spectrum are represented on the curved side of the colored shape.

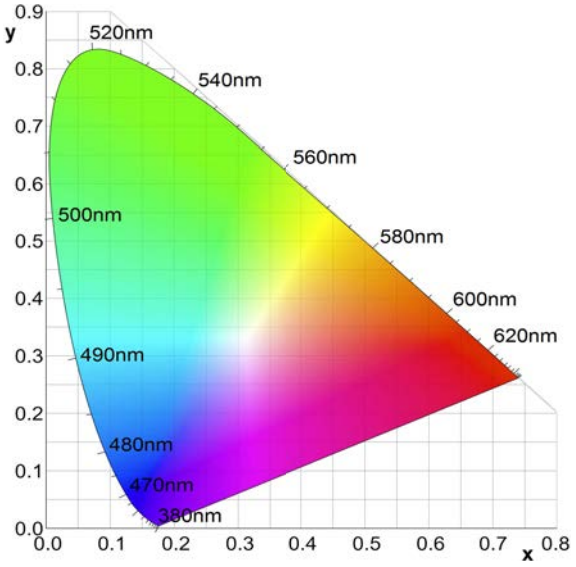



Figure 1.8 – CIE 1931 chromacity diagram - Note that colors near the edges cannot be represented on classical displays

Although the CIE 1931 color spaces were standardized long time ago, they are still widely used even though many color spaces emerged later. They usually serve as a reference for color space conversions.

### 1.1.5 Color Spaces

In the previous section, we described the techniques to create color spaces using colorimetry. There are a large number of color spaces, each with a specific purpose. The RGB color spaces are usually associated with displays because of the technique they used to reproduce colors, similarly the CMYK color space is used for printers as it mimics the way impression is performed. The CIELAB and CIELUV color spaces allow a more perceptually uniform representation of colors and luminance. In this section, we describe some color spaces and their interest for image/video compression.

Color spaces can be classified in two groups: device dependent and device independent. A large majority of capture and display devices have their own color space which is a subspace of the diagram of Figure 1.8. However, this subspace is not exactly the same for all devices, explaining why there are called device-dependent. The problem with such color spaces is that the captured/reproduced colors will be different for each device. Conversely, device-independent color spaces do not depend on the characteristics of a specific device. Therefore, these color spaces are considered more accurate for color representation, however there primaries are usually not reproducible by any devices. The CIE XYZ color space is a perfect example of device-independent color space.

#### 1.1.5.1 RGB Color Spaces

There are many RGB color spaces in addition to the CIE 1931 RGB color space. Indeed, the CIE RGB primaries were not reproducible due to technological limitations of the capture and display devices. For these reasons, several RGB color spaces were defined. The RGB system is very common and is used in almost every computer screen and TV. Any RGB color space is defined by three additive primaries: one located in the Red area, one in the Green area and one in the Blue area (see Fig.1.8). A fourth additional point, called the white point, is also added. It is the color represented when each RGB component is at its maximum (R=G=B=1). The three primaries define a triangle on the CIE chromacity diagram, as illustrated on Figure 1.9. This triangle is called the color gamut and it represents all the reproducible colors by additive combination of the three primaries. Thus, RGB color spaces do not represent all colors but only a subset of the chromacity diagram.

For explanation purposes, we focus only on the most common RGB color space, named sRGB. This color space is used in many devices and is one of the most popular for SDR imaging [SACM96]. Its gamut is represented on Figure 1.9 and is defined with the three following primaries and their chromaticity coordinates $\mathcal{R}$ ($x_R = 0.64$, $y_R = 0,33$), $\mathcal{G}$ ($x_G = 0.3$, $y_G = 0.6$), $\mathcal{B}$ ($x_B = 0.15$, $y_B = 0.06$). All colors outside this gamut can not be captured/displayed with sRGB devices. The white point used in the sRGB color space is the CIE standard illuminant D65 white point ($x_W = 0.3127$, $y_W = 0.329$). It is a commonly used white point standardized by the CIE representing the average daylight.

A linear relationship exists between the XYZ and RGB color spaces. Indeed, any

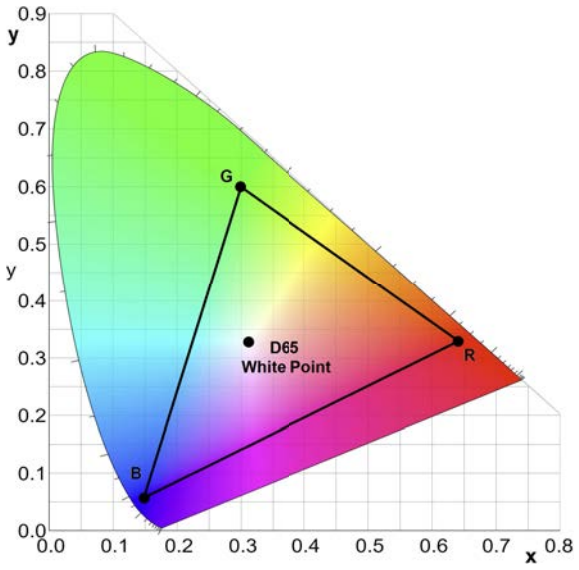



Figure 1.9 – sRGB color gamut and D65 white point

tristimulus color space can be converted to another tristimulus color space using a 3X3 transform matrix. It is defined with the three primaries and the white point coordinates as following [RGB93]:

$$\begin{bmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \end{bmatrix} = \begin{bmatrix} x_R & x_G & x_B \\ y_R & y_G & y_B \\ z_R & z_G & z_B \end{bmatrix} \cdot \begin{bmatrix} C_R & 0 & 0 \\ 0 & C_G & 0 \\ 0 & 0 & C_B \end{bmatrix} \cdot \begin{bmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{bmatrix} \tag{1.6}$$

with $z_i = 1 - x_i - y_i$, as explained in (1.5), and where $C_R$, $C_G$ and $C_B$ are computed such that the white point $(x_w, y_w)$ and the maximum luminance value $(\mathrm{Y} = 1)$ are associated to the RGB coordinate $(1, 1, 1)$:

$$\begin{bmatrix} C_R \\ C_B \\ C_G \end{bmatrix} = \begin{bmatrix} x_R & x_G & x_B \\ y_R & y_G & y_B \\ z_R & z_G & z_B \end{bmatrix}^{-1} \cdot \begin{bmatrix} x_w/y_w \\ 1 \\ z_w/y_w \end{bmatrix} \tag{1.7}$$

In the sRGB color space case, we obtain the following matrices:

$$\begin{bmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \cdot \begin{bmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{bmatrix} \tag{1.8}$$

$$\begin{bmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{bmatrix} = \begin{bmatrix} 3.2405 & -1.5371 & -0.4985 \\ -0.9693 & 1.8760 & 0.0416 \\ 0.0556 & -0.2040 & 1.0572 \end{bmatrix} \cdot \begin{bmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \end{bmatrix} \tag{1.9}$$

Note that the luminance Y is reproduced with a larger amount of green than red or blue. Indeed, Figure 1.7 shows that the spectral distribution for the luminance $\bar{y}$ looks a lot more like the green spectral distribution $\bar{g}$, therefore the luminance depends more on the green primary.

More recent RGB primaries have been defined for HDR imaging. These primaries and their color gamuts are described in Section 1.3.

#### 1.1.5.2 Color Opponent Spaces

Despite the fact that RGB is preferred for capture and display devices, these color spaces are rarely used for image/video compression. One issue of this color space is the high correlation between R, G and B values. For natural images, if a pixel has a large value for one component, there is a high probability that the other two components have a large value. This means that the three components have a lot of redundant information.

XYZ or $xy$Y color spaces seems more appropriate for compression since they decorrelate luminance and chrominance information but these color spaces also are inappropriate. Indeed, in the XYZ chromaticity diagram the colors are not uniformly distributed. Two equidistant colors in the green area or in the blue area will not be discriminated in the same way, the differences are going up to 80:1 [Poy12]. This aspect is well described with the MacAdam ellipses [Mac42]. For compression, perceptually uniform representations are usually preferred. With such representations, two equidistant colors in any area of the diagram will provide the same visual difference.

There is a lot of evidence that the LMS cone responses also carry a lot of redundant information, as in the RGB color space [Man06]. Thus, to reduce the amount of information transmitted to the brain, the HVS decorrelates the color information in two opponent channels: a red-green one and a blue-yellow one. This property of the HVS allows us to perceive color mixtures from these two channels like red-yellow (orange) or red-blue (purple) but not mixtures of red-green or blue-yellow [Wan95]. Several color spaces are based on this opponent property and proposed as an alternative to RGB and XYZ color spaces: CIELAB, CIELUV, IPT or Y'CbCr.

In 1976, the CIE standardized two new color spaces: the CIELAB [LAB76] or L*a*b* and the CIELUV [LUV76] or L*u*v* after successive improvements. Each one has small advantages over the other one but both of them reduce the non-uniform color distribution of XYZ from 80:1 to approximately 6:1 [Poy12]. However, these color spaces still provide some non-uniformity, which is why the most recent IPT color space [EF98] was developed. It is inspired from the HVS since IPT consists in an intensity channel I, a red-green channel P and a blue-yellow channel T directly derived from the LMS primaries (Figure 1.6). These three color spaces are appreciated for their perceptually uniform distribution of colors. However, among all color opponent spaces, the Y'CbCr one is always preferred for real-time compression or display (RGB conversion). The Y'CbCr color space is the less computationally expensive and therefore usually recommended in television standards [Poy12].

#### 1.1.5.3 Y'CbCr Color Space

The Y'CbCr color space is different from the previous color opponent spaces. It is directly derived from RGB color spaces and so, it does not contain all the color of the chromacity diagram but only a subset of color defined by the RBG primaries (see Fig.1.9). The Y'CbCr is the standard color space used in television ITU-R Recommendations. This color space has the advantage of providing backward compatibility with old black and white television and it allows a simple transition between capture and display devices. Indeed, this color space is made up of simple operations. First, a TVI function is applied to each RGB components to account for the perceptual non uniformity of color and lumi-

nance, as described in Section 1.1.3:

$$\mathrm{R}' = TVI(\mathrm{R}) \qquad \mathrm{G}' = TVI(\mathrm{G}) \qquad \mathrm{B}' = TVI(\mathrm{B}) \tag{1.10}$$

The brightness information Y' is then computed with a weighted combination of R', G' and B' components. Similarly, the color difference components Cb and Cr that are computed using a weighted subtraction:

$$\begin{aligned} \mathrm{Y}' &= k_R.\mathrm{R}' + k_G.\mathrm{G}' + k_B.\mathrm{B}' \\ \mathrm{Cb} &= \frac{1}{2.(1-k_B)} \cdot \left(\mathrm{B}' - \mathrm{Y}'\right) + O \\ \mathrm{Cr} &= \frac{1}{2.(1-k_R)} \cdot \left(\mathrm{R}' - \mathrm{Y}'\right) + O \end{aligned} \tag{1.11}$$

where $O$ is an offset value depending on the range of component values ($O = 0.5$ for values between $[0, 1]$) and with $k_R$, $k_G$ and $k_B$ values depending on the ITU-R Recommendations. This color space is widely used for SDR and HDR compression but not with the same TVI functions or $k$ values. Some examples are provided in Section 1.3.3.3.
This approach allows to easily reconstruct the R'G'B' components at the display side, using simple weighted additions and subtractions. The prime symbol in Y' is here to avoid confusion with the "true luminance" Y (see Eq(1.8)). Y is obtained according to a process called constant luminance whereas Y' is obtained according to a process called non-constant luminance [Poy12]. Indeed, Y is directly computed with the linear-light RGB components (1.8), then transformed with a TVI function. For Y', the TVI is directly applied to the RGB components. In this way, some luminance information remain in the Cb and Cr components. Constant and non-constant luminance is further explained in Section 1.3.1.2. For this reason, the component Y' is called "luma" (Frequently called luminance, wrongly) and the color components Cb and Cr are called Chroma.

The term YUV is widely used in video compression to describe the Y'CbCr color space. Rigorously, the terms YUV and Y'UV were used for analog encoding in television systems, while Y'CbCr is used for digital encoding [Poy12]. Note also that the U and V components in YUV are not related to the CIELUV components.

#### 1.1.5.4 Chrominance Subsampling

CIELAB, CIELUV, IPT and Y'CbCr consist in a luminance channel and two chrominance ones (Chroma for Y'CbCr). This particularity allows to take advantage of another HVS property which is its low chromatic sensitivity. The human eye perceives more spatial details with brightness discrimination than with color discrimination [vdBL01]. Therefore, the spatial resolution of chrominance channels can be reduced along with the carried amount of information. This is very useful in image/video storing/compression were several types of subsampling exist.

Considering a reference block with 2 lines, the subsampling types are described as a three parts ratio: $J$:$A$:$B$. With $J$ the number of luminance pixels by line in the reference block, usually 4. With $A$, the number of chrominance pixels in the first line and $B$ the number of chrominance pixels in the second line. The most common formats are the following ones:

— **4:4:4 format**: No subsampling is applied. All pixels are kept in each component.
— **4:2:2 format**: An horizontal subsampling of factor 2 is applied, 2 chrominance pixels are kept in each line.
— **4:2:0 format**: An horizontal and vertical subsampling of factor of 2 are applied, 2 chrominance pixels are kept only in the first line
— **4:1:1 format**: An horizontal subsampling of factor 4 is applied, 1 chrominance pixel is kept in each line.
— **4:0:0 format**: The chrominance components are discarded.

## 1.2 HDR Content Acquisition

The first step in the HDR pipeline (Fig.1.1) is to create the content. The acquisition of SDR content is rather limited in terms of luminance range due to legacy capture and display technologies. These devices only capture a limited dynamic range and limited color gamut in a scene and therefore lead to several artifacts such as underexposure, overexposure, color saturation, ... Furthermore, SDR images are represented with a limited number of values, so potentially with high contrast steps which results in quantization artifacts. The ultimate HDR acquisition goal is to generate/capture all the luminance and color information of a scene. In this way, we can obtain a physical description of the scene and, despite HDR display limitations, HDR imaging can benefit from such information. Indeed, several HDR applications like tone mapping or color processing are easier when all the scene information are known.

This section describes different techniques for the creation of HDR contents. The first method to create HDR contents is to generate them using computer graphics imaging (CGI) techniques. This method is still widely used in animation or video games and the CGI techniques keep evolving. The second method consists in capturing the wide luminance range of real-world scenes. Due to limitations of SDR capture devices, several bracketing algorithms were proposed to overcome these limitations and to build HDR contents using SDR devices. More recently, HDR cameras were introduced. They allow native HDR capture and remove the defects of the previous algorithms. Another method to create HDR contents is to generate it by expanding legacy SDR contents. This method has inevitable flaws and is usually described as a post-processing algorithm for HDR displays, for these reasons, we describe it in Section 1.4.3.

### 1.2.1 Computer Graphics Contents

The field of CGI consists in rendering images/videos from a virtual camera looking at a 3D virtual scene. This virtual scene is usually composed of different light sources, different reflecting materials and objects of different shapes. Thus, the virtual scene can simulate many light effects such as reflections, indirect lightning, shadows or even refractions. Based on such simulations, it is possible to compute the contribution of all these effects for each pixel of a rendered image.

The rendering of 3D imagery with computer graphics is an old and well studied subject but still evolving. In the mid-eighties, the rendering of 3D imagery was improved by combining realistic models with physically-based lighting simulation [PJH16]. This

improvement allowed a representation of virtual scenes with radiometric or photometric values (see Sec.1.1). In this way, we can obtain rendered images with pixel values expressed in Luminance unit, as for HDR imaging. The most used algorithms for CGI are Radiosity [CW12], Ray tracing [LS03] or Rasterization [AMHH16].

CGI allows the creation of many contents, furthermore without any constraint on dynamic range, color gamut or contrast step. It is therefore possible to "capture" all levels of luminance or color.

### 1.2.2 HDR Capture - Bracketing Techniques

Current SDR sensors can detect a wide dynamic range of luminance but can only capture a limited range in one shot. The lower and upper bound of this dynamic range is defined by the exposure parameter. Values outside the defined range are clipped, thus generating underexposed and overexposed areas. Figure 1.10 illustrates photographs of the same scene taken with different Exposure Values (EV). On the leftmost pictures, we

Figure 1.10 – Different photographs of the same scene with different exposure values

can see many underexposed areas but with a lot of details in the areas with high luminance levels, and conversely on the rightmost pictures, there are overexposed areas but we have many details in the previously underexposed areas.

Another issue with SDR cameras is that they apply a nonlinear function to the luminance hitting the sensor, for storing or subjective purposes. This nonlinear response is called the camera response function (CRF) and it differs depending on the brand or model of the camera and is usually kept private. Therefore, it is difficult to recover the physical luminance values. The best way to solve this problem is to measure the CRF with calibrated tests but this is complicated and time consuming. Automatic methods are usually prefer such as [MP95], [DM97] or [RBS03].

Once the CRF is known, it's possible to create High Dynamic Range images with SDR sensors using the bracketing technique. This technique is used in photography to overcome the limits of single shot captures. It simply consists in combining several shots taken with different capture parameters. This technique is applied to different types of parameters like Focus, ISO, or Exposure. In the latter case, one solution proposed by the authors in [MP95], consists in combining several shots (with different EV) of the same scene, like the ones in Figure 1.10. In this way, we obtain an HDR image containing

details in the darkest and brightest areas of the scene. Well detailed equations and an implementation of [MP95] is available in [BADC17]. The number of required shots with different EVs depends on the dynamic range of the captured scene. For a scene with a small dynamic range, one shot can be enough but this is not the case for most scenes. The bracketing technique can be set up in several ways, as described in the following sections.

#### 1.2.2.1 Temporal Bracketing

The temporal bracketing is the simplest method and achievable with only one SDR capture device. It simply consists in taking shots of the same scene one after another. This method is effective under two conditions, if the camera is motionless and if the scene has no moving objects. Indeed, the merging of these shots consists in a weighting of each shot and therefore misaligned structures or objects will appear on different places of the final HDR image. This effect is known as ghosting artifacts. While it is easy to keep a camera steady using tripods, it's difficult to avoid moving objects. Even in static scenes of nature, the wind moves all types of object like grass or branches. This problem is particularly important for high resolution images. For the moving camera, the usual solution is to use alignment methods. However, it is not adapted to moving objects. For this reason, several deghosting algorithms were proposed, some based on optical flow computation [GGC+09] and others based on local motion detection [KAR06, SKY+12].

Temporal bracketing is the method used in most smartphones, and labeled as "HDR mode". Two exposures are automatically captured successively and combined. However, the result is not the reconstruction of an HDR image but a tone mapped SDR image. Tone mapping algorithms are described in Section 1.4.2.

This bracketing method can be extended to video but with some difficulties. One solution consists in capturing video frames with alternate exposures and combining them into an HDR frame. In this case, it requires a high frame rate since multiple SDR frames are combined into one HDR frame. For low framerate, [KSB+13] proposed a solution where each exposure frame is transformed into one HDR frame. In both cases, the misalignment between frames and objects is corrected with deghosting algorithms but artifacts remains easily noticeable, especially on videos.

#### 1.2.2.2 Spatial Bracketing

Spatial bracketing avoids the temporal mismatch and can be achieved in several ways. One method consists to trade spatial resolution against a wider dynamic range. Indeed, using filters placed on the sensor, it is possible to capture different exposures on neighboring pixels [NM00, SBS+12]. However, on top of decreasing the frame resolution, this method requires to make costly modifications to the sensor.

Another method of spatial bracketing consists in using multiple cameras or multiple sensors embedded in the same camera. By placing them on a rig, each sensor can capture the scene at a different exposure but with geometrical disparities, thus requiring post processing as for temporal bracketing. Another solution is to use a light beam splitter [TKTS11]. As illustrated in Figure 1.11, the beam splitter divides the light in different exposure levels and redirects it on multiple sensors/cameras. With high precision, this solution removes all problems of disparity between exposure frames and preserves the

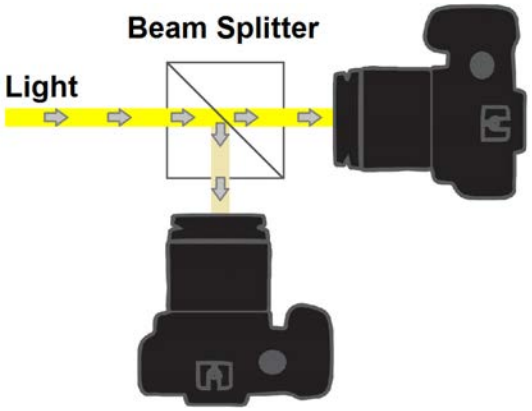



Figure 1.11 – Different photographs of the same scene with different exposure values

spatial resolution. Methods using several sensors remain costly and the non-portability of such systems limits their applications.

The dynamic range achievable with bracketing methods is theoretically unlimited since it depends on the number of frames with different EVs. In practice, it's difficult to use more than 3 sensors for spatial bracketing and increasing the number of frames in temporal bracketing enhances the ghosting artifacts.

### 1.2.3 HDR Cameras

The CCD and CMOS sensors used in SDR cameras respond linearly to the light [RWP+10]. As explain in Section 1.1.3, the HVS does not respond linearly to luminance but roughly logarithmically. Therefore, to extend the dynamic range, several CCD sensors with a logarithmic response were introduced by different companies. Such sensors allow the capture of a scene with strong lightning changes since they don't require controlled exposure levels, as for bracketing techniques. However, these sensors still use low resolutions and produce visible noise in dark regions. Furthermore, their effective dynamic range is inferior to the HVS dynamic range. For example, the SpheronVR camera [CBB+09] can capture a dynamic range of 6 magnitude orders of luminance.

All these HDR acquisition methods allow to create contents with a wider dynamic range. However, to represent all these luminance levels, it is necessary to use a large range of values. The following section presents several methods to store HDR contents.

## 1.3 HDR Content Storage

Most of SDR image formats (BMP, PNG, JPEG,...) use 24 bits per pixel (8 bits per component). Since HDR imaging is intended to represent much more data, 24 bits per pixel become insufficient. Indeed, real-world scenes can contain up to 14 orders of luminance magnitude. To avoid loss of information, HDR images are usually stored using floating-point values, thus resulting in 96 bits per pixel (32 bits per component). Thus, the amount of space needed for an image is increased by a factor four. For this reason, several solutions were proposed to improve the storage of HDR images.

The next section, we first explain how SDR storage works to help the understanding of HDR storage. In a second part, we present several early HDR image formats and finally, we present the most recent ones.

### 1.3.1 Legacy SDR Storage

One problem with SDR storing is the loss of all luminance information of the captured scene. Indeed, since SDR cameras can only capture around 3 orders of luminance magnitude among the 14 perceivable by the HVS, and since SDR displays can only reproduce around 3 orders of luminance magnitude, it was more convenient to scale the captured luminance values to the ones reproducible by the display. Therefore, SDR image formats only retain relative luminance information. For example, when capturing an outdoor scene in bright sunlight or under moonlight, the maximum luminance value in each case will addresses the same value in the SDR image format. However, building standards based on the characteristics of particular devices is not appropriate when technologies evolves quickly, such as currently. In contrast, HDR storing is intended to preserve the absolute luminance values of the captured scene. The maximum luminance value in each scenes is different and therefore they address different values in the HDR image format.

#### 1.3.1.1 Luminance Storage

Luminance storage on digital support requires quantization but this process induces an inevitable loss of data. We usually use 8 bits or 256 integer values (=$2^8$) to quantize the approximate three luminance orders of SDR images. Thus, the lower luminance level of a display is addressed with the integer value 0 and the highest luminance level with the integer value 255.

As seen on Section 1.1.3, the luminance perception is not linear and can be described using a TVI function. Therefore, it is advised to perform the quantization of luminance after applying a TVI function. In the context of storing and quantization, this TVI function is refered to as an OETF for Opto-Electronic Transfer Function. As stated in Section 1.1.3.1 and based on [SS63], the power law is a good OETF function for the luminance range of SDR displays. It is widely used for luminance quantization in SDR imaging and called the gamma-correction or inverse gamma function. Figure 1.12 illustrates the differences between a quantization with and without using such inverse-gamma function. In the linear case, the quantization is made directly on the luminance Y but in the gamma case, the quantization is performed after applying the inverse-gamma OETF: $\mathrm{Y}^{1/\gamma}$ with Y normalized between $[0; 1]$ and $\gamma \in [2; 3]$ [SS63]. As we can see, on the left side of Figure 1.12, linear uniform quantization does not result in uniform perceptual error. Quantization errors are more visible in the dark areas. However, with the inverse-gamma OETF, we obtain a uniform perceptual error for the SDR luminance range.

At the display side, we apply an EOTF (Electro Optical Transfer Function $= \mathrm{OETF}^{-1}$) to retrieve and display the relative luminance values. Note that, in practice, inverse-gamma OETFs usually include a short linear part for the darker values to avoid amplifying the capture noise.

Since quantization inevitably includes loss of data, the best choice is to keep quantization errors below the visible threshold of the HVS. As you can see on Figure 1.13, for the

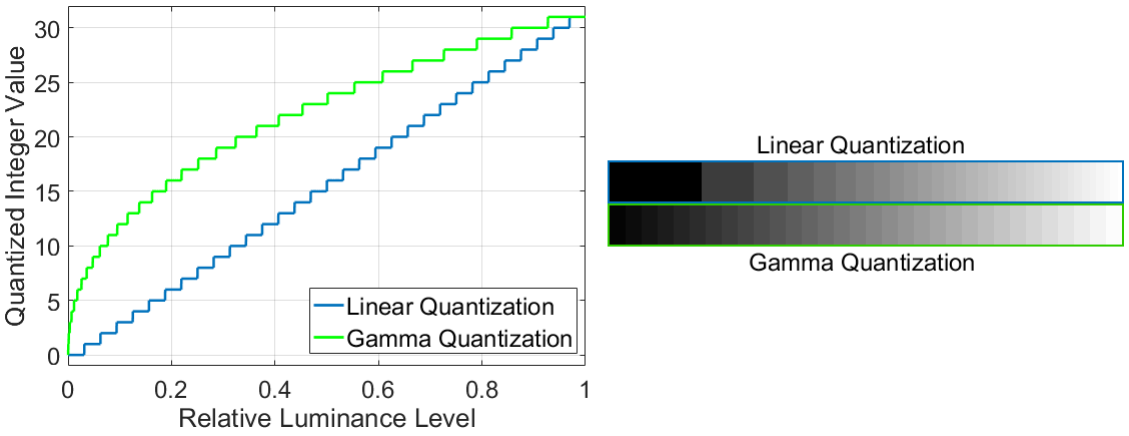



Figure 1.12 – Luminance quantization differences in Linear and Gamma domain ($\gamma = 2.4$). The quantization is done using 5 Bits (32 values) to enhance the quantization errors.

SDR luminance range, this is achieved for approximately 8 bits. For lower bit depth, quan-

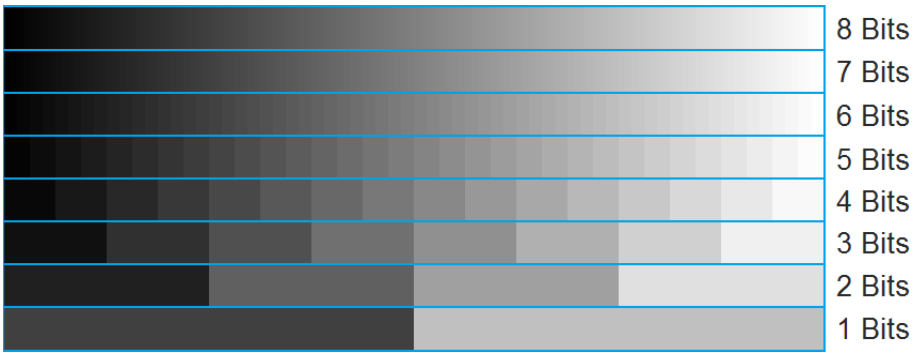



Figure 1.13 – Luminance Quantization with a power law ($\gamma = 2.4$) using different number of bits. Note that, based on your display capacity in luminance range, the luminance steps will be more or less visible.

tization artifacts known as "banding" or "contouring" appear. For the luminance range of HDR images, much bigger than the SDR ones, it is rather evident that 8 bits is not sufficient. Furthermore, the gamma OETF is rather efficient for the SDR case but not adapted to luminance values outside its range.

#### 1.3.1.2 Gamma Correction - Constant vs. Non-Constant Luminance

Historically, the inverse-gamma OETF was imposed by a technological constraint of cathode ray tube (CRT) displays. Indeed, due to physical characteristics, CRTs convert the R'G'B' input voltage into an RGB output intensity using a "gamma" function approximated by a power law somewhere between 2.2 and 2.8. Thus, the RGB components of an image need to be gamma-corrected prior to be displayed.

The optimal method for storing is to apply the inverse-gamma OETF only on the luminance since it accounts for its nonlinear perception by the HVS. Thus, as illustrated in the upper part of Figure 1.14, the luminance is retrieved with the gamma EOTF at the display side and converted into RGB components. Then these components are gamma

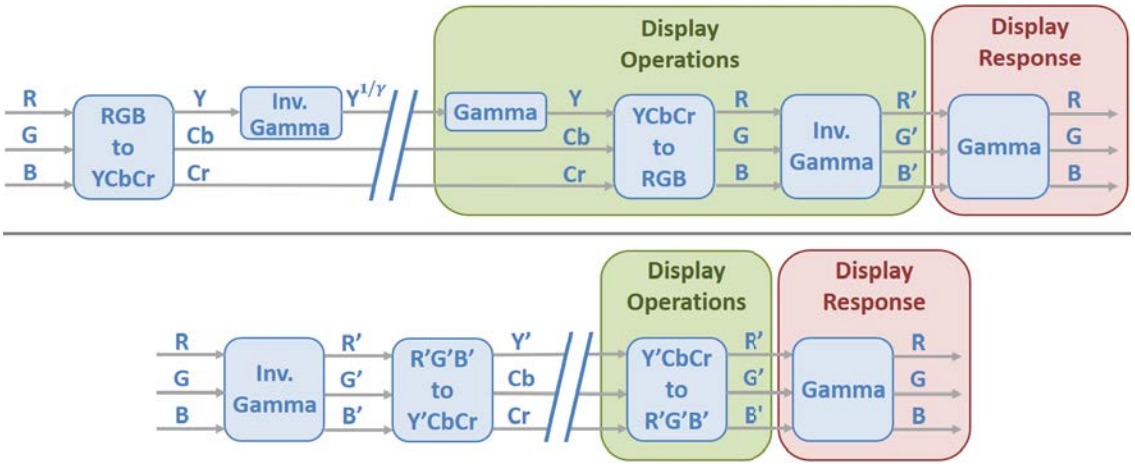



Figure 1.14 – Constant Luminance (upper) vs Non-Constant Luminance (lower) diagram

corrected, R'G'B', to compensate for the display response (in the red box of Figure 1.14). This encoding and transmission method is called constant luminance. With this method, the display need to perform two power functions: the EOTF to compensate for the inverse-gamma OETF applied prior to quantization, and a conversion to compensate for the display response. Since these two operations are approximately the inverse of each other and since the two operations are computationally expensive, the non-constant luminance process was adopted. In this case, as illustrated in the lower part of Figure 1.14, the RGB components are directly transformed with the inverse-gamma OETF or gamma correction, prior to quantization and to compensate for the display response. Note that gamma correction and OETF-EOTF are separate issues, however it is a coincidence that both problems can be addressed with similar functions.

If the chroma components Cb and Cr are transmitted with the same resolution as the luma Y', then the non-constant luminance process poses no problem. But since chroma components are usually subsampled (see Sec.1.1.5.4), it impacts the reconstruction of RGB components and generate minor artifacts. Some alternative solutions were proposed to correct these artifacts [SSD16].

To provide backward compatibility between CRT displays and more recent LCD displays, the non-constant luminance process was kept and is still used at this time. Meaning that LCD displays mimics the CRT displays gamma response. The gamma EOTF is included in the definition of several standards such as the sRGB one [SACM96], the television ITU-R Recommendation 709 [ITU14] and the ITU-R Rec 1886 [ITU11a] with slightly different gamma values. Usually, the OETF is not standardized, what matters is the method used by the display to reconstruct the image, so the EOTF. For subjective purposes, content producers can apply a slightly different OETF ($\text{OETF} \neq \text{EOTF}^{-1}$). Indeed, the CRF introduced in Section 1.2.2 modifies the OETF.

#### 1.3.1.3 SDR Image Formats and Color Spaces

In image formats, not only the luminance needs to be quantized but also the color components. In terms of perceptual uniformity, since color and luminance information is not distinguished in the same way, the best color spaces for image formats seems to be

color opponent spaces (see Sec.1.1.5.2). Indeed, such color spaces separate luminance and chrominance information and furthermore, they propose a perceptually uniform representation of color. Thus, chrominance components can be quantized uniformly and luminance can be quantized using an OETF. In fact, the CIELAB, CIELUV and IPT color spaces, introduced in Section 1.1.5.2, transform the luminance using an inverse-gamma OETF ($\gamma$ = 3 for CIELAB or CIELUV and $\gamma$ = 1/0.43 for IPT). However, these three color spaces are derived from the XYZ color space and designed to represent all perceivable colors. Therefore, there are many values that do not correspond to a physical color (for ex. $x = 0.7$, $y = 0.2$ in Fig 1.8). This is not optimal for storage because some quantized values will refer to "imaginary" colors. In addition, they are computationally expensive.

In practice, all perceivable colors of the chromacity diagram are not captured due to technical limitations of cameras. Only the subset gamut (see Sec.1.1.5.1) is represented in the resulting RGB color space. For this reason and for their low complexity, R'G'B' color spaces and their derived color space Y'CbCr are usually preferred. In this way, the quantized values only represent the limited acquirable range of colors and not the entire chromacity diagram plus imaginary values. These color spaces are quite perceptually uniform (not on the same level as CIELUV or IPT) and can be uniformly quantized. The typical SDR color gamut is equivalent to the one defined for the sRGB color space [SACM96] such as in the ITU-R Rec 709 [ITU14] or the ITU-R Rec 601 [ITU11b].

## 1.3.2 Early HDR Image Formats

As seen in Section 1.1.3, the luminance perception by the HVS is roughly proportional to logarithm of luminance. Based on this observation, early HDR image formats represented the pixel value with three RGB floating point values using 32 bits (96 bits per pixel). Indeed, the integer representation of floating point values is roughly comparable to a quantization in the logarithm domain. It can encode 79 orders of luminance magnitude [RWP+10], which is way larger than the 14 orders perceived by the HVS. This representation of HDR pixels is clearly not optimal and better alternatives were proposed.

### 1.3.2.1 The Radiance Format / RGBE-XYZE

One of the first common HDR image format was the Radiance picture format [War91] or RGBE/XYZE. This is one of the first HDR image formats and it was mostly used for HDR image rendering. It uses the extension *.hdr* or *.pic*. Its pixel representation is a kind of floating representation using 32 bits per pixel (bpp), 8 bits for each of the three components and a shared exponent using 8 bits. Due to that property, it performs a quantization in a roughly logarithmic domain. This format can be used with the RGB color space or the XYZ one and is thus refereed to as RGBE or XYZE. The RGBE format only improves the dynamic range storing since it proposes the same color gamut than SDR standards, such as [ITU14]. For this reason, the XYZE format was introduced allowing to represent all colors in the chromacity diagram (see Sec.1.1.4.2). However, as for the XYZ color space and its imaginary primaries, several XYZE values correspond to imaginary colors with no physical values. Thus, this format is not optimal for storing. The conversion from the

RGBE format to the captured/displayed RGB values is made with the following formula:

$$\mathrm{R} = \frac{R_M + 0.5}{256}.2^{E_M - 128} \quad \mathrm{G} = \frac{G_M + 0.5}{256}.2^{E_M - 128} \quad \mathrm{B} = \frac{B_M + 0.5}{256}.2^{E_M - 128} \tag{1.12}$$

with $R_M, G_M, B_M$ and $E_M$ the components of the RGBE format. Note that the conversion for the XYZE format is the same with the XYZ primaries.

This image format can represent around 76 orders of luminance magnitude [RWP+10], still way more than the range perceived by the HVS, thanks to its shared exponent. It results in high quantization steps since we keep the same dynamic range than the 96 bpp representation. Note that the exponent of RGBE format was defined by taking advantage of the correlation between R,G and B values in natural images (see Sec.1.1.5.2). However, this property is not applicable to the XYZ values, which makes the XYZE format less interesting.

#### 1.3.2.2 EXR Format / OpenEXR

The OpenEXR or EXtended Range format is very popular in the movie industry. It uses the extension *.exr* and was introduced in 2002 by Industrial Light & Magic (ILM) with an open-source C++ implementation [BKH03, LM]. This format is commonly used in the special-effect industry. It offers several types of representation but the most popular one consists in representing each RGB components with a 16 bits half precision floating point value, or half float [Hou81]. Thus, each pixel is represented with 48 bits. As for the RGBE format, the floating point representation allows a quantization in a roughly logarithmic domain. But unlike RGBE, the OpenEXR format covers a more restricted dynamic range, around 11 orders of luminance magnitude. This is a little less than the HVS range but, in practice, this range is lower for some people and extreme values are not frequently captured. Using a reduced range allows a finer quantization, making the quantization steps smaller than the visible threshold of the HVS (see Fig.1.15).

Concerning the color gamut, the OpenEXR format can be used with any RGB primaries and is thus opened to wider color gamut, such as the recently introduced ones: The P3 gamut [P311] or the REC.2020 [ITU15] (see Fig.1.17). Furthermore, this format allows the use of negative values and therefore allows to consider negative amounts of RGB primaries (see Sec.1.1.4.1). In this way, all colors of the chromacity diagram (Fig 1.8) can be represented.

#### 1.3.2.3 LogLuv Format

Since the logarithm was considered to be a good TVI model (see Sec.1.1.3), the authors of [Lar98] proposed two image formats based on the CIELUV color space where the luminance is transformed using a logarithm function rather than an inverse-gamma one (see Sec.1.3.1.3). The two image formats have different representations using either 32 bits per pixel (16 bits for the luminance including a sign bit and 8 bits for each chrominance component) or 24 bpp (10 bits for the luminance and 7 bits for each chrominance component). The 32 bpp format covers 38 orders of luminance magnitude, more than the HVS range, and the 24 bpp format covers only 4.8 orders of luminance magnitude. Unlike SDR image formats using the CIELUV colorspace and including "imaginary" colors (see

Sec.1.3.1.3), the LogLuv image format quantized only the perceivable colors of the chromacity diagram using a 2D look-up table. However, using only 7 or 8 bits for the entire chromacity diagram is not enough and will result in color banding/contouring.

### 1.3.3 HDR Image Formats - Optimal OETF and Wider Color Gamut

One shortcoming of RGBE and OpenEXR formats is the use of floating point numbers which roughly account for the HVS perception of luminance. A better alternative is to quantize the values after applying an OETF function adapted to HDR luminance levels, as for the SDR case in Section 1.3.1.1. In the HDR case, to avoid banding/contouring artifacts, we need a higher bit depth than 8 to quantize the covered luminance range. This can be achieved using the TIFF image format [Ado92] allowing the support of advanced pixel data types, including integer with more than 8 bits per component or floating point images. This format uses the extension *.tif* or *.tiff*.

The LogLuv image format, described in Section 1.3.2.3, uses the TIFF format with a logarithmic OETF. But as seen on Section 1.1.3, the logarithm is not a good approximation of the HVS sensibility to luminance. Based on TVI and CSF experiments, several OETF/EOTF were proposed [MKMS04, MDMS05, MND12]. These proposals were made in the context of HDR compression, however they are also used for HDR image storing. The authors in [MKMS04] proposed an OETF based on TVI experiments, covering the luminance range $10^{-4}$ to $10^{8}$ nits, followed by a quantization using 10 or 12 bits. However, as explained in Section 1.1.3.1, TVI experiments are suboptimal compared to CSF experiments. Therefore, the authors in [MKMS04] improved their model [MDMS05] using the Daly's CSF [Dal93]. More recently, based on the more complete Barten's CSF [Bar99], a similar but more popular OETF was proposed [MND12]. This OETF and its effectiveness are discussed in the next section.

#### 1.3.3.1 ST2084 EOTF - Perceptual Quantizer (PQ-ST2084)

The perceptual quantizer (PQ) or ST2084 EOTF was proposed by Dolby and standardized by the SMPTE [MND12]. The corresponding OETF is directly based on the Barten threshold ramp, a TVI function derived from CSFs using different background luminance levels, as explained in Section 1.1.3.2. The Barten ramp is based on the peak of each CSF and thus represents the contrast threshold of the HVS in its conditions of highest sensitivity. Figure 1.15 illustrates this ramp in dotted lines. In practice, this means that it's not possible to see a difference between two luminance patches with a contrast step (or quantization error) below the Barten ramp. Just above that ramp, banding and contouring artifacts may be visible and it gets worse when you move away up from that ramp.

However, since the Barten ramp takes into account the cases of higher sensitivity and is computed on calibrated patches, it may not be a good representation of HVS sensitivity for normal images. Another alternative is the Schreiber ramp which is a coarse approximation of Schreiber contrast threshold experiment's [Sch93]. This ramp is less optimal than the Barten one since it doesn't account for spatial frequency sensitivity of the HVS (see Sec.1.1.3). It uses the Weber-Fechner law above 1 nits and an inverse-gamma law below. This ramp is also illustrated in Figure 1.15. Note that the Schreiber ramp shape is similar to the Barten one but results in higher quantization errors.

The PQ-ST2084 OETF was designed to directly match the Barten ramp. Instead of a look-up table (LUT) giving the contrast threshold for each luminance level, they proposed the following equations:

$$\text{OETF}: \quad V = \left( \frac{c_1 + c_2 . J^n}{1 + c_3 . J^n} \right)^{1/m} \quad \text{with} \quad J = \text{clip}(\ \text{Y}/10000,\ 0,\ 1) \tag{1.13}$$

$$\text{EOTF}: \quad \text{Y} = \left( \frac{V^{1/m} - c_1}{c_2 - c_3 . V^{1/m}} \right)^{1/n} \cdot 10000 \tag{1.14}$$

$$m = 78.8438; \quad n = 0.1593; \quad c_1 = 0.8359; \quad c_2 = 18.8516; \quad c_3 = 18.6875$$

Using an equation is helpful for standardization and simpler for image and video processing. This OETF covers 7 orders of luminance magnitude from $10^{-3}$ to $10^4$ nits, less than the HVS range but enough to represent the future displays luminance range. Several examples are provided in Figure 1.15 using different bit depths. Image formats using 12 bits for the luminance would be totally lossless since the 12 bits PQ-ST2084 is under the Barten ramp for all luminance levels. Indeed, using 11 bits, the PQ-ST2084 is slightly above the Barten ramp but under the Schreiber ramp, as for the 10 bits PQ-ST2084.

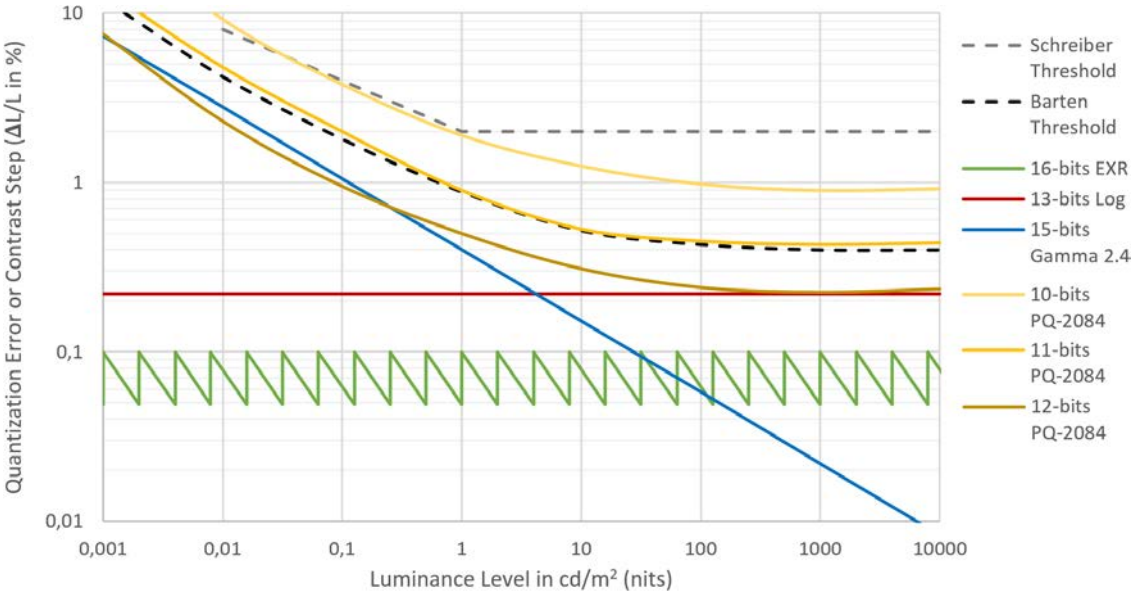



Figure 1.15 – Evaluation of several OETFs compared to the Schreiber and Barten threshold ramp

For comparison purposes, Figure 1.15 includes other previously introduced OETFs. We can see that, to be under the Barten ramp with the inverse-gamma OETF, you need 15 bits. And with a logarithm OETF, you need 13 bits. The inverse-gamma ($\gamma = 2.4$) OETF is clearly not adapted to high luminance levels and conversely the logarithm is not adapted to low luminance levels. Figure 1.15 also includes the quantization errors obtained with the OpenEXR image format. As you can see, as stated in Section 1.3.2, using floating point values is roughly similar to using a logarithmic OETF. Note that, the range of the EXR format goes beyond the represented one.

The PQ-ST2084 is included in television ITU-R Recommendations 2100 [ITU17] and is used in the reference compression standard for HDR, as further explained in Chapter 2. The Rec 2100 covers the support of 10 and 12 bit depth but note that, currently and in the near future, broadcast infrastructure and displays are going to operate with 10 bits. Therefore, many HDR contents only uses a 10 bit representation.

The dynamic range covered by the PQ-ST2084 can be easily increased with higher bit depth. However, for a given bit depth, you have to compromise between the covered dynamic range and the quantization errors. Indeed, it would be easy to extend the 10 bits PQ-ST2084 to higher dynamic range but it would result in higher quantization errors. Therefore, a 10 bits PQ-2084 covering 7 orders of luminance magnitude is a good compromise since it is under the Schreiber ramp. To have a 10 bits PQ-ST2084 below the Barten ramp, it is necessary to diminish the addressed dynamic range or conversely you need to increase the bitdepth.

#### 1.3.3.2 Other EOTFs - Hybrid Log-Gamma

Several new OETFs were proposed by several companies like Sony [GDY+], Philips [BV15] or the BBC [BC15a] standardized in [BC15a]. Most of them were proposed in the context of HDR compression [TSS+14], as an alternative to the PQ-ST2084. Only the BBC one, the Hybrid Log-Gamma (HLG), was kept [KYH16] and included in the Rec 2100 [ITU17].

The HLG is based on the observation that inverse-gamma OETFs are efficient with low luminance levels and that logarithm OETFs are efficient with high luminance levels. As indicated by its name, the proposed OETF is a composition of inverse-gamma and logarithm functions:

$$\text{OETF}: \quad V = \begin{cases} (3.J)^{0.5} & 0 \leq J \leq \frac{1}{12} \\ a.ln(12.J - b) + c & \frac{1}{12} < J \leq 1 \end{cases} \tag{1.15}$$

$$J = \text{clip}(\ \text{Y}/1000,\ 0,\ 1); \quad a = 0,17883277; \quad b = 1 - 4a; \quad c = 0,5 - a.ln(4a)$$

In some definition, the OETF includes a short linear part for the darker values to avoid amplifying the capture noise. Unlike the PQ-ST2084, the HLG only covers the range $10^{-2}$ to $10^{3}$ nits, as visible on Figure 1.16. Even on this limited range, the PQ-ST2084 gives the most efficient use of bits throughout the entire range. The benefit spotlighted by the authors is that the HLG has a behavior closer to the inverse-gamma one for low luminance levels. This property is supposed to provide some compatibility between SDR displays and HDR contents or between SDR EOTF and HLG OETF. In other terms, HDR contents corrected with the HLG can be interpreted as SDR contents. However, in practice, directly displaying HDR contents transformed with HDR-OETFs is clearly not optimal in terms of subjective quality.

#### 1.3.3.3 Wide Color Gamut - DCI-P3 and Rec.2020

"High Dynamic Range" is the commonly used expression to describe future standards and technologies. However, High dynamic range only refers to luminance. The increased capacity in colors is called Wide Color Gamut (WCG). As explain in Section 1.3.1.3,

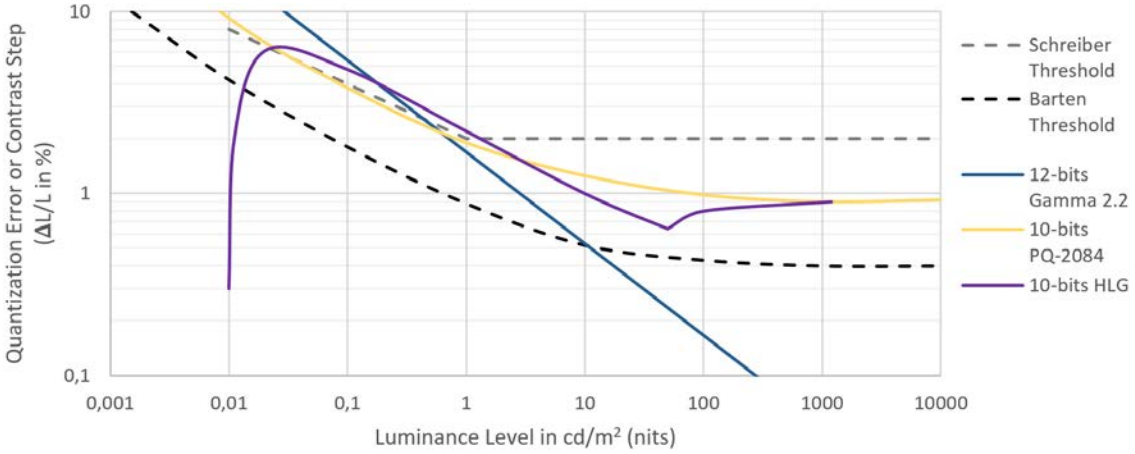
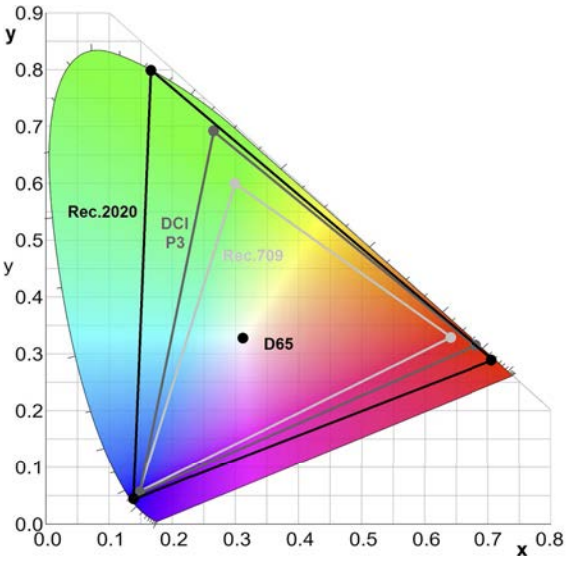



Figure 1.16 – Evaluation of the HLG and PQ-2084 compared to the Schreiber and Barten threshold ramp

color opponent spaces (see Sec.1.1.5.2) covering more than the visible chromacity diagram are not optimal for storing. To represent only the capturable color gamut and for its simplicity, the Y'CbCr color space is preferred. However, in the context of HDR and especially WCG, two wider color gamuts than the typical SDR one (Rec 709 [ITU14], see Sec.1.3.1.3) have been standardized: the DCI-P3 [P311] and the Rec 2020 [ITU15]. All these color gamuts are illustrated in Figure 1.17 with their coordinates. The values




| Color Gamuts | Rec.709 | DCI P3 | Rec.2020 |
|---|---|---|---|
| $x_W$ | 0.3127 | 0.3127 | 0.3127 |
| $y_W$ | 0.3290 | 0.3290 | 0.3290 |
| $x_R$ | 0.6400 | 0.6800 | 0.7080 |
| $y_R$ | 0.3300 | 0.3200 | 0.2920 |
| $x_G$ | 0.3000 | 0.2650 | 0.1700 |
| $y_G$ | 0.6000 | 0.6900 | 0.7970 |
| $x_B$ | 0.1500 | 0.1500 | 0.1310 |
| $y_B$ | 0.0600 | 0.0600 | 0.0460 |
| $K_R$ | 0.2126 | ~ | 0.2627 |
| $K_B$ | 0.0722 | ~ | 0.0593 |

Figure 1.17 – Color gamuts and corresponding coordinates

$K_R$ and $K_B$ are for Y'CbCr conversion (see Eq(1.11)). The Rec 709 color gamut covers only 35.9% of the chromacity diagram while the DCI-P3 and Rec 2020 ones respectly cover 45.5% and 75.8%. The Rec 2020 color gamut is also part of the Rec 2100 [ITU17] but, note that, this color gamut is currently beyond the capabilities of any consumer display. Still, recent HDR contents are usually stored using the Rec 2020 color gamut. Even though, they contain only color fitting in the DCI-P3 color gamut.

The Rec 2100 also includes a newly introduced color space, called ICtCp [Lab16], which improves the original IPT color space by using the PQ-ST2084 OETF instead of an inverse-gamma OETF. Furthermore, this color space is derived from RGB primaries and thus only represents the capturable color gamut. It is intended as a replacement for Y'CbCr color space. However, Y'CbCr is still currently preferred for its computational simplicity (see Sec.1.1.5.2).

## 1.4 HDR Content Display

SDR displays can usually reproduce 3 orders of luminance magnitude with a peak luminance around 300/400 nits. This is less than the instantaneous luminance range perceived by the HVS. The introduction of the term HDR is not recent and is sometimes confusing. Indeed, HDR displays are more recent than HDR capture techniques. Therefore, Tone Mapping Operators (TMO) were proposed to provide compatibility between HDR contents and SDR displays. For example, the HDR mode in many smartphones is not about HDR display but only about HDR capture. The displayed image is an SDR one. More recently, HDR displays allowing higher luminance range than SDR ones were introduced. Indeed, they can reproduce up to 5 or 6 orders of luminance magnitude with a peak luminance around 10000 nits. For some displays, this increased luminance range is mostly made on the black levels. Meaning that these HDR displays are slightly brighter than the SDR ones. Indeed, very bright HDR screen requires a high power consumption and they are therefore limited to professional uses.

In this section, we start by describing the different HDR display technologies and available displays for professional or private use. Secondly, we present the conversion from HDR to SDR contents using TMOs. Several types of TMO exists as well as Inverse TMOs, to create HDR contents from SDR ones.

### 1.4.1 High Dynamic Range Displays

#### 1.4.1.1 Professional Displays

In recent LCD SDR displays, the LCD panel is backlit with a uniform light over the entire surface of the screen. Thus, the dynamic range using such technology is rather limited. The first HDR displays proposed by [SHS+04] were based on two different technologies. The first one consists in using a Digital Light Processing (DLP) projector behind an LCD screen to enhance its dynamic range, as illustrated in Figure 1.18(a). The DLP projector usually has a low resolution and allows to boost the dynamic range of the LCD panel, whereas the LCD panel displays the details with its higher resolution and the color information. The achievable contrast is a multiplication of each contrast of the two devices (DLP and LCD). However, this technology has many flaws. Indeed, using a projector requires a display with a large depth (around one meter for a 15 inch display [BADC17]). Furthermore, the projector is prone to significant heat generation and high power consumption. For these reasons, the second method proposed by [SHS+04] consists in replacing the DLP projector by an array of ultra bright Light Emitting Diodes (LED), as illustrated in Figure 1.18(b). In this way, the display remains relatively thin however, LEDs reduce the energy requirements but still have a high power consumption

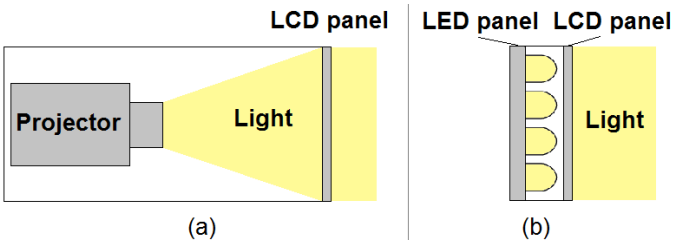



Figure 1.18 – HDR displays using dual modulation

and require cooling. These LEDs can improve the luminance peak in small area without impacting the rendering of black areas, thus improving the dynamic range of the display. As for the first method, the achievable contrast is a multiplication of each contrast of the two devices (LEDs and LCD). HDR displays using these technologies are referred to as dual-modulation, local dimming or two-layer display.

Several prototypes were developed however the first commercially available display and the most used one is the SIM2HDR47 [Mul]. This display was first introduced with a peak luminance of 4000nits, then 6000nits, and more recently a display with a peak of 10000 nits was announced [WB16]. This display is able to reproduce 5 orders of luminance magnitude, the LED panel is composed of 2202 LED and the LCD panel has a resolution of 1920x1080 pixels.

On top of improving the dynamic range of displays, improving the color gamut also has technical challenges. A promising new technology is the organic LED (OLED) technology [MWDG13] made with organic compounds that light up when fed electricity. Each pixel of the display has its own OLED which can be independently turned on or shut off. Thanks to this flexibility, the pixels of OLED displays can appear completely black. Therefore, unlike dual-modulation displays, OLED displays mostly increase the dynamic range with lower luminance levels but also with more saturated colors. Indeed, using the LED panel in dual modulation displays tends to desaturate the color [BPNS15]. However, OLED displays can hardly reproduce the high luminance peak of dual modulation displays, well appreciated in HDR experience by viewers. The reference OLED display is the Sony BVM-X300 OLED [SON] with a luminance peak of 1000 nits and 6 orders of luminance magnitude.

#### 1.4.1.2 Consumer Displays

Consumer displays usually have slightly lower performances than professional ones. As for professional displays, there are two types of consumer displays but with lowest capacities. For luminance, the most powerful consumer displays only achieve a luminance peak around 1100 nits and for colors, there isn't yet displays able to reproduce all color within the gamut of Rec 2020 (see Fig.1.17), and there likely won't be until few years. Indeed, current consumer displays struggle to reproduce all color of the DCI-P3 gamut. The Rec 2020 color gamut and OETFs with luminance peak around 10000 nits were standardized to support the arrival of future displays.

Recently, the UHD alliance agreed on a label, "Ultra HD Premium", certifying the new generation of displays [All17]. This label includes guidelines for supported color gamuts, peak luminance and black levels. For OLED TVs, the requirements are to have

a black level under 0.0005nits and a peak luminance over 540 nits, resulting in 6 orders of luminance magnitude. For LED TVs, the requirements are to have a black level under 0.05nits and a peak luminance over 1000 nits, resulting in 5 orders of luminance magnitude. However, for colors, the requirement is the same for both technology: at least 90% of the DCI-P3 color gamut (see Fig.1.17).

### 1.4.2 Tone Mapping Operator (TMO)

Currently, the vast majority of consumers is still equipped with traditional SDR displays. However, HDR signals contain a much bigger dynamic range than the one compatible with SDR displays. In fact, HDR contents also contain luminance levels outside the capacity of all current HDR displays. Therefore current displays are unable to represent real-world luminance levels. To provide a compatibility between HDR contents and SDR displays, we use an algorithm which remaps the wide luminance of an HDR image to an SDR one. This algorithm is referred to as Tone Mapping Operator or TMO. For HDR displays, the process is also called display retargeting [Bis17].

A simple solution to create an SDR image could be to mimics the behavior of SDR cameras and select an exposition in an HDR image (see Fig.1.10). However, using this method consists in deliberately introducing over/under exposed areas. TMOs are a better alternative since they can create SDR images with the advantage of knowing the entire dynamic range.

The development of TMOs is an active field since many years and many TMOs have been proposed. All of them give very different results since they can have different intents. Some try to mimics the HVS while others are designed for subjective purposes. However, TMOs can be approximately classified in two types, Global TMOs and Local TMOs, as described in the following sections. For a more detailed classification of TMOs, refer to [BADC17].

Usually the TMO is only applied on the luminance channel Y while colors remain unprocessed. The resulting SDR luminance channel $L_S$ is then applied to each RGB components, however this process can create saturated colors. For this reason, a desaturation is usually applied using a coefficient $C \in [0, 1]$. This process is described with the following formulas:

$$L_S = TMO(\mathrm{Y}) \tag{1.16}$$

$$\begin{bmatrix} \mathrm{R}_S \\ \mathrm{G}_S \\ \mathrm{B}_S \end{bmatrix} = L_S \cdot \left( \frac{1}{\mathrm{Y}} \cdot \begin{bmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{bmatrix} \right)^C \tag{1.17}$$

with $\mathrm{R}_S$, $\mathrm{G}_S$, and $\mathrm{B}_S$ the SDR RGB components. The desaturation method is rather simple and can cause minor luminance and color artifacts. [MTH09] proposed an alternative formulation preserving the color appearance of HDR images in SDR images. Another alternative is to use more decorrelated color spaces such as IPT [PAB+13].

The SDR RGB components are usually transformed with a gamma correction to account for the display response (see Sec.1.3.1.2) and quantized on 8 bits to match SDR image formats (see Sec.1.3.1.1). Note that, due to this quantization step, the TMO is a lossy process. A tone-mapped image may look like all luminance levels are kept but it's

not the case. Some luminance information are lost and cannot be retrieve using only the SDR image. Furthermore, SDR image formats only contain relative luminance levels and not absolute values. Therefore, the resulting SDR image loses the absolute luminance values of the real scene.

#### 1.4.2.1 Global TMOs

Global TMOs are the simplest ones, since they consist in applying the same operator to all pixels of the HDR image. It usually consists in an increasing monotonous function between HDR and SDR pixel values. The operator is generally based on global statistics of the HDR image, like the average, min or max values. It is designed to preserve global contrasts in an image to the detriment of local ones: Large luminance variations are kept while small ones are discarded. Therefore, the resulting SDR image usually lacks finer details.

The simplest TMO is a linear scaling, where the minimum and maximum values of the HDR image are mapped to the minimum and maximum values of the SDR one. However, this is not well adapted to the HVS which perceives luminance nonlinearly, as illustrated in Figure 1.19(a). Global TMOs usually consists in a non-linear mapping of luminance levels. Either using power laws [TR93], logarithmic functions [FPSG96, DMAC03] or sigmoid functions [Sch95]. Two global TMOs higlighted by subjective experiments [LCTS05, SbMEMS12] are the photographic tone reproducer (PTR) [RSSF02] and the display adaptive tone mapping (DATM) [MDK08], as illustrated in Figure 1.19. The DATM is based on an histogram analysis and is designed to minimize the contrast artifacts inherent to TMO. Furthermore, it can be adjusted to the targeted SDR display. The PTR is inspired from photographic techniques such as dodging and burning effects and based on a sigmoid function. This TMO also exists in a local version, as described in the following section, also well appreciated in subjective experiments.

#### 1.4.2.2 Local TMOs

Local TMOs are more complex than the global ones since they apply different treatments to pixels with the same value, based on their location on the image. Two pixels with the same luminance level in the HDR image can have different values in the SDR one. The operator is based on local statistics of the HDR image to preserve the local contrasts, generally to the detriment of the global ones. Bright and dark areas in the HDR image can have the same average luminance in the SDR image. Enhancing the local contrast can produce artifacts around edges such as halos.

Local TMOs can be designed to account for the spatial neighborhood of each pixel, for example using a Gaussian filter [JRW97] or Bilateral filter [DD02]. This approach is improved by using Gaussian or Laplacian pyramids [PFFG98, Ash02]. Other types of local TMOs are based on segmentation, such as [MKR07, RC09]. A different mapping is applied in segmented areas. In this case, it consists in selecting the best looking areas in different exposures of the HDR image. These operators are also known as exposures fusion.

Regarding the global TMO previously introduced, the PTR [RSSF02], it was extended to a local version by using a Gaussian pyramid. By avoiding sharp edge contrasts, this operator has the advantages to avoid halo artifacts. Figure 1.19 shows an HDR image

tone mapped with the PTR in its global and local version. As visible, the local version


Figure 1.19 – HDR image Memorial, from [War03], tone mapped with different TMOs: (a) Linear TMO - (b) [RSSF02] Global Version - (c) [RSSF02] Local Version - (d) [MDK08]

enhances the local contrasts. Especially in the highest circular window, where the frame is more visible.

#### 1.4.2.3 Video TMOs

The previously cited TMOs were mostly designed for image tone mapping. Indeed, using them directly on HDR videos can lead to different artifacts such as temporal incoherency or flickering [BCTB14a]. Global TMOs are usually preferred for videos since they are less prone to such artifacts and more suitable to real-time applications compared to local TMOs. Indeed, in most cases, a simple temporal filter on a global TMO is enough [PTYG00, MDK08] to remove temporal artifacts such as flickering. Besides, [ASC+14] proposed a local video TMO, eligible to real time processing, based on an optical flow and spatio-temporal filtering. Such processings remove the flickering artifacts usually brought by local TMOs. In a recent PhD thesis [Boi14], several solutions were proposed with techniques that adapts to any TMO. An experiment proposed by [EUWM13] provides an overview of the recently proposed video TMOs and evaluates them with subjective tests. This experiment also provides hints for future improvements in the field of video TMO.

### 1.4.3 Inverse TMOs / Expansion Operator

In opposition to TMO which provides compatibility between HDR contents and SDR displays, Expansion Operator (EO) provides compatibility between SDR contents and HDR displays. In this case, the problem is more complicated since it involves a reconstruction of lost information. Simple solutions propose to simply invert the gamma correction of SDR content [Far01, AFR+07], but these solutions do not account for the CRF [LGYS04] (See Sec.1.2.2). Furthermore, several problems prevent the reconstruction of HDR contents. First, since SDR contents only contain relative luminance values, the absolute ones need to be estimated. Secondly, over/under exposed areas in the SDR image

need to be filled. Thirdly, artifacts caused by the quantization to 8 bits are enhanced on a wider dynamic range (see Sec.1.3.1.1). And finally, conversely to TMOs, colors in the resulting HDR image can be desaturated.

Several EOs [MDS06, RTS+07, BLDC10] attempt to solve these problems, however EOs need to make strong assumptions about the initial contents and therefore the result may be not subjectively pleasant, depending on the input. Semi automatic methods are also proposed [DMHS08], however the computation becomes relatively slow since user intervention is required.

The terms "Inverse TMO" (ITMO) and "Expansion Operator" are usually used without distinctions but they describe different operations. Indeed, EOs, as defined in [BADC17], expand SDR images without information about the scene, while ITMOs reconstruct an HDR image previously tone mapped with a known TMO. SDR tone mapped images can be stored with metadata and so solve many of the previously introduced problems. Indeed, they usually do not contain over/under exposed areas, the absolute min and max luminance values can be saved as for the color desaturation parameter. For global operators, the only information lost is due to the quantization on 8 bits. However, for local TMOs the reconstruction is more difficult since the operator is not invertible.

## 1.5 Summary

In this chapter, we provided background information on High Dynamic Range and its pipeline. In Section 1.1, we presented how light and color are perceived by the human eye with the basic concept of Radiometry, Photometry and Colorimetry. Section 1.2 detailed the different techniques and devices used for the acquisition of HDR content. Then, Section 1.3 addresses the issues to store these acquired HDR contents by explaining the limitations of SDR image storing and the alternatives for HDR images. And finally, Section 1.4 describes the different available HDR displays along with their technologies and limitations. Besides, the methods used to provide compatibility between HDR/SDR content with SDR/HDR display are also introduced.

# Chapter 2

# High Dynamic Range Compression and Evaluation

While image formats reduce the amount of information by selecting the optimal quantization step and by restricting the represented range of luminance or color, compression is about removing the redundant and imperceptible information in an image/video with a lot of different techniques. The goal is to drastically reduce the content size, especially for videos. Indeed, 10 minutes of SDR video, with a high definition and using 60 frames per second (fps), requires an approximate space of 45 Gigabyte (24bpp x 1920 x 1080 x 60fps x 60sec x 10min). A just 10 min video is already more than a single-layer Blu-ray Disc capacity (25 Gigabyte), thus a strong size reduction is required to store full movies. The same conclusion can be drawn for broadcast applications with the limited bandwidth of current video transmission pipelines.

To fit the capacity of these different distribution media, advanced compression techniques and, usually, additional quantization are essential. In this case, the quantization is made after several steps of prediction. Compression algorithms using the quantization step are referred to as "lossy" while compression algorithms without this step are named "lossless". In this chapter, we start by explaining the basic concepts of SDR image/video compression and the different existing standards in Section 2.1. Then, the High Efficiency Video Coding (HEVC/H.265) standard is further described in Section 2.2. Since many compression standards exist, we focus only on this one which is, currently, the most efficient for both image and video [EFP+16, LHVA16].

As seen in the previous chapter, HDR content needs to be represented with higher bit-depth. This new type of content bring new challenges regarding its distribution. The HDR image formats described in Section 1.3.2 also proposed some simple compression algorithms which can divide the file size by two. However, as stated above, this is highly insufficient for Blu-ray storage or broadcast applications. New tools and several adaptations were introduced to reuse legacy SDR compression standards, such as HEVC. These compression solutions are described in Section 2.3.

Due to the additional quantization step, lossy compression leads to visible distortion of the compressed content. The stronger the quantization, the more artifacts appear but the smaller is the content size. To assess the efficiency of compression standards and for algorithms validation, we perform a quality evaluation. It usually consists in creating curves measuring the resulting distortion as a function of the content size/rate. By com-

paring these curves, we can identify the most efficient compression methods. Section 2.4 further describes this process along with different methods to measure the distortion of compressed content.

## 2.1 SDR Image/Video compression

### 2.1.1 Motivations and Standards

The motivation to further improve video compression is due to the ever growing size of contents. Indeed, image and video contents keep evolving with higher bit depth, higher resolution or higher frame rate. As seen in Section 1.3, HDR and WCG contents represent a much larger range of light or color and need higher quantization steps, thus higher bit depth, to avoid banding artifacts. The size of images and videos evolve fast. Indeed, in few years, the video resolution went from Standard Definition (SD $\approx$ 720x480 pixels), to High Definition (HD = 1280x720 pixels), to Full HD (= 1920x1080 pixels), and finally to Ultra HD (UHD = 3840x2160 pixels). Figure 2.1 illustrates the relative size of these resolutions.

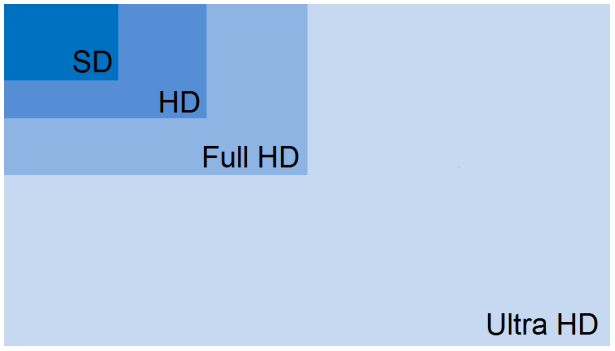



Figure 2.1 – Evolution of television resolutions

The number of represented pixel is almost multiplied by 4 in each new resolution. Even though a higher resolution than UHD does not seem to be useful for conventional TVs [CV16], other devices/technologies could benefit from it such as virtual reality devices.

For video content only, another increasing parameter is the framerate or the number of displayed frames per second (fps). Indeed, the common assumption that, using more than 24 fps is unnecessary for the HVS, is false. Increasing the framerate to 50 fps, 100 fps and even 200 fps provides an improvement in perceived quality [CV16].

Compared to this fast content evolution, the broadcast capacities are evolving slowly. Furthermore, the amount of transmitted videos has grown enormously with the multiplication of video channels and the advent of internet video platforms such as YouTube or Netflix. Therefore, better and better compression algorithms are required to fit the rate capacity of broadcast pipelines. Over the years, numerous tools have been proposed for image/video compression. And to ensure tool compatibility among devices, standardization is essential. It allows to define a set of mandatory tools used at both the encoding and decoding side.

Note that the standard only compels the decoding step. The encoding step can be performed in different ways, the only requirement is to provide a compressed stream which

can be decoded by the defined standard. In this way, an image/video compressed with a specific standard can be decoded by all compatible devices.

Several standards dedicated to image compression have been proposed, the most popular ones being JPEG [Wal92] and JPEG2000 [CSE00], but most standards are now focused on video compression. Anyway, video compression always start by image compression to encode the first frame, as explained in Section 2.2.1. Concerning video compression standardization, the most popular standards come from two bodies who started working separately and now jointly. The first one is the Video Coding Expert Group (VCEG) from the International Telecommunication Union (ITU-T) and the second one is the Moving Picture Experts Group (MPEG), part of the ISO/IEC (Organization for Standardization / International Electrotechnical Commission). Over the years, these bodies proposed several video compression standards, as illustrated in Figure 2.2. The standard evolution is a direct response to the content evolution explained above. Thanks to the increased computational performances of encoding and decoding devices, standards can use more and more complex algorithms and therefore improve the compression performances.

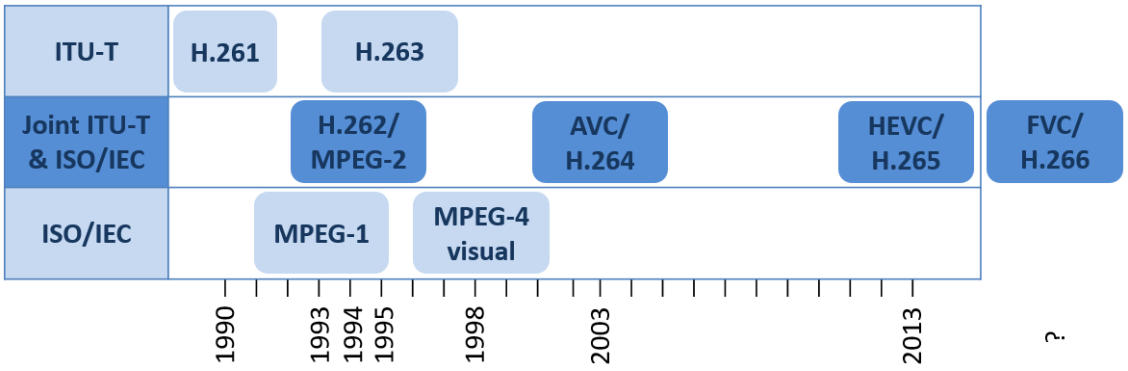



Figure 2.2 – Timeline for video compression standards

The first proposed and widely used standard introduced in 1990 was the Recommendation H.261 [H2693]. In this standard, the main steps of video compression such as prediction, transform, quantization and entropy coding were defined. The most recent standards still lie on these main steps, described in the following sections. In 1993, the ISO/IEC also presents their first standard, MPEG-1 [MPE93], very similar to H.261 but with additional algorithms. Then, a first collaboration between these two bodies was made to propose, in 1994, the joint standard known as MPEG-2/H.262 [H2612, MPE03]. It was massively adopted through DVD applications or TV broadcast. Currently, this standard is still in use for TV broadcast in the USA. After this proposal, both standardization bodies worked independently to propose the Recommendation H.263 [H2605] in 1995 and the MPEG-4 Visual or MPEG-4 part2 [MPE04] in 1998. These very similar standards both introduced interesting tools such as the frame type in H.263 (see Sec.2.2.1) or the support of high bit depth [1] in MPEG-4 Visual. They both were more efficient than MPEG-2/H.262, however they did not replace it in TV broadcast or for DVD. Indeed, many consumers were recently equipped with MPEG-2/H.262 compatible devices and content encoded with a new standard would have involved to purchase new TVs or DVD players. The MPEG-4

1. more than 8 bits per component

Visual standard was more popular for web diffusion, especially with the Xvid implementation of this standard. In 2003, a second collaborative standard was presented, the Recommendation H.264 [H2617] or the MPEG Advanced Video Coding (AVC) [MPE14]. It proposed a significant improvement over MPEG-2/H.262 [WSBL03] and successfully replaced it in TV broadcast and through Blu-Ray applications. AVC/H.264 is currently the standard for TNT diffusion in France. Finally, in 2013, another collaborative standard was presented, the Recommendation H.265 [H2618] or the MPEG High Efficiency Video Coding (HEVC) [MPE17]. The HEVC standard improves the compression performances of AVC/H.264 by 50% [OSS+12]. As for the MPEG-4 Visual and AVC/H.264 standards, HEVC supports contents with a bit depth higher than 8 bits, up to 16 bits per component. In practice, these higher bit depths were exclusively used for professional applications since SDR devices only support 8 bits. However, they can be used to compress HDR and WCG content, as explained in Section 2.3.

In Section 2.2, we present only the HEVC standard since it is currently one of the most performing in both image compression [EFP+16] and video compression [SM16] and it contains a majority of tools introduced over past standards. Furthermore, HEVC is currently the only standard supporting HDR content distribution. Note that, a new collaborative standard is already in preparation under the name Versatile Video Coding (VVC) or Recommendation H.266 [SBW17] to further improve the compression performances of HEVC.

Several other video standards were proposed by several companies, however they share many similarities with the ITU-T or ISO/IEC standards. The most popular ones are the Daala codec from Xiph.Org (sponsored by Mozilla), AV1 recently promoted by the Open Media Alliance and the VP codecs from On2 Technology, acquired by Google in 2010.

### 2.1.2 General Compression concepts

The compression goal is to convert an input content into a bit stream containing as little information as possible and allowing the content reconstruction. To make this bit stream lighter than the input content, the techniques mainly consist in exploiting the limits of the HVS and removing the redundant information in the input.

The first method to reduce the image/video size consists to sub-sample the Chroma components as explained in Section 1.1.5.4. In general, the image/video uses the Y'CbCr colorspace sub-sampled with the 4:2:0 format. In this way, an SDR content goes from 24 bpp (8 bits for each component) to about 12 bpp (averaged for 4 pixels: 4x8 bits for Y' and 1x8 bits for Cb or Cr), thus reducing the content size by a factor two. Usually, the input content in compression schemes is considered to be already sub-sampled, meaning that the measured compression efficiency does not account for this prior content size reduction.

The sub-sampled content is then divided into blocks of different size, depending on the content, and each block is fed to the compression scheme as illustrated in Figure 2.3. As explained above, this compression scheme structure is the same for all previously introduced standards, including JPEG, and is referred to as hybrid block-based compression scheme. The only exception is JPEG2000 which is a wavelet-based compression method [CSE00].

The hybrid block-based scheme uses the prefix hybrid since it combines a prediction

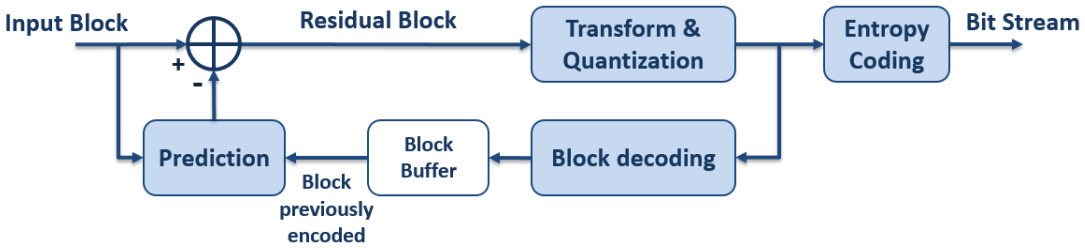



Figure 2.3 – Main operations performed in hybrid block-based compression schemes

step, to remove the redundant information in the signal, and a transform/quantization step, to further reduce the dependencies between pixels and also remove high-frequency signals imperceptible by the HVS. These two steps significantly reduce the content size and each block is thus represented with few bits. These few bits can be further compressed using an entropy coding that will remove their statistical redundancies. All the compression scheme steps are further described in the following sections.

#### 2.1.2.1 Prediction

The prediction is a lossless step where we try to predict the current block only with previously encoded ones, stored as reference blocks. Indeed, at the decoding side, only the previously decoded blocks are available to reconstruct the prediction. The prediction block is then subtracted from the current block, thus providing a residual block with smaller pixel values centered on zero. These values usually have a smaller amplitude and can be quantized with less than 8 bits to reduce the content size. However, applying a transform step before is beneficial, as explained in the next section.

It is possible to distinguish two types of prediction. The first one is the temporal prediction, also called inter prediction, and it is based on a block matching algorithm. It consists in using a reference block from previously encoded frames to predict the current one. This process is very efficient since in many videos the background is usually static, therefore many blocks of the current frame can be predicted with collocated blocks from previously encoded frames. For moving objects, a motion compensation allows to predict the current block with a block from a different location in the previously encoded frames. The most appropriate block in this motion compensation is selected through a motion estimation step. Temporal prediction is present in all video standards previously presented but, obviously, not in the JPEG standard since there is no previous frame. HEVC temporal prediction is further described in Section 2.2.3.2.

The second prediction type is the spatial prediction, also called intra prediction. Conversely to the temporal case, spatial prediction does not require a previously encoded frame. It is designed to predict the current block with parts of the spatially neighboring blocks previously encoded. Only small features of spatial prediction were present in early standards. It was much more enrich in the recent ones such as AVC/H.264 and HEVC/H.265. The spatial prediction from HEVC is further described in Section 2.2.3.1.

#### 2.1.2.2 Transform and Quantization

The transform and quantization are directly applied after the prediction step, to the residual block. The first goal of the transformation is to remove the correlation between pixels. It is implemented by mean of a fixed transform called discrete cosine transform (DCT). It consists in converting the samples of the residual block in a frequency domain, where the different frequency signals are separated. The result is a transformed block of the same size where lowest frequencies are located in the up-left transformed coefficients and highest frequencies in the low-right transformed coefficients.

Note that for most contents, the information is mainly located on low frequency coefficients. This representation allows to reduce the remaining redundancy in the residual block. For example, a residual block with the same pixel values is a very low frequency block where all pixel values are redundant. After transformation, only one pixel will be non-zero: the upper-left one. Since the information is located on fewer coefficients in the transform block than in the residual one, the transform step improves the compression performances. Indeed, thanks to the entropy coding details in the following section, the zero coefficients of the transform block have a very low bit cost.

As demonstrated in Figure 1.5, the HVS is much less sensitive to high frequencies than low ones. Thanks to this property, the quantization is not made uniformly in the transform domain but with higher consideration for low frequency coefficients. In this way, small quantization steps only remove imperceptible data. However, depending on the rate requirement or storage capacity, the quantization steps usually need to be higher. In this case, noticeable distortion is introduced. However, it still remains more efficient than performing a uniform quantization of the residual block.

The quantization step is defined by the quantization parameter (QP). This QP regulates the rate/size of the content and therefore the distortion level. Indeed, for small rate, the QP needs to be higher and the distortion artifacts become increasingly visible. On the contrary, small QP values limit the distortion but increase the rate/size of the content.

Note that only the quantization stage makes a compression scheme lossy, as opposed to lossless compression schemes where the quantization is discarded and no distortion is introduced. Indeed, quantization is the only non-reversible stage since original pixel values can not be retrieved after applying it.

#### 2.1.2.3 Entropy Coding

The Shannon entropy is a measure to quantify the amount of information in a signal. For a random block $X$ with $n$ possible values and where $p_i$ represents the occurrence probability of each pixel value $i$, the entropy $H$ is defined as follows:

$$H(X) = -\sum_{i=0}^{n} p_i.log_2(p_i) \tag{2.1}$$

$H(X)$ provides the averaged necessary number of bits to represent each pixel value $i$ of the signal $X$. This entropy value $H(X)$ is the theoretical low limit that entropy coders aim to achieve. To do so, entropy coders can use Variable Length Codes (VLC) to represent all the pixel values. Short binary codes represent the most occurring pixel values (with high $p_i$), while the less occurring ones (with low $p_i$) are represented with longest binary codes.

In this way, the averaged number of bits used to represent each pixel of the random block $X$ is reduced and converges to $H(X)$.

To have a block represented with a minimum of bits, it must have a low entropy. Since the redundant information in the input block is removed by the prediction and the transform step, the transformed residual block has a lower entropy than the original input one, furthermore after the quantization step. Indeed, the distribution of the values from the transformed and quantized residual block is generally narrower than the input block one. Thus, the pixel values in this transformed and quantized residual block can be represented with much less bits than the pixel values from the input block.

## 2.2 HEVC overview

The previous section presented the main stages of the typical hybrid block-based compression scheme in a very general way. Indeed, HEVC is a gigantic tool box with many algorithms newly introduced or inherited from older standards. Its superior compression efficiency over earlier standards comes from the assembly of these numerous algorithms, each providing a small improvement.

A reference software is usually developed along with the video standard definition. For HEVC, this software is the result of a collaborative work and is called the HM test model [HM]. Note that, the HEVC standard does not specify the encoding process. The HM software only provides an example of encoding algorithms and specify the decoding process.

The HEVC/HM codec structure is illustrated in Figure 2.4. Its structure is based on the same principles than the ones introduced in Figure 2.3. First, the encoder splits the video in different Group Of Pictures (GOP) and assignes a type to each of these pictures, as explained in Section 2.2.1. Once a frame is selected, it is split into square blocks using a newly introduced partitioning method based on a quadtree, detailed in Section 2.2.2. Each of these blocks is then predicted with either Intra or Inter prediction. Section 2.2.3 presents the new prediction features introduced in HEVC. The residual block obtained after the prediction step is then transformed, scaled and quantized, as explained in Section 2.2.4. To finish the encoding step, Section 2.2.6 details how the remaining quantized transform coefficients are entropy coded with other signaling information. Note that, the encoder also duplicates the decoder to ensure that the prediction process is identical in both cases. After inverse transformation and prediction reconstruction, the reconstructed images are filtered to remove commonly appearing artifacts. These filters are described in Section 2.2.5. The decoded pictures are then stored in a buffer for the prediction of future frames.

### 2.2.1 GOP and Frame Types

Due to the inter prediction, there are dependencies between the encoded frames. Therefore, to decode the current frame and reconstruct its prediction, it is necessary to decode all the previously encoded frames. However, when you watch TV or jump to a specific time in your Blu-ray, the video must start right away. It cannot depends on all the previous frames. To overcome this problem, the video is decomposed in several identical Groups Of Pictures (GOP) mostly independent from each other. In this way, the decoding process

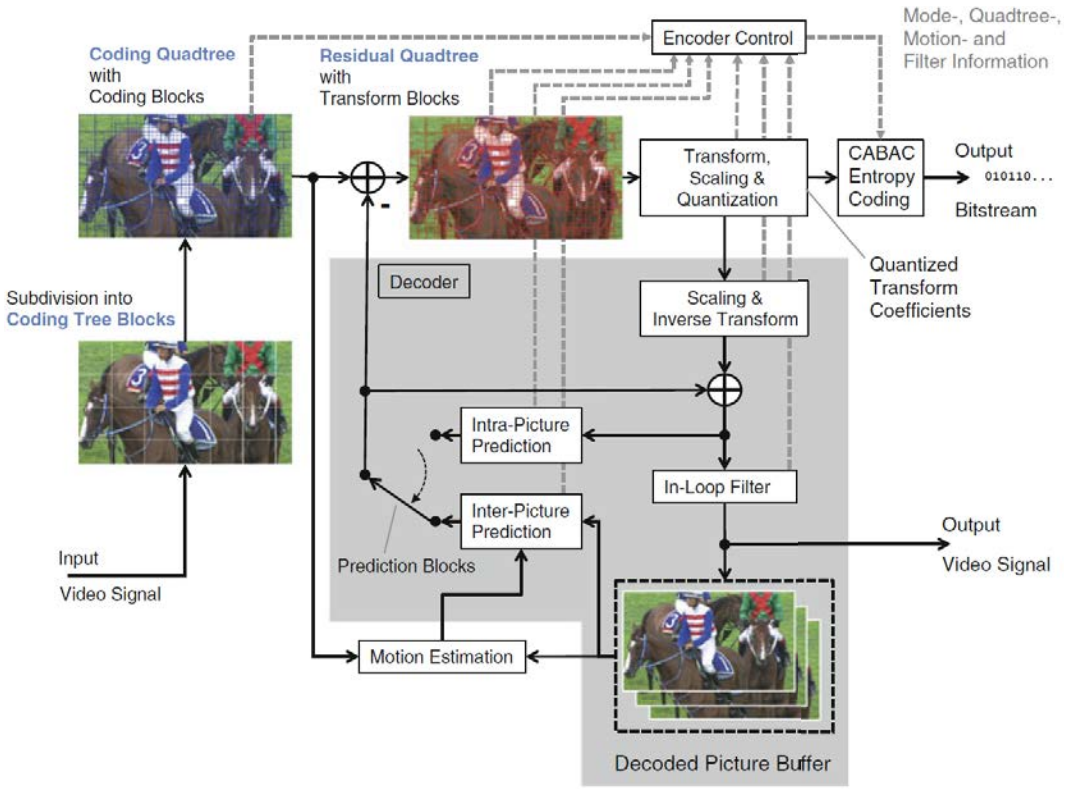



Figure 2.4 – Block diagram of an HEVC encoder with built-in decoder. Source: [SBS14]

can be started from any one of these GOPs. Inside the same video, these GOPs use the same pattern and thus define a coding period.

Inside these GOPs, the frames can be encoded/decoded in a different order than the display order, as illustrated in Figure 2.5. The coding order is called Picture Order Count

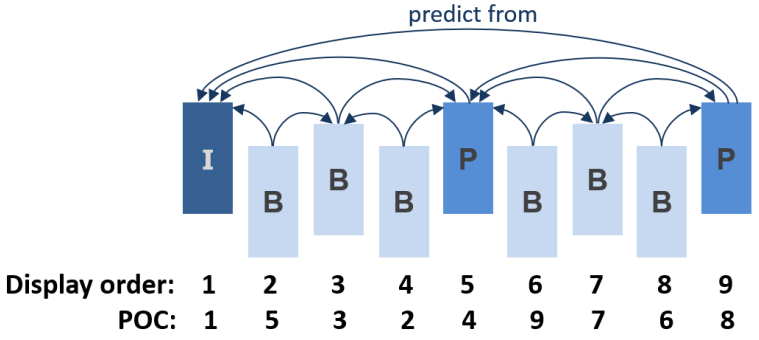



Figure 2.5 – Example of hierarchical GOP structure

(POC) [SOHW12]. This hierarchical arrangement allows to have some frames predicted from past and future frames (in display order) depending on their type. Indeed, each frame is associated with one of the following frame type. For sake of simplification, we explain the behavior in the case of non-multiple reference frames (which is an extension of what is presented here).

— **Frame Type I**: Each GOP start with an intra coded frame which is coded independently from all other frames. All blocks inside are predicted using only intra

prediction. In this way, the frame can be decoded on its own since it does not depend on other frames. However, preventing the use of inter prediction usually results in poorer prediction of the frame. Thus, I-frames usually have an higher bit cost than the other frame types.

— **Frame Type P**: The predictive coded frame encodes blocks using intra and inter prediction, but only with previously decoded frames. P-frames can only be predicted from past I-frames or P-frames. The size cost of P-frames is usually lower than the I-Frames one, thanks to the inter prediction.

— **Frame Type B**: The bidirectional predictive coded frame also encodes blocks using intra prediction and inter prediction but with forward and backward frames. B-frames can thus be predicted from previously decoded I-frames, P-frames and B-frames. Another advantage of B-frames is the possibility to generate prediction combined from different reference blocks. This prediction mode is called bi-prediction and it allows the current block to be predicted with two blocks from different frames. Thanks to these features, the size cost of B-frames is usually rather small. This frame type significantly improves the compression performances.

Note that these frame types were firsly defined in the H.263 standard [H2605]. I-frame encoding can be used for image compression since the prediction is only performed within the frame [LHVA16]. Conversely, P and B-frames are typical of video compression since they require previously encoded frames.

The size of the GOP influences the compression performances. Indeed, the higher the GOP size, the smaller the number of I-frame, however, the longer the coding period. For broadcast applications, it is generally advised to use a GOP size that does not exceed 1 second (depend on the GOP size and the frame rate).

Several GOP configurations are proposed in the HEVC reference software [HM]: All-Intra, Low-Delay and Random-Access [SBS14]. The All-Intra configuration consists in using only I-frames and therefore provides a video of independently decodable frames. The Low-Delay configuration allows the use of all frame types, however the encoding order is the same as the display order. Therefore, no prediction from future frames. This configuration is useful in low-delay scenarios where decoding the multiple frames of a GOP, before displaying them, is critical. The Random-Access configuration on the other hand allows a hierarchical prediction of the frames as depicted in Figure 2.5. This last configuration is the most efficient in terms of compression performances.

### 2.2.2 Block Partitioning Structures

Once the frame is selected, it is split into blocks of various sizes. One of the most efficient tool introduced in HEVC is its block partitioning structure based on a recursive quadtree [KML+12]. Indeed, the HEVC partitioning has a greater flexibility than the macro-block approach of H.264/AVC and therefore it better adapts to the input frame. The HEVC partitioning is made with the four following block units:

— **The Coding Tree Unit (CTU)**: It is the largest block unit. The first partitioning operation consists in splitting the frame into CTUs of the same size. This size is defined during the encoding process and can take the following values: 64x64, 32x32 or 16x16 pixels [SBS14]. Note that, the largest block size in H.264/AVC is 16x16 pixels. Indeed, HEVC targets larger video resolutions such as UHD and

therefore requires a larger block size.
Once the frame is split into CTUs, they are processed iteratively in raster scan order (from left to right then from top to bottom).

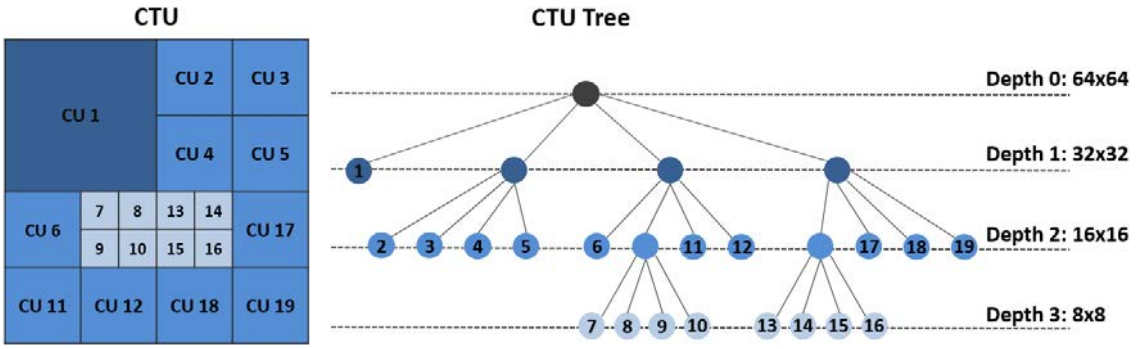



Figure 2.6 – Example of CTU quadtree partitioning into CUs

— **The Coding Unit (CU)**: Each CTU is further split into CUs using a quadtree structure, as illustrated in Figure 2.6. The CTU can first be divided into one or four CUs. Then, if the CTU is divided in four CUs, each of them can be recursively divided into four smaller CUs. In this quadtree, the size of the CTU is the root node and the leaves are the CUs. When the depth increases, the CUs size decreases. The largest CU size is equal to the selected CTU size and the smallest CU size is 8x8 pixels thus, for a 64x64 CTU, the maximum depth is 3. The large CU sizes are usually selected for uniform areas of the frame while the small ones are more appropriate to the frame areas with high spatial activity. Each CU leaf is the base unit for the prediction and transform step. They are processed in a Z-scanning order, as numbered in Figure 2.6.

— **The Prediction Unit (PU)**: Each CU is then predicted with either Intra or Inter prediction. Depending on the prediction type, the CU can be further split into PUs, as illustrated in Figure 2.7. The PU partitioning can be done in 8 ways (two classical method 2Nx2N and NxN, two symmetrical methods 2NxN and Nx2N, and four asymmetrical methods 2NxnU, 2NxnD, nLx2N and nRx2N [KML+12]). In the 2Nx2N case, the PU size is the same as the CU size. Note that inside a CU the prediction type is the same for all PUs (either Intra or Inter), however the prediction mode can be different for each PU (see Sec.2.2.3).

— **The Transform Unit (TU)**: After the prediction step, each CU is then transformed and quantized. As for the PUs, the CU can be further split into TUs. Similarly to the CU partitioning process in Figure 2.6, the TU partitioning is made with a quadtree structure where the maximum size of the TU is 32x32 pixels and its minimum size is 4x4 pixels. Note that, the TU partitioning is independent from the PU partitioning.

### 2.2.3 HEVC Prediction

As explained in the previous section, each CU can be predicted with either intra-picture or inter-picture prediction and each PU can be predicted with several intra or inter prediction modes. This section details the different prediction types and modes in HEVC.

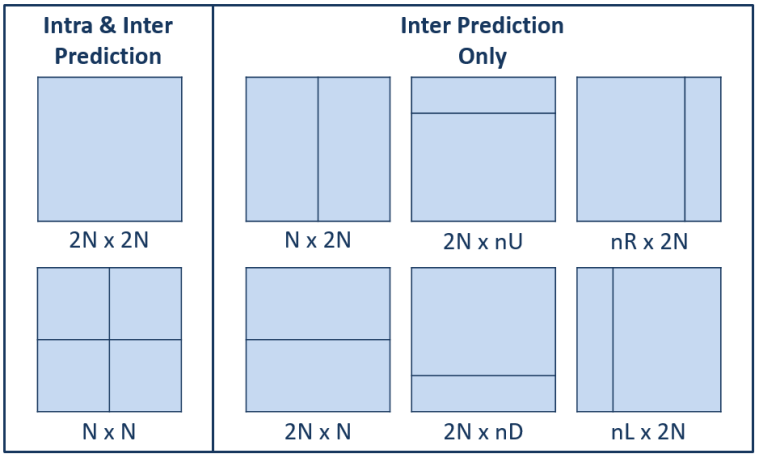



Figure 2.7 – CU partitionning into PUs depending on the prediction type

#### 2.2.3.1 Intra-Picture Prediction

Intra-picture prediction is briefly introduced in Section 2.1.2.1. It consists in creating a prediction for the current PU using previously encoded CUs from the same frame. Indeed, intra prediction exploits the spatial correlations that can occur in a frame. To construct this prediction, the neighboring decoded pixels of the current PU are used. Due to the Z-scanning order, the neighboring pixels are located in the top and left CUs and, due to the quadtree approach, these neighboring CUs can be bigger than the current one, as for the CU 13 in Figure 2.6. Therefore, some neighboring pixels can be located under or to the right of the current PU. Figure 2.8(b) illustrates the neighboring pixels used to create the intra prediction. Note that when these right and bottom pixels are not available, they are padded from the available ones [SBS14]. In this way, all the intra prediction modes can be tested.

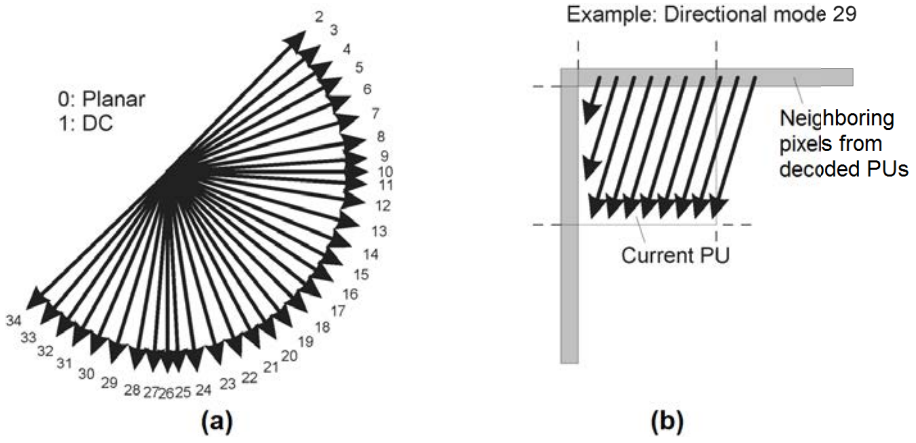



Figure 2.8 – Intra prediction modes in HEVC - (a) 35 modes - (b) Example for mode 29. Source: [SOHW12]

Indeed, the intra prediction can be derived with several modes. HEVC disposes of 35 intra prediction modes against a maximum of 9 for H.264/AVC: one DC mode, one planar mode and 33 directional modes. These directional modes, illustrated in Figure 2.8(a),

consist in propagating the neighboring pixels values in a given direction. An example for the mode 29 is provided in Figure 2.8(b). The DC mode simply consists in predicting all the pixels of the current block with the average value of all neighboring pixels. And finally, the planar mode, designed to prevent the discontinuities along the block boundaries that can occur with the DC mode, consists in averaging the horizontal (mode 10) and vertical (mode 26) prediction modes [SBS14].

For each PU, the best [2] intra prediction is selected at the encoding side and signaled in the bit stream using only the chosen mode index. At the decoding side, the prediction reconstruction is determined by this index and reconstructed with the previously decoded PUs. Note that, unlike the 35 available modes for luma component, only 5 modes can be used for the chroma components: DC, planar, horizontal (mode 10), vertical (mode 26), and the direct mode. The direct mode consists in using the same mode as for the luma. Indeed, the spatial correlation between luma and chroma components is reflected in the choice of the prediction mode for each component.

#### 2.2.3.2 Inter-Picture Prediction

Conversely to intra-picture prediction which exploits the spatial correlation in still image, inter-picture prediction takes advantage of the temporal correlations in video. In HEVC, the inter prediction consists in performing a motion estimation and compensation at the PU level. The currently encoded PU is motion compensated with a reference block from a previously encoded frame selected with the motion estimation step. This motion estimation step is usually performed with a block matching algorithm where several reference blocks are evaluated and the best one is selected. The chosen reference block is then signaled in the bit stream by a motion vector and by its frame number to allow the same motion compensation at the decoding side. In the case of bi-prediction, as introduced in Section 2.2.1, the prediction is performed using two blocks potentially coming from different frames. Therefore, bi-prediction is signaled with two motion vectors and two frame numbers.

The motion vector describes the horizontal and vertical displacements of the reference block according to the position of the current PU with a quarter pixel precision. For full pixel precision, the reference block is directly available in previously encoded/decoded frames. However, in half or quarter pixel precision, the reference block needs to be interpolated with eight and seven-tap filters for the luma component and four-tap filters for the chroma components [SOHW12].

As in AVC/H.264, HEVC proposes a Weighted Prediction (WP) algorithm to compensate for strong light changes that can occur in videos. Indeed, this functionality improves the temporal correlation between the different frames of the video and thus, improves the inter prediction. Note that, Chapter 5 focuses on the HEVC Weighted Prediction and describes it with more details.

The main advancement of HEVC for inter prediction, is the method used to predict the motion vectors. Indeed, the Advanced Motion Vector Prediction (AMVP) of HEVC consists in predicting the current motion vector with the ones from previously encoded neighboring PUs (inter-predicted only). Additionally to this AMVP mode, HEVC also defines two specific other modes: merge and skip. For these ones, no motion vector is

2. given a dedicated criterion

signaled in the bit stream. It is directly derived from the previously encoded/decoded neighboring PUs (inter-predicted only). A list of motion vectors is constructed and only the index of the chosen one is sent in the bit stream. The main difference of the skip mode, compared to the merge one, is to skip the transform and quantization part. The reference block is consider sufficient to reconstruct the current PU and no residual data is sent in the bit stream. The skip mode is very efficient for scenes with static background, where the collocated blocks are identical.

#### 2.2.3.3 Mode Selection

Either for intra-picture or inter-picture prediction, it is not easy to guess the most efficient mode among the available ones. Especially since it should be jointly chosen with the block partitioning structure. The best combination of partitioning and modes is not the one providing the lower distortion but rather the one optimzing a Rate-Distortion (RD or RDO for Rate Distortion Optimization) cost [SW98]. This RD cost allows to obtain the best trade-off between rate and distortion. It is usually noted $J$ and computed as a Lagrangian function:

$$J = D + \lambda.R \tag{2.2}$$

where $D$ is the distortion between the current block and the decoded one (after the transform/quantization step), usually computed with the mean square error (MSE). $R$ is the rate/size of the block. And $\lambda$ is a Lagrangian multiplier depending on the quantization parameter.

In the HM software implementation, an exhaustive evaluation among all available modes and block partitioning structures is performed. Starting from the largest CU size (same size as the CTU) to the lowest one, the encoder checks all the available intra/inter modes for each partitioning and select the combination partitioning/modes bringing the smallest RD cost. However, for the inter prediction, a huge amount of reference block is available. All of them can not be evaluated, especially with a RD cost. The motion estimation, which consists to find the best motion vector for the current PU, is usually performed with a block-matching or template-matching algorithm [HCT+06] that evaluates a subset of the available blocks.

Note that the mode selection is not defined by the HEVC standard and any other method than a RDO can be used. Indeed, the mode selection is only performed at the encoding side since the final choice of partitioning and modes is signaled in the bit stream. At the decoding side, the prediction reconstruction is straightforward and therefore faster.

### 2.2.4 HEVC Transform and Quantization

After the prediction step, we obtain a residual CU which can be further split in TUs, as explained in Section 2.2.2. These TUs are then transformed with a Discrete Cosine Transform (DCT) in order to decorrelate the coefficients of the residual CU. The DCT is usually preferred in compression standards since it has the advantage of being separable. It can be computed independently on each line and then on each column of the TUs. However, in the particular case of 4x4 PU predicted with intra modes, the Discrete Sine Transform (DST) is preferred since it provides better results in that specific case.

In comparison to previous standards, the best HEVC improvement in the transform part, is due to the TU partionning. However, several new tools have also been introduced to treat and signal the residual information such as mode dependent scanning, significance map, coefficient level and sign coding [SJN+12]. In AVC/H.264, only the zigzag scan was available to scan the transformed TU, but in HEVC, several scan patterns are available: diagonal, horizontal and vertical. While the diagonal one is used for all block types, the horizontal and vertical mode are restricted to intra predicted block.

After being transformed, the TU is scaled to give more weight to lower frequency coefficients, as advised for the HVS (see Sec.2.1.2.2), and quantized. A fixed quantization step $Q_{step}$ is used over these transform values and is defined as a function of a Quantization Parameter (QP) ranging from 0 to 51:

$$Q_{step} = 2^{\frac{QP-4}{6}} \tag{2.3}$$

The quantization step is designed such that every 6 QP values, the quantization step size doubles. The amount of distortion and the resulting rate/size of the block is controlled by this QP value. Indeed, as explained in Section 2.1.2.2, only the quantization step introduces the distortion.

### 2.2.5 In-Loop Filtering

In most cases, the rate/size constraint of broadcast network is rather low and this requires the use of high QP. Therefore, a strong quantization is applied to the content and it can generate several artifacts such as banding in uniform areas or ringing around strong edges in the frame [ZZRW14]. These artifacts are characterized by the appearance of lines that are not present in the original content. Due to the block structure of compression schemes and that each block is treated differently, the block boundaries are also very prone to the appearance of artifacts, commonly referred to as blocking since it is possible to distinguish the blocks in the image. Some examples of these artifacts are illustrated in Figure 2.9.


Figure 2.9 – Compression Artifacts - (a) Blocking - (b) Ringing. Source: [ZZRW14]

To attenuate these banding, ringing and blocking artifacts, two filters are performed during the HEVC encoding: the DeBlocking Filter (DBF) [NBF+12] applied first, and the Sample Adaptive Offset (SAO) [FAA+12]. The DBF, already present in H.264/AVC is designed to reduce the blocking artifacts. It simply consists in performing an adaptive

filtering to the vertical then to the horizontal block boundaries. Indeed, depending on the smoothness of these boundaries, there is either no filter, a normal filter or a strong filter. The SAO, firstly introduced in HEVC, attenuates the banding and ringing artifacts. It consists in selecting regions of the frame containing one or multiple CTUs where two SAO types can be applied: Edge Offset (EO) or Band Offset (BO). In both cases, each sample/pixel is classified into a category (5 for EO and 32 for BO) and a common offset is then added to each sample depending on its category. The offset is then signaled in the bit stream for the decoder.

Since SAO and DBF focus on different artifacts, their benefits are additive and they are usually both applied. They are referred to as In-Loop filters since they are included in the encoding and decoding loop (See Fig.2.4). Indeed, the reference frames are filtered before performing the inter prediction.

### 2.2.6 HEVC Entropy coding

Once a CTU is predicted, transformed and quantized, all the signaling information (Partitioning, prediction modes, transform, SAO, ...) and the remaining coefficients of the CTU (residual transformed and quantized) are converted into binary data and fed into an entropy coder.

In HEVC, several entropy coders are used to compress and transmit the binary data such as Variable Length Coding (VLC) or Context-adaptive binary arithmetic coding (CABAC). All information at the CTU level is entropy coded with the CABAC while the high level syntax (HLS) information (for the entire frame, GOP or Sequence) are entropy coded with a VLC. The amount of information of the HLS is much lower than the CTUs one. For this reason, the CABAC is usually considered as the only entropy coder of HEVC.

The CABAC was firstly introduced in AVC/H.264 and kept in HEVC since it provides a higher coding efficiency than a VLC. On the other hand, CABAC requires a larger amount of processing. It consists in three main steps: binarization, context modeling and arithmetic coding [SB12]. The binarization is a step where all the information is converted to binary symbol using different methods. Then for a sequence of these binary symbols, the context modeling keeps track of the probability of each symbol, depending on a specific context. And finally, the arithmetic coding [WNC87] compresses these binary symbols using the probability information.

## 2.3 Encoding techniques for HDR content

All previously presented standards were developed to compress SDR content. Instead of creating new compression standards to compress HDR content, the main strategies has been to adapt these content to the legacy compression standards such as JPEG or HEVC. To do so, several methods were proposed and can be classified in two categories. The first category, detailed in Section 2.3.1, includes high bit-depth solutions that focus only on HDR and are therefore only compatible with new decoders. The second category, detailed in Section 2.3.2, are the backward compatible approaches where the HDR content is transformed into an SDR one and thus compatible with legacy decoders and SDR displays.

### 2.3.1 High Bit-Depth HDR Compression

SDR content are always gamma corrected (See Sec.1.3.1.1), while HDR content are usually considered in linear-light values (in nits). As explained in Section 1.3, an efficient storage of HDR content requires to be transformed with an OETF. The high Bit-Depth HDR compression methods share many similarities with HDR image storing since they also consists in converting the linear-light HDR content into an uniform integer representation using an OETF. Once the HDR content is represented with 10, 12 or even 14 bits, it can be compressed with legacy SDR codecs supporting high bit-depth, such as AVC/H.264 or HEVC (See Sec.2.1.1).

Multiple solutions based on this concept were proposed such as [MKMS04] where the HDR content is transformed on a 11 bits representation and encoded with AVC/H.264. Another example is the solution provided by [MT10] which proposes to use the logluv OETF [Lar98], described in Section 1.3.2.3, to represent the HDR content with 14 bits and also encode it with AVC/H.264. These solutions are not optimal since the chosen OETF is not optimal. Moreover, in [MT10], the HDR content is represented with 14 bits while in Section 1.3.3 we demonstrated that 12 bits is enough with other OETFs.

More recently, an ad-hoc group was created by the ITU-T and the ISO/IEC to evaluate the possible extensions of HEVC for HDR and WCG compression. Many issues were addressed in this group and a Call for Evidence (CfE) was presented [LFH15] to determine if the current compression standards were sufficient for HDR content. In this CFE, the linear-light HDR content is transformed as described in Figure 2.10. The process is very

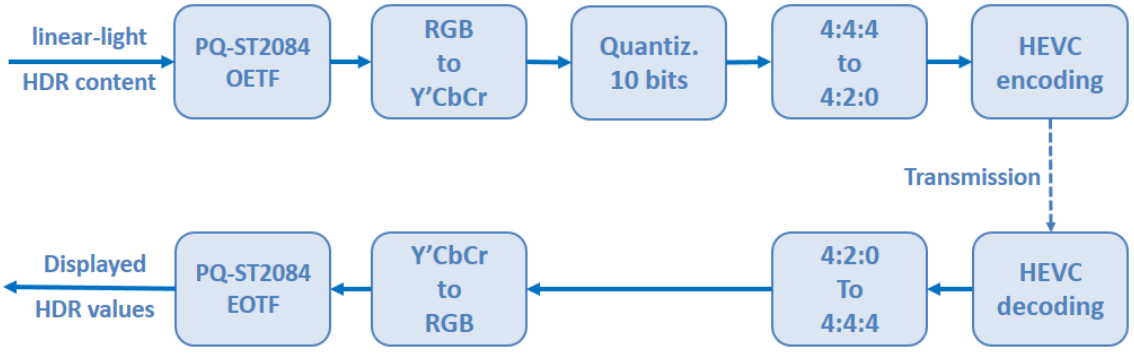



Figure 2.10 – Standardized HDR compression scheme

similar to the one for SDR compression. The quantization is made with 10 bits instead of 8 and the gamma correction is replaced by the PQ-ST2084 OETF [MND12] presented in Section 1.3.3.1. As a response to the CfE, many other OETFs were proposed, as explained in Section 1.3.3.2. Only the HLG [BC15b] was retained and standardized in the television Recommendation 2100 [ITU17].

These methods efficiently compress the HDR content however, they do not address backward compatibility with legacy SDR displays. Indeed, most consumers are equipped with SDR displays, that cannot decod HDR content compressed with new standards. Since the consumers will slowly renew their display to an HDR one, it is of great importance to deliver an HDR content that can be decoded by both SDR and HDR displays/decoders. The following section described the backward compatible HDR compression schemes.

### 2.3.2 Scalable and Backward Compatible HDR Compression

To address both SDR and HDR displays the simplest solution consists in performing an independent coding of both SDR and HDR contents. This method is called simulcast but it is clearly not optimal in terms of bandwidth utilization. Indeed, the SDR and HDR content share many similarities and therefore the two compressed contents contain many redundant information.

A better solution consists to use backward compatible or scalable compression solutions, as illustrated in Figure 2.11. The HDR and SDR contents are jointly compressed,

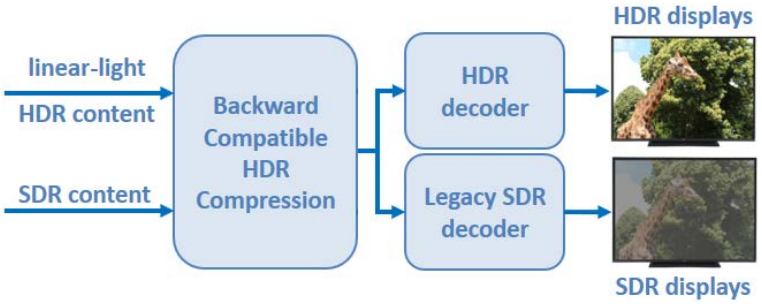



Figure 2.11 – HDR backward compatible compression

thus resulting in better compression performances. At the decoding side, both HDR and SDR displays/decoders can be addressed with the same bitstream. Note that with this method, the HDR and SDR contents come from different input, however the SDR content is usually created from the HDR one with TMOs (See Sec.1.4.2) or by manual grading.

Many backward compatible or scalable compression scenarios exist. Some solutions consider that the SDR and HDR input are supplied separately, such as in [MEMS06], while other methods [OA07, BDLC08] only use an HDR input which is tone mapped to generate the SDR content as illustrated in Figure 2.12. This latter approach is the most commonly studied in the literature, especially because the TMO can be designed for different purposes. The compression scheme in Figure 2.12 summarizes the typical backward compatible approach. After the TMO, the SDR content is encoded with a legacy SDR encoder to generate an SDR base layer compatible with SDR displays. To predict the HDR content, the SDR content is decoded, inverse tone mapped and subtracted from the HDR one. Note that to be invertible the TMO needs to be a global one (See Sec.1.4.2.1). We obtained a residual layer which can be encoded and send along with metadata containing the inverse TMO information. At the HDR decoding side, the SDR content is decoded and inverse tone mapped with the metadata, then the enhancement layer is decoded and used to reconstruct the HDR content.

Other backward compatible methods replace the inverse tone mapping box by a division like in JPEG-HDR [WS06] or in [LK12]. In this method based on the JPEG standard, each pixel value of the HDR image is divided by the pixel value of the SDR image. The enhancement layer thus contains a ratio image instead of a residual. This approach has the advantage to be compatible with non invertible TMOs.

Several solutions only encode and send the base layer, thus discarding the residual. In this case, the enhancement layer only contains the inverse TMO metadata. All these solutions differs by the used TMO. Indeed, some can be designed to preserve an artistic

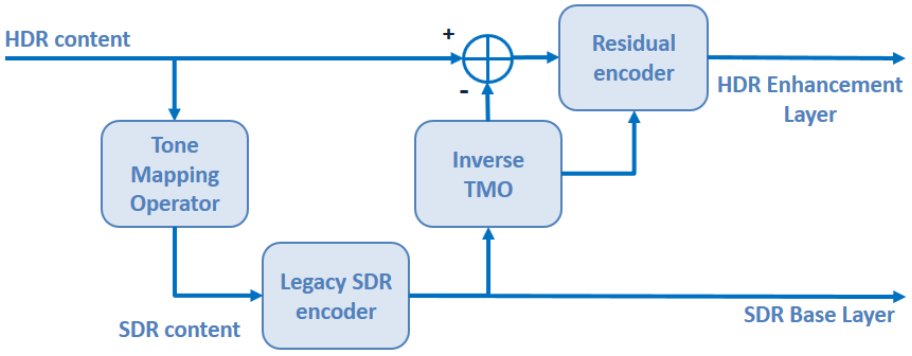



Figure 2.12 – Scalable and backward compatible compression scheme

intent producing a visually pleasing image [RSSF02, MDK08], while other can be designed to improve the compression performances [MMM+11]. These two goals are not always compatible and a compromise must be made.

Note that it's difficult to define a clear difference between HDR scalable compression and HDR backward compatible compression. The term scalable is usually reserved to the solutions considering the residual/ratio image in the enhancement layer.

## 2.4 Quality Evaluation

The quality evaluation is necessary to measure the amount of distortion brought by an encoder and its algorithms. It allows to assess the efficiency of compression standards or more simply the compression performances of a specific algorithm. One video compression standard/algorithm can be considered better than another one, if it provides a better quality for the same compressed rate/size or conversely a lower rate for the same quality.

The most evident way to evaluate image or video is to use subjective experiments as defined in [ITU07, ITU08]. In these methods, a panel of human observers assess compressed contents then the results are averaged in the form of a Mean Opinion Score (MOS). In this way, we obtain an estimation of the perceived distortion. These methods remains the most reliable, however they are generally impractical for multiple reasons.

Indeed, performing a subjective experiment implies to have a significant number of viewers and it takes a long time in set up and realization. In practice, new algorithms are tested with many parameters variations and on large video sets, thus needing many of these experiments. Therefore quality metrics quickly calculable are necessary. Such metrics are also referred to as objective metrics since they are more determinist than the subjective ones. Indeed MOS can slightly change depending on the viewer panel.

Objective quality metrics can be based on simple image difference or simulate the behavior of the HVS to detect the perceptual differences between the compressed and uncompressed content. The latter ones have different levels of precision depending on how the HVS is modeled. Note that, all these metrics only modelize some aspects of the HVS and not its entirety.

In this section, we start by a brief overview of the most used SDR quality metrics in Section 2.4.1. Then, we explain how SDR quality metrics can adapt to HDR contents and we also present a popular metric dedicated to HDR in Section 2.4.2.

## 2.4.1 SDR Metrics

### 2.4.1.1 PSNR

The most straightforward metrics, to evaluate the differences between the compressed and uncompressed image, are the pixel-based ones such as the MSE or its derivative the Peak Signal to Noise Ratio (PSNR). The MSE is defined as follow:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{N} \Big( I_C(i) - I_S(i) \Big)^2}{N} \tag{2.4}$$

where $N$ is the number of pixels, $I_S$ the source image and $I_C$ the decoded compressed one. The PSNR, expressed in decibels (dB), is then derived with the following formula:

$$\mathrm{PSNR} = 10 \cdot log_{10} \left( \frac{(2^{nb} - 1)^2}{\mathrm{MSE}} \right) \tag{2.5}$$

where $nb$ is the bit depth of the SDR signal, so usually $nb = 8$. In this case, the PSNR value are typically around 30 dB and 50 dB where the higher one is the better. Indeed, if the MSE tends to 0, the PSNR tends towards infinity.

The PSNR is the most commonly used metric to measure the distortion brought by lossy compression. However, the MSE and PSNR metrics have rather limited consideration for the HVS. The principal aspect is to consider a perceptually uniform representation such as gamma correction (see Sec.1.3.1.1) and color opponent spaces (see Sec.1.3.1.3). In HEVC and previous standards, the PSNR is computed with the Y'CbCr color space, independently for the three components and averaged with different weights. Since chroma components are subsampled and since the HVS is more sensible to variations in the luma component (see Sec.1.1.5.4), the PSNR is weighted as follows:

$$\mathrm{PSNR}_{\mathrm{Y'CbCr}} = \frac{1}{8} \cdot \Big( 6.\mathrm{PSNR}_{\mathrm{Y'}} + \mathrm{PSNR}_{\mathrm{Cb}} + \mathrm{PSNR}_{\mathrm{Cr}} \Big) \tag{2.6}$$

Note that for videos, the PSNR or other metrics are computed independently for each frame and averaged for the entire sequence. Few metrics are dedicated to video and they are rarely used in practice.

### 2.4.1.2 SSIM

The Structural SIMilarity (SSIM) is one of the most popular and influential quality metrics in recent years. It is a structure based metric designed to provides a better correlation with the MOS results than the PSNR [WB02]. In contrast with the pixel-wise PSNR, the SSIM is computed on blocks of the image displaced pixel by pixel. This method allows to detect structural or contrast changes. Since the HVS is specialized in detecting structural information, this metric can well approximate the perceived image quality.

Considering $x$ a block in the compressed image and $y$ the colocated block in the source image, the SSIM for the block is:

$$\mathrm{SSIM(x, y)} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{2.7}$$

where $\mu_x$ and $\mu_y$ are the average value of the blocks $x$ and $y$, $\sigma_x^2$ and $\sigma_y^2$ the variance of each block, $\sigma_{xy}$ the co-variance, $c1 = \big(0, 01.(2^{nb} - 1)\big)$ and $c2 = \big(0, 03.(2^{nb} - 1)\big)$. The typical size of the block $x$ and $y$ is 8x8. All block SSIM values are then averaged for the entire frame. As for the PSNR, the SSIM is computed independently for the three Y'CbCr components and averaged with the same weights.

#### 2.4.1.3 Bjontegaard Computation

It's difficult to compare compression standards or algorithms using only the distortion. Indeed, it implies to have two images/videos encoded with the same rate and see which encoder introduces the least distortion. In practice, it is extremely difficult to have two contents encoded with different encoder and with the exact same rate.

To overcome this problem, Bjontegaard [Bjo01] proposed to compute an average difference between Rate-Distortion (RD) curves. Instead of comparing two encoders with one rate, the comparison is made with several ones. The idea is to encode the content with both encoder at different rates and to compute the resulting distortions in each case. The RD points are then plotted on a 2D graph, as illustrated in Figure 2.13 and the curves are plotted with a 3rd order polynomial interpolation. The gain from one encoder over

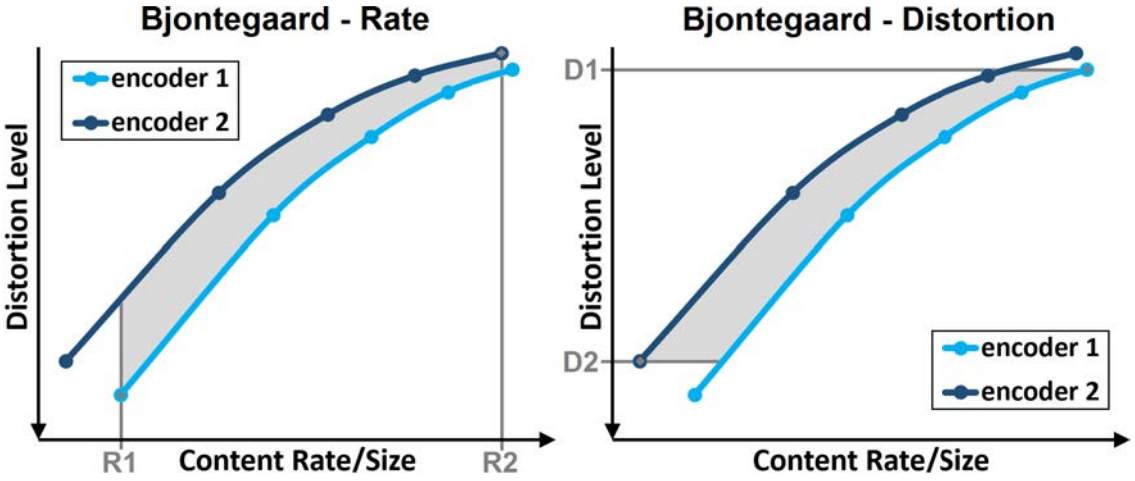



Figure 2.13 – Bjontegaard computation in rate and distortion

the other one is then computed as an averaged distortion or rate based on the area between the two RD curves. This area can be computed in two ways, either at equivalent rate or at equivalent distortion. In the first case, the BjontegaarD rate (BD-rate) gain is computed between the points R1 (maximum of the minimum rates for each encoder) and R2 (minimum of the maximum rates for each encoder) and for the BD-distortion, the gain is computed between the points D1 (maximum of the minimum distortions for each encoder) and D2 (minimum of the maximum distortions for each encoder).

The Bjontegaard computation is widely used for the evaluation of compression standards or algorithms. Indeed, this method allows to make comparisons at different rate levels. In some cases, new algorithms only improve the RD performances at low or high rates. Therefore, computing the distortion at only one rate point does not reflect the impact on the compression. Another advantage of this method is to adapt to every distortion metric such as the PSNR or the SSIM.

## 2.4.2 HDR Metrics

### 2.4.2.1 Based on Legacy SDR Metrics

As stated in Section 2.3.1, the main difference between HDR and SDR content is that the HDR one provides linear-light values (in nits), while the SDR one is always corrected with an OETF (gamma correction). Directly apply SDR metrics to the linear-light HDR content provides distortion values poorly correlated to the human quality perception. The first methods to evaluate the quality for HDR content were based on multiple exposures techniques. It consists in decomposing the HDR images in multiple SDR ones with different exposure values, and averaging the SDR distortions. These methods were more efficient than computing a distortion on the linear-light HDR content but they remains complex and unnecessary.

Indeed, as for HDR compression, legacy SDR metrics can be used on perceptually uniform HDR representations, such as the ones described in Section 1.3.3. This method was proposed in [AMS08] where the compressed and uncompressed HDR images are transformed with an OETF in a Perceptually Uniform (PU) representation and evaluated with the PSNR or the SSIM. These metrics are refereed to as PU-PSNR or PU-SSIM. Other ones based on the same concept exist, such as the logPSNR, the PQ-ST2084 PSNR (refereed to as HDR-PSNR in the following chapters), or the tPSNR advised in the HDR CFE [LFH15]. The tPSNR combines two OETFs, where one of them is the PQ-ST2084 therefore, tPSNR and PQ-ST2084 PSNR are relatively similar.

These metrics are efficient, however they suffer the same defects than SDR metrics. Indeed, PSNR and SSIM do not account for many properties of the HVS. With the advent of HDR, a new metric based on many features of the HVS was put in light, the HDRVDP2, which is discussed below.

### 2.4.2.2 HDRVDP2

The HDRVDP2 [MKRH11] is a quality metric accurately measuring the early stages of the HVS. It is based on a comprehensive model of discrimination and detection, which has been calibrated and validated on datasets of psychophysical experiments [Man16]. Therefore, the metric is rather complex and includes many steps as illustrated in Figure 2.14.

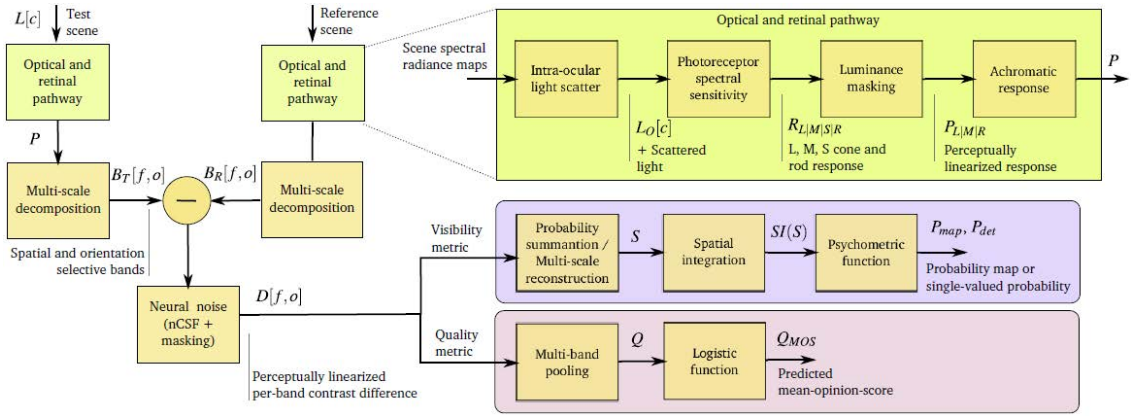



Figure 2.14 – Block diagram of HDRVDP2 process. Source:[MKRH11]

The HDRVDP2 accounts for many vision phenomena, such as scattering of the light in the eye and optics (glare), photoreceptors vision, luminance masking, spatial contrast sensitivity, contrast masking, neural noise in visual channels and contrast consistency [Man16]. This metric allows the detection of many artifacts ignored by most of standard metrics however, it is not computationally efficient and it requires several additional input such as the display brightness, the viewing distance, and the display size. In the end, the HDRVDP2 provides a visibility map ($P_{map}$ and $P_{det}$ in Figure 2.14) where each pixel is associated with the probability of detecting a difference, and a global quality score ($Q_{MOS}$ in Figure 2.14) between 0 and 100 (the higher the better).

Most studies show that the HDRVDP2 outperforms other metrics such as the PU-PSNR or PU-SSIM however, the test set is usually limited and it is therefore difficult to draw general conclusions [Man16].

## 2.5 Summary

This chapter introduced the motivations behind content compression and the basic functioning of popular compression standards. The most recent HEVC standard is then further detailed with some of its major improvements. While the standardized HDR compression scheme remains very similar to legacy SDR compression, backward compatible techniques introduces several interesting concepts. This latter method must be emphasized to allow a smooth transition from SDR to HDR displays. Finally, several solutions are presented to assess the quality of decoded SDR and HDR contents.

The following chapters focus on backward compatible compression solutions. Several works from the literature are more precisely described and compared with the proposed work.

# Chapter 3

# Rate-Distortion Optimization of a Tone Mapping with SDR Quality Constraint for Scalable HDR Compression

As explained in Section 1.4.2, HDR contents need to be tone mapped to be visualized on legacy SDR displays. Many tone mapping operators (TMO) have been developed to convert an HDR content into an SDR one, with the goal of preserving the perception of the content or the artistic intent. One example of TMO is the photographic tone reproducer (PTR) [RSSF02], shown to be one of the two TMO that consistently performs well and to be the best in terms of visual quality and image integrity [LCTS05, SbMEMS12].

In this chapter, we consider the problem of SDR backward compatible coding of HDR content with two layers. In the proposed compression scheme, as introduced in Section 2.3.2, the input HDR content is first tone mapped with a global TMO into an 8 bit SDR version which is encoded with a legacy SDR codec and send as a base layer. The decoded SDR content is then inverse tone mapped and subtracted from the input HDR one to yield an enhancement layer containing the HDR residue which is separately encoded. While the TMOs introduced in Section 2.3.2 aim at preserving the perception and artistic intent when converting the content, one can also try to design the TMO in order to optimize the compression efficiency. The question of optimizing TMOs for compression purposes has already been addressed in the literature [MMM+11, LVCD14, PGT14]. However, in all these works no enhancement layer is considered. However, due to the similarities with our context, we first review this work.

The authors in [MMM+11] design a TMO parameterized by a piecewise linear curve which minimizes the distortion between the HDR content and the inverse tone mapped HDR content. In their work, the HDR distortion is computed as the MSE in a roughly perceptually uniform representation (logarithmic values, See Sec.1.1.3 and Sec.1.3.3) of the HDR luminance values. A closed form expression is then derived for the slopes of the TMO curve as a function of the HDR histogram.

In [LVCD14] and [PGT14], the authors notice that a TMO minimizing only the HDR distortion is not optimal in terms of Rate Distortion (RD) performances. Indeed, the approach of [MMM+11] can increase the spatial complexity of the generated SDR content and thus requires a higher rate. For this reason, the [MMM+11] approach is extended in [LVCD14] and [PGT14] in order to minimize a RD cost, considering the rate cost of the

SDR layer only. In [PGT14], this rate cost is considered proportional to the entropy of the generated SDR content while in [LVCD14], it is considered as a function of the total variation [ROF92] of the SDR content.

Optimizing the TMO only for compression purposes may yield SDR content in which the perception and artistic intent of the HDR content is not preserved. Therefore, [KD13] propose to trade-off between compression performances and perceptual quality of the SDR content. To do so, they extend the HDR distortion minimization of [MMM+11] by adding a constraint on the MSE between the tone mapped SDR signal and a reference SDR content. However, as for [MMM+11], improving the compression performances by considering only the HDR distortion is not optimal in terms of RD performances.

In addition, all the previous works propose TMOs optimized in a backward compatible compression scheme where the HDR residual is discarded. They consider the inverse tone mapped HDR content to be of sufficient quality for HDR display. However, this is not the case when the base SDR layer is represented on only 8 bits. Indeed, the compression artifacts generated in the SDR content may be amplified in the HDR one, depending on the slopes values of the TMO. Therefore, it is important to consider a scalable compression scheme in which the HDR residue is encoded when optimizing the TMO. In [MMNW13], the tone mapping curve is optimized in order to minimize a total (base plus enhancement layers) bit rate with a constraint for the perceptual quality of the SDR content. Unfortunately, their constraint for the base bitrate bring no or negative contribution to the coding performance and is therefore discarded.

The SDR perpceptual quality constraint in [MMNW13] is based on the same principle as in [KD13] and consists in a MSE distance between the uncoded tone mapped SDR signal and a reference SDR content. However, in [MMNW13], they chose a specific weighting in their optimization problem that leads to a varying perceptual quality of the SDR signal depending on the input HDR signal.

In this chapter, we present a new TMO that optimizes the total distortion of the HDR scalable coding scheme by taking the rate of both the base and enhancement layers into account, under a constraint on the perceptual quality of the compressed SDR signal. We first show that, given a constraint on the sum of the SDR and HDR rates, at high rates, the optimal TMO that minimizes the HDR distortion only depends on the SDR signal rate. It is then shown that this optimization problem can be decomposed into two consecutive minimization steps: (i)- a minimization of the enhancement layer variance depending on the TMO and the base layer bitrate; (ii)- and a minimization of the total HDR distortion depending on the rate repartition between the base and the enhancement layers. Part of this chapter has been published in [GRG+16].

The remainder of this chapter is organized as follows. Section 3.1 presents the selected backward compatible HDR compression scheme and the model considered in this Chapter. The tone mapping optimization problem is then formalized in Section 3.2 along with the proposed models for each term. This problem is then simplified using several assumptions detailed in the same section. The TMO is then parameterize as a piecewise linear function to obtain the solution expressed in Section 3.3. Then, Section 3.4 presents the experimental results in comparison with several other TMOs using either a single or dual layer configuration. Finally, the conclusions are drawn in Section 3.5

## 3.1 Proposed Scalable Compression Scheme and Modelisation

### 3.1.1 HDR uniform representation and SDR gamma correction

In this section, we present the chosen scalable or backward-compatible compression scheme. As explained in section 2.3.2, it consists in compressing HDR content with two layers by using a tone mapping and inverse tone mapping process. Indeed, a residual is computed between the input HDR image and the inverse tone mapped one, instead of a ratio image computed by dividing the input HDR image by the SDR one. However, several methods can be used, as explained below.

Since classical TMOs such as [RSSF02] operate on linear light HDR values, the classical backward compatible compression scheme [OA07, BDLC08] treats linear light HDR values. This scheme is illustrated in Figure 3.1. In this case, the TMO process usually in-

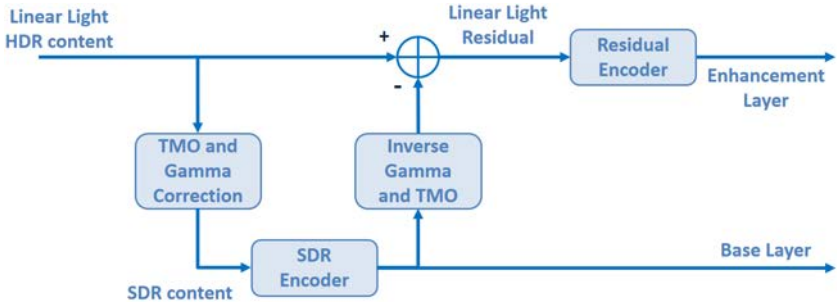



Figure 3.1 – Classical Backward-compatible compression scheme

cludes a gamma correction step to account for the SDR display response (see Sec.1.3.1.2). After the inverse Gamma and inverse TMO step, the residual is computed using linear light values. However, the compression of this residual is really complex. Indeed, since HDR images are not interpreted with perceptually uniform representations, the residual values are not linked to the human eye contrast sensitivity (See Sec.1.1.3 and Sec.1.3.3). For pixels with high luminance values in the HDR images, the corresponding residual values can be more strongly quantized than for low luminance values. Since the luminance level information is lost in the residual image, it is therefore difficult to compress it efficiently. Note that, this problem also transposes to compression schemes using a ratio image instead of a residual.

In contrast, recent TMOs, designed for compression purposes [MMM+11, MMNW13, KD13], start by transforming the linear light HDR content with an OETF (see Sec.1.3.3) into a perceptually uniform integer representation, as illustrated in Figure 3.2. In this case, since an OETF correction is already applied to the HDR content prior to the TMO, the classical gamma correction for the SDR content is no longer needed. But more importantly, after the inverse TMO, the residual computation is made in perceptually uniform representation. Therefore, this residual can be quantized uniformly and compress more efficiently.

However, the OETF used in [MMM+11, KD13] is a logarithm ($log_{10}$) and the one used in [MMNW13] is a gamma function. As demonstrated in Section 1.3.3.1, these OETFs are

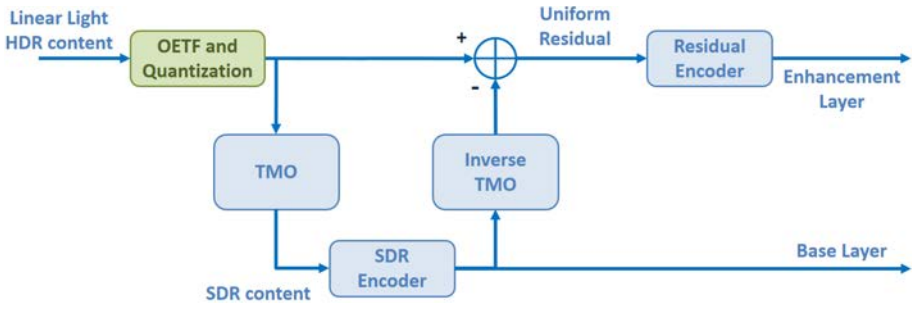



Figure 3.2 – Backward-compatible compression scheme for optimized TMO

clearly not perceptually uniform. Instead, the proposed compression scheme uses the PQ-ST2084 OETF [MND12] and a quantization on 12 bits. As explained in Section 1.3.3.1, this prior quantization has no perceptual impact on the HDR content.

A simple alternative to the compression scheme illustrated in Figure 3.1 is depicted in 3.3. This solution allows to use classical TMO, that operates on linear light HDR values,

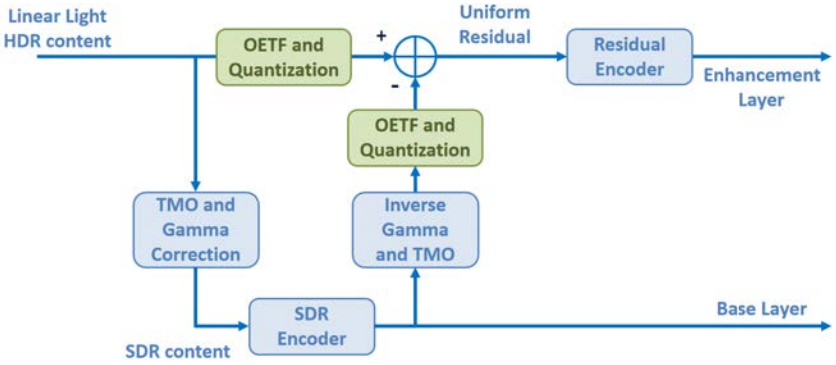



Figure 3.3 – Adaptation of the Classical Backward-compatible compression scheme

and to have a uniform representation of the residual.

In the following, HDR content is assumed to be in a perceptually uniform integer representation using the [MND12] OETF. For the sake of equality, the TMOs proposed in [MMM+11, MMNW13, KD13] are computed with this same OETF instead of the original non-optimal one they used.

### 3.1.2 Compression Scheme Model

This section presents the compression scheme model illustrated in Figure 3.4. The HDR input content consists of a sequence of luminance (or chrominance) pixel values. Therefore, we model the HDR signal as a random process denoted $X$ that generates sequences of luminance (or chrominance) values. For ease of presentation, and when there is no ambiguity, we omit the pixel index and use the same notation for a random process and one random variable of the process.

Then, the HDR signal $X$ is tone mapped (3.1) and quantized on 8 bits. It leads to a SDR signal with smaller dynamic range (See Fig.3.4), also modeled as a random process, denoted $Y$. The TMO, denoted $F$ is global (one HDR value $X$ leads to one SDR value

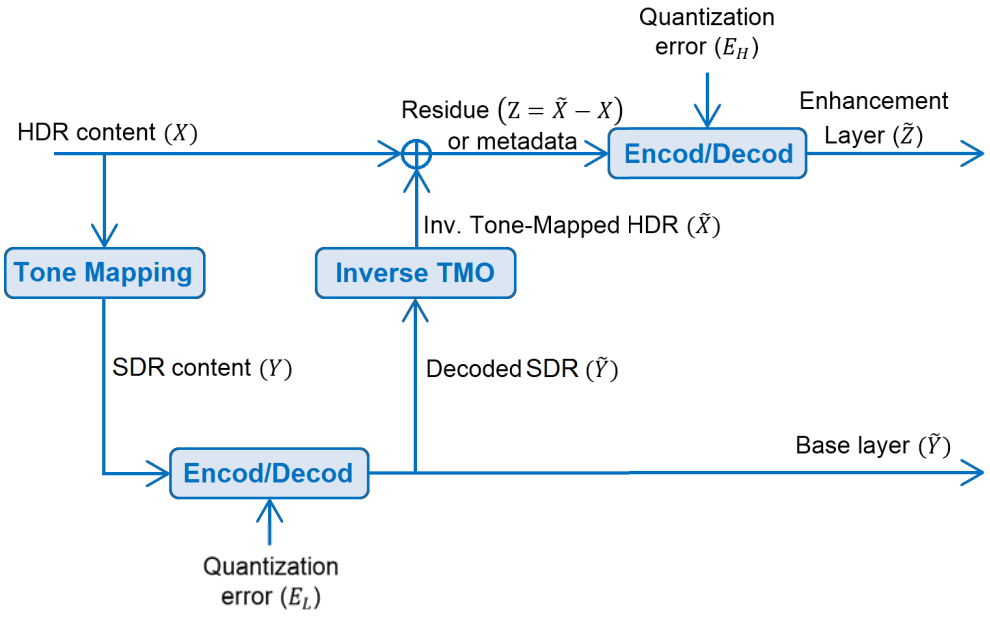



Figure 3.4 – Model of the HDR Backward-compatible compression scheme

$Y$) and is invertible such that $F^{-1}$ exists. The SDR is then compressed with a legacy SDR encoder, send as a base layer, and decoded to reconstruct a distorted SDR signal denoted $\tilde{Y}$. $E_L$ represents the quantization error brought by the SDR encoder and is defined by (3.2).

To compute the enhancement layer, the decoded SDR signal $\tilde{Y}$ is first inverse tone mapped to yield a distorted HDR signal $\tilde{X}$ (3.3), and a residue $Z$ is computed (3.4). Finally, this residue is compressed and the quantization error on the residue is $E_H$, (3.5). Therefore, the encoding process can be summarized as follows:

$$Y = F(X) \tag{3.1}$$
$$E_L = Y - \tilde{Y} \tag{3.2}$$
$$\tilde{X} = F^{-1}(\tilde{Y}) \tag{3.3}$$
$$Z = X - \tilde{X} \tag{3.4}$$
$$E_H = Z - \tilde{Z} \tag{3.5}$$

where (3.1) corresponds to the Tone mapping, (3.2) to the compression error on the SDR signal, (3.3) to the inverse tone mapping, (3.4) to the HDR residue computation and (3.5) to the encoding error on the residue.

The displayed content depends on which screen is available at the receiver and on which bitstreams are received. If the available screen is an SDR one, then only the base layer bitstream is decoded and the displayed video is $\tilde{Y}$. If the available screen is HDR, there are two use cases. If only the base layer bitstream has been received (due to bandwidth limitation for instance), then the displayed video is $\tilde{X}$. And if the viewer receives both the base and enhancement layer bitstreams, then the displayed video is:

$$\hat{X} = \tilde{X} + \tilde{Z} \tag{3.6}$$

## 3.2 Problem Statement and assumptions

### 3.2.1 Problem Statement

Knowing the compression, the goal is now to find the TMO $F$ that minimizes the distortion of the HDR signal $D_{HDR}$ for a given rate budget. Indeed, the base and enhancement layers are encoded at rate $R_{SDR}$ and $R_{HDR}$ and the optimization is performed under a global rate budget $R_T$ i.e. under the constraint that $R_{SDR} + R_{HDR} \leq R_T$. The RD function being decreasing [CT06, Ex10.5] with the rate $R$, the minimum distortion is obtained when the rate is maximal i.e. when the rate budget is fully utilized. Therefore, the rate inequality constraint can be turned into an equality constraint.

Most of the algorithms in the literature assume that the optimal TMO does not depend on the encoding distortions. The originality of this work considers these distortions in the global minimization problem. $E_L$ and $E_H$ values depend on the rate repartition among the two layers, impacting the global R-D performance. Depending on the total rate and chosen distribution, the distortions $E_L$ and $E_H$ take on different values and impact the R-D performances. Their influence is demonstrated in the proposed optimization problem in Section 3.2.4.

As in [KD13] and [MMNW13], this tone mapping optimization is performed while computing the perceptual quality $D_{SDR\_REF}$ of the resulting SDR image such that it does not exceed a maximum value $D_0$. This SDR quality constraint is computed as a distortion between the compressed SDR signal $\tilde{Y}$ and a provided SDR signal obtained with another TMO $G$ that preserves the perception and artistic intent of the HDR content. On the whole the optimization problem can be stated as:

$$\min_{F, R_{SDR}, R_{HDR}} \quad D_{HDR} \qquad (3.7)$$
$$\text{s.t.} \quad \begin{cases} D_{SDR\_REF} \leq D_0 \\ R_{SDR} + R_{HDR} = R_T \\ R_{SDR} \geq 0, R_{HDR} \geq 0 \end{cases}$$

To solve this optimization problems, we now express the distortions $D_{HDR}$ and $D_{SDR\_REF}$ as a function of the TMO curve $F$ in the following sections.

### 3.2.2 Proposed HDR distortion model

As explained in Section 3.1.1, X is considered as a uniformly perceptual representation. Thus, the Mean Square Error (MSE) is a relevant metric to estimate the quality of the HDR signal and therefore:

$$D_{HDR} = \mathbb{E}[(X - \hat{X})^2] \qquad (3.8)$$

One originality of the proposed work lies in the fact that (3.7) explicitly takes into account the compression of the enhancement layer. Indeed, this encoding is neglected in [MMM+11, KD13, MMNW13] since the HDR distortion is measured between the original ($X$) and inverse tone mapped ($\tilde{X}$) HDR signal. First, this causes a mismatch since the TMO is optimized in order to have the best inverse tone mapped HDR $\tilde{X}$, which differs from the HDR signal that is available at the receiver $\hat{X}$. Second, the rate to encode the

enhancement layer is not used as a parameter that can be optimized in order to perform a rate allocation between the base and enhancement layer.

To further detail the HDR distortion term (3.8), note that, the RD performances for the compression of the SDR signal (3.2) and of the HDR residue (3.5) is well approximated by its Shannon bound. Therefore:

$$\mathbb{E}[(Y-\tilde{Y})^2] = \alpha \cdot \sigma_Y^2 \cdot 2^{-2R_{SDR}} \tag{3.9}$$

$$\mathbb{E}[(Z-\tilde{Z})^2] = \beta \cdot \sigma_Z^2 \cdot 2^{-2R_{HDR}} \tag{3.10}$$

where $\sigma_Y^2$ and $\sigma_Z^2$ are the variances of the SDR signal $Y$ and of the residue $Z$, and where $\alpha$ and $\beta$ are constants depending on the type of distribution and quantizer. Indeed, this model holds for many quantizers and in particular for high rate scalar quantization with variable length codes of i.i.d. but also Markovian processes [WS11, Chap. 5]. Therefore, we have used this model as an approximation for the compression of data with HEVC although HEVC is not strictly a scalar quantizer.

Now, by construction of the two-layer compression scheme, we have that the quantization error on the residue equals the error on the HDR signal. Indeed, from (3.4) and (3.6), we have $X - \hat{X} = Z - \tilde{Z}$. Therefore, from (3.8) and (3.10):

$$D_{HDR} = \mathbb{E}[(Z-\tilde{Z})^2] = \beta \cdot \sigma_Z^2 \cdot 2^{-2R_{HDR}} \tag{3.11}$$

Using equation (3.1) to (3.4):

$$Z = X - F^{-1}\big(\tilde{Y}\big) \tag{3.12}$$

$$Z = X - F^{-1}\big(F(X) + E_L\big) \tag{3.13}$$

Following the first order Taylor expansion of the function $F$ as in [PGT14], we have at high rates (i.e. when the error $E_L$ takes small values):

$$Z \approx X - \Big(F^{-1}\big(F(X)\big) + E_L \cdot F^{-1\prime}\big(F(X)\big)\Big) \tag{3.14}$$

$$Z \approx -E_L \cdot F^{-1\prime}\big(F(X)\big) = \frac{-E_L}{F'(X)} \tag{3.15}$$

where $'$ stands for the first derivative.
We further assume that the quantization error is independent of the compressed signal. This is a usual assumption, merely made for convenience. However, this can be justified by the fact that the optimal quantization of an infinite length Gaussian signal is achieved when the quantization noise is Gaussian and independent of the compressed signal [CT06, Chap. 10]. From the independence between the residue $E_L$ and $X$, and assuming that $Z$ has zero mean, we have:

$$\sigma_Z^2 = \mathbb{E}[Z^2] = \frac{\mathbb{E}[E_L^2]}{\mathbb{E}\left[F'(X)^2\right]} \tag{3.16}$$

Another originality of the proposed work lies in the fact that the quantization error $E_L$, that depends on the rate, has a significant impact on the overall R-D performances.

State-of-the-art algorithms do not account for this distortion and therefore are sub-optimal. Indeed, the encoding of $Y$ (3.9) yields:

$$\mathbb{E}[E_L^2] = \alpha \cdot \sigma_Y^2 \cdot 2^{-2R_{SDR}} \tag{3.17}$$

$$= \alpha \cdot \mathbb{E}\left[\left(F(X) - \mathbb{E}[F(X)]\right)^2\right] \cdot 2^{-2R_{SDR}} \tag{3.18}$$

Finally, from (3.16) and (3.18), the HDR distortion (3.11) can be rewritten as a function of only the HDR image $X$, the tone mapping $F$ and the rates of both layers $R_{SDR}$ and $R_{HDR}$:

$$D_{HDR} = \alpha\,\beta \cdot \mathbb{E}\left[\left(F(X) - \mathbb{E}[F(X)]\right)^2\right] \cdot \mathbb{E}[F'(X)^{-2}] \cdot 2^{-2(R_{SDR}+R_{HDR})} \tag{3.19}$$

### 3.2.3 Proposed SDR quality constraint model

As for [KD13] and [MMNW13], we chose to measure the SDR perceptual quality constraint $D_{SDR\_REF}$ of the resulting SDR signal using a MSE with a reference SDR signal $G(X)$ obtained from the HDR image $X$ with $G$ a TMO preserving the perception and artistic intent. However, we choose to compute this MSE with the coded SDR signal $\tilde{Y}$ while previous works use the uncoded SDR signal $Y$. Indeed, we noticed that the SDR quality constraint is usually not respected after the encoding step in [KD13] and [MMNW13], especially at low bitrates. Therefore:

$$D_{SDR\_REF} = \mathbb{E}[(G(X) - \tilde{Y})^2] \tag{3.20}$$

Using equation (3.1) and (3.2):

$$D_{SDR\_REF} = \mathbb{E}[(G(X) - F(X) - E_L)^2] \tag{3.21}$$

Reusing the independence assumption introduced in (3.16), we obtain:

$$D_{SDR\_REF} = \mathbb{E}[(G(X) - F(X))^2] + \mathbb{E}[(-E_L)^2] \tag{3.22}$$

And finally, with (3.18), the distortion between the SDR signal and the reference can be derived as:

$$D_{SDR\_REF} = \mathbb{E}\left[\left(G(X) - F(X)\right)^2\right] + \alpha\mathbb{E}\left[\left(F(X) - \mathbb{E}[F(X)]\right)^2\right]\ 2^{-2R_{SDR}} \tag{3.23}$$

### 3.2.4 Optimization Problem

Using the proposed models for the HDR distortion (3.19) and for the SDR quality constraint (3.23), the optimization problem (3.7) becomes:

$$\min_{F,R_{SDR},R_{HDR}} \alpha\,\beta \cdot \mathbb{E}\left[\left(F(X) - \mathbb{E}[F(X)]\right)^2\right] \cdot \mathbb{E}[F'(X)^{-2}] \cdot 2^{-2(R_{SDR}+R_{HDR})} \tag{3.24}$$

$$\text{s.t.} \quad \begin{cases} \mathbb{E}\left[\left(G(X) - F(X)\right)^2\right] + \alpha\mathbb{E}\left[\left(F(X) - \mathbb{E}[F(X)]\right)^2\right] \cdot 2^{-2R_{SDR}} \leq D_0 \\ R_{SDR} + R_{HDR} = R_T \\ R_{SDR} \geq 0, R_{HDR} \geq 0 \end{cases}$$

**Lemma 1.** Consider a backward compatible compression scheme to encode a HDR signal into two layers (SDR and HDR), while preserving a small distortion between the SDR signal and a reference. A rate constraint is added on the sum of the rates for the SDR and HDR layers. At high rates, the optimal tone mapping $\tilde{F}$ that minimizes the HDR distortion only depends on the rate needed to encode the SDR signal.

**Proof.** From (3.19) and the sum rate equality constraint $R_{SDR}$+$R_{HDR}$=$R_T$, the optimization problem (3.24) is equivalent to first optimizing the tone mapping for each $R_{SDR}$:

$$\begin{aligned} F^*_{R_{SDR}} = \arg\min_F \mathbb{E}\left[\big(F(X) - \mathbb{E}[F(X)]\big)^2\right] \cdot \mathbb{E}\big[F'(X)^{-2}\big] \\ \text{s.t. } \mathbb{E}\left[\big(G(X) - F(X)\big)^2\right] + \alpha\mathbb{E}\left[\big(F(X) - \mathbb{E}[F(X)]\big)^2\right] \cdot 2^{-2R_{SDR}} \leq D_0 \end{aligned} \tag{3.25}$$

and then finding the optimal rate allocation between the base and enhancement layers:

$$\begin{aligned} \min_{R_{SDR}} \mathbb{E}\left[\big(F^*_{R_{SDR}}(X) - \mathbb{E}[F^*_{R_{SDR}}(X)]\big)^2\right] \cdot \mathbb{E}[F^{*'}_{R_{SDR}}(X)^{-2}] \\ \text{s.t. } 0 \leq R_{SDR} \leq R_T \end{aligned} \tag{3.26}$$

The first optimization problem (3.25) consists in minimizing the distortion between the HDR content and the inverse tone mapped HDR content under a constraint for the SDR perceptual quality. This problem is equivalent to the ones solved in [MMM+11, KD13, MMNW13], i.e. it searches to minimize the variance $\sigma_Z^2$, but with a different model. For clarification, a comparison is made in the following section. The second optimization problem (3.26) minimizes the distortion between the HDR content and the reconstructed HDR content using two layers.

### 3.2.5 Comparison with previous related works

As highlighted in previous section, the competitive methods are the following ones [MMM+11, KD13, MMNW13]. None of these previous method considered the separation in consecutive steps and they didn't perform the second one (3.26). However, the optimization problem they solved is equivalent to the first step (3.25) of the optimization problem (3.24).

Indeed, for [MMM+11], the TMO is designed to minimize the distortion between the HDR frame and it's inverse tone mapped version:

$$F^* = \arg\min_F D_{HDR}(X - \tilde{X}) \tag{3.27}$$

They do not account for the SDR perceptual quality nor an enhancement layer. And whatever the targeted rate for the base layer, they use the same Tone mapping in their compression scheme.

The second method [KD13] proposes to minimize the exact same HDR distortion than [MMM+11] under a quality constraint for the uncoded SDR content:

$$\begin{aligned} F^* = \arg\min_F D_{HDR}(X - \tilde{X}) \\ \text{s.t. } D_{SDR\_REF}\big(G(X) - Y\big) \leq D_0 \end{aligned} \tag{3.28}$$

The resulting solution does not account for an enhancement layer and as previously the same TMO is used for all targeted base layer rates. Furthermore, since they did not end up with an analytical solution to this problem, they considered a sub-optimal solution.

The third method [MMNW13] improves the [MMM+11] optimization problem by accounting for the SDR quality constraint. However, they simplify the constrained problem into an unconstrained one with lagrangian tuning and therefore choose an operating point:

$$F^* = \arg\min_F 5 \cdot D_{HDR}(X - \tilde{X}) + 1 \cdot D_{SDR\_REF}\big(G(X) - Y\big) \tag{3.29}$$

Section 3.4 shows that this simplification can lead to various results for the SDR perceptual quality. In their model, they considered an enhancement layer, but the expression of the HDR distortion remained the same, and the solution consists only in performing the TMO optimization, like in the proposed first step (3.25). As for the previous methods, they use the same TMO whatever the targeted base layer rate in the compression scheme.

In the first step of the proposed method (3.25), the expressions for the HDR distortion and SDR quality constraint differ from the previous methods since they account for the error $E_L$ introduced by the SDR encoder and adapt to the base layer rate.

As demonstrated in Lemma 1, the consideration of an enhancement layer consists in adding a rate allocation step to the minimization problem. None of the previous method considered the separation in consecutive steps but in the in the experimental tests (see Sec.3.4), for fair comparison, we performed the best rate allocation for all methods.

## 3.3 Proposed Solution and TMO parameters

### 3.3.1 Piecewise Linear Tone Mapping

To solve the optimization problem (3.25), the TMO $F(X)$ is approximated as a continuous piecewise linear function (3.30), as in [MMM+11, MMNW13, KD13]. Let $x$ denote a realization of $X$. Then $\forall x \in [x_k, x_{k+1}[$

$$F(x) = y = (x - x_k)s_k + y_k \tag{3.30}$$

where $k \in [0, n-1[$, $n$ being the number of chosen linear pieces. And where $s_k$ and $y_k$ are respectively the slope and the intercept of the linear piece corresponding to the current value $x$.

To ease the optimization, we aim at reducing the number of unknowns. For this reason, the interval between $x_k$ and $x_{k+1}$ is considered the same and is noted $\delta$, as originally proposed in [MMM+11]. The values $x_0, x_n$ are chosen to adaptively cover the entire range from the minimum to the maximum values of the dynamic range of the input image. Knowing $x_0$ and $x_n$, $\delta$ is therefore fully determined with $n$ the number of linear pieces. The minimum and maximum values of $y$ are chosen to span the whole range of the SDR output. Therefore, $y \in [0, 2^{n_b}-1]$ (with $n_b$ the number of SDR bits) and finally, $\forall x \in [(x_0 + \delta.k), (x_0 + \delta.(k+1))[$ the tone mapping (3.30) is:

$$F(x) = \Big(x - (x_0 + \delta.k)\Big).s_k + \delta \sum_{j=0}^{k-1} s_j \tag{3.31}$$

In this new representation (3.31), the only unknowns are the slopes $\{s_k\}_{k=0...n-1}$.

Moreover, we seek for a global invertible TMO. The parameterization in (3.31) leads to a continuous function. To add the invertibility constraint, it is further assume that the slopes $s_k$ are strictly positive. Therefore, the optimization will be performed under the constraint that $s_k > 0$. Finally, the tone mapping curve can be assumed to be rather smooth. Therefore, the piecewise linear approximation is very accurate, even with a small number of slopes.

Figure 3.5 illustrates an approximation of the TMO [RSSF02] using a piecewise linear function with either 5 or 20 slopes. As visible, using only 5 slopes allows to have a cor-

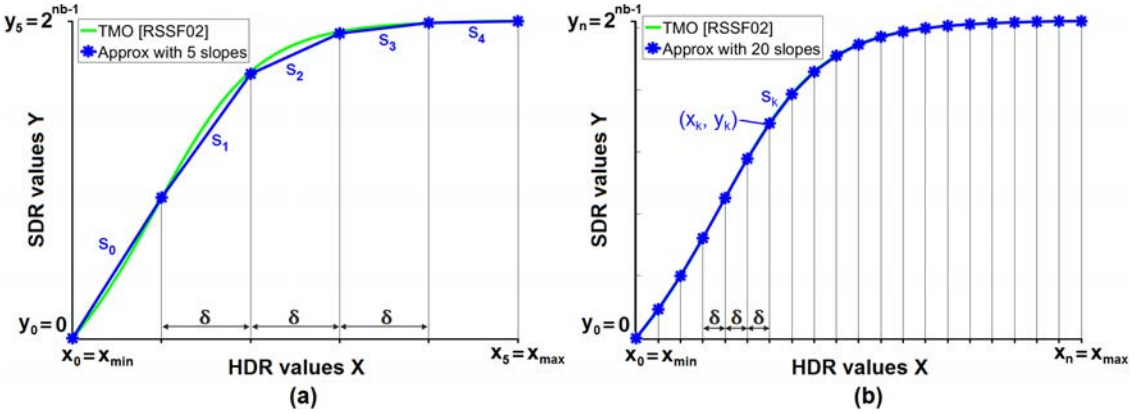



Figure 3.5 – Piecewise linear function approximation of the TMO [RSSF02] for the Tibul image (see Fig 3.6) - (a) Using 5 slopes - (b) Using 20 slopes

rect approximation while using 20 slopes allows to reproduce the TMO curve with great precision. In the following, we use 20 slopes for all the tested TMOs ($n = 20$). Tests were made with more slopes however, it does not noticeably impact the performances.

### 3.3.2 Parameterization of the Optimization Problem

Using the parameterization (3.30-3.31), the optimization problem (3.25-3.26) can be redefined as follows. First, let's express $\mathbb{E}[F(X)]$ with the expression 3.30:

$$\mathbb{E}[F(X)] = \int_{x_{min}}^{x_{max}} F(t).p(t).dt \tag{3.32}$$

$$= \sum_{k=0}^{n-1} \int_{x_k}^{x_{k+1}} \Big((t - x_k)s_k + y_k\Big).p(t).dt \tag{3.33}$$

If we assume that for each slope $\forall t \in [x_k,\ x_{k+1}[\ ,\ \mathbb{E}[t] \approx \frac{x_{k+1}+x_k}{2}$, then:

$$\mathbb{E}[F(X)] = \sum_{k=0}^{n-1} \Big(\frac{x_{k+1} - x_k}{2}.s_k + y_k\Big). \int_{x_k}^{x_{k+1}} p(t).dt \tag{3.34}$$

$$= \sum_{k=0}^{n-1} \Big(\frac{\delta}{2}.s_k + y_k\Big).p_k \tag{3.35}$$

where $p_k = \mathbb{P}(x_k \leq t < x_{k+1}) = \int_{x_k}^{x_{k+1}} p(t).dt$ is easily computed from the HDR image histogram and where $x_{k+1} - x_k = \delta$ as explained in Section 3.3.1. Since $y_k$ can be expressed with the slopes $s_0$ to $s_{k-1}$ then (3.35) can be rewritten:

$$\mathbb{E}[F(X)] = \sum_{k=0}^{n-1} c(s_{0..k}).p_k \tag{3.36}$$

$$\text{with } c(s_{0..k}) = \frac{\delta}{2}.s_k + y_k = \frac{\delta}{2}.s_k + \delta \sum_{j=0}^{k-1} s_j \tag{3.37}$$

In the same way, we can express the different terms [1] of (3.25):

$$\mathbb{E}\left[\left(F(X) - \mathbb{E}[F(X)]\right)^2\right] = \sum_{k=0}^{n-1} \left(c(s_{0..k}) - \sum_{k=0}^{n-1} c(s_{0..k}).p_k\right)^2 p_k \tag{3.38}$$

$$\mathbb{E}\left[F'(X)^{-2}\right] = \sum_{k=0}^{n-1} \left(\frac{p_k}{s_k^2}\right) \tag{3.39}$$

$$\mathbb{E}\left[\left(G(X) - F(X)\right)^2\right] = \sum_{k=0}^{n-1} \left(c(r_{0..k}) - c(s_{0..k})\right)^2 p_k \tag{3.40}$$

where $r_{0..(n-1)}$ are the slopes of the reference TMO $G(X)$.
Therefore, the optimization problem (3.25-3.26) becomes (3.41) where the only unknown parameters are the slopes $s_{0..(n-1)}$:

$$\min_{R_{SDR}} \min_{\{s_k\}} \sum_{k=0}^{n-1} \left(c(s_{0..k}) - \sum_{k=0}^{n-1} c(s_{0..k}).p_k\right)^2 p_k \cdot \sum_{k=0}^{n-1} \left(\frac{p_k}{s_k^2}\right) \tag{3.41}$$

$$\text{s.t.} \begin{cases} \alpha.2^{-2R_{SDR}} \sum_{k=0}^{n-1} \left(c(s_{0..k}) - \sum_{k=0}^{n-1} c(s_{0..k}).p_k\right)^2 p_k + \\ \quad \sum_{k=0}^{n-1} \left(c(r_{0..k}) - c(s_{0..k})\right)^2 p_k \leq D_0 \\ 0 \leq R_{SDR} \leq R_T \\ \sum_{k=0}^{n-1}(s_k) = \frac{2^{n_b}-1}{\delta} \end{cases}$$

In this work, a closed form solution for the $s_k$ has not been identified. Instead, the optimization problem can be solved using numerical approach. In this work, we use the interior-point optimization method.

## 3.4 Experimental Results

In this section, we show and explain the results obtained with the previously described solution (3.41).

As explained in Section 3.2.4, solving the optimization problem in a dual layer scenario is equivalent to first finding the optimal TMO, knowing $R_{SDR}$, in a single layer scenario (3.25), and then optimizing the rate allocation (3.26) between the base layer $R_{SDR}$

1. The derivation of these terms is detailed in Appendix A

and the enhancement layer $R_{SDR}$ for a given rate budget $R_T$. Therefore, the experimental results are presented in two parts. In Section 3.4.1, we present the results of the TMO optimization in a single layer scenario and in Section 3.4.2, we present the results of the TMO optimization combined to the rate allocation optimization in a dual layer scenario.

### 3.4.1 Single Layer Results

For sake of equality, the experimental process is the same for all tested TMOs. First, the original linear light HDR content is transformed such that the quantization becomes uniform with respect to the human eye contrast sensitivity as explained in Section 3.1.1. More precisely, the HDR signal is perceptually transformed with the PQ-2084 OETF [MND12] and uniformly quantized to 12 bits. This builds the input HDR signal $X$.

For each HDR image $X$, the pixel distribution is computed to obtain an histogram with $n$ bins ($n = 20$ since we use 20 slopes for the TMO). In this way, we obtain the $p_k$ values in (3.41). For each SDR rate value $R_{SDR}$, a TMO is computed and then, the image $X$ is tone-mapped to produce the 8 bit base layer $Y$. The SDR image is encoded with the HEVC reference software (HM 16.2) using several rates/QPs (= 18, 22, 26, 30, 34) and finally inverse tone-mapped to obtain the reconstructed HDR image $\tilde{X}$.

Figure 3.6 – SDR reference images tone mapped with [RSSF02]. Original HDR images from [MPII, War03] or the MPEG test set. From left to right and top to bottom: Atrium-Night, Nancy Cathedral, FireEater, Tibul.

In this experiment, the chosen SDR reference signal is generated using the PTR TMO proposed in [RSSF02]. Indeed, it is known to give good visual quality and image integrity [LCTS05, SbMEMS12]. Furthermore, this TMO is one of the most efficient in terms of RD performances [MS08, MMM+11]. Thus, it seems to be the most appropriate TMO for comparison since we want to optimize a tradeoff between RD performances and SDR perceptual quality. Several HDR images from [MPII, War03] of different sizes and dynamic ranges have been tested. They are illustrated in Figure 3.6. The (a) graphs in Figure 3.7 to Fig 3.10 plot the PSNR of the inverse tone mapped HDR image $D_{HDR}(X - \tilde{X})$ as a function of the base layer bitrate $R_{SDR}$, while the (b) graphs plot the distortion $D_{SDR\_REF}$ of the decoded SDR frame $\tilde{Y}$ with respect to the SDR reference $G(X)$ versus the base layer

bitrate. For all images, the $D_0$ constraint is adjusted such that the PSNR value between the SDR reference and the tone mapped image equals to 34.2dB (MSE=25). The same $D_0$ value is used for [KD13], while we kept the proposed weighting (3.29) for [MMNW13].

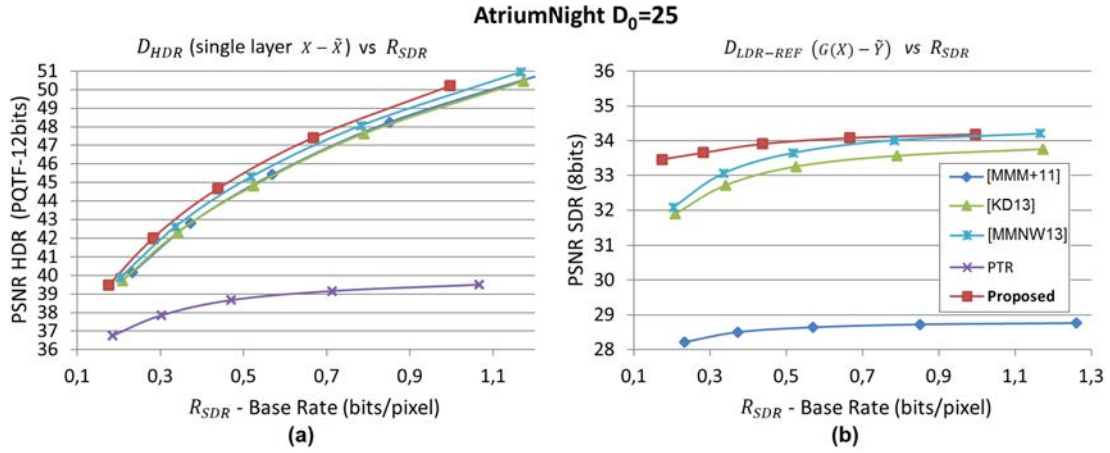



Figure 3.7 – RD performances for AtriumNight and for $D_0$ set to a MSE of 25 (a): Single layer distortion $(X - \tilde{X})$ vs rate of the base layer $R_{SDR}$ - (b): Distortion $D_{SDR\_REF}$ between SDR reference and tone-mapped images $(G(X) - \tilde{Y})$ vs $R_{SDR}$

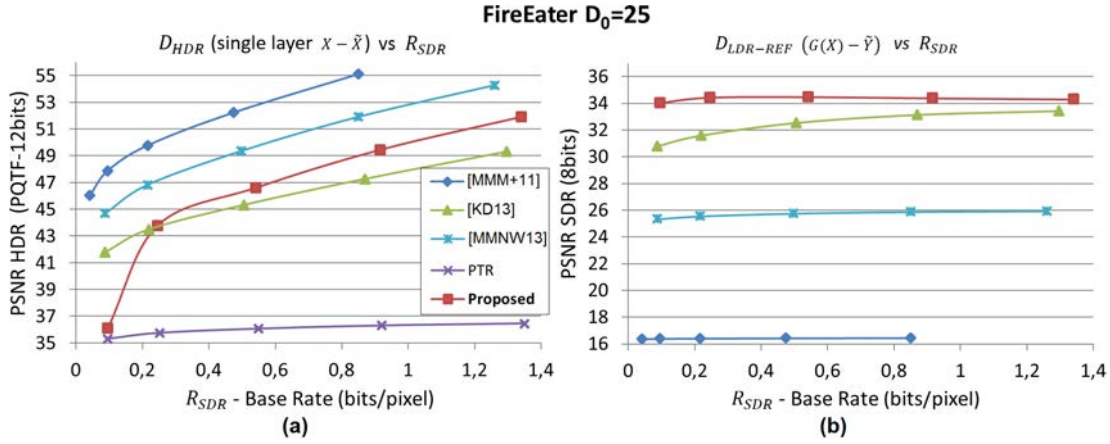



Figure 3.8 – RD performances for FireEater image and for $D_0$ set to a MSE of 25 (a) and (b): same as Figure 3.7

The proposed TMO gives the best tradeoff between RD performances and distortion with respect to the SDR reference for the AtriumNight (Fig.3.7) and Nancy Cathedral (Fig.3.9) images, especially at low bitrates where $D_{SDR\_REF}$ gets worse for [KD13] and [MMNW13]. Surprisingly, the proposed TMO provides better RD performances than the [MMM+11] TMO which is not constrained for the SDR perceptual quality. These examples clearly show that minimizing only the HDR distortion is not optimal in terms of RD performances.

In contrast, for FireEater in Fig.3.8(a) and Tibul in Fig.3.10(a), the unconstrained TMO [MMM+11] outperforms all other tested TMOs including the proposed one. However, it provides an SDR content very different from the reference one, as depicted in

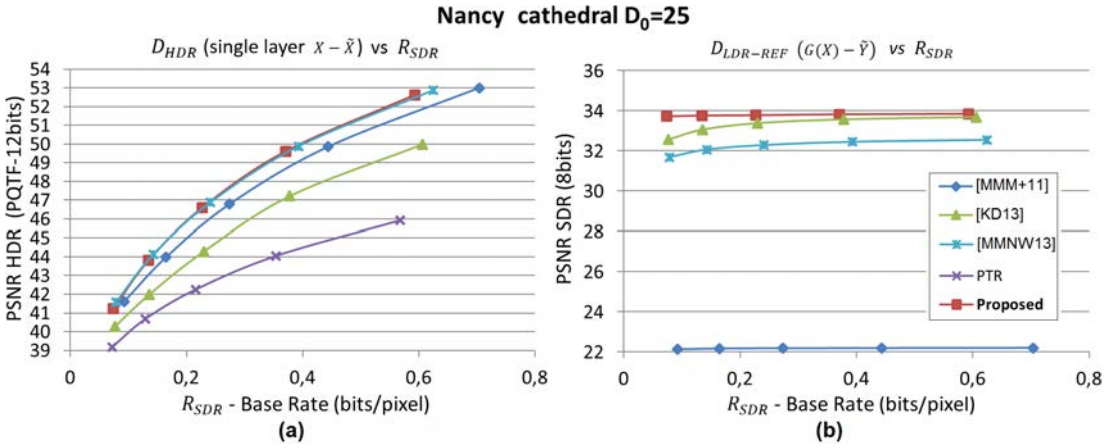



Figure 3.9 – RD performances for Nancy Cathedral image and for $D_0$ set to a MSE of 25 (a) and (b): same as Figure 3.7

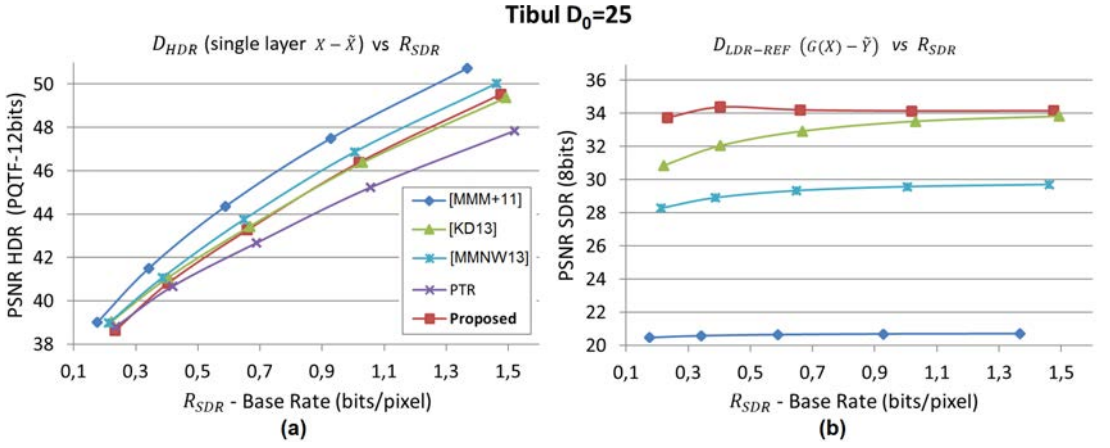



Figure 3.10 – RD performances for Tibul image and for $D_0$ set to a MSE of 25 (a) and (b): same as Figure 3.7

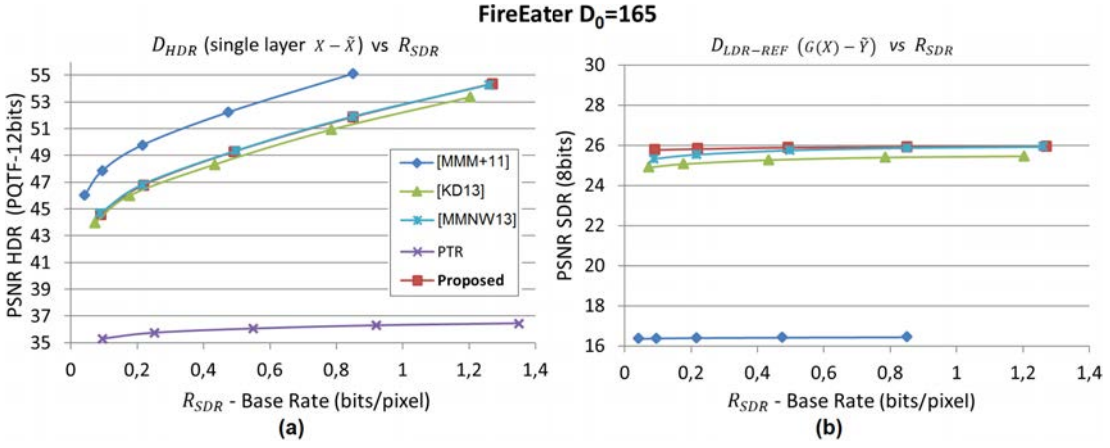



Figure 3.11 – RD performances for FireEater image and for $D_0$ set to a MSE of 165 (a) and (b): same as Figure 3.7

Fig.3.8(b) and Fig.3.10(b). In these cases, the SDR content can have a poor perceptual quality.

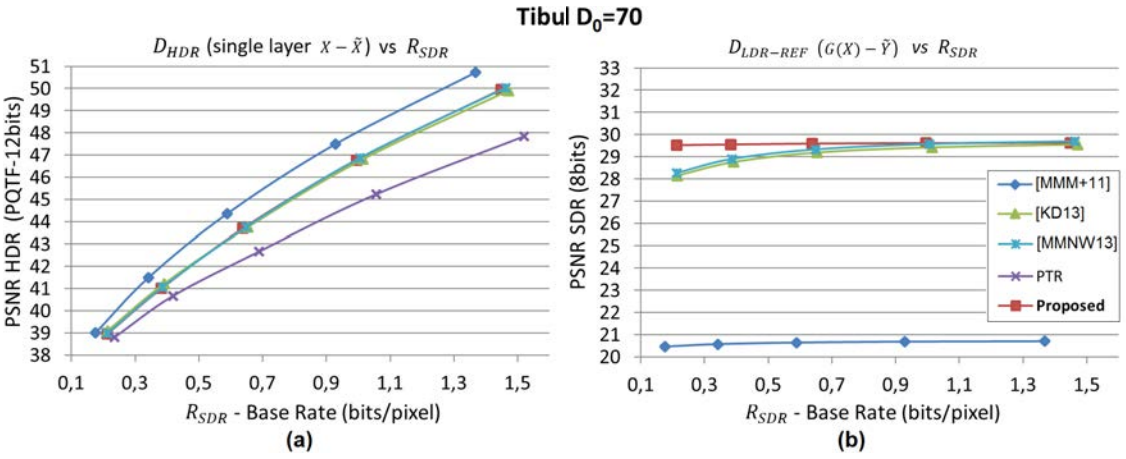



Figure 3.12 – RD performances for Tibul image and for $D_0$ set to a MSE of 70 (a) and (b): same as Figure 3.7

FireEater and Tibul RD results also illustrate the disadvantage of the [MMNW13] unconstrained optimization problem. Since they use an operating point, the distortion with the reference $D_{SDR\_REF}$ does not depend on the constraint $D_0$. The proposed optimization and the one from [KD13] is easier to tune in order to meet the constraint on the SDR signal. As we can see on Fig.3.8(b) and Fig.3.10(b), using [MMNW13], the distortion $D_{SDR\_REF}$ falls over 26dB for FireEater and 29.5dB for Tibul. For this reason, they achieve better RD performances in Fig.3.8(a) and Fig.3.10(a) than the proposed TMO. However, in this case, the trade off between RD performances and quality preservation of the SDR content is difficult to evaluate.

For a fair comparison, we relax the SDR constraint $D_0$ to 26dB (MSE=165) for FireEater and 29.5dB (MSE=70) for Tibul. The results are plotted in Figure 3.11 and 3.12. We can see that the RD performances of [KD13] and the proposed TMO improve, almost aligning the 3 methods on the same level of RD performances, as expected. In this case, the trade off is easier to evaluate. As for the AtriumNight and Nancy Cathedral images, we can see that the proposed TMO keeps the best trade off between minimization of the HDR distortion without sending an enhancement layer and quality preservation of the SDR content, especially at low bitrates. Indeed, [KD13] and [MMNW13] solve a similar optimization problem, with however a simplified model of the variance of the residue, which leads to suboptimality and worst results.

The presented results only assess the SDR perceptual quality with an objective measure $D_{SDR\_REF}$. All resulting SDR images, tone mapped with the tested TMOs, are available in [GRG+b] for subjective evaluation. They can be compared with the reference ones tone mapped with the PTR.

Note that, the convergence time of the proposed algorithm is about 2s (Matlab implementation) i.e. of the same order of magnitude as [MMNW13]. Initializing the algorithm with the PTR or [MMM+11] speeds up the convergence.

### 3.4.2 Dual Layer Results

The second minimization aims to find the best rate allocation for the base and enhancement layers. The quality of the inverse tone mapped HDR image is improved by adding the enhancement layer $Z$, that is the residue represented on 12 bits. This residue is also encoded with the HM 16.2. Multiple encodings have been performed, with different combinations of base and enhancement QPs to test different rate allocations. For each value of the sum rate ($R_{SDR} + R_{HDR}$), the pair ($QP_{SDR}, QP_{HDR}$) leading to the best PSNR is retained, thus yielding the best RD curve for each TMO.

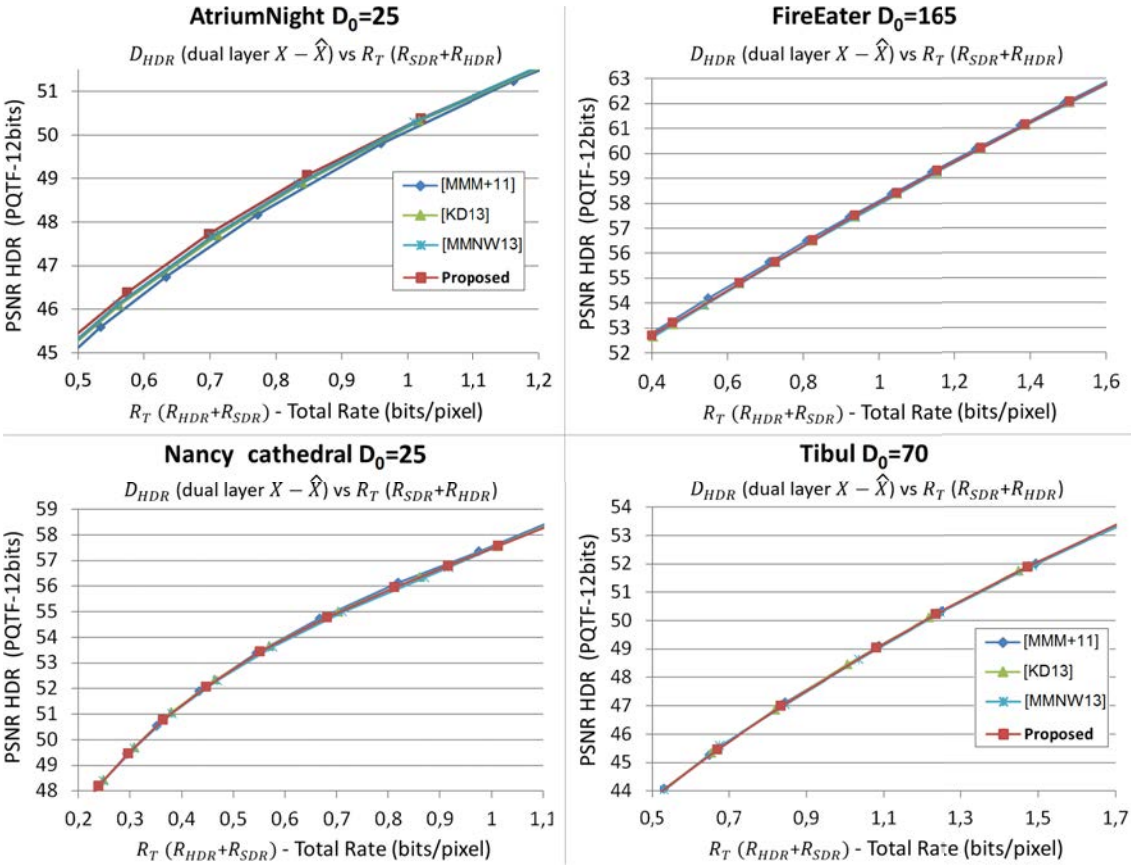



Figure 3.13 – Dual layer distortion ($X - \hat{X}$) vs rate of rate of both layers $R_T$ for, from left to right and top to bottom: AtriumNight with $D_0$ = 25, FireEater with $D_0$ = 165, Nancy Cathedral with $D_0$ = 25, Tibul with $D_0$ = 70.

The plots in Figure 3.13 show the distortion between the HDR frame and the reconstructed HDR frame as a function of the base and enhancement layer bitrates. As we can see, performing the best rate allocation for all TMO reduces the gap betwen the RD curves, compared to the previous results where only the base layer is encoded. The enhancement layer helps the least efficient TMO to improve the HDR reconstruction. However, the proposed TMO still keeps the best trade off since it respects a stricter constraint on the SDR perceptual quality, especially at low bitrates. Note that, if we did not perform the rate allocation step on the state-of-the-art algorithms, their RD performances would have been worse.

## 3.5 Conclusion

To conclude, we demonstrated that the optimal solution for the minimization of the HDR distortion, in a scalable compression scheme with two layers, can be split in two consecutive optimization steps. First, a tone mapping optimization step in a single layer HDR compression scenario and secondly, if we consider an enhancement layer, a rate allocation optimization between the base and enhancement layer.

In this chapter, we also proposed a new TMO that maximizes the RD performance of an HDR coding scheme, while preserving a good quality for the SDR signal. The proposed method leads to the best trade off, in comparison to state-of-the-art methods that solve approximations of the original optimization problem.

Since we demonstrated in this paper that optimizing a TMO for dual layer compression scheme is equivalent to optimizing a TMO in a single layer one, the next chapter focuses on TMO optimization with single layer schemes. The TMOs proposed in the following chapter can be easily combined with a rate allocation optimization step.

# Chapter 4

# Gradient-Based Tone Mapping for Rate Distortion Optimization in Backward Compatible HDR Compression

As explained in Section 2.3.2, the TMO is an essential component for backward compatible HDR compression and it can be designed for different purposes. First, it can be designed to preserve the HDR artistic intent in the SDR content and many TMOs have been developed for this purpose (See Sec.1.4.2). Some examples that consistently perform well [LCTS05, SbMEMS12] are the photographic tone reproducer (PTR) [RSSF02] and the display adaptive tone mapping (DATM) [MDK08]. Secondly, in the context of compression, the tone mapping can be designed to optimize the compression performances [MMM+11, LVCD14, PGT14].

In [MMM+11] the TMO is designed to minimize the distortion of the decoded HDR content. However, this distortion model, also used in [LVCD14, PGT14], assumes that (i) the quantization error is independent of the signal and that (ii) the variance of this quantization error does not depend on the statistics of the signal. In the previous chapter (Chap. 3), we proposed a distortion model only considering the assumption (i) and based on the variance of the SDR signal. In this chapter, we propose a new distortion model based on the image gradient and that better reflects the rate-distortion performance of actual encoders compared with state of the art models. Closed form expressions are derived for this model which relies on the statistics of the HDR image and on the applied TMO.

Since the approach in [MMM+11] does not consider the rate cost of the SDR content, the obtained TMO may lead to SDR data with increased spatial complexity, hence with high rate cost. The authors in [LVCD14] and [PGT14] cope with this limitation by minimizing the HDR distortion under a constraint for the SDR content rate. In [LVCD14], the SDR rate is modelized as a function of the total variation of the SDR signal and in [PGT14] the SDR rate is modelized as a function of the SDR entropy. In this chapter, we address the same minimization problem however, we propose a new rate model based on the image gradient, as for the proposed distortion model. The rate gradient model is shown to provide the highest correlation with the actual SDR rate. Note that, these new gradient-based rate and distortion models hold when an image is first transformed with an invertible piecewise affine function and then compressed with a predictive encoder such as the intra mode of HEVC.

Optimizing the TMO for compression purposes as in [MMM+11], where the authors minimize only the distortion on the inverse tone mapped HDR, may yield SDR content which does not preserve the artistic intent of the HDR. Thus the approach of [MMM+11] is extended in [MMNW10, MMNW13, KD13, KD14] by adding a new constraint on the SDR perceptual quality. It consists of a MSE between the tone mapped signal and a reference SDR version. This problem is also address in the previous chapter (Chap. 3). However, in all these works this SDR quality constraint is usually computed based on coarse assumptions. Instead, we propose a new model for the SDR quality constraint using weaker assumptions.

The influence of the rate is neglected in [KD13, KD14] and thus results in a suboptimal solution. In [MMNW10, MMNW13], the influence of the rate is studied but set aside since it only worsens the RD performances in their implementation. Furthermore, the authors simplified the constrained problem into an unconstrained one, where the value of the lagrangian multiplier is fixed. This leads to choosing an operating point, which may not be the optimal one. Therefore, all these simplifications affect the SDR perceptual quality. In the proposed modelization, we keep and solve the constrained problem which allows to obtain the optimal solution. In the previous chapter (Chap. 3), we study a scalable compression scheme with two layers and therefore only consider the total rate of both layers. The SDR rate is thus not express as a function of the TMO.

In this chapter, we address two use cases. First, we search for a new TMO, using new gradient-based models for the HDR distortion and SDR rate, that optimizes the compression performances of the HDR backward compatible scheme. In a second time, we propose a new problem that optimizes the HDR distortion under a SDR rate constraint and under a SDR perceptual quality constraint. In Chapter 3, we demonstrated that the optimal tone mapping does not depend on the enhancement layer, therefore in this work we only consider a single layer scheme. As explained in the previous chapter, the TMOs proposed in this chapter can be used for dual layer compression optimization by performing an optimal rate allocation step.

Finally, the two tone mapping optimizations lead us to two different conclusions: (i) with the first tone mapping optimization we demonstrate that the proposed solution achieves the optimal rate distortion performance; (ii) we demonstrate that the second tone mapping optimization always provides the best trade-off between the overall rate distortion performance and the quality of the SDR content. Part of this chapter has been published in [GRG+17].

The remainder of this chapter is organized as follows. Section 4.1 presents the compression scheme for the two use cases and the corresponding optimization problems. The models for the rate, distortion and SDR quality constraint are developed in section 4.2 as well as the corresponding tone mapping minimizations. Finally, Section 4.3 presents the results obtained with the two TMOs.

## 4.1 Problem statement and TMO parameters

### 4.1.1 Compression Scheme

In this section, two different optimization problems are introduced for backward compatible HDR compression using a single layer with metadata. The first problem is the

minimization of the distortion on the reconstructed HDR signal under a rate constraint. The second minimization involves the same aspects with an additional quality constraint for the SDR signal. In both cases, the compression scheme is given in Figure 4.1.

To perform the R-D optimization, we model the HDR image as a random process denoted $X$. $X$ generates sequences of random variable, where each random variable models a pixel value. $X$ is tone-mapped to generate an SDR image $Y$ encoded at a given rate $R_{SDR}$ such that the decoded SDR image $\tilde{Y}$ is compatible with legacy SDR displays. $E_Y$ is the error between $Y$ and $\tilde{Y}$. The MSE between $Y$ and $\tilde{Y}$ is the distortion introduced on $Y$ by the encoder at a given rate $R_{SDR}$. To reconstruct the HDR image $\tilde{X}$, from the decoded SDR image $\tilde{Y}$, an inverse tone mapping is performed. This decoded HDR image $\tilde{X}$ is compatible with new HDR displays. Since no enhancement layer is considered, the HDR distortion $D_{HDR}$, generated by the encoding process, is now computed between $X$ and $\tilde{X}$.

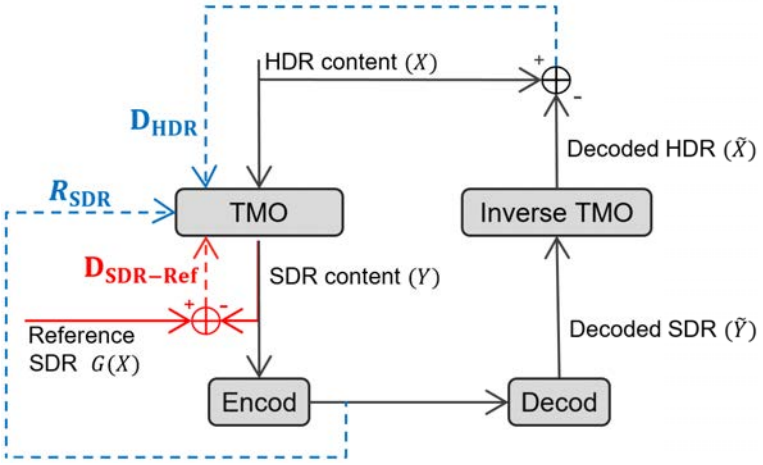



Figure 4.1 – Backward-compatible HDR compression scheme

The HDR image $X$ is considered to be in a uniformly perceptual integer representation. So in practice:

$$X = \lfloor TF(X_{LL}, n_b) \rceil \tag{4.1}$$

with $X_{LL}$ the linear-light HDR signal, $n_b$ the number of bits to represent $X$ and $TF$ the chosen transfer function. In this Chapter, as in Chapter 3, $X_{LL}$ is quantized on 12 bits with the PQ-2084 OETF [MND12], the MPEG recommended transfer function [LFH15].

As $X$ is considered as a uniformly perceptual representation, the Mean Square Error (MSE) is a relevant metric to estimate HDR signals quality [MMM+11, KD14]. Therefore:

$$D_{HDR} = \mathbb{E}\left[(\tilde{X} - X)^2\right] \tag{4.2}$$

### 4.1.2 Rate-Distortion Optimization of the HDR

The goal is to find the tone mapping that minimizes $D_{HDR}$ for a given rate budget $R_0$. Therefore the minimization problem is:

$$\begin{aligned} \min_{TMO} \quad & D_{HDR} \\ \text{s.t.} \quad & R_{SDR} = R_0 \end{aligned} \tag{4.3}$$

We use an equality constraint instead of an inequality constraint since the R-D function is decreasing [CT06, Ex10.5] with the rate R, therefore the minimum distortion is obtained when the rate is maximal i.e. when $R_{SDR} = R_0$.

Problem (4.3) aims at optimizing the RD performances only. Therefore, there is no guarantee that the SDR content will have a good perceptual quality.

### 4.1.3 Rate-Distortion Optimization of the HDR under a SDR quality constraint

To improve the SDR perceptual quality, in the second optimization problem, the TMO is designed to minimize the HDR distortion under a rate constraint but also under a quality constraint for the SDR signal. This constraint $D_{SDRREF}$ is modeled as in [MMNW13] and [KD13]. It is a distortion between $Y$ and an SDR image that preserves the perception and artistic intent of the HDR image. This SDR reference image is noted $G(X)$ since it is related to the HDR image by a tone mapping curve $G$ that has been specially designed to preserve the perception of the HDR content. Therefore:

$$D_{SDRREF} = \mathbb{E}\Big[(G(X) - Y)^2\Big] \tag{4.4}$$

With this new constraint, the problem (4.3) becomes:

$$\begin{aligned} \min_{TMO} \quad & D_{HDR} \\ \text{s.t.} \quad & \begin{cases} R_{SDR} \leq R_0 \\ D_{SDRREF} \leq D_0 \end{cases} \end{aligned} \tag{4.5}$$

The contributions in this chapter are twofold. First, we propose novel rate and distortion models based on the gradient of the HDR image. We show in the following that these models lead to a good approximation of the rate and of the distortion measured with a predictive codec such as HEVC [MPE17]. Second, we derive the optimal solution of the global optimization problems (4.3) and (4.5).

To solve these optimization problems (4.3) and (4.5), the TMO is approximated by a piecewise affine function (3.30) as in [MMM+11, MMNW13, KD13] and as in Chapter 3. A more detailed presentation of this parameterization is made in Section 3.3.1.

In the following sections, only the luminance compression is considered but it could be extended to color by computing in a similar manner a TMO for the U and V components or separately for each RGB color channel.

## 4.2 Gradient-Based Models And Proposed Solution

### 4.2.1 Gradient-Based SDR Bit-Rate Model

Different studies for Rate Control compare the H.264/ AVC Intra-Frame rate to different estimators based on either the gradient, the entropy, the variance and other measures of the HDR image [KYK99, CMZ09]. In these studies, the gradient is shown to be the most reliable estimator.

Similarly, in [TZC14] it is shown that the HEVC Intra-Frame rate and the spatial gradient of the SDR image are highly correlated. For these reasons, we choose to model the SDR rate as a function of the SDR gradient $\nabla_Y$ as:

$$R_{SDR} = f\Big(\mathbb{E}[\nabla_Y]\Big) \tag{4.6}$$

In the sequel, we consider the following estimation for the spatial gradient of the SDR image which are then shown to accurately predict the actual rate:

$$\nabla_Y = \big(|Y - Y'_H| + |Y - Y'_V|\big) \tag{4.7}$$

$$\nabla_Y = \min\big(|Y - Y'_H|,\ |Y - Y'_V|\big) \tag{4.8}$$

where $Y'_H$ and $Y'_V$ are the shifted versions of the image $Y$ in the horizontal and vertical directions respectively. In (4.7), the sum of the horizontal and vertical gradient absolute values is computed. In (4.8) instead, the minimum between the vertical and horizontal gradients is computed.

To solve the optimization problems (4.3) and (4.5), we now need to express the SDR gradient (4.6) as a function of the parameters of the TMO. Since the TMO is related to the SDR luminance values $Y$, we first introduce $Y$ in the gradient expression:

$$\mathbb{E}[\nabla_Y] = \sum_{y=0}^{y_{max}} \mathbb{P}(Y = y) \cdot \mathbb{E}[\nabla_Y | Y = y] \tag{4.9}$$

$$\mathbb{E}[\nabla_Y] = \sum_{y=0}^{y_{max}} \mathbb{P}(Y = y) \cdot \sum_{\nabla y=0}^{\nabla y_{max}} \nabla_y \cdot \mathbb{P}(\nabla_Y = \nabla_y | Y = y) \tag{4.10}$$

$$\mathbb{E}[\nabla_Y] = \sum_{y=0}^{y_{max}} \sum_{\nabla y=0}^{\nabla y_{max}} \nabla_y(y) \cdot \mathbb{P}(\nabla_Y = \nabla_y, Y = y) \tag{4.11}$$

where $\nabla_Y$ stands for the random variable and $\nabla_y$ its realization. $\nabla_y(y)$ stands for the gradient value $\nabla_y$ at a given $y$ SDR luminance value.

Let us assume that a pixel value $X$ and its neighbor $X'$ have similar values. Therefore, the same slope $s_k$ is used to tone map these two values and:

$$\nabla_Y = Y - Y' = X.s_k - X'.s_k = \nabla_X.s_k \tag{4.12}$$

Therefore, $\forall y \in [y_k, y_{k+1}[\ ,\ \forall x \in [x_k, x_{k+1}[$

$$\sum_{\nabla y=0}^{\nabla y_{max}} \nabla_y(y) \cdot \mathbb{P}(\nabla_Y = \nabla_y, Y = y) = \sum_{\nabla x=0}^{\nabla x_{max}} \nabla_x(x) \cdot s_k \cdot \mathbb{P}(\nabla_X = \nabla_x, X = x) \tag{4.13}$$

This approximation holds when two neighboring pixels belong to the same bin (i.e same slope). This is in general true except for highly textured image region. In addition, it is interesting to note that the number of slopes used for the approximation of the TMO has little influence on the accuracy of the gradient computation (4.13). Indeed, reducing the number of slopes, will increase the chances that neighboring pixels are transformed with the same slope. Thus, it reduces the number of occurrence of approximation errors, but the value of these errors is potentially larger. On the other hand, increasing the number of slopes reduces the chances that neighboring pixels are transformed by the same slope but the neighboring slopes values are increasingly closer, hence the approximation errors have a small amplitude.

Using the equations (4.11) and (4.13), the SDR gradient is now approximated by:

$$\mathbb{E}[\nabla_Y] = \sum_{k=0}^{n-1} \sum_{x=x_k}^{x_{k+1}} \sum_{\nabla x=0}^{\nabla x_{max}} \nabla_x(x) \cdot s_k \cdot \mathbb{P}(\nabla_X = \nabla_x, X = x) \tag{4.14}$$

$$\mathbb{E}[\nabla_Y] = \sum_{k=0}^{n-1} s_k \sum_{x=x_k}^{x_{k+1}} \sum_{\nabla x=0}^{\nabla x_{max}} \nabla_x(x) \cdot \mathbb{P}(\nabla_X = \nabla_x, X = x) \tag{4.15}$$

$$\mathbb{E}[\nabla_Y] = \sum_{k=0}^{n-1} s_k \cdot g_k \tag{4.16}$$

where $g_k$ only depends on the pdf of the HDR image and its gradients.

This model is written for a generic formulation of the gradient and it is valid for different expressions of the gradient, as (4.7) and (4.8). More generally, the model is valid as long as the gradient is a sum of differences (as in (4.7)) or a minimum of differences (as in (4.8)). We studied different expressions for the gradient but (4.7) and (4.8) seemed to be the most reliable to the actual rate.

To validate the model (4.16), we compare its value to the SDR gradient directly computed on the SDR image. This test has been carried out with 26 different HDR images and 3 different TMOs $F(s_{0..k})$, where the slopes in $\{s_{0..k}\}$ are randomly drawn to yield a strictly increasing TMO that maps the whole range of the input HDR image. Each point is computed with a pair (image, TMO). The HDR images used for this test are very diverse and are all depicted in [GRG+a]. They have different resolutions, up to 4K, different dynamic ranges, and issued from various sets (natural/animations, high/low spatial complexity).

Figure 4.2(a), (b) and (c) show the estimated gradient with two estimators (the proposed gradient estimator given in (4.7) and (4.16) and the one proposed in [MMNW13] against the actual gradient values per pixel. The proposed model is more reliable whatever the number of slopes. As mentioned by the authors of [MMNW13], their estimator becomes worse when the number of slopes decreases. This aspect is confirmed in Fig.4.2. Indeed, the estimator in [MMNW13] used a totally different model and is based on the assumption that, inside a given bin, the values of all the pixels are the same as the one of the centers of the bins. This approximation is only valid if the bin size is sufficiently small i.e. if the number of slopes/bins is high. Therefore, their model is only accurate when using more than 50 slopes for the TMO. Figure 4.2(d) further shows the gradient estimated as the minimum of the horizontal and vertical gradient (4.8) as a function of the true gradient.

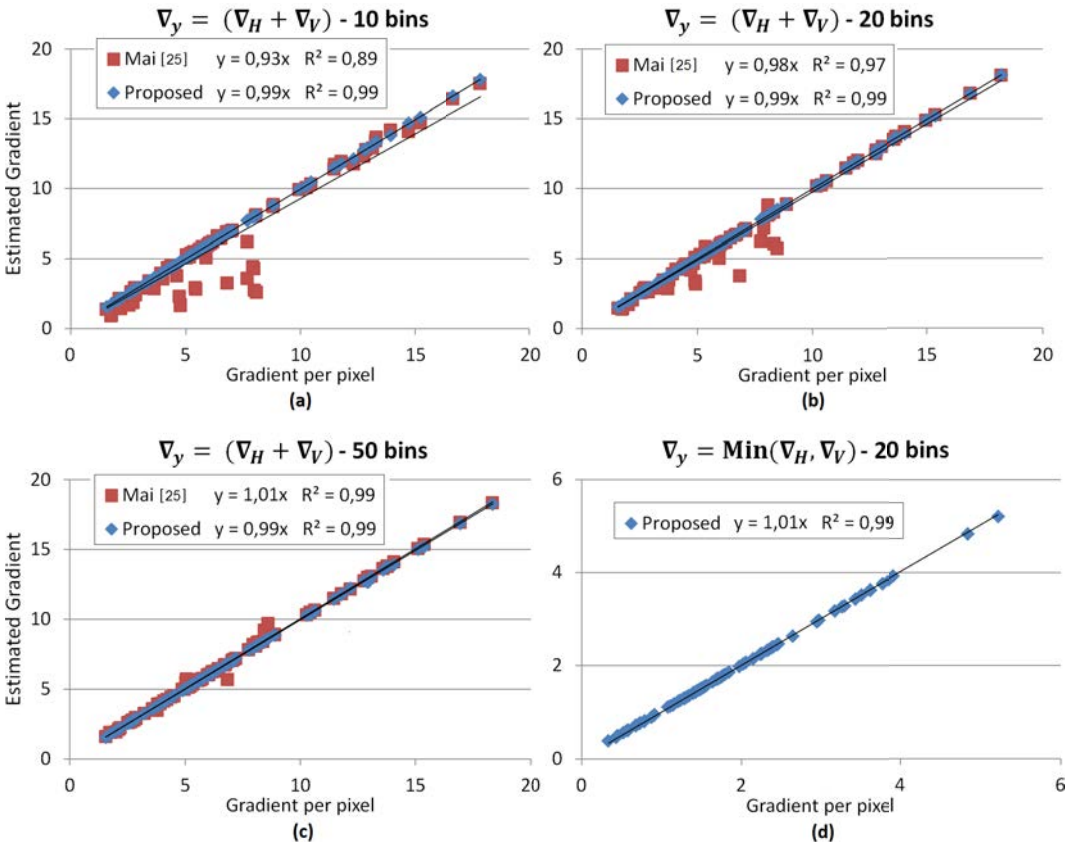



Figure 4.2 – Relation between SDR gradient measured on the SDR image and the values given by the proposed SDR gradient estimator (4.16) and the estimator used in [MMNW13] - (a) sum of the vertical and horizontal gradients (4.7) for 10 slopes - (b) for 20 slopes - (c) for 50 slopes - (d) min of the vertical and horizontal gradients (4.8) for 20 slopes

By comparing Figure 4.2(b) and Figure 4.2(d), we conclude that both models (4.7) and (4.8) provide a very accurate estimation of the SDR bitrate. In the following, we decide to use the (4.8) model as it mimics the selection of an intra prediction mode in a predictive coding scheme (as in HEVC-Intra). Indeed, the Intra coding of a block (see Sec.2.2.3.1) in HEVC or H.264/AVC consists in computing the difference between the current block and several shifted version of the causal neighborhood. The shifted version minimizing the RD cost is then selected. This selection process is similar to the gradient (4.8) where we select the best gradient direction (in a predictive sense) for each pixel, either horizontal or vertical.

To ensure that the proposed model (4.16) is well correllated to the actual SDR rate, the scatter plots in Figure 4.3 compare the Intra-Frame HEVC rate of the SDR content and its estimate. The SDR rate estimate is based on the gradient (4.16) in Figure 4.3(a) and on the entropy of the SDR signal [PGT14] in Figure 4.3(b). We use the same test set used in Figure 4.2 (26 images combined with 3 TMOs) using 20 slopes for the TMOs. As for Figure 4.2, tests were made with different numbers of slopes but since the results were similar, we only keep the ones with 20 slopes. Note that, each point represents a pair (image, TMO).

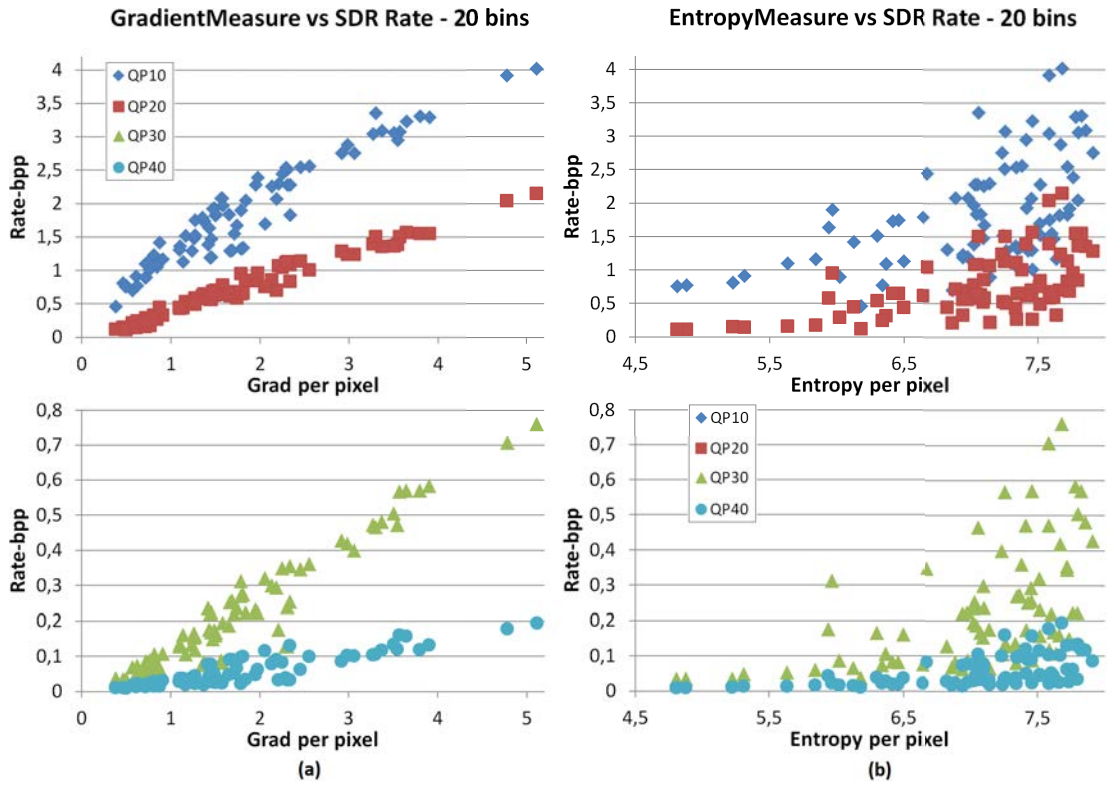



Figure 4.3 – Experimental validation of the proposed rate model - (a) Cloud point of the rate as a function of the gradient based estimator for QP = [10, 20, 30, 40] - (b) Cloud point of the rate as a function of the Entropy-based estimator for QP = [10, 20, 30, 40]

Figure 4.3(a) clearly shows less dispersion than the scatter plots in Figure 4.3(b). Therefore, we can conclude that the HDR image gradient provides a more accurate estimate of the actual rate, for all QP (Quantization parameter in HEVC), than the entropy. Moreover, an affine function seems sufficient to model the relationship between the HDR image gradient and the rate of any images. For this reason, the SDR rate function (4.6) can be rewritten as:

$$R_{SDR} = a \cdot \mathbb{E}[\nabla_Y] + b \tag{4.17}$$

$$= a \cdot \sum_{k=0}^{n-1} \Big( s_k \cdot g_k \Big) + b \tag{4.18}$$

where the coefficients $a$ and $b$ are different for each QP since the affine function varies in Figure 4.3(a).

Using least squares fitting, we search for the most appropriate $a$ and $b$ values for (4.18) to approximate the cloud points in Figure 4.3(a). The obtained values are represented in Figure 4.4. Using these values allows us to find the SDR rate of any image, knowing the HDR image gradient $g_k$ and the applied TMO $s_k$. Moreover, a relationship between the QPs and the coefficients '$a$' and '$b$' have been obtained through nonlinear least squares

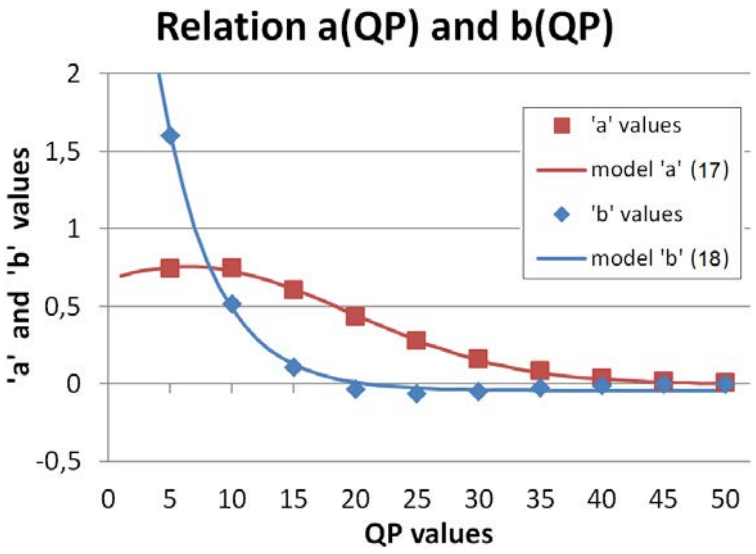



Figure 4.4 – Values of coefficients 'a' and 'b' versus QP and related models

fitting (Trust-Region algorithm [MS83]):

$$\hat{a} = 0,7567 \cdot \mathrm{e}^{-\left(\frac{QP-6,337}{18,72}\right)^2} \tag{4.19}$$

$$\hat{b} = 5,161 \cdot \mathrm{e}^{-0,228.QP} - 0,044 \tag{4.20}$$

The model (4.18) and the coefficients (4.19) and (4.20) allow to estimate the rate of any HDR image using only its gradient and the chosen TMO.

### 4.2.2 Classical HDR Distortion Model

Considering the MSE in the chosen transfer domain, the HDR distortion in [MMM+11, PGT14, MMNW13, KD13] and in the precedent chapter is:

$$D_{HDR} = \mathbb{E}\Big[(\tilde{X} - X)^2\Big] \tag{4.21}$$

In [MMM+11, MMNW13, KD13], using the parameterization (3.30), the HDR distortion becomes

$$D_{HDR} = \sum_{k=0}^{n-1} \sum_{x=x_k}^{x_{k+1}} \sum_{\tilde{y}=\tilde{y}_k}^{\tilde{y}_{k+1}} \Big(\tilde{x}(\tilde{y}, s_k) - x\Big)^2 \cdot \mathbb{P}\Big(E_Y = (\tilde{y} - y),\ X = x\Big) \tag{4.22}$$

In the previous chapter as in [MMM+11, PGT14, MMNW13, KD13], it is assumed that the quantization error on the SDR signal $E_Y = \tilde{Y} - Y$ is independent of the SDR image $Y$ (called *independence assumption*). Under this assumption, the HDR distortion becomes:

$$D_{HDR} = \sigma^2_{E_Y} \cdot \mathbb{E}\Big[F'(X)^{-2}\Big] \tag{4.23}$$

where $F$ is the TMO, $'$ stands for the first derivative and where $\sigma^2_{E_Y}$ is the variance of the SDR signal error $E_Y$. Using the parameterization (3.30) of [MMM+11, MMNW13, KD13], the HDR distortion (4.23) becomes:

$$D_{HDR} = \sigma^2_{E_Y} \cdot \sum_{\forall k} s_k^{-2} \cdot p_k \tag{4.24}$$

where $p_k = \mathbb{P}(x_k \leq X < x_{k+1})$.

The equation (4.24) leads to an accurate estimation as shown in Figure 4.5. Indeed, the scatter plots shows a high correlation between the actual HDR distortion and the estimated one (4.24). This experimental validation uses the same test set as in Figure 4.2 (26 images combined with 3 TMOs = 78 points). However, equation (4.24) is intractable in practice

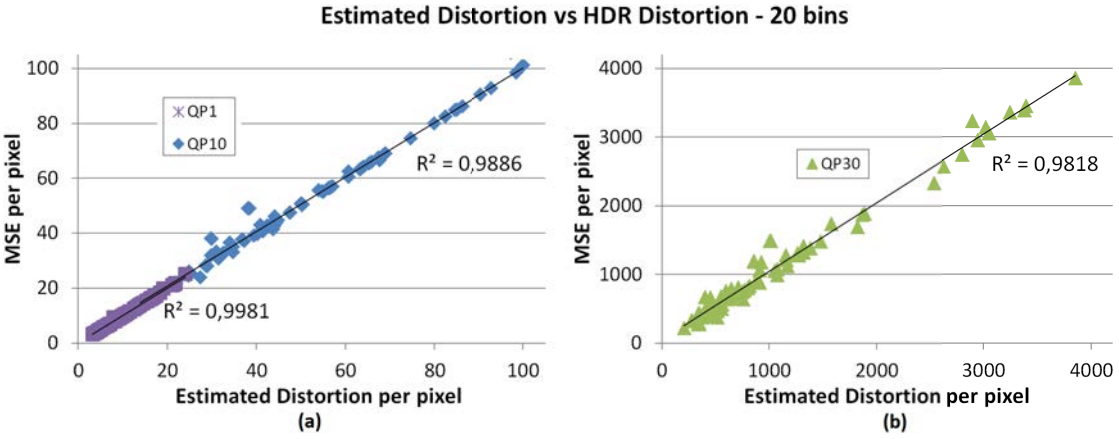



Figure 4.5 – Experimental validation of the estimated distortion (4.24) - (a) At very-high rate (QP1) and high rate (QP10) - (b) At medium-low rate (QP30)

since to evaluate the variance $\sigma^2_{E_Y}$, one needs to compress the data and get $\tilde{Y}$: it's a vicious circle. In order to overcome this difficulty, [PGT14, MMNW13, KD13] assume that $\sigma^2_{E_Y}$ depends on the QP factor only but does not depend on the statistics of the SDR image (called *invariance assumption*). In [MMM+11], they assume that $\sigma^2_{E_Y}$ is multiplicative constant and neglect it. Thus, in both cases, $D_{HDR}$ is rewritten as:

$$D_{HDR} \propto \mathbb{E}\Big[F'(X)^{-2}\Big] \tag{4.25}$$

$$D_{HDR} \propto \sum_{\forall k} s_k^{-2}.p_k \tag{4.26}$$

To test the validity of the independence and the statistics invariance assumptions, we compute the estimated distortion expressed in (4.26) and compare it to the corresponding HDR distortion at different QPs, as in Figure 4.5. Figure 4.6 shows that the distortion $D_{HDR}$ scales linearly with the estimated $D_{HDR}$ (4.26) at low QP only. At QP=10, the correlation coefficient is much lower and it decreases as the QP increases. At QP=30, there is little correlation between the measured and estimated values. Figure 4.6 shows the distortion when the TMO use 20 slopes but, as in Figure 4.2, the results were the same with 50 slopes. All these observations are mainly due to the assumed independence between the SDR image $Y$ and its compression error $E_Y$, which is not valid as shown in Figure 4.7. Indeed, the error $E_Y$ still contains details of the original image $Y$, especially for high QP.

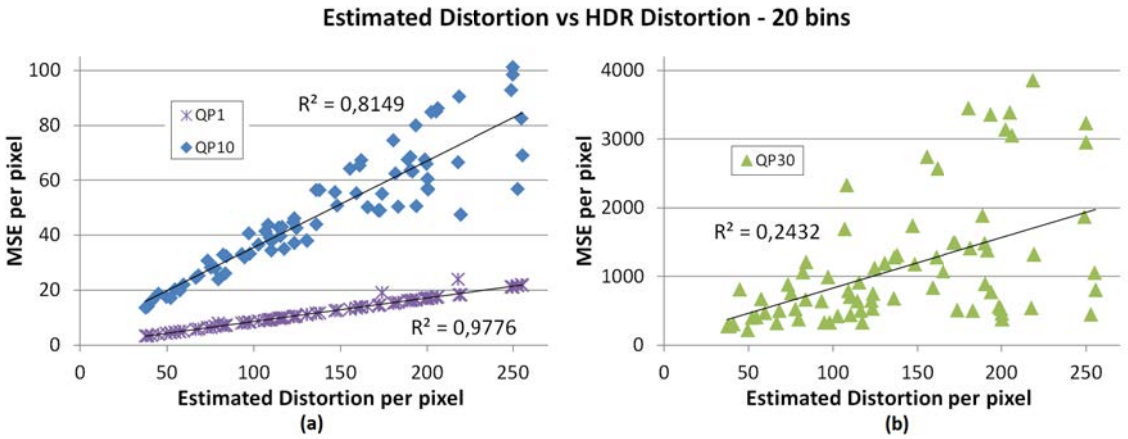



Figure 4.6 – Experimental validation of the estimated distortion (4.26) - (a) At very-high rate (QP1) and high rate (QP10) - (b) At medium-low rate (QP30)

### 4.2.3 Proposed Gradient-Based HDR Distortion Model

As in the previous chapter and in [MMM+11, PGT14, MMNW13, KD13], the proposed HDR distortion model is a MSE in the chosen transfer domain. Therefore, using the parameterization (3.30), we use the expression (4.22) for the HDR distortion. Here we neither use the independence nor the invariance assumption, instead, the proposed formulation takes into account the joint distribution of the HDR image $X$ and the compression error $E_Y$.

Considering that the luminance value of the original and reconstructed SDR image ($y$ and $\tilde{y}$ respectively) belong to the same bin $[x_k, x_{k+1}[$ and therefore the same slope $s_k$, the HDR distortion (4.22) becomes:

$$D_{HDR} = \sum_{k=0}^{n-1} \sum_{x=x_k}^{x_{k+1}} \sum_{\tilde{y}=\tilde{y}_k}^{\tilde{y}_{k+1}} \frac{\Big(\tilde{y} - y(x, s_k)\Big)^2}{s_k^2} \cdot \mathbb{P}\Big(E_Y = (\tilde{y} - y),\ X = x\Big) \tag{4.27}$$

$$D_{HDR} = \sum_{k=0}^{n-1} \frac{1}{s_k^2} \cdot \sum_{x=x_k}^{x_{k+1}} \sum_{\tilde{y}=\tilde{y}_k}^{\tilde{y}_{k+1}} E_y^2\Big(\tilde{y}, x, s_k\Big) \cdot \mathbb{P}\Big(E_Y = (\tilde{y} - y),\ X = x\Big) \tag{4.28}$$

where $E_y(.)$ stands for the function that computes the SDR error from the HDR content $x$, the reconstructed SDR $\tilde{y}$ and the TMO slope $s_k$.

Figure 4.7 shows that the compression error $E_Y$ depends on the gradient of the SDR signal and that this dependence increases with the QP value. We propose to model $E_Y^2$ as an affine function of $(\nabla_Y)^\gamma$, where the coefficients of the affine function and $\gamma$ depend on the QP value only. This leads to a new model for $D_{HDR}$ where (4.28) becomes:

$$D_{HDR} \propto \sum_{k=0}^{n-1} \frac{1}{s_k^2} \cdot \sum_{x=x_k}^{x_{k+1}} \sum_{\nabla_y=0}^{\nabla_{ymax}} \Big(\nabla_y\big(y(x, s_k)\big)\Big)^\gamma \cdot \mathbb{P}\Big(\nabla_Y = \nabla_y,\ X = x\Big) \tag{4.29}$$

Following the same assumption as in (4.12) and (4.13), we obtain:

$$D_{HDR} \propto \sum_{k=0}^{n-1} \frac{g_k(\gamma)}{s_k^{2-\gamma}} \tag{4.30}$$

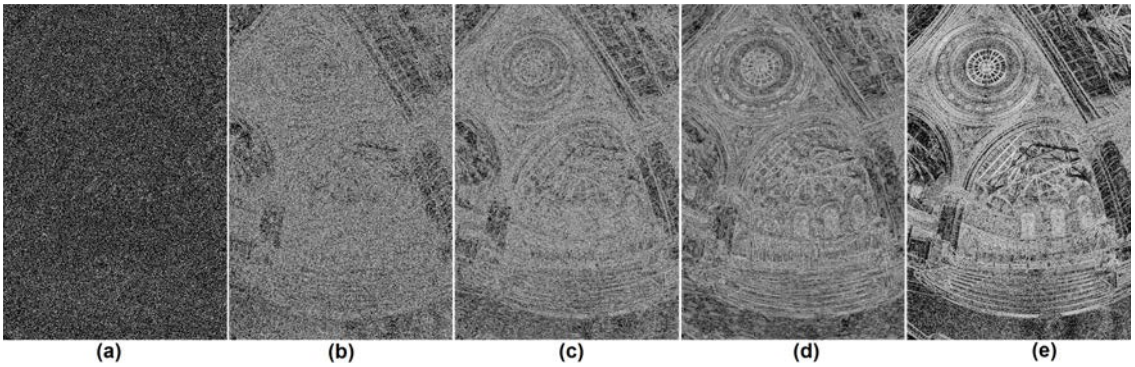



Figure 4.7 – Error $E_Y$ between SDR source and reconstructed SDR compared to SDR source gradient $\nabla_Y$ - (a) error at QP=10 - (b) error at QP=20 - (c) error at QP=30 - (d) error at QP=40 - (e) Gradient – For display and print purposes, the displayed images are modified as follows: $(E_Y^2)^{1/8}$ and $(\nabla_Y)^{1/8}$

where:

$$g_k(\gamma) = \sum_{x=x_k}^{x_{k+1}} \sum_{\nabla_x=0}^{\nabla_{xmax}} \Big(\nabla_x(x)\Big)^{\gamma} \cdot \mathbb{P}\Big(\nabla_X = \nabla_x,\ X = x\Big) \tag{4.31}$$

Note that the $g_k(\gamma)$ values depend on the HDR image and $\gamma$ only.

To complete the HDR distortion model, we need to estimate the relationship between $\gamma$, the affine coefficients and QP. We first consider the estimation of $\gamma$. With sufficient image tests (same test set as in Figure 4.2), we compare the estimated distortion (4.30), computed with different $\gamma$ values, to the corresponding HDR distortion at a given QP. For each QP value, we retain the $\gamma$ value which maximizes the correlation coefficient between the distortion and its estimation. The best $\gamma$ values are plotted in Figure 4.8(a). In the worst case, the correlation coefficient is $0.945$ at QP16.

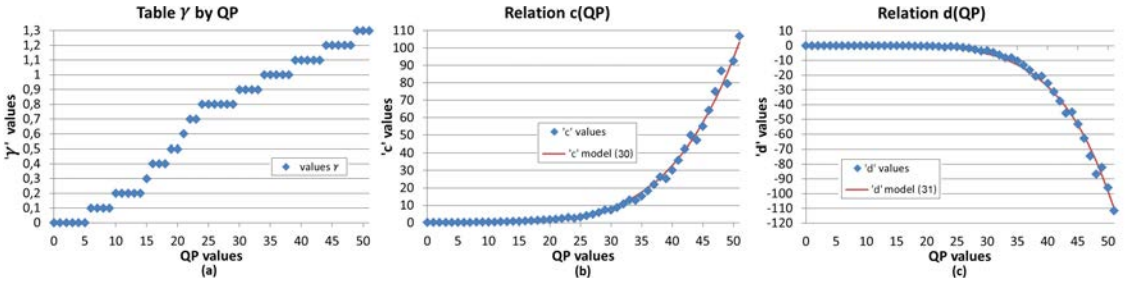



Figure 4.8 – (a) $\gamma$ values versus QP - (b) values of coefficient c versus QP and related model - (c) values of coefficient d versus QP and related model

We now establish a relationship between the affine coefficients $c$ and $d$ of the HDR distortion model (4.32) and QP:

$$D_{HDR} = c \cdot \sum_{k=0}^{n-1} \left( \frac{g_k(\gamma)}{s_k^{2-\gamma}} \right) + d \tag{4.32}$$

where the best $c$ and $d$ values are plotted in Figure 4.8(b) and (c). As shown in this figure, the relationship between $c$, $d$ and the QP can be modeled using a nonlinear least squares fitting (Trust-Region algorithm [MS83]) as:

$$\hat{c} = (\quad 8.939.10^{-7}) \cdot QP^{4.722} + 0.124 \tag{4.33}$$

$$\hat{d} = (-2,223.10^{-8}) \cdot QP^{5,677} + 0.034 \tag{4.34}$$

Finally, the parameters of the functions (4.33-4.34) and a table for $\gamma$ (one value per QP is stored leading to 51 entries) are stored at the encoder to perform the TMO. Since these values are used at the encoder side only, there is no need to send them.

As we made this learning process with different numbers of slopes and that the resulting values were the same, we conclude that $\gamma$, $c$ and $d$ relations are valid for any number of slopes. Figure 4.9 shows that the proposed distortion estimator (4.30) is better than the classical model (4.26) represented in Figure 4.6. Based on the well correlated offline

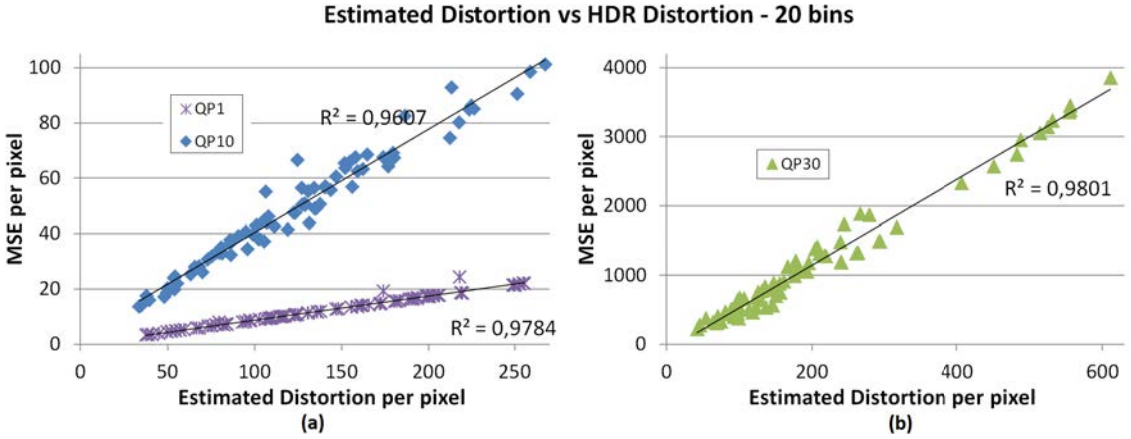



Figure 4.9 – Experimental validation of the distortion estimated with the model proposed in (4.32) - (a) At very-high rate (QP1) and high rate (QP10) - (b) At medium-low rate (QP30)

models for $c$, $d$ and $\gamma$ parameters, the proposed estimator (4.32) has the advantage to estimate the distortion for any input HDR image; while in the classical estimator (4.24), $\sigma^2_{E_Y}$ remains unknown and dependent of the image.

Note that, although the proposed distortion model is here applied in the context of HDR encoding, it could be used to predict the encoder distortion at a given rate using only the SDR image and its gradient.

### 4.2.4 Rate-Distortion Optimization

We now aim at solving the optimization problem (4.3). As explained in Section 3.3.1, the TMO is chosen to span the whole range of the SDR output. Therefore, we need to add a new constraint to the minimization problem to obtain an image in the specified SDR range. Moreover, the model for $R_{SDR}$ (4.18) and $D_{HDR}$ (4.32) introduce the hidden parameter $QP$. Therefore, (4.3) requires an additional optimization over this hidden parameter $QP$

and becomes:

$$\min_{QP} \min_{s_1 \dots s_n} \quad c \cdot \sum_{k=0}^{n-1} \left( \frac{g_k(\gamma)}{s_k^{2-\gamma}} \right) + d \tag{4.35a}$$

$$\text{s.t.} \quad \begin{cases} \sum_{k=0}^{n-1} s_k \cdot g_k(1) = \dfrac{R_0 - b}{a} & (4.35b) \\ \sum_{k=0}^{n-1} s_k = \dfrac{2^{n_b-1}}{\delta} & (4.35c) \end{cases}$$

where $s_k$ represents the slopes of the TMO, $g_k(\gamma)$ is defined in (4.31), a and b in (4.19-4.20), c and d in (4.33-4.34), $n_b$ and $\delta$ in Section 3.3.1.

The optimization over the QP parameter is either solved exhaustively or by learning a relationship between QP and the rate constraint $R_0$, see Section 4.3.1 for details. We now solve the second minimization problem over the TMO parameters $\{s_{0..k}\}$ and assume that the QP parameter is fixed. The Karush-Kuhn-Tucker necessary conditions for optimality [BV04, Chap5.5.3] on the Lagrangian expression of (4.35) lead to:

$$\begin{cases} \dfrac{-(2-\gamma) g_j(\gamma).c}{s_j^{(3-\gamma)}} + \lambda.g_j(1) + \mu = 0 \quad \forall j \in [0, n] & (4.36a) \\ \sum_{k=0}^{n-1} \left( s_k.g_k(1) \right) - \dfrac{R_0 - b}{a} = 0 & (4.36b) \\ \sum_{k=0}^{n-1} \left( s_k \right) - \dfrac{2^{n_b-1}}{\delta} = 0 & (4.36c) \end{cases}$$

Fortunately, (4.36a) depends on one $s_j$ value only. Therefore, the optimal solution of (4.35) is:

$$s_k = \left( \frac{(2-\gamma).g_k(\gamma).c}{\mu + \lambda.g_k(1)} \right)^{\frac{1}{3-\gamma}} \tag{4.37}$$

where $\mu$ and $\lambda$ are the solutions of:

$$\begin{cases} \sum_{k=0}^{n-1} \left( \dfrac{(2-\gamma).g_k(\gamma).c}{\mu + \lambda.g_k(1)} \right)^{\frac{1}{3-\gamma}} \cdot g_k(1) = \dfrac{R_0 - b}{a} & (4.38a) \\ \sum_{k=0}^{n-1} \left( \dfrac{(2-\gamma).g_k(\gamma).c}{\mu + \lambda.g_k(1)} \right)^{\frac{1}{3-\gamma}} = \dfrac{2^{n_b-1}}{\delta} & (4.38b) \end{cases}$$

where $g_k(\gamma)$ is defined in (4.31), a and b in (4.19-4.20), c and d in (4.33-4.34), $n_b$ and $\delta$ in Section 3.3.1. The system (4.38) is solved numerically. The results of this minimization are discussed in Section 4.3

Note that, one can find a suboptimal but analytical solution to (4.35) by omitting the rate constraint (4.35b), we only minimize the distortion and find the following analytical

solution:

$$s_k^{\circ} = \frac{2^{n_b-1} \cdot g_k(\gamma)^{\frac{1}{3-\gamma}}}{\delta \cdot \sum_{k=0}^{n-1} g_k(\gamma)^{\frac{1}{3-\gamma}}} \tag{4.39}$$

The solution proposed in [MMM+11] is a particular case of equation (4.39). More precisely, the TMO in [MMM+11] is the same as our high rate regime TMO, computed for $\gamma = 0$ (and thus $\nabla_x^{\gamma} = 1$, $g_k(0) = \sum_{x=x_k}^{x_{k+1}} \mathbb{P}(X = x)$).

On the other hand, if we remove the spanning constraint (4.35c), we obtain:

$$s_k^{\dagger} = \alpha . \frac{g_k(\gamma)^{\frac{1}{3-\gamma}}}{g_k(1)^{\frac{1}{3-\gamma}} \cdot \sum_{k=0}^{n-1} \left( g_k(1)^{\frac{2-\gamma}{3-\gamma}} \cdot g_k(\gamma)^{\frac{1}{3-\gamma}} \right)} \tag{4.40}$$

where $\alpha$ is adjusted to respect the removed spanning constraint. The derivation of these two TMOs (4.39) and (4.40) is detailed in Appendix B.

**Remark.** Experimentally, we observe that the TMO curve defined by (4.37) is always between the TMO curves defined by $\{s_k^{\dagger}\}$ and $\{s_k^{\circ}\}$ respectively. This observation also holds when the TMO curves defined by $\{s_k^{\dagger}\}$ and $\{s_k^{\circ}\}$ intersect.

### 4.2.5 R-D Optimization with SDR quality constraint

We now consider the use of an additional SDR quality constraint, i.e, the optimization problem given by (4.5). The models for $D_{HDR}$ and $R_{SDR}$, expressed in the previous section, remain the same, however we also need to model the SDR constraint quality $D_{SDRREF}$.

As in [MMNW13, KD13], the SDR quality compared with a reference tone-mapped SDR is computed as:

$$D_{SDRREF} = \mathbb{E}\Big[ (G(X) - Y)^2 \Big] \tag{4.41}$$

$$D_{SDRREF} = \sum_{k=0}^{n-1} \sum_{x=x_k}^{x_{k+1}} \Big( G(x) - F(x) \Big)^2 \cdot \mathbb{P}(X = x) \tag{4.42}$$

with $G$ a global TMO optimized for the SDR perceptual quality and $F$ the optimized TMO. In [MMNW13, KD13], a coarse assumption is made:

$$\forall x \in [x_k, x_{k+1}[, \qquad x = \frac{x_{k+1} - x_k}{2} \tag{4.43}$$

which yields a correct result if the function to integrate $\big(G(x) - F(x)\big)^2$ is linear in $x$ and if the distribution is uniform on $[x_k, x_{k+1}[$. This leads to a piecewise constant distortion measure. Instead, we propose to keep only the uniform distribution assumption (4.44):

$$\forall x \in [x_k, x_{k+1}[, \quad \mathbb{P}(X = x) = \frac{\mathbb{P}(x_k \leq X < x_{k+1})}{\delta} \tag{4.44}$$

Then (4.42) becomes:

$$D_{SDRREF} = \sum_{k=0}^{n-1} \frac{p_k}{\delta} \cdot \sum_{x=x_k}^{x_{k+1}} \Big( G(x) - F(x) \Big)^2 \tag{4.45}$$

where $p_k = \mathbb{P}(x_k \leq X < x_{k+1})$.
If we parameterize the function $F$ with (3.30) and the function $G$ with:

$$G(x) = \Big((x - x_k).t_k + z_k\Big) \qquad \forall x \in [x_k, x_{k+1}[ \tag{4.46}$$

Then, the expression (4.45) becomes:

$$D_{SDRREF} = \sum_{k=0}^{n-1} p_k \cdot h\Big(t_0 ... t_k,\ s_0 ... s_k,\ \delta\Big) \tag{4.47}$$

with:

$$h\Big(t_0..t_k,\ s_0..s_k, \delta\Big) = \frac{\Big(\delta.t_k + z_k - \delta.s_k - y_k\Big)^3 - \Big(z_k - y_k\Big)^3}{\delta \cdot 3 \cdot (t_k - s_k)} \tag{4.48}$$

The derivation of (4.48) is detailed in Appendix C.

To test the validity of the assumption (4.44), we compare the MSE described in (4.42) and the estimated distortion expressed in (4.47) in Figure 4.10. This test has been carried out with 26 different images and 3 different TMOs $F(s_{0..k})$ (same test set as in Figure 4.2), where the slopes in $\{s_{0..k}\}$ are randomly drawn to yield a strictly increasing TMO that maps the whole range of the input HDR image. In this test, the reference TMO $G(t_{0..k})$ is always the same global TM0 [RSSF02] approximated with $n$ slopes. The correlation be-

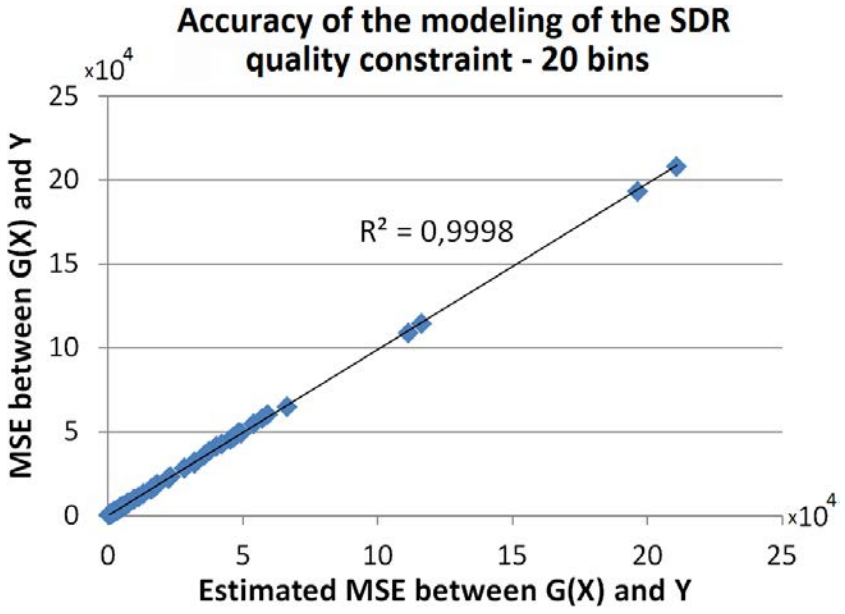



Figure 4.10 – Experimental validation of the estimated SDR quality constraint (4.47)

tween the estimated and real values (4.42) is very high, demonstrating that the assumption made is very acceptable. Figure 4.10 shows the distortion when we use 20 slopes for both TMOs. As in Figure 4.2, the results were the same with 50 slopes. This test is also performed with the SDR quality constraint proposed in [MMNW13], the results are the same. One possible explanation of the close results obtained for both $D_{SDRREF}$ expression is

the use of very different tone mapping curves for the comparison. We suppose that with the use of closer tone mapping curves, expression in [MMNW13] leads to approximation errors for $D_{SDRREF}$.

With the developed models for $R_{SDR}$ (4.18), $D_{HDR}$ (4.32), $D_{SDRREF}$ (4.47) and with the spanning constraint, the minimization problem (4.5) becomes:

$$\min_{QP} \min_{s_1 \ldots s_n} \quad c \cdot \sum_{k=0}^{n-1} \left( \frac{g_k(\gamma)}{s_k^{2-\gamma}} \right) + d \tag{4.49}$$

$$\text{s.t.} \quad \begin{cases} \sum_{k=0}^{n-1} s_k \cdot g_k(1) \leq \frac{R_0 - b}{a} \\ \sum_{k=0}^{n-1} s_k = \frac{2^{n_b}-1}{\delta} \\ \sum_{k=0}^{n-1} p_k \cdot h\big(t_0 \ldots t_k, s_0 \ldots s_k, \delta\big) \leq D_0 \end{cases}$$

This problem can be solved numerically with the interior-point optimization method. The results of this minimization are discussed in Section 4.3.

## 4.3 Experimental Results

### 4.3.1 R-D Optimization

In this section, we show the RD performances achieved with the backward compatible HDR compression scheme of Figure 4.1 with the TMO curve obtained in Section 4.2.4. As in the previous experiments, the TMOs are computed with 20 slopes. Tests were made with 20 and 50 slopes and the RD performances were the same.

The proposed TMO design considers one TMO per frame hence assumes the TMO parameters to be sent for each frame. This cost has been counted in the rate cost. However, the rate cost for transmitting the TMO parameters is negligible. Using 16 bits for each of our 20 slopes leads to an extra rate of 0.0002 bits/pixel for a 1920x1080 image ($\frac{16*20}{1920*1080}$). This cost can be further reduced by considering entropy coding of the slopes.

The original HDR image is first perceptually transformed using the PQ-2084 OETF to follow the contrast sensitivity function of the human eye, and uniformly quantized to 12 bits. This provides the HDR content denoted $X$. $X$ is then tone mapped using the TMO that minimizes (4.35) and the resulting SDR content is encoded with the HEVC reference software (HM 16.2).

First, the optimal TMO is derived. Rather than solving Problem (4.35), where an exhaustive search over the QP parameter is performed for a given rate constraint $R_0$, we solve an equivalent problem, where an exhaustive search over the rate constraint $R_0$ is performed for a given QP parameter, as described in Algorithm 1.

**Algorithm 1** RD simulation: convex hull

**Require:** HDR image $X$
  **for all** QP **do**
    Compute $\nabla_X$, then $p(X, \nabla_X)$
    Compute $a$, $b$, $c$, $d$, $\gamma$ and $g_k(\gamma)$ from (4.19), (4.20), (4.33), (4.34), Fig-4.8(a) and (4.31) respectively.
    Compute the two TMO bounds $\{s_k^\dagger\}$ (4.39) and $\{s_k^\circ\}$ (4.40)
    For each bound, compute the necessary rate: $R_{SDR}^\dagger$ and $R_{SDR}^\circ$ from (4.18)
    Quantize the interval $[R_{SDR}^\dagger, R_{SDR}^\circ]$ to get 20 values.
    **for all** $R_0 \in [R_{SDR}^\dagger, R_{SDR}^\circ]$ **do**
      Solve numerically (4.38) to get $\lambda$, $\mu$
      Compute $\{s_k\}$ with (4.37)
      Apply the TMO (defined by $s_k$) to $X \rightarrow$ SDR image
      Encode the SDR image $\rightarrow R_{SDR}$
      Inverse Tone Mapping of the decoded SDR image
      Compute $D_{HDR}$
    **end for**
  **end for**
  Choose the best $R_{SDR} - D_{HDR}$ point

Note that in Algorithm 1, the search over the optimal $R_0$ is performed in a reduced interval $[R_{SDR}^\dagger, R_{SDR}^\circ]$. This is possible because the TMO curves (4.39) and (4.40) are obtained without the spanning and rate constraints respectively and therefore lead to lower and upper bounds on the required encoding rate.

Figure 4.11 shows the RD performance obtained with Algorithm 1. For each QP value, a RD curve parameterized by $R_0$ shows the RD performance obtained with the optimized TMO. Note that the left-most and right-most points of each curve at QP fixed correspond

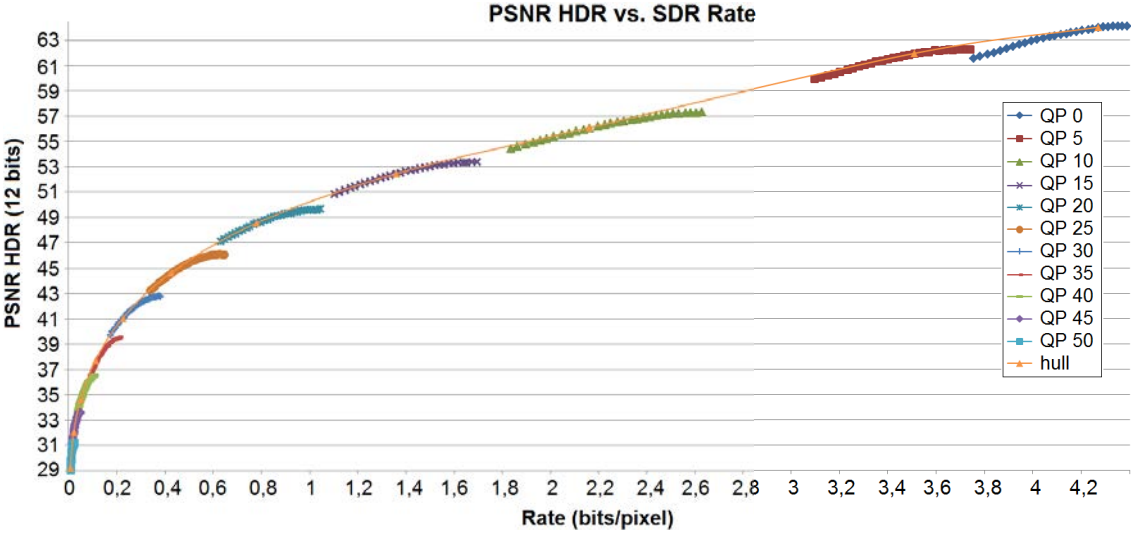



Figure 4.11 – RD optimization without SDR quality constraint Algorithm-1. Given a QP value, different $R_0$ values are tested. The best RD performance is given by the convex envelop of all the points.

to the rates $R^{\dagger}_{SDR}$ and $R^{\circ}_{SDR}$. Then the convex hull of all these curves is drawn. The optimal rate constraints $R_0$ correspond to the case, where each individual curve (QP fixed) is tangent to the convex hull.

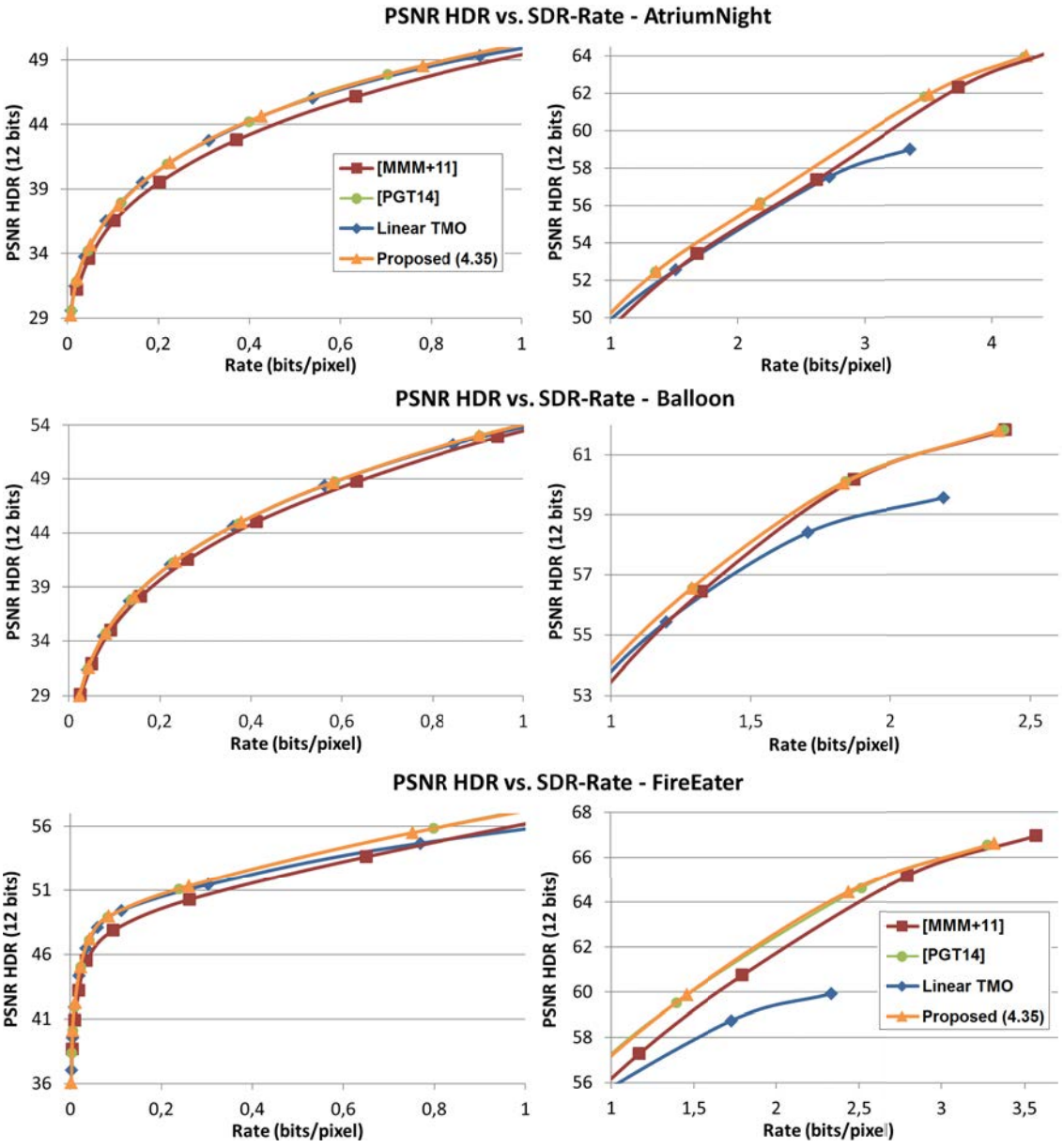



Figure 4.12 – Distortion between the original HDR image $X$ and the reconstructed image $\tilde{X}$ vs SDR rate. From top to bottom: Results for image AtriumNight at low and high rates - Results for image Balloon at low and high rates - Results for image FireEater at low and high rates

The exhaustive search in Algorithm 1 is rather complex as it requires encoding the SDR image for each $R_0$ value. However, the loop over $R_0$ can be avoided by learning a correspondence between this rate constraint $R_0$ and the QP parameter. To learn such a relation, an exhaustive search is performed for multiple images such that we obtain the optimal $R_0$ values for each QP. Then, a law is fitted to these optimal $R_0$ points. Given a QP

value, at run time, this law allows us to have directly the optimal $R_0$ value. Deducing the optimal TMO then becomes feasible in real-time. Fig 4.12 shows the RD performances obtained with this straightforward method, in comparison with several existing TMOs: a simple linear TMO, a TMO that minimizes the distortion only [MMM+11] and a TMO optimized according to a RD criterion, but with a RD model valid in the high rate regime only [PGT14]. The experiment is done with images from the MPEG test set and the one proposed in [MPII, War03]. Figure 4.13 shows different tested images. More images are tested in the supplementary materials [GRG+a]. The upper charts show the PSNR of the

Figure 4.13 – Sample of tested HDR images. For display purposes, the HDR images are tone-mapped with [RSSF02]. From left to right: AtriumNight, Balloon, FireEater

reconstructed HDR image versus the SDR image bitrate for high QPs (0 to 1 bit/pixel) while the lower ones show the same results for lower QPs.

First, we notice that, as expected, the proposed algorithm, performing a RD optimization, allows to outperform the linear and [MMM+11] TMOs at any rate. More precisely, the Bjontegaard [Bjo01] rate gains between the proposed and the linear TMOs are 3.0% for AtriumNight, 11.4% for FireEater and 1.4% for Balloon. The Bjontegaard rate gains between the proposed TMO and [MMM+11] are 15.5% for AtriumNight, 26.9% for FireEater and 7.5% for Balloon. At high rates, the TMO always seems to converge towards the same R-D performances as [MMM+11] and towards the same R-D performances as the linear TMO at low rates. For each image, the R-D performances are different but the behavior seems to be the same regardless of the dynamic range or spatial activity (defined in [VSLD14]), as one can see in Table 4.1.

| | **AtriumNight** | **Balloon** | **FireEater** |
|---|---|---|---|
| **RD Gain to [MMM+11]** | 15.5% | 7.5% | 26.9% |
| **RD Gain to Linear** | 3.0% | 1.4% | 11.4% |
| **Dynamic Range** | $4.10^9$ | $9.10^7$ | $7.10^6$ |
| **Spatial Activity** | 76 | 82 | 38 |

Table 4.1 – RD Gains, Dynamic Range and Spatial Activity of the tested images as defined in [VSLD14]

Regarding [PGT14], the performances are very close to the proposed method for all the tested images. This result is surprising since we demonstrated the better accuracy of the proposed model for the rate and distortion in Section 4.2.1 and Section 4.2.3. However, [PGT14] learns the Lagrangian multiplier in the RD optimization. This learning compensates for the independence and invariance assumptions (see Section 4.2.2) made in their

distortion model and leads to a corrected RD model, which is as accurate as our. This shows indeed how important the accuracy of the RD model is.

### 4.3.2 R-D Optimization with SDR quality constraint

We now analyze the RD performance obtained with the TMO solution of (4.49) with a new constraint on the SDR perceptual quality. The problem is solved with the interior point algorithm (optimization toolbox from Matlab). The resulting TMO depends on two parameters: the $D_{SDRREF}$ constraint $D_0$ and the rate constraint $R_0$. $D_0$ is adjusted to 34.2dB (MSE=25) to insure sufficient similarity with the SDR reference image.

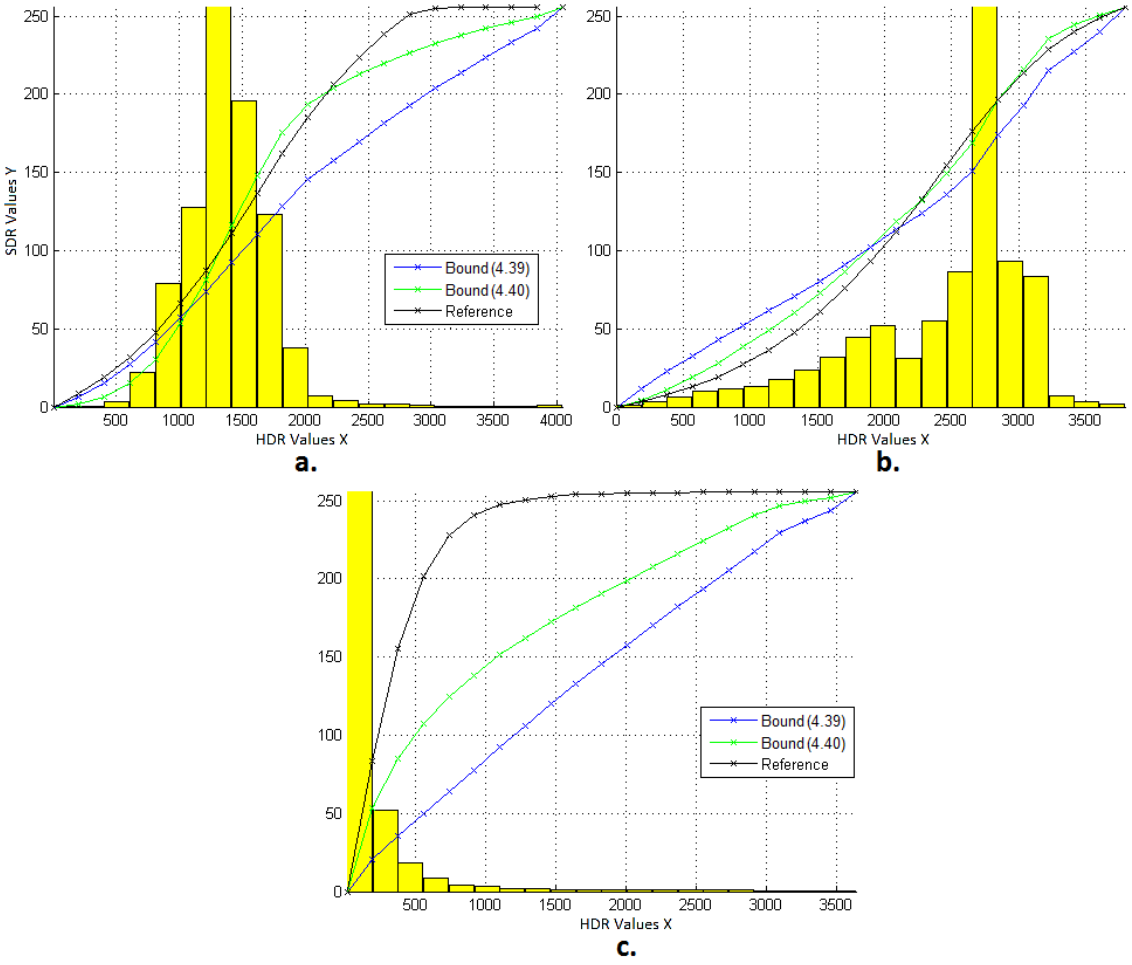



Figure 4.14 – Different tone mapping curves for the following images - (a) AtriumNight - (b) Ballon - (c) Fireater. The (4.39) bound by the slopes $s_k^{\circ}$, the (4.40) bound is defined by the slopes $s_k^{\dagger}$ and the SDR perceptual reference TMO is [RSSF02]

The rate constraint $R_0$ needs to be higher than in the previous Section 4.3.1 since adding a new constraint will necessarily increase the required rate. Second, the rate constraint will depend on the reference Tone-Mapping chosen. For instance, Figure 4.14 plots the bounding TMOs described in Section 4.3.1 and the chosen reference Tone-Mapping [RSSF02], the PTR (Photographic tone reproductor). For AtriumNight and Ballon images, the PTR is close to the two bounding TMOs, whereas the PTR is quite far from

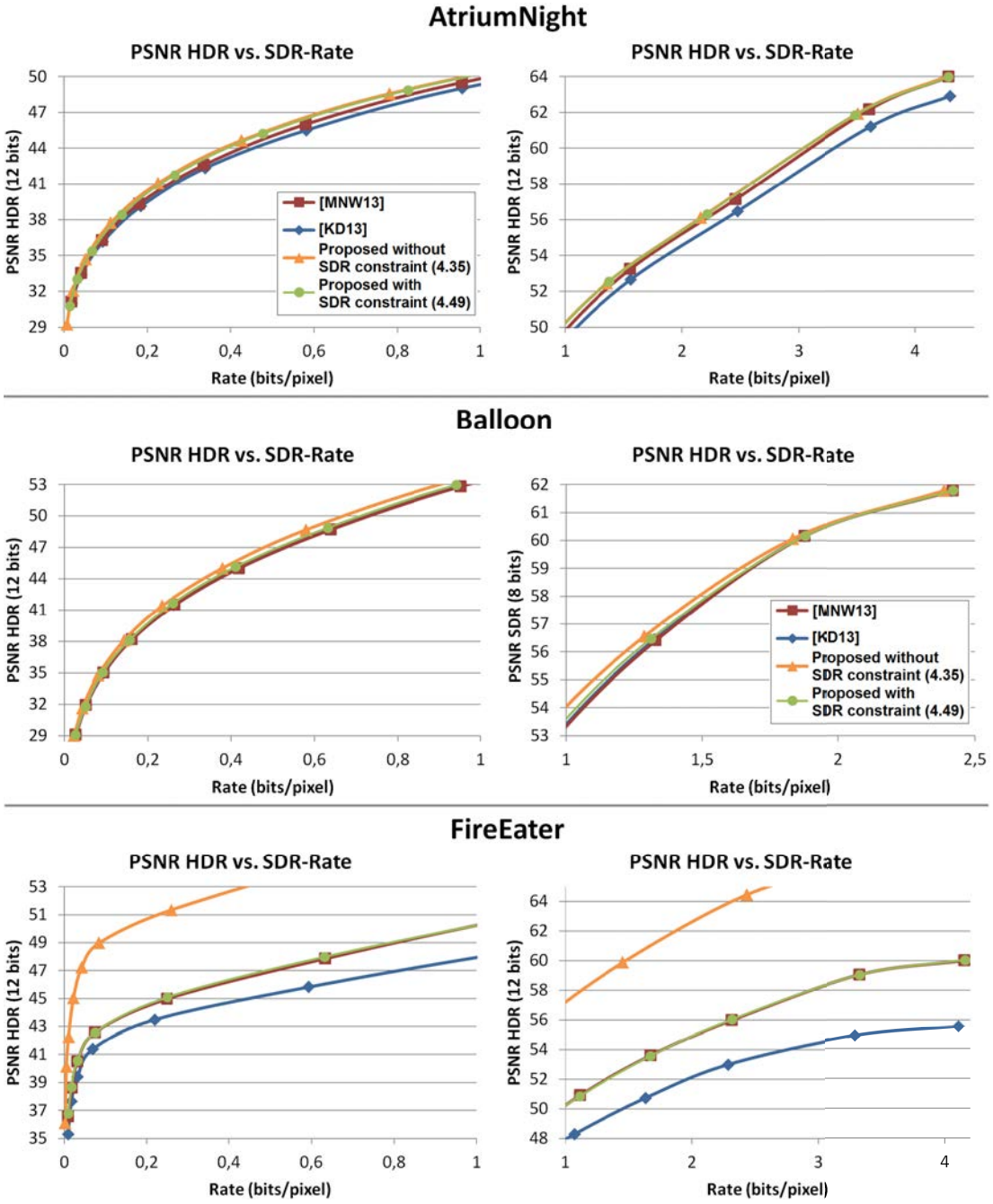



Figure 4.15 – Results for the images AtriumNight Ballon and FireEater. Distortion between the original HDR image $X$ and the reconstructed image $\tilde{X}$ vs SDR rate - Left: at low rates - Right: at high rates

these bounds for FireEater and this will significantly increase the rate (since we observed that the optimal TMO with best RD trade off lies in between the two bounds). Therefore, to find the optimal TMO for a given QP parameter, a first rate constraint $R_{max}$ is computed as the maximum between the rate needed with the PTR TMO and the rate needed for $\{s_k^\dagger\}_{k=0..n}$ (4.39), i.e. the upper bound on the rate constraint computed in the previous Section 4.3.1. $R_{max}$ ensure to find a solution to (4.49). Using the previously learnt

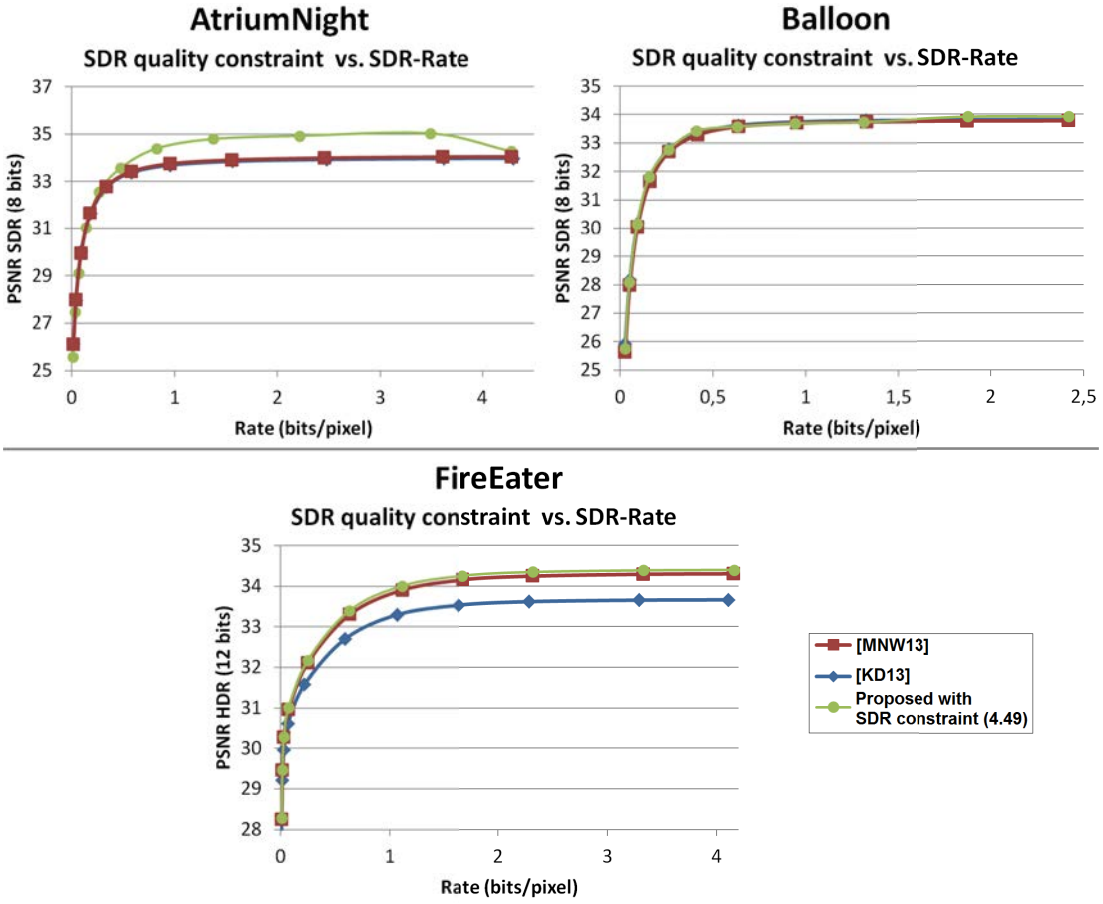



Figure 4.16 – Results for the images AtriumNight Ballon and FireEater - SDR quality constraint vs SDR rate

relationship between $R_0$ and the QP parameter for the RD optimization without SDR constraint (Section 4.3.1), we can compute the optimal rate constraint $R_0$. If the optimal $R_0$ does not satisfy the SDR quality constraint when solving (4.49), $R_0$ is increased until we reach $R_{max}$. The loop stops when we find the smallest $R_0$ value that satisfies the SDR constraint. The complexity of this solution is fast: under 1 sec with the Matlab optimization toolbox.

The R-D performance of this TMO is compared with two state of the art TMO designed for R-D performances under a quality constraint for the SDR image [MMNW13, KD13] and with the unconstrained one defined in the previous section. With this test, the R-D losses induced by the SDR quality constraint can be measured. Experimental test conditions are identical to the previous section. The results are shown on Figure 4.15 and Figure 4.16.

The left and right charts in Figure 4.15 plot the PSNR of the reconstructed HDR image versus the SDR image bitrate at low and high rate respectively. The charts in Figure 4.16 plot the PSNR of the SDR constraint quality $D_{SDRREF}$ versus the SDR bitrate. The proposed tone mapping saves 11.78% 39.17% and 2% of bitrate for the image AtriumNight, FireEater and Ballon compared to [KD13] and 5.44%, 1.95% and 2.66% compared to [MMNW13] for the same or even better SDR quality constraint. This can be explained by the fact that our proposed RD model is more accurate than those in [MMNW13] and

[KD13]. Moreover, as expected from Figure 4.14, the SDR quality constraint has a much bigger impact on the image FireEater.

## 4.4 Conclusion

This chapter presented two new TMOs for HDR backward compatible compression. Using the HDR image gradient, we provided new statistical models for estimating $D_{HDR}$, $R_{SDR}$ and $D_{SDRREF}$ and showed their accuracy to the real data. The first TMO minimizes the distortion of the HDR image under a rate constraint on the SDR layer. The second TMO remains the same minimization with an additional constraint to preserve the SDR perceptual quality.

Experimental results show that the first TMO achieves the optimal rate-distortion performances with a global tone mapping and can be computed in real-time. As expected, the second TMO leads to the best trade-off between rate-distortion performances and quality preservation of the SDR signal, in comparison to state of the art methods that solve approximations of the original optimization problem. Besides, we showed that spatial gradient is an accurate estimator of both rate and distortion.

The TMOs presented in this chapter only encode still image, but they can be easily extended to video by computing them for each frame. However, one could argue about the formation of temporal artifacts like flickering, or about the optimality of the TMOs for RD performances. Indeed, the proposed TMO minimizations could be reformulate to account for a temporal gradient instead or along with the spatial one. The following chapter will consider the extension to video compression, taking into account a better utilization of the temporal axis for additional RD gains and an improved temporal consistency.

# Chapter 5

# Optical-Flow Based Weighted Prediction for SDR and Backward Compatible HDR Video Coding

In the previous chapter, we have developed efficient TMOs to compress HDR images and videos using only spatial information. For videos, the TMO is independently computed for each frame and its parameters are sent as metadata. However, one can argue that the temporal coherence on the resulting SDR video is not preserved and thus has an impact on the compression performances. In this chapter, we focus on backward-compatible compression of HDR videos. Combining spatial and temporal information to compute the TMO is expected to further improve the coding performances.

SDR contents generated with a TMO usually present more temporal brightness variations than manually-graded or recorded SDR contents. Two major problems can arise from these observations. The first problem is a poor SDR temporal quality. Indeed, as explained in Section 1.4.2.3, when compared to HDR video, SDR content can present different temporal artifacts [BCTB14a]. The second problem of TMO-generated SDR videos is their compression. More temporal variations almost always means worse rate-distortion (RD) performances. In backward-compatible schemes, this problem also impacts the HDR RD performances since the total rate also includes the SDR rate.

Improving the temporal quality of tone mapped SDR videos has been extensively addressed in the literature. Some works proposed to apply a temporal filtering on the TMO curve of each frame [PTYG00, MDK08] and others proposed more elaborate operators [IFM05, BCTB14b]. These works were evaluated in terms of subjective temporal quality [EUWM13] but hardly ever in terms of RD performances. To the best of our knowledge, only [BTCB13] studied the RD impact of their temporal-quality designed TMO. They concluded that temporally filtered TMOs, such as [PTYG00, MDK08], have no impact on either the SDR and HDR RD performances whereas their TMO diminishes the HDR RD performances but improves the SDR ones. However, the SDR RD performances should be treated with caution since the SDR reference differs for each evaluated TMO, especially since their TMO reduces the dynamic range of each SDR frame.

Some solutions were proposed to improve the HDR RD performances of backward compatible schemes by modifying the SDR signals in order to reduce the temporal discontinuities in the resulting SDR video. A first set of propositions consists in applying a

temporal filter over the TMO curves of each frame. However, this solution experimented in [MMM+11] and [KD14], as in [BTCB13], did not improve the HDR RD performances and sometimes even worsened them. Probably because the temporal filtering modifies the optimal tone curve for each frame. Another contribution [OLVD16] consists in compensating the temporal variations in the TMO optimization process. The authors in [OLVD16] introduced a new TMO minimizing the HDR distortion under a temporal constraint. This TMO is only optimized to improve the HDR RD performances but on the other hand the SDR video is modified, which may strongly alter the original artistic intent.

In this chapter, we propose novel weighted prediction modes and algorithm to compensate for illumination variations in the SDR signals that allows preserving the artistic intent present in the HDR signal rather than altering the temporal effect. In this way, we improve the SDR RD performances and therefore the HDR ones, but also preserve the SDR temporal quality, as well as the original artistic intent. Indeed, by improving the coding efficiency of the SDR content, we improve the overall coding efficiency of backward compatible compression schemes. And since the processing is directly applied into the video codec, the SDR content is not impacted and therefore its temporal quality is preserved.

Weighted prediction (WP) is already available in the H264 and HEVC codec and has already been addressed in many works. In the HEVC reference software, the WP process is rather simple. It consists of computing a simple linear transformation between the source image and a reference image without motion compensation [SBS14, Chap. 5]. Using a global WP is efficient to predict global temporal brightness variations but several solutions were proposed to improve the WP for local temporal brightness variations. In [RdSdF01], they proposed to compute a linear WP for each block and send along the WP parameters with each block. The authors in [YTB05] noticed that the overhead of sending the WP parameters for each block is non negligible and therefore impacts the RD performances in H264. Instead, they proposed to compute the local linear WP for each block using the neighboring pixels. In this way, no overhead cost is added since the WP can be recovered at the decoding process with the previously decoded pixels. However, the WP parameters will lack of precision since they do not use the current block values but the neighboring ones. This method was also studied in HEVC [ZLF+15] and proposed as a new tool [LCC+15] for future video codec standardization [SBB+17]. Another solution is to take advantage of the multiple reference frames in H264 or HEVC [KK11, TLCS13, TCS13]. Each reference frame contains the same frame with different WP parameters and each block can point to the most appropriate one. However, the number of different WP is limited by the number of reference frames. The common characteristic of all these works is to use a block-based motion compensation for the computation of their local WPs and to always use linear functions, for global WP or local WP. Indeed, the block-based motion compensation often fails to recover the real motion between two frames, especially with temporal brightness variations.

In this chapter, we propose both a new algorithm to compute the WP mode parameters and different WP models. Instead of computing the WP directly between the consecutive images, the proposed WP algorithm uses an optical-flow based motion compensation. Furthermore, to predict the complex temporal variations between SDR frames, two new WP models, implying syntax and decoder modifications, are introduced: a global non-linear WP and a local non-linear WP. The chapter is organized as follows. Section 5.1 demon-

strates the equivalence between the proposed global non-linear WP and the temporally constrained TMO [OLVD16]. An overview of the proposed WP algorithm is proposed in section 5.2 with further details on the HEVC implementation in section 5.3. Finally, experimental results are presented in section 5.4.

## 5.1 Temporally constrained TMO and Weighted Prediction

### 5.1.1 Temporally constrained TMO

In [OLVD16], the authors proposed to improve the TMO optimization described in [MMM+11] by accounting for the rate of the tone mapped SDR video. They expressed this rate as a function of the temporal activity $C(S_t)$ of the SDR content, i.e. as:

$$R \propto C(S_t) = \sum_{i,j} \Big( S_t(i,j) - M\big(S_{t-1}(i,j)\big) \Big)^2 \tag{5.1}$$

with $S_t$ the current tone mapped SDR frame, $S_{t-1}$ the previous tone mapped SDR frame, $(i,j)$ the pixel position in the respective frame, and $M$ the motion compensation between $S_{t-1}$ and $S_t$. However, when computing the current TMO, called $F_t$, the SDR frame $S_t$ is unknown and the motion compensation cannot be computed. This is illustrated in Figure 5.1. To overcome the motion compensation problem, the authors estimated $M$ between the

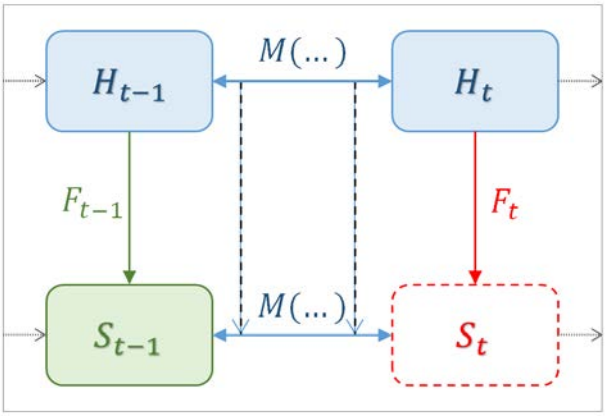



Figure 5.1 – Tone Mapping computation for two consecutive HDR frames.

HDR frames, $H_{t-1}$ and $H_t$, and applied it to the SDR frame $S_{t-1}$. However, the constraint (5.1) still needs to be expressed as a function of the TMO $F_t$. To do so, they parameterize $F_t$ as a piecewise linear tone curve (see Sec.3.3.1), but also make a coarse assumption over the distribution of the HDR values in each bin of the HDR histogram (same assumption as in (4.43)). We do not detail the obtained expression here since it has no interest in the proposed development, we only wanted to show the approximations in their solution.

Adding the rate constraint (5.1) to the distortion optimization problem proposed in [MMM+11], already defined in equation (4.26), the optimization problem in [OLVD16]

is:

$$\min_{s_k} \sum_{k=1}^{N} \left(p_k.s_k^{-2}\right) + \lambda.C(S_t) \tag{5.2}$$
$$s.t. \quad \sum_{k=1}^{N} s_k = \frac{2^{nb}-1}{\delta}$$
$$\forall k,\ s_k > 0$$

with $nb$ the bit-depth for the SDR frame, $\delta$ the selected interval for the HDR frame histogram, $p_k$ the probability of the k-th histogram bin, $s_k$ the TMO ($F_t$ in Fig 5.1) slope for the $k$-th bin and $\lambda$ a Lagrangian multiplier to simplify the constrained problem into an unconstrained one (For more details about the TMO parameterization see Sec.3.3.1). The remaining constraint ensures that the TMO curve covers all the SDR dynamic.

Focusing only on the minimization of the rate constraint (5.1) (i.e. the specific case where $\lambda = \infty$ in (5.2)), and ignoring the simplifications made in [OLVD16], then (5.1) becomes:

$$C(S_t) = \sum_{i,j} \Big(S_t - M(S_{t-1})\Big)^2 \tag{5.3}$$

Note that, in this equation and the following ones, we removed the pixel index $(i, j)$ for a simpler notation.

$$C(S_t) = \sum_{i,j} \Big(F_t(H_t) - M(F_{t-1}\big(H_{t-1}\big))\Big)^2 \tag{5.4}$$

With $F_t$, the TMO applied at the $t$-th HDR frame $H_t$ therefore providing the SDR frame $S_t$. If one defines $BV$ the brightness variations from $M(H_{t-1})$ the motion compensated HDR frame $H_{t-1}$, to $H_t$, then $H_t = BV\big(M(H_{t-1})\big)$ and:

$$C(S_t) = \sum_{i,j} \Big(F_t(H_t) - M(F_{t-1}\big(M^{-1}(BV^{-1}(H_t))\big)\Big)^2 \tag{5.5}$$

$$C(S_t) = \sum_{i,j} \Big(F_t(H_t) - F_{t-1}\big(BV^{-1}(H_t)\big)\Big)^2 \tag{5.6}$$

Thus, the optimal solution becomes obvious:

$$F_t(...) = F_{t-1}\big(BV^{-1}(...)\big) \tag{5.7}$$

The TMO $F_{t-1}$ is a global invertible function and $BV^{-1}$ usually is a pixelwise non-invertible function. The challenge here, is to reproduce $BV^{-1}$ with the global invertible function $F_t$.

However, optimizing (5.2) with $\lambda = \infty$ is clearly not optimal in the context of backward compatible compression. First, the term considering the HDR distortion is then completely ignored which is clearly not optimal for RD performances. Secondly, the visual quality of the resulting SDR video will be strongly impacted. Using the optimal solution, all global luminance variations over an HDR video will disappear in the SDR video. It is therefore necessary to compromise between distortion and rate minimization but also temporal regularization. In practice the authors in [OLVD16] empirically chose $\lambda = 0.1$ for their experiments.

### 5.1.2 Weighted Prediction vs. Temporally constrained TMO

As explained previously, the aim of the proposed weighted prediction is to improve the RD performances of the temporally constrained TMO proposed in [OLVD16] but with an additional constraint namely to improve the SDR temporal quality. The proposed method shares many similarities with this TMO but also differs on key points, as illustrated in Figure 5.2. Indeed, using a WP avoids to compromise between distortion minimization and

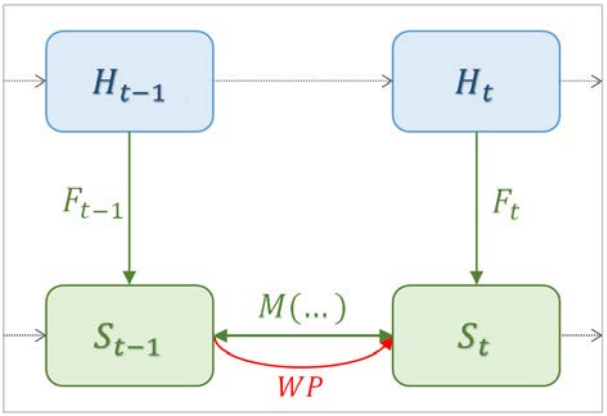



Figure 5.2 – Weighted prediction computation for two consecutive HDR frames

temporal regularization. It transforms the constrained problem (5.2) into two consecutive problems, first minimizing the HDR distortion of $H_t$ using the TMO [MMM+11] and in a second step, inside the video codec, minimizing the brightness variations between $S_{t-1}$ and $S_t$ with a WP. Consequently, the proposed strategy is easily adaptable to any TMO, which is not the case for [OLVD16]. One can use a TMO that preserves the SDR perceptual quality or a TMO that minimizes the distortion under a rate constraint for the still image, like in Chapter 4, and thus benefit from a reduction of spatial and temporal complexity. Having two steps also allows to know the two SDR frames before computing the WP and therefore allows to directly compute the optical flow between $S_{t-1}$ and $S_t$. One could argue that the computation of the optical flow is more accurate on the HDR level however, experimentally, we didn't found noticeable differences between the computed motions.

Another advantage of the proposed WP is to preserve the temporal consistency of the SDR content. Indeed, in [OLVD16] the frame $S_t$ is deliberately modified for compression purposes, therefore impacting the original temporal consistency when compared to HDR content. Using a weighted prediction, the frame $S_{t-1}$ is used to predict $S_t$ directly inside the encoding loop. Therefore, the original input SDR video is preserved.

Focusing on the second step, minimizing the brightness variations with a WP is really similar to minimizing (5.3):

$$C(S_t) = \sum_{i,j} \Big( S_t - WP\big(M(S_{t-1})\big) \Big)^2 \tag{5.8}$$

$$C(S_t) = \sum_{i,j} \Big( S_t - WP\big(M(F_{t-1}(H_{t-1}))\big) \Big)^2 \tag{5.9}$$

With $BV$ the brightness variations from $M(H_{t-1})$ to $H_t$:

$$C(S_t) = \sum_{i,j} \Big( S_t - WP\big(M(F_{t-1}(M^{-1}(BV^{-1}(H_t))))\big) \Big)^2 \tag{5.10}$$

$$C(S_t) = \sum_{i,j} \Big( S_t - WP\big(F_{t-1}(BV^{-1}(H_t))\big) \Big)^2 \tag{5.11}$$

$$C(S_t) = \sum_{i,j} \Big( S_t - WP\big(F_{t-1}(BV^{-1}(F_t^{-1}(S_t)))\big) \Big)^2 \tag{5.12}$$

where (5.11) follows from the fact that $M^{-1}$ can be permuted with $F_{t-1}$ since $M^{-1}$ is a pixels displacement and $F_{t-1}$ a global function over the pixel values. And therefore, the optimal solution is to find:

$$WP^{-1}(...) = F_{t-1}\Big(BV^{-1}\big(F_t^{-1}(...)\big)\Big) \tag{5.13}$$

By comparing (5.7) and (5.13), one realizes that the proposed WP needs to reproduce a composition of three functions instead of two for (5.7). As explained previously, the most difficult part is to estimate the pixelwise non-invertible function $BV$ since $F_{t-1}$ and $F_t$ are non-linear invertible functions. Approximating $WP$ by a linear function, as currently done in HEVC, seems suboptimal as it is a composition of three functions, which are each of them non linear. Therefore, our first proposal is to use a non-linear function as $WP$ model in HEVC. To the best of our knowledge, global non-linear WP has never been proposed so far. As demonstrated in this section, this WP is supposed to provide better RD performances than the ones obtained by [OLVD16]. Indeed, we have removed the compromise between HDR distortion and SDR temporal regularization and in addition, the proposed non-linear WP is not limited to invertible functions since it is not needed to reconstruct the HDR content.

Regarding the second proposed WP model, to better estimate the function $BV$, we extend our global approach to non-linear local WP. As explained previously, global and local linear WP have already been studied however, for comparison purposes, we also implemented these solutions in our framework. In this way, we can better assess the benefits of non-linear WPs.

## 5.2 New models of Weighted Prediction

As explained in the introduction, we present a new WP algorithm and new WP modes. To clarify, the WP modes describe the functions $WP$ used to create the frame $P_t = WP(S_{t-1})$ which predicts $S_t$ and the WP algorithm is the method used to estimate the WP model parameters. Note that, the proposed WP algorithm only implies encoder modifications and is therefore compatible with the current HEVC standard. In contrast, the new WP modes require decoder modifications and thus, a modification of the HEVC standard.

Figure 5.3 summarizes the WP algorithm and WP modes presented in the following sections. The HEVC standard algorithm is improved with an optical flow based algorithm and the WP modes are extended from one type to four.

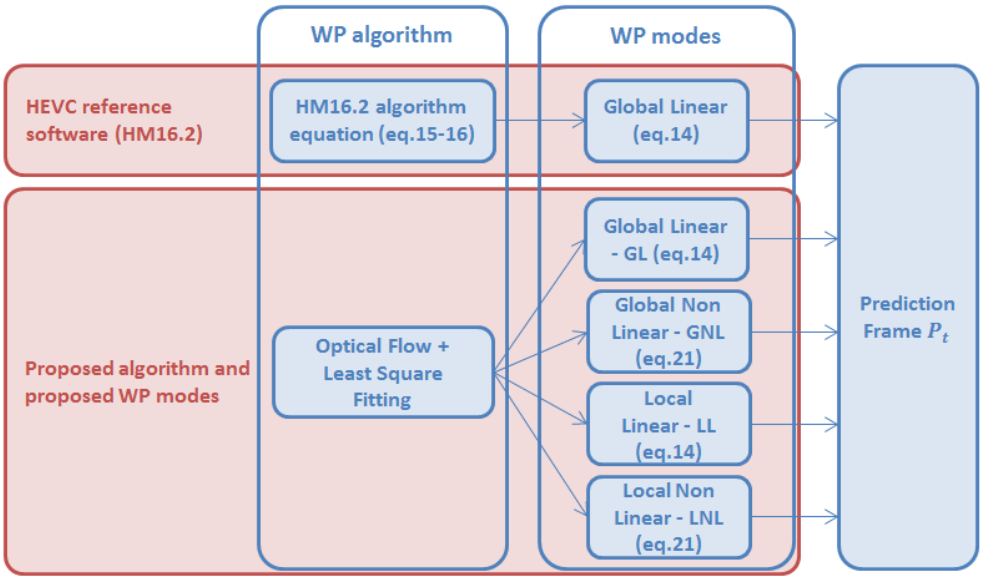



Figure 5.3 – Weighted Prediction Scheme - WP algorithms and WP modes.

### 5.2.1 HEVC Weighted prediction

In the HEVC reference software (HM16.2), the WP is a global linear function directly computed between $S_{t-1}$ and $S_t$ using the mean and variance of each frame. The prediction frame for $S_t$ is obtained with:

$$P_t(i,j) = W \cdot S_{t-1}(i,j) + O \tag{5.14}$$

with $(i,j)$ the pixel position in the frame and where:

$$W = \frac{\sum_{\forall i,j}\left(S_t(i,j) - \sum_{\forall i,j}\frac{S_t(i,j)}{N}\right)}{\sum_{\forall i,j}\left(S_{t-1}(i,j) - \sum_{\forall i,j}\frac{S_{t-1}(i,j)}{N}\right)} \tag{5.15}$$

$$O = \sum_{\forall i,j}\frac{S_t(i,j)}{N} - W * \sum_{\forall i,j}\frac{S_{t-1}(i,j)}{N} \tag{5.16}$$

with $N$ the number of pixels. Then the sum of absolute differences (SAD) is estimated between $S_t$ and $S_{t-1}$ and between $S_t$ and $P_t$. If:

$$\frac{SAD(S_t,P_t)}{SAD(S_t,S_{t-1})} \leq 0.99 \tag{5.17}$$

the WP is applied in the encoding process. This method is simple and fast but provides an imprecise WP and thus an imprecise luminance compensation.

### 5.2.2 Proposed Weighted Prediction

For fair comparison with [OLVD16], we choose to compute the proposed WPs between $M(S_{t-1})$ and $S_t$ using the same optical-flow $M$ proposed by [CP11]. This optical-flow was successfully used in real time for 30 fps videos with 640x480 pixels, thus eligible for video compression. As in the HM16.2, the WP activation in the proposed encoding

process is determined by a threshold:

$$\frac{SAD(S_t, M(P_t))}{SAD(S_t, M(S_{t-1}))} \leq 0.95 \tag{5.18}$$

As opposed to (5.17), the SADs are computed between $S_t$ and the motion compensated frame $S_{t-1}$, thus providing much lower values. In practice, a threshold of $0.99$ was leading to an over-activation of the WP with (5.18). For these reasons, we decrease the threshold to $0.95$. The same activation threshold is used for all the following WP modes.

#### 5.2.2.1 Global and Local WP

Using the same WP for the entire frame results in a rather small rate cost since we send only few coefficients by frame. However, using a global WP suggests that the brightness variations between two frames is global. For many contents, this assumption is false hence the use of local WP. As explained in the introduction, the local WP can be implemented with different methods. In this case, we choose to send the local WP as metadata within each CTU (Coding tree unit/maximum block size) in HEVC. Some could argue about the overhead cost of local solutions. Indeed, as explained in the introduction, [YTB05] noticed that sending the WP parameters for each block in the AVC/H.264 standard results in a significant overhead cost and wrongly impacts the RD performances. However, the maximum block size in AVC/H.264 is rather small (16x16) when compared to HEVC (64x64). For a sequence with a 1920x1080 resolution, supposing that each block uses the WP, it results in sending 8160 WP for AVC/H.264 against 510 WP for HEVC. It can be concluded that the overhead cost is significantly reduced for the particular case of HEVC.

As previously mentioned, we implemented 4 WP modes: global linear (GL), global non-linear (GNL), local linear (LL), local non-linear (LNL). GL and LL WPs have already been proposed in the literature, however we implemented them to compare their RD performances to the ones obtained with the GNL and LNL WPs. Furthermore the GL WP allows us to assess the performance of the proposed WP algorithm based on an optical flow. Indeed, since the GL WP mode is the same as in HEVC, the RD performance gains of the GL WP will only be due to the luminance compensation estimation with proposed WP algorithm.

Global methods and local ones are computed using the same process, only the input pixels change. Indeed, for global methods, all matching pixels of $S_t$ and $M(S_{t-1})$ are considered, but for local methods, we consider the matching pixels in each matching 64x64 block of $S_t$ and $M(S_{t-1})$. The next sections detail the linear and non-linear WPs.

#### 5.2.2.2 Linear WP

For the linear WP modes, the prediction frame is computed using (5.14). All matching pixels of the input are plotted on a 2D graph, as illustrated in Figure 5.4, and the parameters $W$ and $O$ are obtained with linear least square fitting. In practice, we chose to constrain the minimum and maximum values of $W$ and $O$:

$$-128 \leq W \leq 127 \tag{5.19}$$

$$-256 \leq O \leq 255 \tag{5.20}$$

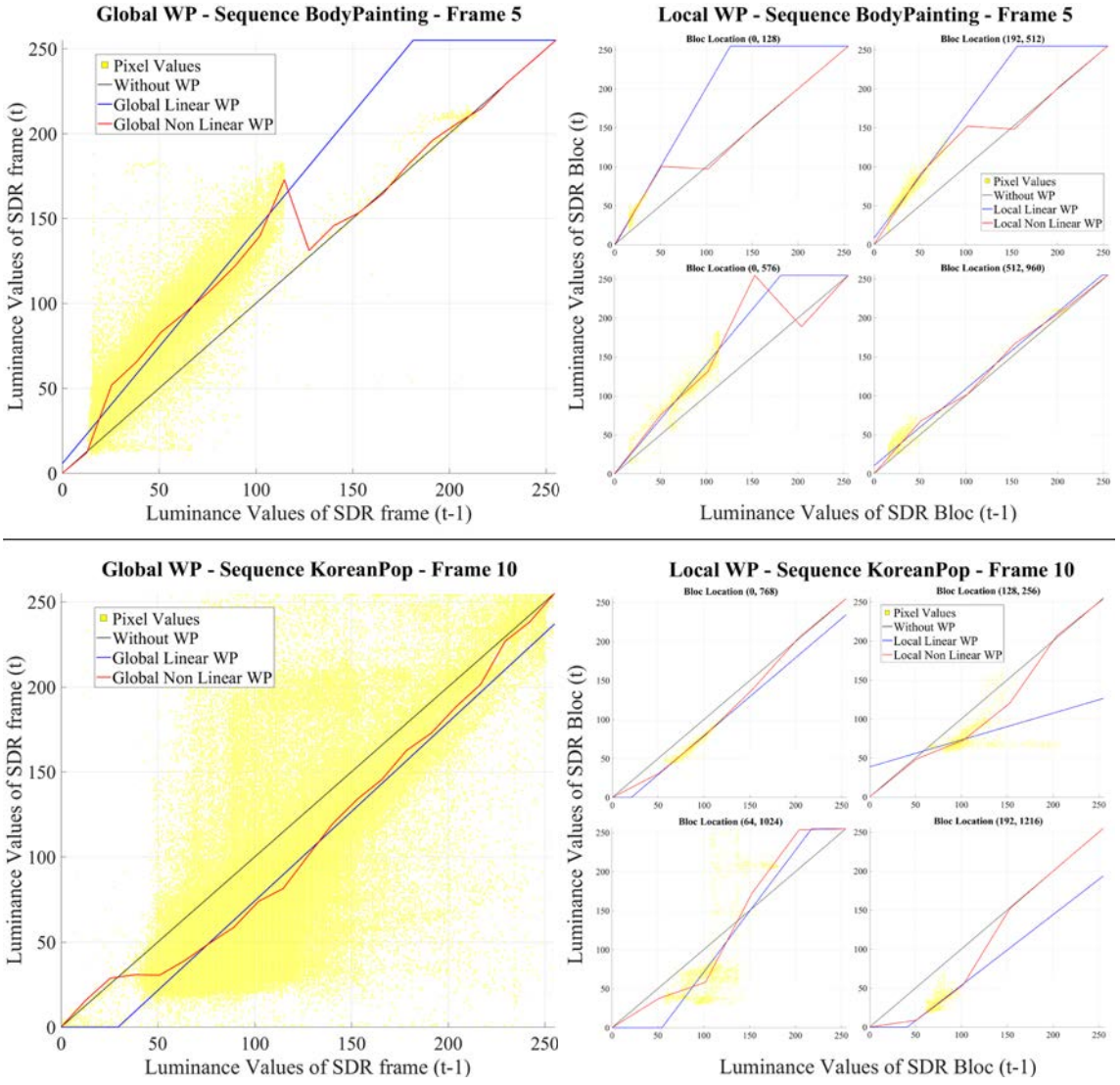



Figure 5.4 – Computed WP for two images (top: BodyPainting, bottom: KoreanPop). The left charts show the global linear and global non-linear WP (using 20 slopes) for the entire frame. The right charts show the local linear and local non-linear WP (using 5 slopes) for four different 64x64 blocks

Figure 5.4 also shows some examples of GL and LL WPs. Some samples of the corresponding SDR videos are visible in Figure 5.5. For the GL WP, the metadata for each frame only consists in the two coefficients $W$ and $O$. On the other hand, the metadata for the LL WP depend on the video size but also on the activation frequency of the WP. The coefficients $W$ and $O$ are sent for each CTU where the WP improves the RD performances.

#### 5.2.2.3 Non-linear WP

For the non-linear WP modes, as in [OLVD16], we chose to parameterize $WP$ as a piece wise linear function. Therefore, the prediction frame is computed using the follow-

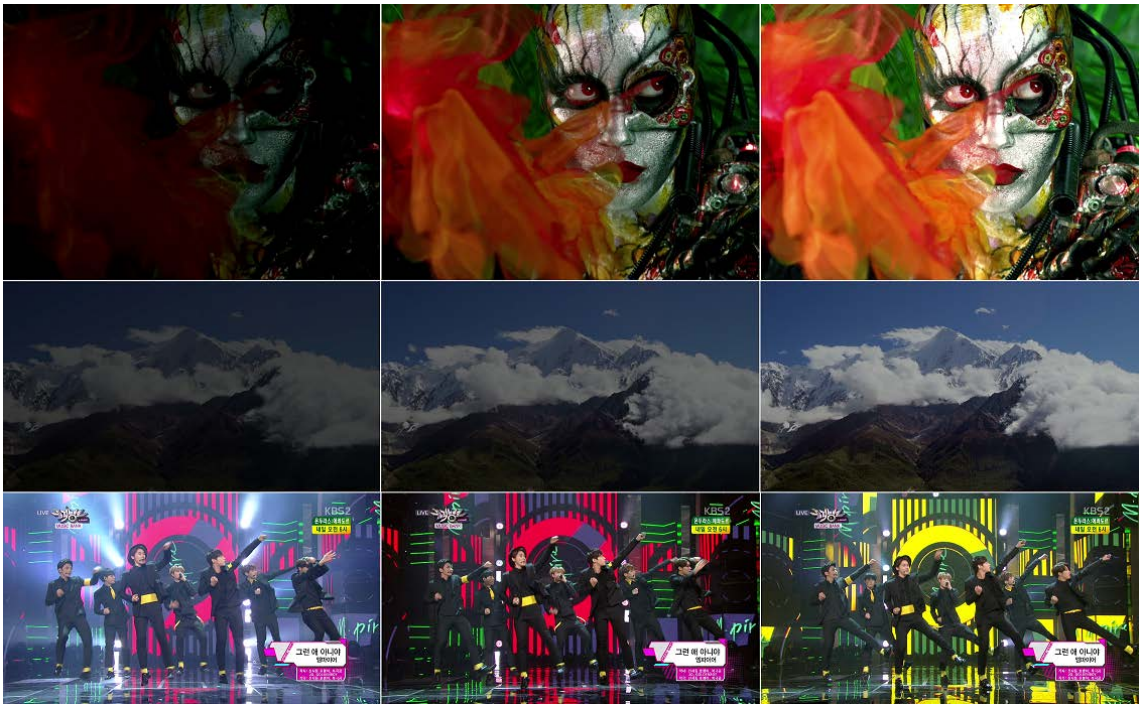


Figure 5.5 – Samples of tested SDR sequences. Each row represents three selected frames of each sequence. From top to bottom: BodyPainting Frame 4-6-8, Tangerine Frame 1-10-20, KoreanPop Frame 1-4-18

ing equation:
$\forall i,j \ \cap \ \forall S_{t-1}(i,j) \in [\delta.k, \delta(k+1)[$

$$P_t(i,j) = (S_{t-1}(i,j) - \delta.k) \cdot s_k + \delta \cdot \sum_{j=0}^{k-1} s_j \tag{5.21}$$

with $s_k$ the $k$-th slope in the piece wise linear function and $\delta$ the interval where each slope applies. In practice, $WP$ always spans all the possible values of $S_{t-1}$ (0 to 255). Therefore, $\delta = \frac{255}{n}$ with $n$ the chosen number of slopes.

As for the linear methods, all matching pixels of the input are plotted on a 2D graph and the slopes $s_k$ are obtained with piecewise linear least square fitting. An example of GNL and LNL WP is visible in Figure 5.4. The value of $n$ was fixed empirically for both methods, we chose $n = 20$ for the GNL WP and $n = 5$ for the LNL WP.

Knowing $n$, the $WP$ function is fully determined with the slopes $s_k$. These slopes are represented with real values and therefore need many bits for their representation. In practice, we chose to represent the $n$ slopes with $n + 1$ ordinates. Indeed, the ordinates can be easily rounded without much precision loss. Therefore, using the GNL WP, the metadata consists in 19 ordinates. The first and the last ones are always the same, 0 and 255, and are therefore ignored. As for the LL WP, the metadata of the LNL WP depend on the input size and efficiency of the WP. Each CTU where the WP improves the RD performances is sent with 4 ordinates.

## 5.3 HEVC Implementation

The WP algorithm and WP modes are computed once, as a first step, on the uncompressed video using a Matlab implementation. Then, in a second time, all these WP modes are parsed and used in the HEVC reference software (HM16.2) to encode the video. In this way, it allows us to test multiple strategies in HEVC without recomputing all the WP modes parameters and optical flows. This section details the specificity of the HEVC implementation, especially the chosen signaling for each mode.

### 5.3.1 Frame Level Activation

The first strategy used consists in choosing the WP mode at a frame level, as illustrated in Figure 5.3. Each frame meeting the constraint (5.18) is compressed twice, one time without WP, and a second time with the current WP (GL, GNL, LL, or LNL). The frame with the best RD performances is then encoded along with its metadata.

For the linear WPs, the coefficients $W$ and $O$ use a real representation requiring many bits, in practice, we choose to round up these values. Empirically, we found that the most efficient way was to round them after a multiplication by $2^5$. The coefficient $W$ then needs 13 bits and the coefficient $O$ 14 bits. For the non-linear WPs, each ordinate value requires 8 bits but we choose a more efficient way to compress these ordinates using the VLC (Variable Length entropy Coding) in the HM16.2. Instead, we keep the difference between each ordinate to which we subtract the values $\delta$ defined in (5.21). This process gives us values closer to $0$, therefore costing less than 8 bits per slopes.

For the global WPs and local WPs, we just signal the metadata for each frame where the WP applies. Regarding the local WPs, one could simply send the WP parameters for all the CTU in the frame, however we found that it was much efficient to signal a one bit flag for each CTU then sending the WP when it improves the CTU RD performances. Therefore, each CTU is compressed with and without the WP then encoded with the corresponding flag and potentially with the WP parameters.

As mentioned previously, the brightness between two frames varies greatly depending on the input, even in the same video. To further improve the RD performances, we also choose to put all these WP modes in competition since each of them outperforms the other ones for specific frames. To do so, each frame is compressed with each WP mode and the best one is encoded along with the frame. This competition mode needs an additional signaling at frame level, indeed an index is sent to signal the chosen WP mode. When using four competing WP modes the index uses two bits. We also put only two modes in competition, in this case the index uses only one bit.

The algorithm 2 details the HEVC signaling for all WPs and for the competition modes when we choose the WP at a frame level.

### 5.3.2 CTU Level Activation

For the frame level activation, only the local WPs were decided at a CTU level. In this case, this strategy is applied to all WP modes. If a frame meets the constraint (5.18) each CTU is compressed twice, one time without WP, and a second time with the current WP (GL, GNL, LL, or LNL). The frame with the best RD performances is then encoded along

**Algorithm 2** Frame Level Activation - HEVC Signaling at the encoder side

```
for all Frame do
  Signal if the WP is activated, using 1 bit
  if WP is activated then
    if Competition between modes is used then
      Signal the WP mode for the current frame
      1 bit if two competing modes, 2 bits if four competing modes
    end if
    if WP == GL then
      Write W using 13 bits and O using 14 bits
    else if WP == GNL then
      Write the 19 ordinates values using the VLC
    else
      for all CTU do
        Signal if the WP is used for the current CTU, using 1 bit
        if WP CTU turn on then
          if WP == LL then
            Write W using 13 bits and O using 14 bits
          else if WP == LNL then
            Write the 4 ordinates values using the VLC
          end if
        end if
      end for
    end if
  end if
end for
```

with its metadata. As for section 5.3.1, we put the different WP modes in competition using different combinations. Each competing modes is then used to compress the CTU and the best one in terms of RD performance is selected.

All WPs use the same representation as the one presented in section 5.3.1, however the signaling syntax changes in some cases, as visible in Algorithm 3. When evaluating only one WP mode, the process is almost similar. For the global WPs, we only add one flag by CTU to signal the activation of the WP, and for the local WPs only the frame decision changes. Indeed, in section 5.3.1 the decision was made after encoding the frame twice, in this case we only evaluate (5.18). The biggest difference concerns the competition of WP modes, indeed the chosen mode needs to be signaled for each CTU against only once with frame level competition. Since the signaling cost is larger, we evaluated several combinations of WP modes using two, three or the four of them. The detailed signaling for each case is explained in Algorithm 3.

Using 4 WP modes, two levels of activation and different modes of WP competition yield many test cases. The RD performances of all these different strategies are presented in the following section.

**Algorithm 3** CTU Level Activation - HEVC Signaling at the encoder side

```
for all Frame do
  Signal the result of (5.18) using 1 bit
  if (5.18) then
    if WP == GL then
      Write the GL WP, W using 13 bits and O using 14 bits
    end if
    if WP == GNL then
      Write the GNL WP, 19 ordinates values using the VLC
    end if
    for all CTU do
      if Competition between modes is used then
        Signal the WP mode for the current CTU
        if 1 mode, 1 bit to signal if the WP is used
        if 2 competing modes, 1 bit for no WP or 2 bits for other modes ( 0, 10, 11)
        if 3 competing modes, 2 bits by mode, including no WP (00, 01, 10, 11)
        if 4 competing modes, 1 bit for no WP or 3 bits for other ones (0, 100, 101,
        110, 111)
      end if
      if WP local turn on for the CTU then
        if WP == LL then
          Write W using 13 bits and O using 14 bits
        else if WP == LNL then
          Write the 19 ordinates values using the VLC
        end if
      end if
    end for
  end if
end for
```

## 5.4 Experimental Results

To compare the proposed WP methods with the HEVC WP, we used different test sets. The first one consists of 12 SDR video sequences of about 20 frames with different types of brightness variations (see figure 5.5). These contents include quite common brightness variations like fades, concert spot lights or TV shows and could benefit from a better WP. The second test set consists of 6 HDR videos tone-mapped using 6 different TMOs from the literature, therefore providing 36 SDR sequences. We point out that the selected HDR test set is less representative of the possible illumination changes you may encounter in TV broadcasting for instance, with some sequences presenting very few brightness variations. The motivation behind this test set selection is to see if some TMOs create brightness variations at the SDR level that can be compensated by one proposed WP modes thus providing better RD performances for these TMOs.

Only the luminance compression is considered in the following results, however it could be easily extended to the chrominance components by computing the WP in a similar manner. As said in the previous section, the WPs were implemented in the HM16.2

using the default lowdelay_P_Main configuration file [Sul13]. However, for sake of simplicity, we reduce the number of reference frames from 4 to 1. In this way, each P frame only use one reference frame. The encoding were performed using four quantization parameters (QP) (i.e. QP=22, 27, 32, and 37) and therefore the Bjontegaard rate (BD-rate) gains [Bjo01] were computed on the 4 resulting RD points. All the following BD-rate gains are computed using the HM16.2 without WP as a reference. The RD performances of the HEVC reference software WP are reported in each table for comparison purposes.

### 5.4.1 Rate-distortion performances using frame level activation

| Sequences | HEVC WP | GL | GNL | LL | LNL | 4 WP | GNL LL |
|---|---|---|---|---|---|---|---|
| **BodyPainting** | -9,15% | -10,26% | -11,13% | -10,65% | -9,99% | -12,76% | -13,22% |
| **Concert** | -2,35% | -2,38% | -2,44% | -0,59% | -1,48% | -2,96% | -2,49% |
| **Eurovision** | 0,01% | 0,01% | 0,01% | 0,01% | -0,15% | 0,01% | 0,01% |
| **KoreanPop** | 8,17% | -3,02% | -0,79% | -4,49% | -3,30% | -4,62% | -4,51% |
| **LeMatch** | 0,04% | 0,04% | 0,04% | 0,12% | -0,06% | -0,15% | 0,12% |
| **Guetta-0** | 3,65% | -0,30% | -0,27% | -0,26% | -0,06% | -0,52% | -0,44% |
| **Guetta-2** | -0,37% | -0,06% | -0,10% | -0,16% | 0,02% | -0,18% | -0,29% |
| **RedBull** | -37,08% | -36,20% | -40,91% | -4,03% | -6,86% | -40,65% | -40,70% |
| **PopIdol** | 0,61% | -1,34% | -0,54% | -2,07% | -1,35% | -2,07% | -2,23% |
| **SkyFade** | -12,60% | -13,63% | -13,66% | -0,06% | -2,26% | -13,81% | -12,88% |
| **SunFade** | -23,00% | -24,17% | -25,72% | 5,89% | 4,63% | -24,73% | -24,59% |
| **Tangerine** | -10,27% | -10,30% | -16,64% | 3,51% | 0,74% | -17,55% | -16,72% |
| **Average** | **-6,86%** | **-8,47%** | **-9,34%** | **-1,07%** | **-1,68%** | **-9,99%** | **-9,83%** |
| **Min** | **8,17%** | **0,04%** | **0,04%** | **5,89%** | **4,63%** | **0,01%** | **0,12%** |
| **Max** | **-37,08%** | **-36,20%** | **-40,91%** | **-10,65%** | **-9,99%** | **-40,65%** | **-40,70%** |

Table 5.1 – BD-Rate gains of the weighted prediction of HEVC (column 1) and of the proposed WP using Frame activation with respect to the HEVC standard without any weighted prediction. SDR test sequences

Table 5.1 shows the results using the frame level activation presented in section 5.3. The first column represents the BD-rate gains [1] of the HEVC reference software WP, therefore with a global linear function. The second one represents the BD-rate gains of the GL WP. In these two cases, the WP uses the same model, therefore, the gains are due to the proposed optical flow based algorithm for WP parameter estimation. For almost each sequence the proposed algorithm outperforms the existing one and on average it provides around 1.6% BD-rate gains. It can be noticed that the current HEVC WP is sometimes responsible of significant loss, especially for the sequences KoreanPop, PopIdol and Guetta0. Actually, these three sequences present mainly local variations which may explain why the HEVC WP fails. The first new WP model, GNL, further improves the RD performances, around 2.5% BD-rate gains. However, for the three same sequences with local variations, we observe some losses compared to the GL algorithm. Indeed, predict a local transformation with a global one can be tricky and it's more difficult to find

1. BD-rate gains correspond to the negative values

| Sequences | TMO | HEVC WP | GL | GNL | LL | LNL | 4 WP |
|---|---|---|---|---|---|---|---|
| Balloon | Mai | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| | Reinhard | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| | Ferwerda | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| | Mertens | 0.01 % | 0.02 % | 0.24 % | -0.04 % | -0.15 % | -0.12 % |
| | Raman | -5.95 % | -5.74 % | -7.88 % | -2.30 % | -3.97 % | -7.52 % |
| | Schlick | 0.05 % | 0.05 % | 0.05 % | 0.05 % | 0.05 % | 0.05 % |
| Carousel3 | Mai | -0.03 % | -0.13 % | -0.01 % | 0.25 % | -0.17 % | -0.13 % |
| | Reinhard | -0.32 % | -0.40 % | -0.09 % | -0.43 % | -0.49 % | -0.61 % |
| | Ferwerda | -0.07 % | -0.31 % | -0.03 % | -0.15 % | -0.40 % | -0.28 % |
| | Mertens | -0.07 % | 0.01 % | -0.01 % | -0.33 % | -0.23 % | -0.28 % |
| | Raman | -4.00 % | -3.58 % | -3.53 % | -2.11 % | -2.15 % | -3.52 % |
| | Schlick | -0.07 % | 0.02 % | 0.01 % | -0.37 % | -0.34 % | -0.43 % |
| Carousel4 | Mai | 0.00 % | -0.22 % | -0.14 % | -0.33 % | -0.53 % | -0.60 % |
| | Reinhard | -0.04 % | -0.47 % | -0.37 % | -0.53 % | -0.80 % | -0.86 % |
| | Ferwerda | -0.22 % | -0.79 % | -0.47 % | -1.06 % | -1.12 % | -1.28 % |
| | Mertens | 0.00 % | -0.02 % | 0.04 % | -0.46 % | -0.42 % | -0.56 % |
| | Raman | -3.00 % | -2.94 % | -3.26 % | -1.80 % | -2.09 % | -3.06 % |
| | Schlick | -0.06 % | -0.36 % | -0.28 % | -0.65 % | -0.99 % | -1.00 % |
| FireEater | Mai | -0.10 % | -0.12 % | -0.10 % | 0.06 % | 0.01 % | -0.37 % |
| | Reinhard | 0.06 % | 0.01 % | 0.14 % | -0.19 % | -0.24 % | -0.18 % |
| | Ferwerda | -0.67 % | -0.69 % | -0.97 % | -1.97 % | -0.64 % | -2.16 % |
| | Mertens | -0.01 % | 0.00 % | -0.01 % | -0.24 % | -0.11 % | -0.25 % |
| | Raman | -5.80 % | -5.34 % | -4.92 % | -2.82 % | -2.63 % | -5.28 % |
| | Schlick | -2.42 % | -2.53 % | -5.25 % | -3.15 % | -2.86 % | -5.41 % |
| Market | Mai | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| | Reinhard | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| | Ferwerda | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| | Mertens | -0.05 % | -0.04 % | 0.05 % | -0.27 % | -0.45 % | -0.42 % |
| | Raman | -30.22 % | -25.90 % | -32.84 % | -17.25 % | -21.83 % | -31.51 % |
| | Schlick | -4.84 % | -13.79 % | -31.02 % | -16.01 % | -27.68 % | -37.33 % |
| Tibul | Mai | -0.01 % | -1.62 % | -1.29 % | -1.58 % | -3.13 % | -3.33 % |
| | Reinhard | -0.15 % | -2.03 % | -2.53 % | -2.92 % | -4.50 % | -4.71 % |
| | Ferwerda | -0.38 % | -1.45 % | -2.67 % | -4.75 % | -5.20 % | -5.76 % |
| | Mertens | -0.25 % | -0.46 % | -0.31 % | -2.78 % | -2.88 % | -3.06 % |
| | Raman | -6.12 % | -6.12 % | -9.77 % | -6.03 % | -7.36 % | -9.17 % |
| | Schlick | -1.51 % | -2.28 % | -5.28 % | -5.67 % | -7.59 % | -7.60 % |
| **Average** | | **-1.85 %** | **-2.14 %** | **-3.12 %** | **-2.11 %** | **-2.80 %** | **-3.80 %** |
| **Min** | | **0.06 %** | **0.05 %** | **0.14 %** | **0.25 %** | **0.05 %** | **0.05 %** |
| **Max** | | **-30.22 %** | **-25.90 %** | **-32.84 %** | **-17.25 %** | **-27.68 %** | **-37.33 %** |

Table 5.2 – BD-Rate gains of the weighted prediction of HEVC (column 1) and of the proposed WP using Frame activation with respect to the HEVC standard without any weighted prediction. SDR tone-mapped test sequences

the optimal solution. Figure 5.4 perfectly illustrates this problem. For the KoreanPop sequence, we can see a large point cloud which is difficult to predict using a global WP.

However, in the local case, the point clouds are narrowed and the local WPs better fit them. These observations are confirmed by the RD results using local WPs for these three sequences. On the other hand, the average RD results are much worse than the ones using global WPs. For the other sequences with global brightness variations, the local WPs are not efficient since it requires a non negligible overhead cost for each CTU.

This first result analysis shows that the global and local WPs are complementary and not competing. Using all the WPs in the same encoder could only provide better RD performances, as visible in Table 5.1. The combination of the 4 WP modes provides around 3.13% BD-rate gains compared to the HEVC WP. We tested different combinations using one global and one local WP and the most efficient match is GNL-LL which provides around 2.97% BD-rate gains compared to the HEVC WP. Limiting the frame competition to 2 modes isn't more efficient since the overhead cost savings is negligible. Therefore, for the second test set, we only tested the combination of the 4 WP modes.

Table 5.2 shows the results using the second test set, the 36 SDR tone-mapped sequences. The TMOs mentionned in Table 5.2 are the following ones: Mai [MMM+11], Reinhard [RSSF02], Ferwerda [FPSG96], Mertens [MKR07], Raman [RC09] and Schlick [Sch95]. On average, the observations are the same than for Table 5.1. The non-linear WPs improve the RD performances and the combination of the 4 WP modes provides the best RD performances. Some HDR sequences, like Balloon or Market, present very few brightness variations and therefore presents no improvement in RD performances using the TMOs [MMM+11, RSSF02, FPSG96, MKR07]. However, for the TMOs [Sch95, RC09], these sequences present high RD gains. For the other sequences also, these TMOs are usually presenting the higher RD gains. Indeed, these TMOs can generate temporal flickering on the resulting SDR sequences therefore increasing the SDR temporal variations. The proposed WP modes can predict efficiently these flickering artifacts, therefore compensate it and improve the RD performances.

To conclude on the first results, we can notice that the proposed optical flow based algorithm is better for the estimation of WP model parameters than the actual one in the HM16.2, and that the proposed non-linear WPs outperforms the linear ones. We also demonstrated that global and local weighted prediction are complementary and their combination further improves the RD performances.

### 5.4.2 Rate-distortion performances using CTU level activation

Table 5.3 shows the results using the CTU level activation presented in section 5.3. On average, the observations for the frame level activation remains valid for this case. The principal difference with the previous results of Table 5.1 is that each mode needs an overhead cost for each CTU. Surprisingly, it does not impact the RD performances of global WPs. On the contrary, the GL and GNL WPs using CTU activation present 0.43% BD-rate gains and 0.37% BD-rate gains compared to the ones using frame activation. However, in this case, the overhead cost is relatively small (1 bit by CTU). As explained in section 5.3, the process for local WPs using CTU activation is almost similar as the ones using frame activation. In this case, we rely on the threshold (5.18) to decide the activation of the WP. The RD performances show that this threshold choice is not optimal and could be improved.

Using the combination of 4 WP at CTU level provides the same results as those ob-

| Sequences | HEVC WP | GL | GNL | LL | LNL | 4 WP | GL GNL LL |
|---|---|---|---|---|---|---|---|
| **BodyPainting** | -9,15% | -12,12% | -12,58% | -10,60% | -9,88% | -14,46% | -14,45% |
| **Concert** | -2,35% | -4,48% | -4,38% | -0,60% | -1,52% | -4,88% | -5,17% |
| **Eurovision** | 0,01% | 0,01% | 0,01% | 0,01% | 0,01% | 0,01% | 0,01% |
| **KoreanPop** | 8,17% | -5,54% | -4,54% | -4,49% | -3,30% | -7,35% | -7,28% |
| **LeMatch** | 0,04% | -0,14% | -0,15% | 0,12% | -0,16% | -0,24% | -0,27% |
| **Guetta-0** | 3,65% | -0,64% | -0,61% | -0,26% | -0,08% | -1,22% | -1,12% |
| **Guetta-2** | -0,37% | -0,42% | -0,60% | -0,16% | -0,04% | -0,95% | -0,78% |
| **RedBull** | -37,08% | -35,05% | -39,02% | -3,92% | -6,24% | -36,49% | -38,68% |
| **PopIdol** | 0,61% | -2,44% | -2,24% | -2,07% | -1,36% | -3,99% | -3,85% |
| **SkyFade** | -12,60% | -12,98% | -13,54% | 0,16% | -1,83% | -14,21% | -15,25% |
| **SunFade** | -23,00% | -21,43% | -22,95% | 6,13% | 4,51% | -21,31% | -21,80% |
| **Tangerine** | -10,27% | -11,55% | -15,93% | 4,01% | 2,03% | -14,50% | -14,95% |
| **Average** | **-6,86%** | **-8,90%** | **-9,71%** | **-0,97%** | **-1,49%** | **-9,97%** | **-10,30%** |
| **Min** | **8,17%** | **0,01%** | **0,01%** | **6,13%** | **4,51%** | **0,01%** | **0,01%** |
| **Max** | **-37,08%** | **-35,05%** | **-39,02%** | **-10,60%** | **-9,88%** | **-36,49%** | **-38,68%** |

Table 5.3 – BD-Rate gains of the weighted prediction of HEVC (column 1) and of the proposed WP using CTU activation with respect to the HEVC standard without any weighted prediction. SDR test sequences

tained at frame level. The better accuracy of this competition just counterbalance the overhead cost for each CTU. Also here, we tested several combinations of WP, as explained in section 5.3. Table 5.3 only presents the RD performance of the most efficient combination, which is GL-GNL-LL WPs, providing on average 3.44% BD-rate gains against the classical HEVC WP. In this case, limiting the competition to these three modes is more efficient than using the 4 WP model since we move from a cost of 3 bits by CTU to 2 bits by CTU.

Table 5.4 shows the results using the same test set as in Table 5.2. On average, the observations are the same than for Table 5.2 or 5.3, except that in this case the competition of four WP modes provides the best RD performances instead of GL-GNL-LL. The proposed WP modes outperforms the RD performances of the existing WP modes and the better RD gains are achieved with the TMOs that generate temporal flickering artifacts.

To conclude on these results, we showed that using a CTU competition for WPs modes can be more interesting than the frame level activation except for some marginal cases where the overhead cost limits the BD-rate gains.

### 5.4.3 HDR rate-distortion performances using best WP modes

The previous sections 5.4.1 and 5.4.2 present the RD gains on SDR sequences. In this section, we present the RD performances obtained on the HDR level using the TMO [MMM+11], which is easily invertible, and the best WP modes: the 4 WP modes with frame level activation, the 4WP modes and GL-GNL-LL modes with CTU level activation. The SDR RD performances of this configuration are already presented in Table 5.2 and 5.4. In this case, the SDR videos are inverse tone mapped to reconstruct the HDR sequences and the RD performances are computed using different HDR metrics: the tPSNR

| Sequences | TMO | HEVC WP | GL | GNL | LL | LNL | 4 WP | GL GNL LL |
|---|---|---|---|---|---|---|---|---|
| Balloon | Mai | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| | Reinhard | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| | Ferwerda | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| | Mertens | 0.01 % | -0.09 % | -0.14 % | -0.04 % | -0.12 % | -0.33 % | -0.24 % |
| | Raman | -5.95 % | -5.72 % | -7.59 % | -2.30 % | -4.01 % | -7.74 % | -7.73 % |
| | Schlick | 0.05 % | 0.05 % | 0.05 % | 0.05 % | 0.05 % | 0.05 % | 0.05 % |
| Carousel3 | Mai | -0.03 % | -0.58 % | -0.59 % | 0.15 % | -0.14 % | -1.04 % | -0.84 % |
| | Reinhard | -0.32 % | -0.82 % | -0.88 % | -0.43 % | -0.51 % | -1.67 % | -1.58 % |
| | Ferwerda | -0.07 % | -1.17 % | -0.99 % | -0.16 % | -0.36 % | -1.91 % | -1.55 % |
| | Mertens | -0.07 % | -0.37 % | -0.32 % | -0.33 % | -0.21 % | -0.77 % | -0.73 % |
| | Raman | -4.00 % | -4.06 % | -4.16 % | -2.08 % | -2.04 % | -5.29 % | -5.20 % |
| | Schlick | -0.07 % | -0.60 % | -0.52 % | -0.37 % | -0.36 % | -1.04 % | -1.03 % |
| Carousel4 | Mai | 0.00 % | -0.83 % | -0.76 % | -0.33 % | -0.53 % | -1.57 % | -1.38 % |
| | Reinhard | -0.04 % | -1.16 % | -1.20 % | -0.52 % | -0.79 % | -2.07 % | -1.88 % |
| | Ferwerda | -0.22 % | -1.63 % | -1.52 % | -1.06 % | -1.12 % | -2.80 % | -2.62 % |
| | Mertens | 0.00 % | -0.60 % | -0.57 % | -0.45 % | -0.42 % | -1.36 % | -1.18 % |
| | Raman | -3.00 % | -3.59 % | -4.01 % | -1.79 % | -2.07 % | -4.90 % | -4.76 % |
| | Schlick | -0.06 % | -1.13 % | -1.05 % | -0.65 % | -0.95 % | -2.18 % | -1.89 % |
| FireEater | Mai | -0.10 % | -0.87 % | -0.62 % | 0.06 % | 0.01 % | -1.62 % | -1.22 % |
| | Reinhard | 0.06 % | -0.14 % | -0.23 % | -0.19 % | -0.04 % | -1.25 % | -0.81 % |
| | Ferwerda | -0.67 % | -2.04 % | -2.03 % | -1.97 % | -1.51 % | -4.75 % | -3.85 % |
| | Mertens | -0.01 % | -0.28 % | -0.23 % | -0.24 % | -0.13 % | -0.66 % | -0.59 % |
| | Raman | -5.80 % | -5.90 % | -5.70 % | -2.82 % | -2.62 % | -7.52 % | -7.32 % |
| | Schlick | -2.42 % | -2.78 % | -5.71 % | -3.15 % | -2.86 % | -6.70 % | -6.39 % |
| Market | Mai | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| | Reinhard | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| | Ferwerda | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| | Mertens | -0.05 % | -0.34 % | -0.35 % | -0.27 % | -0.38 % | -0.95 % | -0.70 % |
| | Raman | -30.22 % | -26.35 % | -31.27 % | -17.26 % | -21.32 % | -31.60 % | -31.88 % |
| | Schlick | -4.84 % | -16.10 % | -38.67 % | -16.00 % | -27.92 % | -39.73 % | -39.48 % |
| Tibul | Mai | -0.01 % | -3.93 % | -3.19 % | -1.58 % | -3.11 % | -5.82 % | -5.29 % |
| | Reinhard | -0.15 % | -3.90 % | -4.31 % | -2.92 % | -4.46 % | -7.13 % | -6.49 % |
| | Ferwerda | -0.38 % | -3.27 % | -4.78 % | -4.75 % | -5.20 % | -8.73 % | -8.13 % |
| | Mertens | -0.25 % | -1.71 % | -1.69 % | -2.78 % | -2.88 % | -4.60 % | -4.03 % |
| | Raman | -6.12 % | -7.40 % | -11.04 % | -6.05 % | -7.04 % | -12.21 % | -12.09 % |
| | Schlick | -1.51 % | -3.76 % | -7.14 % | -5.67 % | -7.54 % | -9.94 % | -8.99 % |
| **Average** | | **-1.85 %** | **-2.81 %** | **-3.92 %** | **-2.11 %** | **-2.79 %** | **-4.94 %** | **-4.72 %** |
| **Min** | | **0.06 %** | **0.05 %** | **0.05 %** | **0.15 %** | **0.05 %** | **0.05 %** | **0.05 %** |
| **Max** | | **-30.22 %** | **-26.35 %** | **-38.67 %** | **-17.26 %** | **-27.92 %** | **-39.73 %** | **-39.48 %** |

Table 5.4 – BD-Rate gains of the weighted prediction of HEVC (column 1) and of the proposed WP using CTU activation with respect to the HEVC standard without any weighted prediction. SDR tone-mapped test sequences

[LFH15] in Table 5.5, the PU-PSNR [AMS08] in Table 5.6, the PU-SSIM [AMS08] in Table 5.7 and the HDRVDP [MKRH11] in Table 5.8.

| Sequences | TMO | HEVC WP | Frame / 4WP | CTU / 4WP | CTU / GL GNL LL |
|---|---|---|---|---|---|
| **Balloon** | **Mai** | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| **Carousel3** | **Mai** | -0.10 % | -0.14 % | -0.91 % | 0.39 % |
| **Carousel4** | **Mai** | 0.05 % | -0.63 % | -1.56 % | -1.34 % |
| **FireEater** | **Mai** | -0.15 % | -0.33 % | -1.94 % | -1.57 % |
| **Market** | **Mai** | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| **Tibul** | **Mai** | -0.05 % | -3.24 % | -5.83 % | -5.18 % |
| **Average** | | **-0.04 %** | **-0.72 %** | **-1.70 %** | **-1.41 %** |
| **Min** | | **0.05 %** | **0.01 %** | **0.01 %** | **0.01 %** |
| **Max** | | **-0.15 %** | **-3.24 %** | **-5.83 %** | **-5.18 %** |

Table 5.5 – HDR BD-Rate gains using tPSNR of the weighted prediction of HEVC (column 1) and of the proposed WP with respect to the HEVC standard without any weighted prediction. HDR test sequences

| Sequences | TMO | HEVC WP | Frame / 4WP | CTU / 4WP | CTU / GL GNL LL |
|---|---|---|---|---|---|
| **Balloon** | **Mai** | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| **Carousel3** | **Mai** | -0.08 % | -0.02 % | -0.87 % | -0.34 % |
| **Carousel4** | **Mai** | 0.07 % | -0.62 % | -1.53 % | -1.31 % |
| **FireEater** | **Mai** | -0.12 % | -0.34 % | -1.95 % | -1.63 % |
| **Market** | **Mai** | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| **Tibul** | **Mai** | -0.03 % | -3.25 % | -5.80 % | -5.15 % |
| **Average** | | **-0.02 %** | **-0.70 %** | **-1.69 %** | **-1.40 %** |
| **Min** | | **0.07 %** | **0.01 %** | **0.01 %** | **0.01 %** |
| **Max** | | **-0.12 %** | **-3.25 %** | **-5.80 %** | **-5.15 %** |

Table 5.6 – HDR BD-Rate gains using PU-PSNR of the weighted prediction of HEVC (column 1) and of the proposed WP with respect to the HEVC standard without any weighted prediction. HDR test sequences

| Sequences | TMO | HEVC WP | Frame / 4WP | CTU / 4WP | CTU / GL GNL LL |
|---|---|---|---|---|---|
| **Balloon** | **Mai** | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| **Carousel3** | **Mai** | 0.34 % | 0.22 % | -0.72 % | -0.08 % |
| **Carousel4** | **Mai** | 0.16 % | -0.39 % | -1.45 % | -1.12 % |
| **FireEater** | **Mai** | -0.13 % | -0.21 % | -1.55 % | -1.36 % |
| **Market** | **Mai** | 0.02 % | 0.02 % | 0.02 % | 0.02 % |
| **Tibul** | **Mai** | -0.03 % | -2.76 % | -5.28 % | -4.89 % |
| **Average** | | **0.06 %** | **-0.52 %** | **-1.50 %** | **-1.24 %** |
| **Min** | | **0.34 %** | **0.22 %** | **0.02 %** | **0.02 %** |
| **Max** | | **-0.13 %** | **-2.76 %** | **-5.28 %** | **-4.89 %** |

Table 5.7 – HDR BD-Rate gains using PU-SSIM of the weighted prediction of HEVC (column 1) and of the proposed WP with respect to the HEVC standard without any weighted prediction. HDR test sequences

On average, the RD gains are consistent with all metrics, a little lower for the HDRVDP. These RD performances are consistent with the TMO-generated SDR ones and demon-

| Sequences | TMO | HEVC WP | Frame / 4WP | CTU / 4WP | CTU / GL GNL LL |
|---|---|---|---|---|---|
| **Balloon** | **Mai** | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| **Carousel3** | **Mai** | 0.17 % | 0.45 % | 0.46 % | 1.12 % |
| **Carousel4** | **Mai** | 0.33 % | -0.78 % | -2.04 % | -1.56 % |
| **FireEater** | **Mai** | 0.47 % | -0.29 % | -1.29 % | -1.37 % |
| **Market** | **Mai** | 0.01 % | 0.01 % | 0.01 % | 0.01 % |
| **Tibul** | **Mai** | -0.02 % | -3.02 % | -5.21 % | -4.55 % |
| **Average** | | **0.16 %** | **-0.60 %** | **-1.34 %** | **-1.06 %** |
| **Min** | | **0.47 %** | **0.45 %** | **0.46 %** | **1.12 %** |
| **Max** | | **-0.02 %** | **-3.02 %** | **-5.21 %** | **-4.55 %** |

Table 5.8 – HDR BD-Rate gains using HDRVDP of the weighted prediction of HEVC (column 1) and of the proposed WP with respect to the HEVC standard without any weighted prediction. HDR test sequences

strate that improving the SDR RD performances can improve the HDR ones. Furthermore, the SDR content is not impacted and therefore the temporal quality is preserved.

### 5.4.4 Rate-distortion performances using multiple references

The previous result were presented using the default HEVC lowdelay_P_Main configuration file with only one reference frame. This configuration allows to more accurately assess the contribution of each WP models. Indeed, with multiple references frames, local blocks can be predicted using different reference frames, thus allowing implicit local brightness adaptation when using only global WP. Furthermore, with the use of B frame type, the weighted bi prediction also add another candidate for each local block, again implicitly accounting for local illumination change compensation.

To assess the performance of the proposed optical flow based algorithm and new WP models, we also perform tests using the default HEVC lowdelay_B_Main configuration file while maintaining the 4 reference frames. Table 5.9 shows the results using the frame level activation and Table 5.10, the results with CTU level activation. For these tests, we only focus on the SDR test set.

With the GL results in Table 5.9, we can see that the proposed optical flow based algorithm still outperforms the HEVC algorithm with a BD-rate improvement of 3%. However, unlike the configuration using one reference frame, the non linear WP models do not outperform the linear one. This observation is mainly explained with the overhead difference between the linear WP model and the non linear ones. In this case, the non linear WP overhead is multiplied by four since we use four reference frames. However, we also tested several competitions of WP model and using the GL and GNL models, we obtain a BD-rate improvement of 0.5% when compared to the GL model.

For the CTU level results in Table 5.10, we can noticed that the GNL WP model outperform the GL one with a BD-rate gain around 0.5%. However, for the local WP models, the LL WP model provides better results than the LNL WP model. As in Section 5.4.2, the 4 WP competition is penalized by a signaling cost of 3 bits by CTU while the competition GL, GNL and LL only uses 2 bits by CTU. For these reasons, only the competition of GL, GNL and LL provides BD-rate gains, around 0,8%, when compared

| Sequences | HEVC WP | GL | GNL | LL | LNL | 4 WP | GL GNL |
|---|---|---|---|---|---|---|---|
| **BodyPainting** | -13,58% | -14,37% | -13,60% | -12,14% | -10,97% | -14,87% | -15,15% |
| **Concert** | -18,28% | -18,29% | -17,55% | -5,04% | -3,89% | -18,30% | -18,35% |
| **Eurovision** | 0,02% | 0,01% | -0,13% | 0,10% | -0,40% | -0,68% | -0,14% |
| **KoreanPop** | 2,69% | -4,77% | -3,26% | -3,59% | -2,17% | -3,21% | -4,85% |
| **LeMatch** | 0,12% | -0,11% | -0,11% | 0,10% | 0,19% | 0,15% | -0,10% |
| **Guetta-0** | 2,95% | -0,67% | -0,63% | -0,79% | -0,16% | -0,70% | -0,94% |
| **Guetta-2** | -0,15% | -0,48% | -0,52% | -0,35% | -0,26% | -0,56% | -0,66% |
| **RedBull** | -42,81% | -50,55% | -50,45% | -2,28% | -0,31% | -50,49% | -51,09% |
| **PopIdol** | 0,46% | -2,36% | -2,21% | -2,08% | -1,27% | -2,24% | -2,92% |
| **SkyFade** | -21,30% | -23,37% | -23,13% | -0,34% | 1,09% | -23,66% | -24,34% |
| **SunFade** | -38,17% | -44,89% | -41,28% | 6,98% | 8,64% | -42,95% | -44,46% |
| **Tangerine** | -17,99% | -21,84% | -22,86% | 4,15% | 3,60% | -25,15% | -24,81% |
| **Average** | **-12,17%** | **-15,14%** | **-14,64%** | **-1,27%** | **-0,49%** | **-15,22%** | **-15,66%** |
| **Min** | **2.95%** | **0.01%** | **-0.11%** | **6.98%** | **8.64%** | **0.15%** | **-0.10 %** |
| **Max** | **-42.81%** | **-50.55%** | **-50.45%** | **-12.14%** | **-10.97%** | **-50.49%** | **-51.09%** |

Table 5.9 – BD-Rate gains of the weighted prediction of HEVC (column 1) and of the proposed WP using Frame activation with respect to the HEVC standard without any weighted prediction. SDR test sequences in a low delay configuration using 4B frames

| Sequences | HEVC WP | GL | GNL | LL | LNL | 4 WP | GL GNL LL |
|---|---|---|---|---|---|---|---|
| **BodyPainting** | -13,58% | -13,68% | -13,55% | -12,07% | -10,88% | -14,78% | -15,56% |
| **Concert** | -18,28% | -17,41% | -17,17% | -4,96% | -3,78% | -16,89% | -18,63% |
| **Eurovision** | 0,02% | -0,21% | -0,53% | 0,29% | -0,34% | -1,19% | -0,54% |
| **KoreanPop** | 2,69% | -6,47% | -6,37% | -3,59% | -2,17% | -7,13% | -8,37% |
| **LeMatch** | 0,12% | 0,54% | 0,81% | 0,53% | 0,25% | 0,27% | 1,15% |
| **Guetta-0** | 2,95% | -0,84% | -1,06% | -0,61% | -0,02% | -1,50% | -1,86% |
| **Guetta-2** | -0,15% | -0,67% | -1,20% | -0,24% | -0,05% | -1,13% | -1,33% |
| **RedBull** | -42,81% | -45,22% | -47,14% | -0,17% | -0,20% | -42,56% | -46,83% |
| **PopIdol** | 0,46% | -3,03% | -3,50% | -2,07% | -1,27% | -4,25% | -4,94% |
| **SkyFade** | -21,30% | -19,98% | -21,21% | 0,99% | 1,10% | -17,43% | -21,38% |
| **SunFade** | -38,17% | -34,94% | -33,75% | 9,28% | 9,55% | -26,91% | -31,08% |
| **Tangerine** | -17,99% | -17,62% | -19,81% | 5,60% | 5,19% | -17,19% | -19,13% |
| **Average** | **-12.17%** | **-13.29%** | **-13.71%** | **-0.59%** | **-0.22%** | **-12.56%** | **-14.04%** |
| **Min** | **2.95%** | **0,54%** | **0,81%** | **9.28%** | **9.55%** | **0.27%** | **1.15%** |
| **Max** | **-42.81%** | **-45.22%** | **-47.14%** | **-12.07%** | **-10.88%** | **-42.56%** | **-46.83%** |

Table 5.10 – BD-Rate gains of the weighted prediction of HEVC (column 1) and of the proposed WP using CTU activation with respect to the HEVC standard without any weighted prediction. SDR test sequences in a low delay configuration using 4B frames

to the GL WP model.

### 5.4.5 Localized rate-distortion performances

Evaluating the RD performances of the WP in a fair manner is not obvious. For some sequences, the gains can be localized on a specific frame and therefore minimized when averaging the gains over the entire sequence. To overcome this problem, we computed the BD-rate gains, using 4 RD points (4 QPs), for each frame and plotted them on Figure 5.6. The green curves represent the best competition mode using frame competition and the blue curves represent the best ones using the CTU competition. The abscissa represents the frame index in display order of the sequence and the ordinates, the associated BD-Rate gains. For the BodyPainting sequence, limiting the test sequence to the frame 4 to 9, the BD-rate gains would have been even better. Note that, each gain represented in the figure is associated with a rate value. If the rate value is high the gain has a big impact on the average gain and conversely if the rate value is low the gain has a small impact on the average gain.

On average, the proposed WPs always outperforms the HEVC weighted prediction except in a specific case at the beginning of the Tangerine sequence using the combination of GL-GNL-LL. In this case, the threshold (5.18) wrongly activates the WP for the current frame therefore imposing an overhead cost for all CTUs. Since we are at the beginning of a fade-in, the corresponding rate of the frame is very low therefore the overhead cost result in a large BD-rate loss, while in fact it's negligible compared to the other frame rates. This presentation for the BD-rate gains also allows to appreciate the performance of the WP. Indeed, for some frame, the BD-rate gains go up to 26% for BodyPainting, 14% for KoreanPop and 45% for Tangerine.

Figure 5.7 shows the BD-rate gains computed on the HDR sequences Tibul, FireEater and Caroussel4 using the HDRVDP metric. In this case, we show the results using the 4 WP modes with CTU competition since it provides the best RD performances in Tables 5.2 and 5.4. The HDR sequences contain around 200 frames, however for sake of clarity we only plotted the gains for a subset of frame with much temporal variations. As explain above, encoding all the frame of the sequence can round down the WP gains, thereby Figure 5.7 shows BD-Rate gains up to 14% for Tibul, 21% for FireEater and 7% for Carousel4

### 5.4.6 Discussions

We didn't compare the obtained RD performances in section 5.4.3 to the ones obtained with [OLVD16]. Indeed, we implemented their solution meticulously using the same configuration and same inputs but we were unable to achieve better RD performances than the ones obtained with [MMM+11]. The conclusion was the same using different metrics, different $\lambda$ values for (5.2), or different optimization algorithms. This observation coincides with the one made in the introduction stating that all others temporally constrained TMO failed to improve the HDR RD performances.

As explained in section 5.3 the optical flows and WPs are computed prior to the encoding process. The accuracy of WPs could possibly be improved if done directly in the encoding loop. Indeed, all computation could be made with the reconstructed frame $S_{t-1}$ instead of the source frame. Besides, regarding the optical flow, we tested a few others [LMB+15, QV16] but the preliminary results were better using the presented one. How-

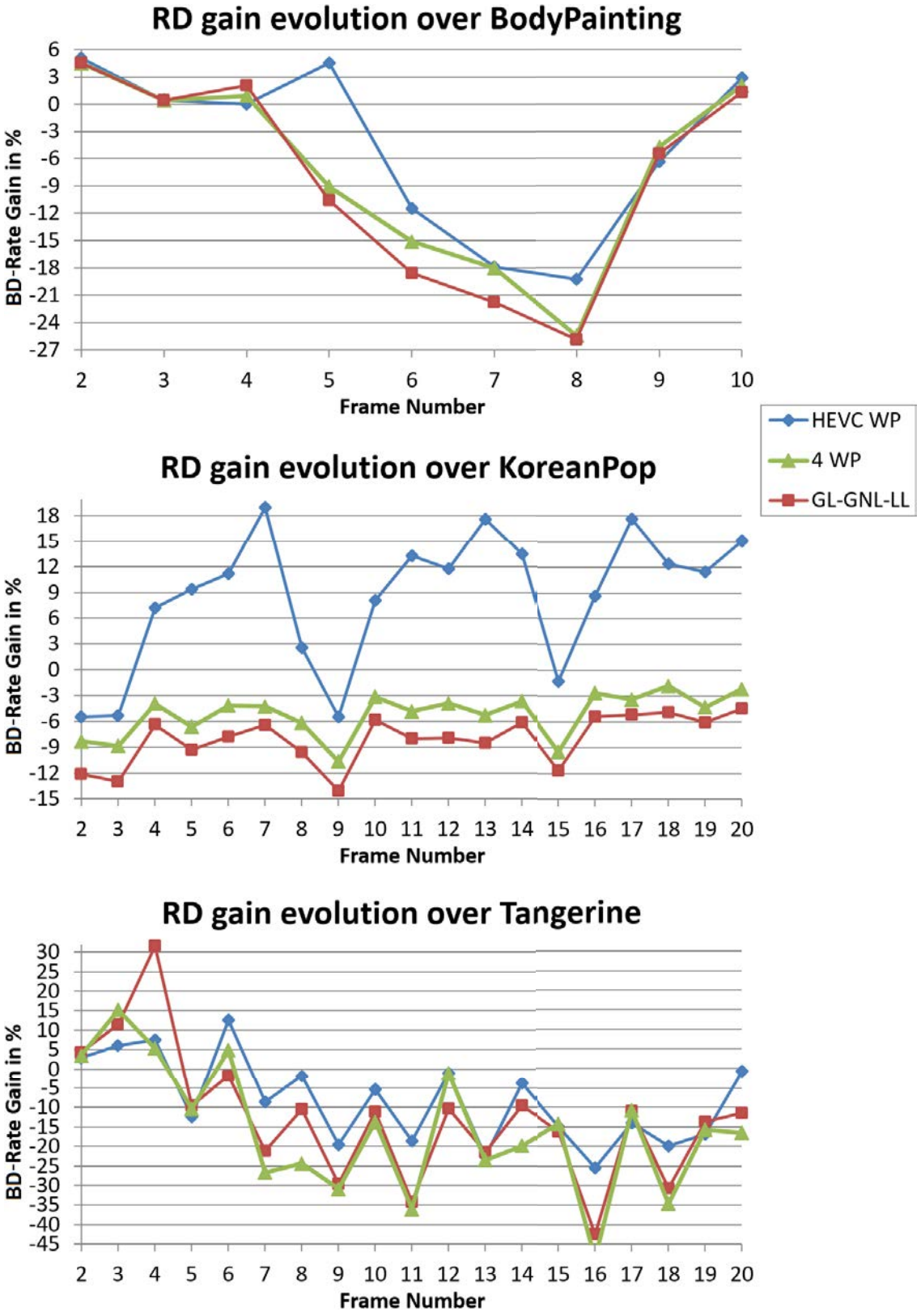



Figure 5.6 – Rate Gain by frame using SDR PSNR and SDR rate. From top to bottom: BodyPainting, Tangerine, KoreanPop

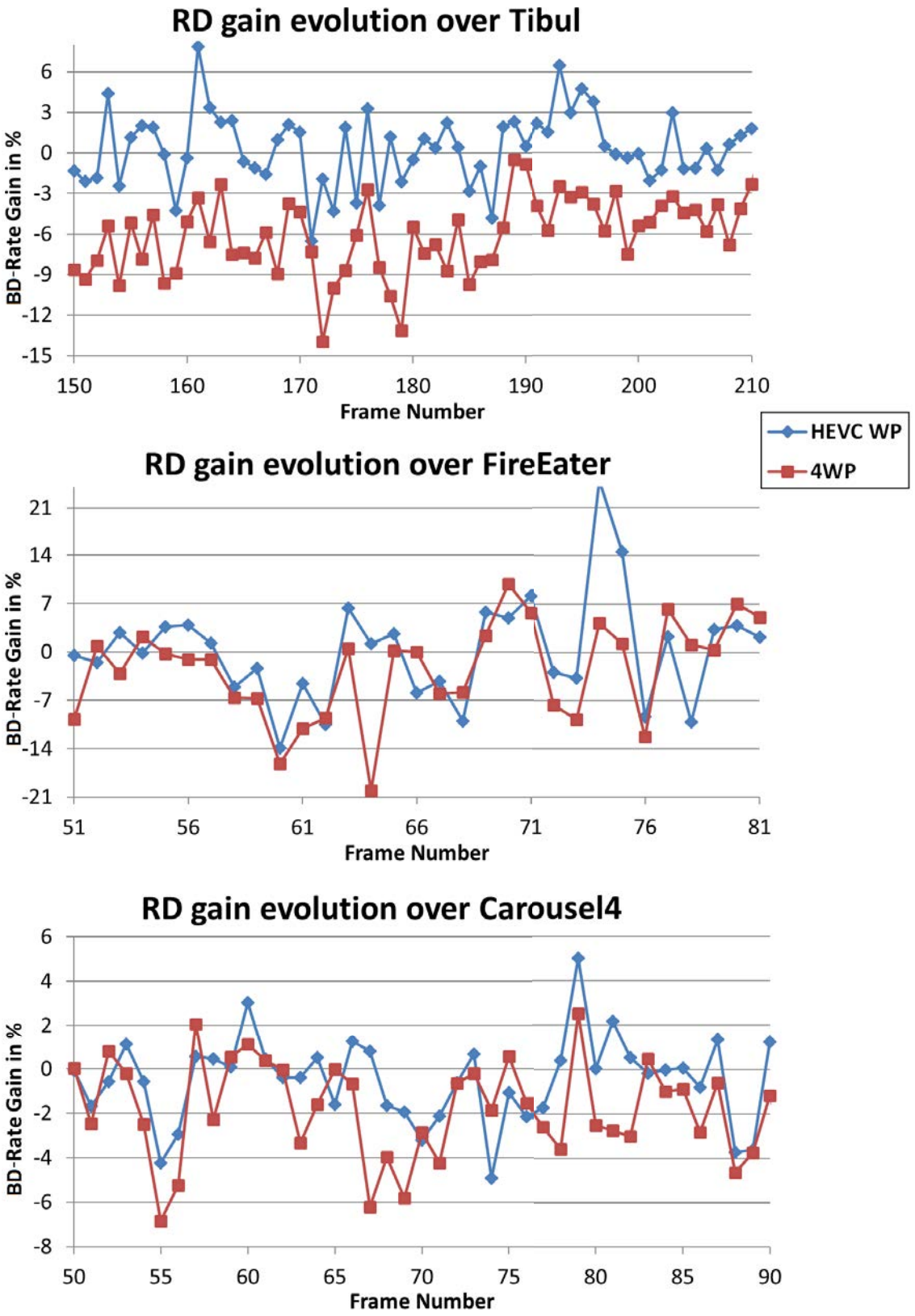



Figure 5.7 – Rate Gain by frame using HDRVDP and SDR rate. From top to bottom: Tibul, FireEater, Carousel4

ever, it's quite likely that more robust optical flows could improve the precision of the weighted prediction and therefore the RD performances.

Another possible improvement could be to reduce the rate by taking advantage of the HEVC CABAC context model [MSW03] to encode the different indexes signaled for each CTU. This solution both applies for the WP using CTU or frame level activation. Indeed, in the CTU level optimal solution (GL-GNL-LL), each 2 bits CTU index is hard-coded. The entropy of all these indexes in the frame could be reduced using the CTU context. In the frame level case, the GNL WP is signaled with 19 ordinates and there entropy is reduced by keeping the differences between each ordinates. However, using the CABAC context model, these ordinates could be more efficiently predict using the ones from previously encoded frames. This improvement should especially benefit to the multiple reference frames configurations, where 19 ordinates are sent for each reference frame.

Regarding the encoding performances, the algorithm and WPs computation is rather straightforward. The optical flow is computed and applied once on the entire frame and the WPs are quickly computed with least square fitting. As explained in section 5.2, the optical flow was performed in real time for 30 fps videos with 640x480 pixels [CP11]. However, performing multiple compression of the frame or the CTU increases the encoder complexity. Instead of multiple compression, some fast pre-decision between WP modes could be easily implemented thus resulting in a complexity equivalent to the original scheme and this probably with limited impact on the RD performances. However, the motivation here was to evaluate the best WP modes and not the efficiency of any fast pre-decision algorithms. Further works will focus on the conception of such algorithms. One option could be to analyze the variance of the point cloud presented in Figure 5.4.

Contrary to the encoder, the decoder complexity is not impacted. Indeed, the decoding process only consists in parsing the WP parameters for the current frame and applying it to the reference one, as in the HEVC reference software.

Finally, the proposed optical flow based WP algorithm can easily be used with the current HEVC standard but the new WP model parameters introduce decoder modifications and are not compatible with HEVC. These models could be suggested for the new standard in preparation VVC/H.266 [SBW17]. However, a new coding tool for luminance compensation has already been introduced in this standard under the name Local Illumination Compensation (LIC) [LCC+15]. Therefore, the contribution of the proposed WP algorithm and models should be compared with it.

## 5.5 Conclusion

In the context of the backward compatible HDR compression of videos, we demonstrated that using a weighted prediction is better than using a temporally optimized TMO in terms of RD performances. This is explained by the fact that, with the weighted prediction, the joint optimization problem of the TMO and the WP is separable into two problems: (i) the TMO optimization, and (ii) the WP optimization. Instead, the temporally optimized TMO requires to jointly optimize the two functions, which requires some simplifying assumptions to compromise between rate and distortion minimization, as well as temporal regularization. Another benefit of the WP approach is that the SDR can be op-

timized with respect to some other criterion, not only for compression but rather aesthetic purpose. Therefore, we proposed a new weighted prediction algorithm and new weighted prediction modes to handle more efficiently the large diversity of brightness variations in video sequences.

The proposed algorithm consists in performing a global motion compensation, based on an optical-flow, followed by the computation of different weighted predictions modes or models. The proposed modes consist in computing a global or local non-linear functions to predict the current frame, unlike state-of-the-art methods always based on linear functions. The performances of the proposed algorithm and each mode is evaluated to identify the best ones, then all modes are put in competition in the same encoding process. This competing strategy is made at two levels, frame and CTU, and different modes combinations are studied.

The several proposed weighted prediction modes range from a complex (many metadata) but accurate prediction model (local non-linear) to a less complex and rough prediction model (global linear). It has been shown that depending on the image, there exists always at least one proposed weighted predictor that outperforms the WP implemented in HEVC. Therefore, we propose to allow competition between all the proposed WP modes, and this brings BD-rate gain of about 10% with respect to the HEVC reference software and about 3.5% with respect to the HEVC software with existing WP enabled. Indeed, this competition allows to compensate more types of brightness variation. The RD results also show the superiority of non-linear functions, either in global or local weighted prediction, compared to linear ones despite a higher signaling overhead. The RD performances are improved on SDR graded content and also with TMO-generated SDR content, especially when the TMO introduces flickering artifacts. The RD performances were also improved on the HDR level and thus shows that the proposed solution can improve the HDR RD performances while maintaining the SDR temporal consistency.

In the proposed method, the compression is performed multiple times in some cases, therefore increasing the complexity at the encoder-side only. Further studies will investigate pre-decision algorithms to determine the weighted prediction activation and to allow for fast decision of the proposed WP modes. Additionally, the signaling of the WP will be implemented with the CABAC entropy coder to further reduce syntax overhead and improve the overall RD performances.

# Conclusion

In this thesis, we investigated the backward compatible compression methods for High Dynamic Range (HDR) content. As detailed in Sections 1.3 and 2.3, the standard HDR compression pipeline consists in using legacy SDR compression standards with new Opto-Electonic Transfer Functions (OETF) and a high bit depth representation instead of the classical SDR format using the gamma OETF and the 3x8 bits per pixel representation. While standard HDR compression is incompatible with legacy SDR decoder/display, the backward compatible HDR compression presented in Section 2.3 allows to address both the new HDR displays and the legacy SDR ones. To do so, the HDR content is first transformed into an 8 bit SDR one with a Tone Mapping Operator (TMO), introduced in Section 1.4, then encoded with a legacy SDR encoder such as HEVC, presented in Section 2.2. This SDR signal is thus compatible with an SDR display/decoder but it can also be inverse tone mapped to yield an HDR reconstructed content. Additionally, the residue between the original HDR content and its reconstructed version can be sent as an enhancement layer to improve the HDR content reconstruction. In this context, we proposed several solutions to improve the Rate Distortion performances (RD) of backward compatible compression schemes.

Initially, we focused on backward compatible compression schemes where both an SDR base layer and an enhancement layer are sent to reconstruct the HDR content. In Chapter 3, we proposed a TMO minimizing the distortion of the HDR signal under a total rate constraint for both layers, while preserving a good perceptual quality for the resulting SDR signal. Indeed, some TMOs only optimized for compression purposes may yield SDR content in which the artistic intent of the HDR content is not preserved. The proposed modelisation explicitly account for both layer rates and brings us to the early conclusion that optimizing a dual layer compression scheme can be split into two consecutive steps: a TMO optimization considering only the base layer rate and a rate allocation to find the best rate repartition between the base and enhancement layers. To solve the TMO optimization, the tone curve is parameterized as a piecewise linear function. Furthermore, this parameterization allows to send the TMO with few coefficients and therefore with a small impact on the base layer rate. In the end, in comparison to state-of-the-art methods, the proposed solution provides the best trade-off between RD performances and quality preservation of the SDR content.

Based on the previous conclusion that optimizing a TMO for a dual layer compression scheme is equivalent to, first, a TMO optimization with only the base layer, and secondly, a rate allocation optimization, then we focused on single layer compression schemes in Chapter 4. In this chapter, we proposed new models for the HDR distortion, the SDR rate and the SDR perceptual quality based on the spatial gradient of the HDR image. These models are compared to state-of-the-art ones and the superior accuracy of the gradient

based models is experimentally demonstrated. These improved models are then used to optimize a single layer compression scheme in two use cases. The first one consists to minimize the HDR distortion under a rate constraint for the SDR content, in this way we optimize the RD performances of the compression and ensure an efficient transmission of the HDR content. In the second use case, we kept the same minimization problem but we add a constraint on the SDR perceptual quality to tradeoff the RD performances with a visually pleasing SDR content. Experimental results demonstrate than both TMO provide either the optimal RD performances or the best tradeoff between RD performances and quality preservation of the SDR content.

Note that these two TMOs and the one proposed in Chapter 3 are computed with still images statistics. They can easily be used for video by computing a TMO for each frame, however one could argue about the optimality of such solution in terms of compression performances. Indeed, TMOs reduce a large dynamic of luminance to a smaller one and performing this reduction on consecutive images can create temporal artifacts such as flickering. Therefore, we chose to focus on this temporal aspect.

In Chapter 5, we studied the TMO specially designed for videos and distinguish two types. One type consists to reduce the temporal discontinuities in the SDR sequence, using techniques such as temporal filtering of the tone curves or more complex ones, to reduce the temporal artifacts and therefore provide a temporally coherent SDR content. However, these solutions have a rather limited and varying impact on the RD performances. In contrast, the other type of video TMO is designed to modify the SDR video in order to optimize the RD performances. These latter methods remove artifacts such as flickering, however, they strongly degrade the temporal coherency of the SDR video. In Chapter 5, we combine the benefits of the two methods by introducing new Weighted Prediction (WP) methods inside the HEVC SDR encoder. As a first step, we demonstrate the interest of the WP methods compared to TMO optimized for RD performances. Then we present the newly introduced WP algorithm and WP modes. The WP algorithm consists in performing a global motion compensation between frames using an optical flow, and the new modes are based on non linear functions in contrast with the literature using only linear functions. The contribution of each novelty is studied independently and in a second time they are all put in competition to maximize the RD performances. Tests were made in backward compatible compression scheme but also on SDR compression only. In both cases, the proposed WP methods improve the RD performances while maintaining the SDR temporal coherency.

To conclude this thesis, all the proposed solutions focused on different aspects of backward compatible HDR compression and successfully improved it. In Chapter 3, we tackle compression scheme using two layers and demonstrate that the TMO optimization only depends on the base layer. Then in Chapter 4, we proposed improved and well refined models of a single layer compression to obtain an optimal TMO providing either the best RD performances or the best tradeoff between RD performances and perceptual quality preservation of the SDR content. Finally, in Chapter 5, we demonstrate that the extension of TMO optimization to video can successfully be replaced by the use of a new weighted prediction algorithm for SDR encoder. However, all these works could be further improved as discussed below.

## Perspectives

First, all the previously presented work applies to the luma channel only, while the chroma channels are discarded. A simple approach is to use RGB color space with a TMO optimization for each channel however, as explained in Chapter 1, this method is not the optimal one for compression applications. The proposed TMOs are all optimized on the statistics of the luma component which are very different from the chroma components. For both compression performances and subjective quality of the SDR content, several adjustments are likely required. For Chapter 5, the benefits of a non linear WP still needs to be demonstrated for the chroma components. Therefore, further research is necessary to determine the best practices for the treatment of color.

Secondly, the quality constraint of the SDR content $D_{SDR\_REF}$ proposed in Chapter 3 and 4 is based on the MSE. It could be argued than using a metric based on a MSE with a reference SDR image is not an optimal indicator of the visual quality of the resulting SDR content. Other quality evaluation algorithms should be studied and integrate in the TMO optimization process. The optimal method would be to use subjective evaluation with several users, however this is a time-consuming process and clearly not integrable within an optimization scheme. Note that, this last observation also applies for the quality evaluation of the HDR content.

Thirdly, for sake of simplicity, the slopes of the non-linear WP in Chapter 5 are predicted by subtracting the same fixed value to all slopes and these residual slopes are then entropy encoded with a VLC. The encoding cost of the slopes could be reduced with more advanced prediction techniques using previously encoded slopes from previous frames or previous block. Furthermore, the entropy coding of the residual slopes could be improved with the CABAC entropy coder. Indeed, as explained in Chapter 2, the CABAC is known to be more efficient than the VLC in terms of compression performances.

Several other approaches are considered to extend the work of this thesis. We studied only global TMO in these thesis since it consists in applying the same invertible function to the entire image and therefore ensure a simple reconstruction of the HDR content at the decoding side. It could be considered to use different TMO curves depending on regions of the image. Since for each region the dynamic range would be reduced, the HDR reconstruction would be improved. This idea is introduced in [LKVD13a, LKVD13b] however with limited regions since they only use two, simply chosen with a luma threshold. The disadvantage of a luma threshold is the loss of spatial coherence in the SDR image since values just below the threshold is mapped to the maximum SDR value while values just above the threshold are mapped to the minimum SDR value. More advanced region selection can be made with segmentation algorithms, such as the SLIC superpixels [ASS+12], to improve the coherency of the SDR content. The main difficulty of these local approaches is to find a RD compromise for the number of selected regions. Indeed, these approaches allow a better reconstruction of the HDR content but each TMO needs to be sent along with the compressed SDR image and therefore, it can result in a significant rate increase.

Another approach concerns the use case where both SDR and HDR contents are available. The two contents are independently graded by an artist and sent as such. In this case, the TMO is a manual operation usually not invertible and the inverse transformation from the SDR to the HDR content is too complex/expensive to be signaled in the bitstream. In

this context, the best strategy is to use dedicated scalable codecs such as SHVC to predict the HDR content with the SDR one using inter layer prediction. This work has already been addressed in [GT12, PGT15].

Note that, the HDR distortion and SDR rate models developed in Chapter 4 could be used for SDR compression. Indeed, the gradient is already mainly used in rate control applications to estimate the rate fluctuation in a video. The proposed model is an extension of this estimator to the HDR image through the TMO. However, the HDR distortion model is a novelty and could easily be adapted to the SDR case. In this way, we have a fast and low-complexity estimator of both the expected average distortion for all the image but also its distribution on the image, as illustrated in Figure 4.7. This distortion estimator could be used in many other compression applications to preserve the quality of some regions of interest.

# Author's Publications

## Articles

### Conference Paper

David Gommelet, Aline Roumy, Christine Guillemot, Michael Ropert, and Julien Le Tanou "Rate-Distortion Optimization of a Tone Mapping with SDR Quality Constraint for Backward-Compatible High Dynamic Range Compression". In 2016 IEEE International Conference on Image Processing (ICIP), pages 1384-1388, September 2016.

### Journal Paper

David Gommelet, Aline Roumy, Christine Guillemot, Michael Ropert, and Julien Le Tanou "Gradient-Based Tone Mapping for Rate-Distortion Optimized Backward-Compatible High Dynamic Range Compression. In IEEE Transactions on Image Processing, Volume 26, Number 12, pages 5936-5949, December 2017.

# Appendix A

# Derivation of the optimization problem terms

## Derivation of Equation 3.38

First, let's express $\mathbb{E}\left[\left(F(X) - \mathbb{E}[F(X)]\right)^2\right]$ with the slope parameterization of $F$ detailed in Section 3.3.1:

$$\mathbb{E}\left[\left(F(X) - \mathbb{E}[F(X)]\right)^2\right] = \int_{x_{min}}^{x_{max}} \left(F(t) - \mathbb{E}[F(X)]\right)^2 .p(t).dt \tag{A.1}$$

$$\mathbb{E}\left[\left(F(X) - \mathbb{E}[F(X)]\right)^2\right] = \sum_{k=0}^{n-1} \int_{x_k}^{x_{k+1}} \left((t - x_k).s_k + \delta.\sum_{j=0}^{k-1}\left(s_j\right) - \mathbb{E}[F(X)]\right)^2 .p(t).dt \tag{A.2}$$

Using the same assumption than for (3.35), that for each slope:

$$\forall t \in [x_k,\ x_{k+1}[\ ,\ \mathbb{E}[t] \approx \frac{x_{k+1} + x_k}{2} \tag{A.3}$$

The equation (A.2) becomes:

$$\mathbb{E}\left[\left(F(X) - \mathbb{E}[F(X)]\right)^2\right] = \sum_{k=0}^{n-1} \int_{x_k}^{x_{k+1}} \left(\frac{\delta.s_k}{2} + \delta.\sum_{j=0}^{k-1}\left(s_j\right) - \mathbb{E}[F(X)]\right)^2 .p(t).dt \tag{A.4}$$

$$\mathbb{E}\left[\left(F(X) - \mathbb{E}[F(X)]\right)^2\right] = \sum_{k=0}^{n-1} \left(c(s_{0..k}) - \mathbb{E}[F(X)]\right)^2 . \int_{x_k}^{x_{k+1}} p(t).dt \tag{A.5}$$

$$\mathbb{E}\left[\left(F(X) - \mathbb{E}[F(X)]\right)^2\right] = \sum_{k=0}^{n-1} \left(c(s_{0..k}) - \mathbb{E}[F(X)]\right)^2 .p_k \tag{A.6}$$

where the function $c$ defined in (3.37) and where:

$$p_k = \mathbb{P}(x_k \leq t < x_{k+1}) = \int_{x_k}^{x_{k+1}} p(t).dt \tag{A.7}$$

Using the expression of $\mathbb{E}[F(X)]$ defined in (3.36), the equation (A.2) can be finally rewritten:

$$\mathbb{E}\left[\left(F(X)-\mathbb{E}[F(X)]\right)^2\right]=\sum_{k=0}^{n-1}\left(c(s_{0..k})-\sum_{k=0}^{n-1}c(s_{0..k}).p_k\right)^2 p_k \tag{A.8}$$

## Derivation of Equation 3.39

Now, we express $\mathbb{E}\left[F'(X)^{-2}\right]$ with the same slope parameterization (see Sec.3.3.1):

$$\mathbb{E}\left[F'(X)^{-2}\right]=\int_{x_{min}}^{x_{max}}\frac{p(t)}{F'(t)^2}.dt \tag{A.9}$$

$$\mathbb{E}\left[F'(X)^{-2}\right]=\sum_{k=0}^{n-1}\int_{x_k}^{x_{k+1}}\frac{p(t)}{s_k^2}.dt \tag{A.10}$$

$$\mathbb{E}\left[F'(X)^{-2}\right]=\sum_{k=0}^{n-1}\frac{\int_{x_k}^{x_{k+1}}p(t).dt}{s_k^2} \tag{A.11}$$

$$\mathbb{E}\left[F'(X)^{-2}\right]=\sum_{k=0}^{n-1}\frac{p_k}{s_k^2} \tag{A.12}$$

where $p_k$ is defined in (A.7).

## Derivation of Equation 3.40

Now, we express $\mathbb{E}\left[\left(G(X)-F(X)\right)^2\right]$ with the same slope parameterization (see Sec.3.3.1):

$$\mathbb{E}\left[\left(G(X)-F(X)\right)^2\right]=\int_{x_{min}}^{x_{max}}\left(G(t)-F(t)\right)^2.p(t).dt \tag{A.13}$$

$$\mathbb{E}\left[\left(G(X)-F(X)\right)^2\right]=\sum_{k=0}^{n-1}\int_{x_k}^{x_{k+1}}\left(G(t)-F(t)\right)^2.p(t).dt \tag{A.14}$$

with:

$$G(t)-F(t)=(t-x_k).r_k+\delta.\sum_{j=0}^{k-1}\left(r_j\right)-(t-x_k).s_k-\delta.\sum_{j=0}^{k-1}\left(s_j\right) \tag{A.15}$$

where $r_{0..(n-1)}$ are the slopes of the reference TMO $G$ and $s_{0..(n-1)}$ are the slopes of the reference TMO $F$.
Using the assumption (A.3):

$$G(t)-F(t)=\frac{\delta.r_k}{2}+\delta.\sum_{j=0}^{k-1}\left(r_j\right)-\frac{\delta.s_k}{2}-\delta.\sum_{j=0}^{k-1}\left(s_j\right) \tag{A.16}$$

$$G(t)-F(t)=c(r_{0..k})-c(s_{0..k}) \tag{A.17}$$

with the function $c$ defined in (3.37). Since $G(t) - F(t)$ is now independent from $t$, the equation (A.14) becomes:

$$\mathbb{E}\left[\big(G(X) - F(X)\big)^2\right] = \sum_{k=0}^{n-1} \Big(c(r_{0..k}) - c(s_{0..k})\Big)^2 . \int_{x_k}^{x_{k+1}} p(t).dt \tag{A.18}$$

$$\mathbb{E}\left[\big(G(X) - F(X)\big)^2\right] = \sum_{k=0}^{n-1} \Big(c(r_{0..k}) - c(s_{0..k})\Big)^2 . p_k \tag{A.19}$$

where $p_k$ is defined in (A.7).

# Appendix B

# Derivation of TMO Bounds

## Derivation of Equation 4.39

Removing the rate constraint from (4.35), the problem becomes:

$$\min_{s_1 \dots s_n} \quad c \cdot \sum_{k=0}^{n-1} \left( \frac{g_k(\gamma)}{s_k^{2-\gamma}} \right) + d \tag{B.1}$$

$$\text{s.t.} \quad \sum_{k=0}^{n-1} s_k = \frac{2^{nb-1}}{\delta}$$

Zeroing partial derivatives with respect to $s_j$ (Karush-Kuhn-Tucker necessary conditions) in (B.1) leads to:

$$\begin{cases} \dfrac{-(2-\gamma) g_j(\gamma).c}{s_j^{(3-\gamma)}} + \mu = 0 \quad \forall j \in [0, n] & \text{(B.2a)} \\ \displaystyle\sum_{k=0}^{n-1} \left( s_k \right) - \frac{2^{nb-1}}{\delta} = 0 & \text{(B.2b)} \end{cases}$$

As (B.2a) depends on $s_j$ only, we can exhibit the expression for each $s_k^\circ$:

$$s_k^\circ = \left( \frac{(2-\gamma).g_k(\gamma).c}{\mu} \right)^{\frac{1}{3-\gamma}} \tag{B.3}$$

To get $\mu$, (B.3) is used in the constraint (B.2b):

$$\sum_{k=0}^{n-1} \left( \frac{(2-\gamma).g_k(\gamma).c}{\mu} \right)^{\frac{1}{3-\gamma}} = \frac{2^{nb-1}}{\delta} \tag{B.4}$$

This leads:

$$\mu = \left( \sum_{k=0}^{n-1} \frac{\delta. \left( (2-\gamma).g_k(\gamma).c \right)^{\frac{1}{3-\gamma}}}{2^{nb-1}} \right)^{3-\gamma} \tag{B.5}$$

Finally (B.5) and (B.3) give:

$$s_k^\circ = \frac{2^{nb-1} \cdot g_k(\gamma)^{\frac{1}{3-\gamma}}}{\delta \cdot \sum_{k=0}^{n-1} g_k(\gamma)^{\frac{1}{3-\gamma}}} \tag{B.6}$$

## Derivation of Equation 4.40

Removing the spanning constraint from (4.35), turns the problem into:

$$\begin{aligned} \min_{s_1 \dots s_n} \quad & c \cdot \sum_{k=0}^{n-1} \left( \frac{g_k(\gamma)}{s_k^{2-\gamma}} \right) + d \\ \text{s.t.} \quad & \sum_{k=0}^{n-1} s_k \cdot g_k(1) = \frac{R_0 - b}{a} \end{aligned} \tag{B.7}$$

The Karush-Kuhn-Tucker necessary conditions for optimality on the Lagrangian expression of (B.7) lead to:

$$\begin{cases} \dfrac{-(2-\gamma) g_j(\gamma).c}{s_j^{(3-\gamma)}} + \lambda.g_j(1) = 0 \quad \forall j \in [0, n] & \text{(B.8a)} \\ \displaystyle\sum_{k=0}^{n-1} \left( s_k.g_k(1) \right) - \frac{R_0 - b}{a} = 0 & \text{(B.8b)} \end{cases}$$

Interestingly (B.8a) depends on one $s_j$ value only. Therefore, we can deduce the same expression for all $s_k^\dagger$:

$$s_k^\dagger = \left( \frac{(2-\gamma).g_k(\gamma).c}{\lambda.g_k(1)} \right)^{\frac{1}{3-\gamma}} \tag{B.9}$$

To get $\lambda$, we use (B.9) in the constraint (B.8b):

$$\sum_{k=0}^{n-1} \left( \frac{(2-\gamma).g_k(\gamma).c}{\mu + \lambda.g_k(1)} \right)^{\frac{1}{3-\gamma}} \cdot g_k(1) = \frac{R_0 - b}{a} \tag{B.10}$$

This leads:

$$\lambda = \left( \sum_{k=0}^{n-1} \frac{a.g_k(1).\left( (2-\gamma).g_k(\gamma).c \right)^{\frac{1}{3-\gamma}}}{(R_0 - b).\left( g_k(1)^{\frac{1}{3-\gamma}} \right)} \right)^{3-\gamma} \tag{B.11}$$

Finally, with (B.11), (B.9) becomes:

$$s_k^\dagger = \frac{(R_0 - b) \cdot g_k(\gamma)^{\frac{1}{3-\gamma}}}{a \cdot g_k(1)^{\frac{1}{3-\gamma}} \cdot \sum_{k=0}^{n-1} \left( g_k(1)^{\frac{2-\gamma}{3-\gamma}} \cdot g_k(\gamma)^{\frac{1}{3-\gamma}} \right)} \tag{B.12}$$

# Appendix C

# Derivation of SDR quality constraint

The derivation from (4.45) to (4.47), requires to express the $\Big(G(x)-F(x)\Big)^2$ term for $x \in [x_k, x_{k+1}[$. By using numerical integration with rectangle rule, we have the following approximation:

$$\int_{x=x_k}^{x_{k+1}} \Big(G(x)-F(x)\Big)^2 dx \approx \frac{x_{k+1}-x_k}{\delta} \cdot \sum_{x=x_k}^{x_{k+1}} \Big(G(x)-F(x)\Big)^2 \tag{C.1}$$

where $\frac{x_{k+1}-x_k}{\delta} = 1$. Integration by parts leads:

$$\int_{x=x_k}^{x_{k+1}} \Big(G(x)-F(x)\Big)^2 dx = \left[\frac{(G(x)-F(x))^3}{3\cdot(G'(x)-F'(x))}\right]_{x_k}^{x_{k+1}} - \int_{x=x_k}^{x_{k+1}} \frac{\Big(G(x)-F(x)\Big)^3.3.\Big(G''(x)-F''(x)\Big)}{9.\Big(G'(x)-F'(x)\Big)^2} \tag{C.2}$$

Since $\Big(G''(x)-F''(x)\Big) = 0$, then (C.2) becomes:

$$\int_{x=x_k}^{x_{k+1}} \Big(G(x)-F(x)\Big)^2 dx = \left[\frac{(G(x)-F(x))^3}{3\cdot(G'(x)-F'(x))}\right]_{x_k}^{x_{k+1}} \tag{C.3}$$

From (C.1) and (C.3), we can deduce that:

$$\sum_{x=x_k}^{x_{k+1}} \Big(G(x)-F(x)\Big)^2 = \left[\frac{(G(x)-F(x))^3}{3\cdot(G'(x)-F'(x))}\right]_{x_k}^{x_{k+1}} \tag{C.4}$$

This can be rewritten with the piecewise parameterization for $F$ and $G$ given in (3.30) and (4.46):

$$\sum_{x=x_k}^{x_{k+1}} \Big(G(x)-F(x)\Big)^2 = \frac{\Big(\delta.t_k + z_k - \delta.s_k - y_k\Big)^3 - \Big(z_k - y_k\Big)^3}{3\cdot(t_k - s_k)} \tag{C.5}$$

with $z_k = \delta \sum_{j=0}^{k-1} t_j$ and $y_k = \delta \sum_{j=0}^{k-1} s_j$. And finally, from (4.45) and (C.5), we get:

$$D_{SDRREF} = \sum_{k=0}^{n-1} p_k \cdot h\Big(t_0 ... t_k,\ s_0 ... s_k, \delta\Big) \tag{C.6}$$

with:

$$h\Big(t_0 .. t_k,\ s_0 .. s_k, \delta\Big) = \frac{\Big(\delta . t_k + z_k - \delta . s_k - y_k\Big)^3 - \Big(z_k - y_k\Big)^3}{\delta \cdot 3 \cdot (t_k - s_k)} \tag{C.7}$$

# List of abbreviations

| | |
|---|---|
| **AMVP** | Advanced Motion Vector Prediction |
| **AVC/H.264** | Advanced Video Coding |
| **BD** | BjontegaarD |
| **bpp** | bits per pixel |
| **CABAC** | Context Adaptive Binary Arithmetic Coding |
| **CfE** | Call for Evidence (refer to the HDR CfE [LFH15]) |
| **CIE** | Commission Internationale de l'Eclairage (International Commission of Lighting) |
| **CGI** | Computer Generated Imagery |
| **CRF** | Camera Response Function |
| **CSF** | Contrast Sensitivity Function |
| **CTU** | Coding Tree Unit (in HEVC) |
| **CU** | Coding Unit (in HEVC) |
| **DATM** | Display Adaptive Tone Mapping [MDK08] |
| **DBF** | DeBlocking Filter |
| **DCT** | Discrete Cosine Transform |
| **DLP** | Digital Light Processing |
| **DST** | Discrete Sine Transform |
| **EO** | Expansion Operator |
| **EOTF** | Electro Optical Transfer Function |
| **EV** | Exposure Values |
| **FPS** | Frames Per Second |
| **GL** | Global Linear (refer to the Weighted Prediction in Chapter.5) |
| **GNL** | Global Non-Linear (refer to the Weighted Prediction in Chapter.5) |
| **GOP** | Group Of Pictures |
| **JND** | Just Noticeable Difference |
| **HD** | High Definition - Resolutions between 1280x720 and 1920x1080 |
| **HDR** | High Dynamic Range |
| **HEVC/H.265** | High Efficiency Video Coding |
| **HLG** | Hybrid Log Gamma transfer function |
| **HVS** | Human Visual System |
| **IR** | Infra Red |
| **ITMO** | Inverse Tone Mapping Operator |
| **IEC** | International Electrotechnical Commission |
| **ISO** | International Organization for Standardization |
| **ITU-T** | International Telecommunication Union |
| **LCU** | Largest Coding Unit (or CTU) |
| **LDR** | Low Dynamic Range (Same as SDR) |

| | |
|---|---|
| **LED** | Light Emitting Diode |
| **LL** | Local Linear (refer to the Weighted Prediction in Chapter.5) |
| **LNL** | Local Non-Linear (refer to the Weighted Prediction in Chapter.5) |
| **LUT** | Look Up Table |
| **MOS** | Mean Opinion Score |
| **MPEG** | Moving Picture Experts Group |
| **MSE** | Mean Square Error |
| **OETF** | Opto Electronic Transfer Function |
| **OLED** | Organic Light Emitting Diode |
| **POC** | Picture Order Count |
| **PQ** | Perceptual Quantizer (usually refer to [MND12]) |
| **PSNR** | Peak to Signal Noise Ratio |
| **PTR** | Photographic Tone Reproductor [RSSF02] |
| **PU** | Prediction Unit (in HEVC) |
| **PU** | Perceptually uniform (refer to the metrics like PU-PSNR or PU-SSIM) |
| **QP** | Quantization Parameter |
| **RD** | Rate-Distortion |
| **SAD** | Sum of Absolute Differences |
| **SAO** | Sample Adaptive Offset |
| **SD** | Standard Definition - Resolutions below than 1280x720 |
| **SDR** | Standard Dynamic Range (Same as LDR) |
| **SSIM** | Structural SIMilarity |
| **TMO** | Tone Mapping Operator |
| **TU** | Transform Unit (in HEVC) |
| **TV** | TeleVision |
| **TVI** | Threshold Versus Intensity |
| **UHD** | Ultra High Definition - Resolutions higher than 1920x1080 |
| **UV** | Ultra Violet |
| **VLC** | Variable Length Coding |
| **VVC/H.266** | Future Video Coding |
| **WCG** | Wide Color Gamut |
| **WP** | Weighted Prediction |

# List of Figures

# Bibliography


[Ado92] Adobe. TIFF 6.0 Specification. [Online]. Available: partners.adobe.com/public/developer/tiff/index.html, 1992.

[AFR+07] A.O. Akyuz, R. Fleming, B.E. Riecke, E. Reinhard, and H.H. Bulthoff. Do HDR Displays Support LDR Content ? A Psychophysical Evaluation. *ACM Trans. Graph.*, 26(3), July 2007.

[All17] Ultra HD Alliance. UHDA Specifications. [Online]. Available: alliance.experienceuhd.com, 2017.

[AMHH16] T. Akenine-Moller, E. Haines, and N. Hoffman. *Real-Time Rendering, Third Edition*. CRC Press, 2016.

[AMS08] T.O. Aydin, R. Mantiuk, and H.P. Seidel. Extending Quality Metrics to Full Dynamic Range Images. In *Human Vision and Electronic Imaging XIII*, Proceedings of SPIE, pages 6806–6810, San Jose, USA, January 2008.

[ASC+14] T.O. Aydin, N. Stefanoski, S. Croci, M. Gross, and A. Smolic. Temporally Coherent Local Tone Mapping of HDR Video. *ACM Trans. Graph.*, 33(6), November 2014.

[Ash02] M. Ashikhmin. A Tone Mapping Algorithm for High Contrast Images. In *Proceedings of the 13th Eurographics Workshop on Rendering*, EGRW, pages 145–156, Aire-la-Ville, Switzerland, 2002. Eurographics Association.

[ASS+12] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, Nov 2012.

[BADC17] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers. *Advanced High Dynamic Range Imaging, Second Edition*. CRC Press, 2017.

[Bar99] P.G. Barten. *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*. Press Monograph Series. SPIE Optical Engineering Press, 1999.

[BB71] O.M. Blackwell and H.R. Blackwell. Visual Performance Data for 156 Normal Observers of Various Ages. *Journal of the Illuminating Engineering Society*, 1(1):3–13, 1971.

[BC15a] T. Borer and A. Cotton. A Display Independent High Dynamic Range Television System - White Paper 309. Technical report, BBC British Broadcasting Corporation, 2015.

[BC15b] T. Borer and A. Cotton. Essential Parameter Values for the Extended Image Dynamic Range Television System for Programme Production - ARIB STD-B67. Technical report, ARIB Standard - Association of Radio Industries and Businesses, 2015.

[BCTB14a] R. Boitard, R. Cozot, D. Thoreau, and K. Bouatouch. Survey of Temporal Brightness Artifacts in Video Tone Mapping. In *HDRi2014 - Second International Conference and SME Workshop on HDR imaging*, Sarajevo, Bosnia and Herzegovina, 2014.

[BCTB14b] R. Boitard, R. Cozot, D. Thoreau, and K. Bouatouch. Zonal Brightness Coherency for Video Tone Mapping. *Image Communications*, 29(2):229–246, February 2014.

[BDLC08] F. Banterle, K. Debattista, P. Ledda, and A. Chalmers. A GPU-Friendly Method for High Dynamic Range Texture Compression Using Inverse Tone Mapping. In *Proceedings of Graphics Interface*, pages 41–48. Canadian Information Processing Society, 2008.

[Bis17] C. Bist. *Combining Aesthetics and Perception for Display Retargeting*. PhD thesis, 2017.

[Bjo01] G. Bjontegaard. Calculation of Average PSNR Differences Between RD-Curves. 01 2001.

[BKH03] R. Bogart, F. Kainz, and D. Hess. OpenEXR Image File Format. *ACM Siggraph, Sketches & Applications*, 2003.

[BLDC10] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers. Expanding Low Dynamic Range Videos for High Dynamic Range Applications. In *Proceedings of the 24th Spring Conference on Computer Graphics*, pages 33–41, New York, NY, USA, 2010. ACM.

[Bod73] H.W. Bodmann. Visibility Assessment in Lighting Engineering. *Journal of the Illuminating Engineering Society*, 2(4):437–444, 1973.

[Boi14] R. Boitard. *Temporal Coherency in Video Tone Mapping*. PhD thesis, 2014.

[BPNS15] R. Boitard, M. T. Pourazad, P. Nasiopoulos, and J. Slevinsky. Demystifying High-Dynamic-Range Technology: A New Evolution in Digital Media. *IEEE Consumer Electronics Magazine*, 4(4):72–86, Oct 2015.

[BTCB13] R. Boitard, D. Thoreau, R. Cozot, and K. Bouatouch. Impact of Temporal Coherence-Based Tone Mapping on Video Compression. In *21st European Signal Processing Conference (EUSIPCO)*, pages 1–5, Sept 2013.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 1st edition, 2004.

[BV15] Philips International BV. Philips HDR technology - White Paper. Technical report, 2015.

[CBB+09] A. Chalmers, G. Bonnet, F. Banterle, P. Dubla, K. Debattista, A. Artusi, and C. Moir. High Dynamic Range Video Solution. In *ACM SIGGRAPH ASIA Art Gallery & Emerging Technologies: Adaptation*, pages 71–71, New York, NY, USA, 2009. ACM.

[CMZ09] K. Chang, A. Men, and W. Zhang. Adaptive Intra-frame Rate Control for H.264/AVC. In *Third International Conference on Multimedia and Ubiquitous Engineering*, pages 54–58, June 2009.

[CP11] A. Chambolle and T. Pock. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, May 2011.

[CSE00] C. Christopoulos, A. Skodras, and T. Ebrahimi. The JPEG2000 Still Image Coding System: an Overview. *IEEE Transactions on Consumer Electronics*, 46(4):1103–1127, November 2000.

[CT06] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley-Interscience, Hoboken, New Jersey, 2nd edition, 2006.

[CV16] M. Clare and J. Vieron. Why and How to Go to Ultra HD phase 2. Technical report, 4Ever-2 project, 2016.

[CW12] M.F. Cohen and J.R. Wallace. *Radiosity and Realistic Image Synthesis*. The Morgan Kaufmann Series in Computer Graphics. Elsevier Science, 2012.

[Dal93] S. Daly. Digital Images and Human Vision. chapter The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity, pages 179–206. MIT Press, Cambridge, MA, USA, 1993.

[DD02] F. Durand and J. Dorsey. Fast Bilateral Filtering for the Display of High Dynamic Range Images. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH, pages 257–266, New York, NY, USA, 2002. ACM.

[DM97] P.E. Debevec and J. Malik. Recovering High Dynamic Range Radiance Maps from Photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH, pages 369–378, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[DMAC03] F. Drago, K. Myszkowski, T. Annen, and N. Chiba. Adaptive Logarithmic Mapping For Displaying High Contrast Scenes. *Computer Graphics Forum*, 22(3):419–426, 2003.

[DMHS08] P. Didyk, R. Mantiuk, M. Hein, and H.P. Seidel. Enhancement of Bright Video Features for HDR Displays. In *Proceedings of the Nineteenth Eurographics Conference on Rendering*, pages 1265–1274. Eurographics Association, 2008.

[EF98] F. Ebner and M.D. Fairchild. Development and Testing of a Color Space (IPT) with Improved Hue Uniformity. pages 8–13, 01 1998.

[EFP+16] T. Ebrahimi, K. Fliegel, A. Pinheiro, M. Rerabek, T. Richter, and T. Skodras. Overview and Benchmarking Summary for the ICIP 2016 Compression Challenge. In *IEEE International Conference on Image Processing (ICIP)*, Sept 2016.

[EUWM13] G. Eilertsen, J. Unger, R. Wanat, and R. Mantiuk. Survey and Evaluation of Tone Mapping Operators for HDR Video. In *ACM SIGGRAPH Talks*, pages 1–1, New York, NY, USA, 2013. ACM.

[FAA+12] C.M. Fu, E. Alshina, A. Alshin, Y.W. Huang, C.Y. Chen, C.Y. Tsai, C.W. Hsu, S.M. Lei, J.H. Park, and W.J. Han. Sample Adaptive Offset in the HEVC Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1755–1764, Dec 2012.

[Fai05] M.D. Fairchild. *Color Appearance Models*. The Wiley-IS&T Series in Imaging Science and Technology, 2nd edition, 2005.

[Far01] H. Farid. Blind Inverse Gamma Correction. *IEEE Transactions on Image Processing*, 10(10):1428–1433, Oct 2001.

[Fec60] G.T. Fechner. *Elemente Der Psychophysik*, volume 1. Breitkopf und Härtel, 1860. [ English translation - Elements of Psychophysics Vol.1 - 1966 - Holt, Rinehart and Winston].

[FPSG96] J.A. Ferwerda, S.N. Pattanaik, P. Shirley, and D.P. Greenberg. A Model of Visual Adaptation for Realistic Image Synthesis. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH, pages 249–258, New York, NY, USA, 1996. ACM.

[FPSPS96] J.A. Ferwerda, S.N. Pattanaik, N. Sumanta, and D.P. Greenberg P. Shirley. A Model of Visual Adaptation for Realistic Image Synthesis. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, pages 249–258. ACM, 1996.

[GDY+] H. Gaggioni, P. Dhanendra, J. Yamashita, N. Kawada, K. Endo, and C.Clark. S-Log: A New LUT for Digital Production Mastering and Interchange Applications.

[GGC+09] O. Gallo, N. Gelfandz, W.C. Chen, M. Tico, and K. Pulli. Artifact-free High Dynamic Range imaging. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–7, April 2009.

[GRG+a] D. Gommelet, A. Roumy, C. Guillemot, M. Ropert, and J. Le Tanou. Appendix and Supplementary Documents, Pictures and Results. [Online]. Available: www.irisa.fr/temics/demos/TMOGradientBased.

[GRG+b] D. Gommelet, A. Roumy, C. Guillemot, M. Ropert, and J. Le Tanou. Supplementary Pictures and Results. [Online]. Available: www.irisa.fr/temics/demos/TMOoptimization.

[GRG+16] D. Gommelet, A. Roumy, C. Guillemot, M. Ropert, and J. LeTanou. Rate-Distortion Optimization of a Tone Mapping with SDR Quality Constraint for Backward-Compatible High Dynamic Range Compression. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 1384–1388, Sept 2016.

[GRG+17] D. Gommelet, A. Roumy, C. Guillemot, M. Ropert, and J. Le Tanou. Gradient-Based Tone Mapping for Rate-Distortion Optimized Backward-Compatible High Dynamic Range Compression. *IEEE Transactions on Image Processing*, 26(12):5936–5949, Dec 2017.

[GT12] J.U. Garbas and H. Thoma. Inter-layer Prediction for Backward Compatible High Dynamic Range Video Coding with SVC. In *Picture Coding Symposium*, pages 285–288, May 2012.

[H2693] Recommendation ITU-R H.261: Video Codec for Audiovisual Services at px64 kbits. Technical report, International Telecommunication Union, 1993.

[H2605] Recommendation ITU-R H.263: Video Coding for Low Bit Rate Communication. Technical report, International Telecommunication Union, 2005.

[H2612] Recommendation ITU-R H.262: Generic Coding of Moving Pictures and Associated Audio Information: Video. Technical report, International Telecommunication Union, 2012.

[H2617] Recommendation ITU-R H.264: Advanced Video Coding for Generic Audiovisual Services. Technical report, International Telecommunication Union, 2017.

[H2618] Recommendation ITU-R H.265: High Efficiency Video Coding. Technical report, International Telecommunication Union, 2018.

[HCT+06] Y.W. Huang, C.Y. Chen, C.H. Tsai, C.F. Shen, and L.G. Chen. Survey on Block Matching Motion Estimation Algorithms and Architectures with New Results. 42:297–320, 03 2006.

[Hec24] S. Hecht. The Visual Discrimination of Intensity and the Weber-Fechner Law. *The Journal of General Physiology*, 7(2):235, 1924.

[HM] HM Software Repository (main at HHI). [Online]. Available: hevc.hhi.fraunhofer. de/svn/svn_HEVCSoftware.

[Hou81] D. Hough. Applications of the Proposed IEEE 754 Standard for Floating-Point Arithmetic. *Computer*, 14(3):70–74, March 1981.

[Hun05] R.W.G. Hunt. *The Reproduction of Colour*. The Wiley-IS&T Series in Imaging Science and Technology. Wiley, 6 edition, 2005.

[IFM05] P. Irawan, J.A. Ferwerda, and S.R. Marschner. Perceptually Based Tone Mapping of High Dynamic Range Image Streams. In *Proceedings of the Sixteenth Eurographics Conference on Rendering Techniques*, EGSR, pages 231–242, Aire-la-Ville, Switzerland, Switzerland, 2005. Eurographics Association.

[ITU07] Recommendation BT.1788: Methodology for the Subjective Assessment of Video Quality in Multimedia Applications. Technical Report 0, International Telecommunication Union, 2007.

[ITU08] Recommendation P.910: Subjective Video Quality Assessment Methods for Multimedia Applications. Technical report, International Telecommunication Union, 2008.

[ITU11a] Recommendation BT.1886: Reference Electro-Optical Transfer Function for Flat Panel Displays Used in HDTV Studio Production. Technical report, International Telecommunication Union, 2011.

[ITU11b] Recommendation BT.601-7: Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-Screen 16:9 Aspect Ratios. Technical report, International Telecommunication Union, 2011.

[ITU14] Recommendation ITU-R BT.709-6: Parameter Values for the HDTV Standards for Production and International Programme Exchange. Technical report, International Telecommunication Union, 2014.

[ITU15] Recommendation ITU-R BT.2020-2: Parameter Values for Ultra-High Definition Television Systems for Production and International Programme Exchange. Technical report, International Telecommunication Union, 2015.

[ITU17] Recommendation ITU-R BT.2100-1: Image Parameter Values for High Dynamic Range Television for Use in Production and International Programme Exchange. Technical report, International Telecommunication Union, 2017.

[JRW97] D.J. Jobson, Z. Rahman, and G.A. Woodell. A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes. *Trans. Img. Proc.*, 6(7):965–976, July 1997.

[KAR06] E. A. Khan, A.O. Akyuz, and E. Reinhard. Ghost Removal in High Dynamic Range Images. In *International Conference on Image Processing*, pages 2005–2008, Oct 2006.

[KD13] A. Koz and F. Dufaux. Optimized Tone Mapping with LDR Image Quality Constraint for Backward-Compatible High Dynamic Range Image and Video Coding. In *IEEE International Conference on Image Processing (ICIP)*, 2013.

[KD14] A. Koz and F. Dufaux. Methods for Improving the Tone Mapping for Backward Compatible High Dynamic Range Image and Video Coding. *Image Communications*, 29(2), February 2014.

[KK11] D. K. Kwon and H. J. Kim. Region Based Weighted Prediction for Real Time H.264 Encoder. In *2011 IEEE International Conference on Consumer Electronics (ICCE)*, pages 47–48, Jan 2011.

[KML+12] I.K. Kim, J. Min, T. Lee, W.J. Han, and J. Park. Block Partitioning Structure in the HEVC Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12):1697–1706, December 2012.

[KSB+13] N.K. Kalantari, E. Shechtman, C. Barnes, S. Darabi, D.B. Goldman, and P. Sen. Patch-Based High Dynamic Range Video. *ACM Trans. Graph.*, 32(6):1–8, November 2013.

[KYH16] L. Kerofsky, Y. Ye, and Y. He. Recent Developments from MPEG in HDR Video Compression. pages 879–883, Sept 2016.

[KYK99] W-J. Kim, J-W. Yi, and S-D. Kim. A Bit Allocation Method Based on Picture Activity for Still Image Coding. *IEEE Transactions on Image Processing*, 8(7):974–977, Jul 1999.

[LAB76] Colorimetry - Part 4: CIE 1976 L*a*b* Colour Space. *ISO 11664-4:2008(E) /CIE S014-4 /E:2007*, 1976.

[Lab16] Dolby Labs. What is ICtCp - Dolby White Paper. Technical report, 2016.

[Lar98] G.W. Larson. LogLuv Encoding for Full Gamut High Dynamic Range Images. *J. Graph. Tools*, pages 15–31, 1998.

[LCC+15] H. Liu, Y. Chen, J. Chen, L. Zhang, and M. Karczewicz. Local Illumination Compensation. Technical Report Doc. VCEG-AZ06, ITU-T SG16/Q6, Jun 2015.

[LCTS05] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen. Evaluation of Tone Mapping Operators Using a High Dynamic Range Display. *ACM Trans. Graph*, 2005.

[LFH15] A. Luthra, E. Francois, and W. Husak. Call for Evidence (CfE) for HDR and WCG Video Coding. Technical Report N15083, ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, February 2015.

[LGYS04] S. Lin, Jinwei Gu, S. Yamazaki, and Heung-Yeung Shum. Radiometric Calibration from a Single Image. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 938–945, June 2004.

[LHVA16] J. Lainema, M.M. Hannuksela, V.K.M. Vadakital, and E.B. Aksu. HEVC Still Image Coding and High Efficiency Image File Format. In *IEEE International Conference on Image Processing (ICIP)*, pages 71–75, Sept 2016.

[LK12] C. Lee and C.S. Kim. Rate Distortion Optimized Layered Coding of High Dynamic Range Videos. *Journal of Visual Communication and Image Representation*, 23(6):908 – 923, 2012.

[LKVD13a] P. Lauga, A. Koz, G. Valenzise, and F. Dufaux. Region-based Tone Mapping for Efficient High Dynamic Range Video Coding. In *European Workshop on Visual Information Processing (EUVIP)*, pages 208–213, June 2013.

[LKVD13b] P. Lauga, A. Koz, G. Valenzise, and F. Dufaux. Segmentation-Based Optimized Tone Mapping for High Dynamic Range Image and Video Coding. In *Picture Coding Symposium (PCS)*, pages 257–260, Dec 2013.

[LM] Industrial Light and Magic. OpenEXR, High Dynamic Range (HDR) Image File Format. [Online]. Available: www.openexr.com/index.html.

[LMB+15] Y. Li, D. Min, M. S. Brown, M. N. Do, and J. Lu. SPM-BP: Speed-Up Patch Match Belief Propagation for Continuous MRFs. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 4006–4014, Dec 2015.

[LS03] G.W. Larson and R.A. Shakespeare. *Rendering with Radiance: The Art and Science of Lighting Visualization*. Space & Light, 2003.

[LUV76] Colorimetry - Part 5: CIE 1976 L*u*v* Colour Space and u', v' Uniform Chromaticity Scale Diagram. *ISO 11664-5:2009 /CIE S014-5 /E:2009*, 1976.

[LVCD14] P. Lauga, G. Valenzise, G. Chierchia, and F. Dufaux. Improved Tone Mapping Operator for HDR Coding Optimizing the Distortion/Spatial Complexity Trade-off. In *Signal Processing Conference (EUSIPCO), Proc. of the European*, Lisbon, Portugal, 2014.

[Mac42] D.L. MacAdam. Visual Sensitivities to Color Differences in Daylight. *Journal of the Optical Society of America*, 32(5):247–274, May 1942.

[Man06] R. Mantiuk. *High-Fidelity Imaging - The Computational Models of the Human Visual System in High Dynamic Range Video Compression, Visible Difference Prediction and Image Processing*. Doctoral dissertation, Universität des Saarlandes, Saarbrücken, December 2006.

[Man16] R. Mantiuk. Practicalities of Predicting Quality of High Dynamic Range Images and Video. In *IEEE International Conference on Image Processing (ICIP)*, pages 904–908, Sept 2016.

[MDK08] R. Mantiuk, S. Daly, and L. Kerofsky. Display Adaptive Tone Mapping. *ACM Trans. Graph.*, 27(3):1–10, August 2008.

[MDMS05] R. Mantiuk, S.J. Daly, K. Myszkowski, and H.P. Seidel. Predicting Visible Differences in High Dynamic Range Images: Model and its Calibration. In *Human Vision and Electronic Imaging X*, volume 5666, pages 204–215. International Society for Optics and Photonics, 2005.

[MDS06] L. Meylan, S. Daly, and S. Susstrunk. The Reproduction of Specular Highlights on High Dynamic Range Displays. *IS&T 14th Color Imaging Conference (CIC)*, 2006.

[MEMS06] R. Mantiuk, A. Efremov, K. Myszkowski, and H.P. Seidel. Backward Compatible High Dynamic Range MPEG Video Compression. *ACM Trans. Graph.*, 25(3):713–723, July 2006.

[MKMS04] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.P. Seidel. Perception Motivated High Dynamic Range Video Encoding. In *ACM SIGGRAPH Papers*, pages 733–741, New York, NY, USA, 2004. ACM.

[MKR07] T. Mertens, J. Kautz, and F. V. Reeth. Exposure Fusion. In *15th Pacific Conference on Computer Graphics and Applications*, pages 382–390, Oct 2007.

[MKRH11] R. Mantiuk, K.J. Kim, A.G. Rempel, and W. Heidrich. HDR-VDP-2: A Calibrated Visual Metric for Visibility and Quality Predictions in All Luminance Conditions. *ACM Trans. Graph.*, 30(4):1–14, July 2011.

[MMM+11] Z. Mai, H. Mansour, R. Mantiuk, R. Ward, P. Nasiopoulos, and W. Heidrich. Optimizing a Tone Curve for Backward Compatible High Dynamic Range Image and Video Compression. *IEEE Trans. on Image Processing*, June 2011.

[MMNW10] Z. Mai, H. Mansour, P. Nasiopoulos, and R. Ward. Visually-Favorable Tone-Mapping with High Compression Performance. In *IEEE International Conference on Image Processing (ICIP)*, pages 1285–1288, Sept 2010.

[MMNW13] Z. Mai, H. Mansour, P. Nasiopoulos, and R.K. Ward. Visually Favorable Tone-Mapping with High Compression Performance in Bit-Depth Scalable Video Coding. *IEEE Transactions on Multimedia*, 2013.

[MND12] S. Miller, M. Nezamabadi, and S. Daly. SMPTE 2084 - Perceptual Signal Coding for More Efficient Usage of Bit Codes. In *SMPTE Conferences*. Society of Motion Picture and Television Engineers, 2012.

[MP95] S. Mann and R. W. Picard. On Being 'Undigital' With Digital Cameras: Extending Dynamic Range By Combining Differently Exposed Pictures. In *Proceedings of IS&T*, pages 442–448, 1995.

[MPE93] Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media At Up to About 1,5 Mbit/s - Part 2: Video. Technical Report 11172, ISO/IEC, 1993.

[MPE03] Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 2: Video. Technical Report 13818, ISO/IEC, 2003.

[MPE04] Information Technology - Coding of Audio-Visual Objects - Part 2: Visual. Technical Report 14496, ISO/IEC, 2004.

[MPE14] Information Technology - Coding of Audio-Visual Objects - Part 10: Advanced video coding. Technical Report 14496, ISO/IEC, 2014.

[MPE17] Information Technology - High Efficiency Coding and Media Delivery in Heterogeneous Environments - Part 2: High Efficiency Video Coding. Technical Report 23008, ISO/IEC, 2017.

[MPII] Max-Planck-Institut-Informatik. HDR Image Gallery. [Online]. Available: resources.mpi-inf.mpg.de/hdr/gallery.html.

[MR11] J.J. McCann and A. Rizzi. *The Art and Science of HDR Imaging*. John Wiley & Sons, Ltd, 2011.

[MS83] J.J. More and D.C. Sorensen. Computing a Trust Region Step. *SIAM Journal on Scientific and Statistical Computing*, 4(3):553–572, 1983.

[MS08] R. Mantiuk and H.P. Seidel. Modeling a Generic Tone Mapping Operator. *Comput. Graph. Forum*, 27(2):699–708, 2008.

[MSW03] D. Marpe, H. Schwarz, and T. Wiegand. Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):620–636, July 2003.

[MT10] A. Motra and H. Thoma. An Adaptive Logluv Transform for High Dynamic Range Video Compression. In *IEEE International Conference on Image Processing*, pages 2061–2064, Sept 2010.

[MTH09] R. Mantiuk, A. Tomaszewska, and W. Heidrich. Color Correction for Tone Mapping. *Computer Graphics Forum*, 2009.

[Mul] SIM2 Multimedia. SIM2 exclusive High Dynamic Range Display Series. [Online]. Available: hdr.sim2.it/productslist.

[MWDG13] B. Masia, G. Wetzstein, P. Didyk, and D. Gutierrez. A Survey on Computational Displays: Pushing the Boundaries of Optics, Computation, and Perception. *Computers & Graphics*, 37(8):1012–1038, 2013.

[NBF+12] A. Norkin, G. Bjontegaard, A. Fuldseth, M. Narroschke, M. Ikeda, K. Andersson, M. Zhou, and G. Van der Auwera. HEVC Deblocking Filter. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1746–1754, Dec 2012.

[NM00] S.K. Nayar and T. Mitsunaga. High Dynamic Range Imaging: Spatially Varying Pixel Exposures. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR*, volume 1, pages 472–479, 2000.

[OA07] M. Okuda and N. Adami. Two-layer Coding Algorithm for High Dynamic Range Images Based on Luminance Compensation. *Journal of Visual Communication and Image Representation*, 2007.

[OLVD16] C. Ozcinar, P. Lauga, G. Valenzise, and F. Dufaux. HDR Video Coding Based on a Temporally Constrained Tone Mapping Operator. In *IEEE Digital Media Industry and Academic Forum*, Santorini, Greece, July 2016.

[OSS+12] J.R. Ohm, G.J. Sullivan, H. Schwarz, T.K. Tan, and T. Wiegand. Comparison of the Coding Efficiency of Video Coding Standards - Including High Efficiency Video Coding (HEVC). *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12):1669–1684, December 2012.

[P311] D-Cinema Quality Reference Projector and Environment for the Display of DCDM in Review Rooms and Theaters. Technical Report 10607-914, Proc. Society of Motion Picture & Television Engineers Reference Projector and Environment (SMPTE RP), 2011.

[PAB+13] T. Pouli, A. Artusi, F. Banterle, A.O. Akyuz, H.P. Seidel, and E. Reinhard. Color Correction for Tone Reproduction. In *CIC21: Twenty-first Color and Imaging Conference*, pages 215–220. Society for Imaging Science and Technology (IS&T), November 2013.

[PFFG98] S.N. Pattanaik, J.A. Ferwerda, M.D. Fairchild, and D.P. Greenberg. A Multiscale Model of Adaptation and Spatial Vision for Realistic Image Display. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH, pages 287–298, New York, NY, USA, 1998. ACM.

[PGT14] M. Le Pendu, C. Guillemot, and D. Thoreau. Rate Distortion Optimized Tone Curve for High Dynamic Range Compression. In *Proc. of the European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, 2014.

[PGT15] M. Le Pendu, C. Guillemot, and D. Thoreau. Local Inverse Tone Curve Learning for High Dynamic Range Image Scalable Compression. *IEEE Transactions on Image Processing*, 24(12):5753–5763, Dec 2015.

[PJH16] M. Pharr, W. Jakob, and G. Humphreys. *Physically Based Rendering: From Theory to Implementation*. Elsevier Science, 2016.

[Poy12] C. Poynton. *Digital Video and HD: Algorithms and Interfaces*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2nd edition, 2012.

[PTYG00] S.N. Pattanaik, J. Tumblin, H. Yee, and D.P. Greenberg. Time-dependent Visual Adaptation for Fast Realistic Image Display. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH, pages 47–54, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[QV16] C. Qifeng and K. Vladlen. Full Flow: Optical Flow Estimation By Global Optimization over Regular Grids. *Computing Research Repository (CoRR)*, 2016.

[RBS03] M.A. Robertson, S. Borman, and R.L. Stevenson. Estimation Theoretic Approach to Dynamic Range Enhancement using Multiple Exposures. *Journal of Electronic Imaging*, 12:219–228, April 2003.

[RC09] S. Raman and S. Chaudhuri. Bilateral Filter Based Compositing for Variable Exposure Photography. In *Eurographics*, pages 1–4, Munich, Germany, 2009.

[RdSdF01] N.M.M. Rodrigues, V.M.M. da Silva, and S.M.M. de Faria. Hierarchical Motion Compensation with Spatial and Luminance Transformations. In *Proceedings of International Conference on Image Processing*, volume 3, pages 518–521, 2001.

[RGB93] SMPTE Recommended Practice - Derivation of Basic Television Color Equations. 177:1–4, Nov 1993.

[RKAJ08] E. Reinhard, E.A. Khan, A.O. Akyz, and G.M. Johnson. *Color Imaging: Fundamentals and Applications*. A. K. Peters, Ltd, Natick, MA, USA, 2008.

[ROF92] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear Total Variation Based Noise Removal Algorithms. *Journal of Physics D*, 60(1-4):259–268, November 1992.

[RSSF02] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic Tone Reproduction for Digital Images. *ACM Trans. Graph*, 2002.

[RTS+07] A.G. Rempel, M. Trentacoste, H. Seetzen, H.D. Young, W. Heidrich, L. Whitehead, and G. Ward. LDR2HDR: On-The-Fly Reverse Tone Mapping of Legacy Video and Photographs. In *ACM transactions on graphics (TOG)*, volume 26, page 39, 2007.

[RWP+10] E. Reinhard, G. Ward, S.N. Pattanaik, P.E. Debevec, and W. Heidrich. *High Dynamic Range Imaging - Acquisition, Display, and Image-Based Lighting*. Academic Press, 2nd edition, 2010.

[SACM96] M. Stokes, M. Anderson, S. Chandrasekar, and R. Motta. A Standard Default Color Space for the Internet - sRGB. [Online]. Available: www.w3.org/Graphics/Color/sRGB.html, 1996.

[SB12] V. Sze and M. Budagavi. High Throughput CABAC Entropy Coding in HEVC. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12):1778–1791, December 2012.

[SBB+17] A. Segall, V. Baroncini, J. Boyce, J. Chen, and T. Suzuki. Joint Call for proposals on Video Compression with Capability beyond HEVC. Technical Report H1002, Joint Video Exploration Team (JVET), Oct 2017.

[SbMEMS12] Y. Salih, W. bt Md-Esa, A.S. Malik, and N. Saad. Tone Mapping of HDR Images: A Review. In *International Conference on Intelligent and Advanced Systems (ICIAS)*, June 2012.

[SBS+12] M. Schoberl, A. Belz, J. Seiler, S. Foessel, and A. Kaup. High Dynamic Range Video by Spatially Non-Regular Optical Filtering. In *19th IEEE International Conference on Image Processing*, pages 2757–2760, Sept 2012.

[SBS14] V. Sze, M. Budagavi, and G.J. Sullivan. *High Efficiency Video Coding (HEVC): Algorithms and Architectures*. Springer International Publishing, 1st edition, 2014.

[SBW17] G.J. Sullivan, J. Boyce, and T. Wiegand. Requirements for Future Video Coding (FVC). Technical Report BD03, ITU-T SG16/Q6 VCEG, July 2017.

[Sch93] W.F. Schreiber. *Fundamentals of Electronic Imaging Systems: Some Aspects of Image Processing*. Springer Series in Information Sciences Series. Springer London, Limited, 1993.

[Sch95] C. Schlick. Quantization Techniques for Visualization of High Dynamic Range Pictures. In *Photorealistic Rendering Techniques*, pages 7–20, Berlin, Heidelberg, 1995. Springer.

[SG31] T. Smith and J. Guild. The C.I.E. Colorimetric Standards and Their Use. *Transactions of the Optical Society*, 33(3):73, 1931.

[SHS+04] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs. High Dynamic Range Display Systems. *ACM Trans. Graph.*, 23(3):760–768, August 2004.

[SJN+12] J. Sole, R. Joshi, N. Nguyen, T. Ji, M. Karczewicz, G. Clare, F. Henry, and A. Duenas. Transform Coefficient Coding in HEVC. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12):1765–1777, December 2012.

[SKY+12] P. Sen, N.K. Kalantari, M. Yaesoubi, S. Darabi, D.B. Goldman, and E. Shechtman. Robust Patch-Based HDR Reconstruction of Dynamic Scenes. *ACM Trans. Graph.*, 31(6):1–11, November 2012.

[SM16] M.P. Sharabayko and N.G. Markov. Contemporary video compression standards: H.265/HEVC, VP9, VP10, Daala. In *International Siberian Conference on Control and Communications (SIBCON)*, pages 1–4, May 2016.

[SOHW12] G.J. Sullivan, J.R. Ohm, W.J. Han, and T. Wiegand. Overview of the High Efficiency Video Coding (HEVC) Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12):1649–1668, Dec 2012.

[SON] SONY. OLED Monitors - The BMV-X300. [Online]. Available: www.sony.co.uk/pro/product/broadcast-products-professional-monitors-oled-monitors/bvm-x300/overview/.

[SS63] J.C. Stevens and S.S. Stevens. Brightness Function: Effects of Adaptation. *Journal of the Optical Society of America.*, 53(3):375–385, 1963.

[SSD16] J. Strom, J. Samuelsson, and K. Dovstam. Luma Adjustment for High Dynamic Range Video. In *Data Compression Conference (DCC)*, pages 319–328, March 2016.

[Sul13] G. Sullivan. Common HM test conditions and software reference configurations. Technical Report L1100, JCT-VC, Geneva, Switzerland, April 2013.

[SW98] G.J. Sullivan and T. Wiegand. Rate Distortion Optimization for Video Compression. *IEEE Signal Processing Magazine*, 15(6):74–90, November 1998.

[TCS13] S. H. Tsang, Y. L. Chan, and W. C. Siu. Region-Based Weighted Prediction for Coding Video With Local Brightness Variations. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(3):549–561, March 2013.

[TKTS11] M.D. Tocci, C. Kiser, N. Tocci, and P. Sen. A Versatile HDR Video Production System. In *ACM SIGGRAPH Papers*, pages 1–10, New York, NY, USA, 2011. ACM.

[TLCS13] S. H. Tsang, T. K. Lee, Y. L. Chan, and W. C. Siu. Region Based Weighted Prediction Algorithm for H.264/AVC Video Coding. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 269–272, May 2013.

[TR93] J. Tumblin and H. Rushmeier. Tone Reproduction for Realistic Images. *IEEE Computer Graphics and Applications*, 13(6):42–48, Nov 1993.

[TSS+14] A.M. Tourapis, Y. Su, D. Singer, C. Foog, R.V. der Vleuten, and E. Francois. Report on the XYZ/HDR Exploratory Experiment 1 (EE1): Electro-Optical Transfer Functions for XYZ/HDR delivery. Technical Report M34165, ISO/IEC JTC1/SC29/WG11 MPEG, 2014.

[TZC14] L. Tian, Y. Zhou, and X. Cao. A New Rate Complexity QP Algorithm (RCQA) for HEVC Intra-Picture Rate Control. In *International Conference on Computing, Networking and Communications (ICNC)*, pages 375–380, Feb 2014.

[vdBL01] C.J. van den Branden Lambrecht. *Vision Models and Applications to Image and Video Processing*. Springer US, 2001.

[VSLD14] G. Valenzise, F. De Simone, P. Lauga, and F. Dufaux. Performance evaluation of objective quality metrics for HDR image compression. In *Applications of Digital Image Processing XXXVII*. SPIE, August 2014.

[VV72] A. VanMeeteren and J.J. Vos. Resolution and Contrast Sensitivity at Low Luminances. *Vision Research*, 12(5), 1972.

[Wal92] G.K. Wallace. The JPEG Still Picture Compression Standard. *IEEE transactions on consumer electronics*, 38(1):xviii–xxxiv, 1992.

[Wan95] B.A. Wandell. *Foundations of Vision*. Sinauer Associates, 1995.

[War91] G. Ward. Real Pixels. In James Arvo, editor, *Graphics Gems II*, pages 80–83. Academic Press, 1991.

[War03] G. Ward. High Dynamic Range Image Examples. [Online]. Available: www.anyhere.com/gward/hdrenc/pages/originals.html, 2003.

[WB02] Z. Wang and A.C. Bovik. A Universal Image Quality Index. *IEEE Signal Processing Letters*, 9(3):81–84, March 2002.

[WB16] L. Walton and L. Barwick. Next Generation Video to Be Unveiled at IBC Amsterdam. [Online]. Available: warwick.ac.uk/newsandevents/pressreleases/scene-referred_next_generation, 2016.

[WNC87] I.H. Witten, R.M. Neal, and J.G. Cleary. Arithmetic Coding for Data Compression. *ACM Communications*, 30(6):520–540, June 1987.

[WS06] G. Ward and M. Simmons. JPEG-HDR: A Backward-Compatible, High Dynamic Range Extension to JPEG. In *ACM Siggraph Courses*, New York, USA, 2006.

[WS11] T. Wiegand and H. Schwarz. Source Coding: Part I of Fundamentals of Source and Video Coding. *Foundations and Trends in Signal Processing*, January 2011.

[WSBL03] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra. Overview of the H.264/AVC Video Coding Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7), July 2003.

[YTB05] P. Yin, A.M. Tourapis, and J. Boyce. Localized Weighted Prediction for Video Coding. In *IEEE International Symposium on Circuits and Systems*, volume 5, pages 4365–4368, May 2005.

[ZLF+15] N. Zhang, Y. Lu, X. Fan, R. Xiong, D. Zhao, and W. Gao. Enhanced Inter Prediction with Localized Weighted Prediction in HEVC. In *Visual Communications and Image Processing (VCIP)*, pages 1–4, Dec 2015.

[ZZRW14] K. Zeng, T. Zhao, A. Rehman, and Z. Wang. Characterizing perceptual artifacts in compressed video streams. In *Human Vision and Electronic Imaging XIX*, volume 9014. International Society for Optics and Photonics, 2014.

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

UNIVERSITÉ DE RENNES 1

**Titre :** Méthodes pour améliorer la compression HDR (High Dynamic Range) rétro compatible.

**Mots clés :** High Dynamic range (HDR), Compression Vidéo, Tone Mapping (TMO), HEVC

**Résumé:** Ces dernières années, les contenus vidéos ont évolué très rapidement. En effet, les télévisions (TV) ont rapidement évolué vers l'Ultra Haute résolution (UHD), la Haute Fréquence d'images (HFR) ou la stéréoscopie (3D). La tendance actuelle est à l'imagerie à Haute Dynamique de luminance (HDR). Ces technologies permettent de reproduire des images beaucoup plus lumineuses que celles des écrans actuels. Chacune de ces améliorations représente une augmentation du coût de stockage et nécessite la création de nouveaux standards de compression vidéo, toujours plus performant. La majorité des consommateurs est actuellement équipé de TV ayant une Dynamique Standard (SDR) qui ne supportent pas les contenus HDR et ils vont lentement renouveler leurs écrans pour un HDR. Il est donc important de délivrer un signal HDR qui puisse être décodé par ces deux types d'écrans. Cette rétro compatibilité est rendue possible par un outil appelé TMO (Tone Mapping Operator) qui transforme un contenu HDR en une version SDR. Au travers de cette thèse, nous explorons de nouvelles méthodes pour améliorer la compression HDR rétro compatible. Premièrement, nous concevons un TMO qui optimise les performances d'un schéma de compression scalable où une couche de base et d'amélioration sont envoyées pour reconstruire les contenus HDR et SDR. Il est démontré que le TMO optimal dépend seulement de la couche SDR de base et que le problème de minimisation peut être séparé en deux étapes consécutives. Pour ces raisons, nous proposons ensuite un autre TMO conçu pour optimiser les performances d'un schéma de compression utilisant uniquement une couche de base mais avec un modèle amélioré et plus précis. Ces deux travaux optimisent des TMO pour images fixes. Par la suite, la thèse se concentre sur l'optimisation de TMO spécifiques à la vidéo. Cependant, on y démontre que l'utilisation d'une prédiction pondérée pour la compression SDR est aussi bon voir meilleur que d'utiliser un TMO optimisé temporellement. Pour ces raisons, un nouvel algorithme et de nouveaux modes de prédictions pondérées sont proposés pour gérer plus efficacement la large diversité des changements lumineux dans les séquences vidéos.

**Title :** Methods for Improving the Backward Compatible High Dynamic Range Compression

**Keywords :** High Dynamic range (HDR), Video Compression, Tone Mapping (TMO), HEVC

**Abstract:** In recent years, video content evolved very quickly. Indeed, televisions (TV) quickly evolved to Ultra High Definition (UHD), High Frame Rate (HFR) or stereoscopy (3D). The recent trend is towards High Dynamic range (HDR). These new technologies allow the reproduction of much brighter images than for actual displays. Each of these improvements represents an increase in storage cost and therefore requires the creation of new video compression standards, always more efficient. The majority of consumers are currently equipped with Standard Dynamic Range (SDR) displays, that cannot handle HDR content. Consumers will slowly renew their display to an HDR one and it is therefore of great importance to deliver an HDR signal that can be decoded by both SDR and HDR displays. Such backward compatibility is provided by a tool called Tone Mapping Operator (TMO) which transforms an HDR content into an SDR version. In this thesis, we explore new methods to improve the backward compatible HDR compression. First, we design a Tone Mapping to optimize scalable compression scheme performances where a base and an enhancement layer are sent to reconstruct the SDR and HDR content. It is demonstrated that the optimum TMO only depends on the SDR base layer and that the minimization problem can be separated in two consecutive minimization steps. Based on these observations, we then propose another TMO designed to optimize the performances of compression schemes using only a base layer but with an enhanced and more precise model. Both of these works optimize TMO for still images. Thereafter, this thesis focuses on the optimization of video-specific TMO. However, we demonstrate that using a weighted prediction for the SDR compression is as good or even better than using a temporally optimized TMO. Therefore, we proposed a new weighted prediction algorithm and new weighted prediction modes to handle more efficiently the large diversity of brightness variations in video sequences.