



**HAL**  
open science

## Schéma de compression d'images intégrales basé sur l'extraction de vues

Antoine Dricot, Joël Jung, Marco Cagnazzo, Beatrice Pesquet-Popescu,  
Frederic Dufaux

► **To cite this version:**

Antoine Dricot, Joël Jung, Marco Cagnazzo, Beatrice Pesquet-Popescu, Frederic Dufaux. Schéma de compression d'images intégrales basé sur l'extraction de vues. Colloque COMpression et REprésentation des Signaux Audiovisuels (CORESA'2016), May 2016, Vandoeuvre-lès-Nancy, France. hal-01436111

**HAL Id: hal-01436111**

**<https://hal.science/hal-01436111>**

Submitted on 10 Jan 2020

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

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

# Schéma de compression d’images intégrales basé sur l’extraction de vues

A. Dricot<sup>1,2</sup>

J. Jung<sup>1</sup>

M. Cagnazzo<sup>2</sup>

B. Pesquet<sup>2</sup>

F. Dufaux<sup>2</sup>

<sup>1</sup> Orange Labs

<sup>2</sup> Institut Mines-Télécom ; Télécom ParisTech ; CNRS LTCI

antoine.dricot;joelb.jung;{@orange.com}

cagnazzo;beatrice.pesquet;frederic.dufaux;{@telecom-paristech.fr}

## Résumé

L’imagerie intégrale est une technologie qui permet de capturer une scène sous la forme d’une représentation dite *light-field*. Cette représentation offre plusieurs points de vue de la scène et permet d’éliminer beaucoup d’inconvénients connus en stéréoscopie et en auto-stéréoscopie par exemple. Les images intégrales ont cependant une résolution élevée et une structure en micro-images difficiles à encoder. Cet article propose un schéma de compression d’images intégrales basé sur l’extraction de vues. Les gains moyens en BD-rate par rapport à la référence HEVC sont de 15.7% (jusqu’à 31.3%). Les paramètres de ce schéma peuvent prendre un nombre étendu de valeurs. Une recherche exhaustive parmi ces valeurs permet d’abord de donner les résultats de la meilleure configuration. Puis un critère d’optimisation débit-distorsion est proposé pour éviter la recherche exhaustive, permettant de réduire le temps d’encodage tout en préservant les gains. Finalement, l’étude plus fine de l’impact des différents paramètres permet de réduire encore davantage le temps de codage.

## Mots clefs

Imagerie Intégrale, Imagerie Plénoptique, Holoscopie, Compression d’Image, Codage Vidéo, Extraction de vues

## 1 Introduction

Le développement technologique lié à la vidéo 3D tend à créer des expériences de visualisation de plus en plus réalistes et immersives. Les technologies vidéo 3D actuellement disponibles sur le marché présentent cependant plusieurs limitations. L’utilisation de lunettes en stéréoscopie crée un manque de confort. De plus, le conflit entre les distances d’accommodation et de convergence n’est pas naturel pour le système visuel humain et peut causer des gênes. Malgré un nombre plus élevé de vues, les systèmes auto-stéréoscopiques sont limités par des éléments de perception non-naturels, comme par exemple le manque d’une parallaxe de mouvement fluide (i.e. une visualisation continue quand l’utilisateur se déplace devant l’écran), qui est un élément clé dans la perception de la profondeur [1]. L’imagerie intégrale est une technologie basée sur la photographie plénoptique [2]. Cette technique permet d’obtenir

une représentation du *light-field* d’une scène [3], qui permet d’éliminer la plupart des inconvénients existants dans les technologies 3D actuelles (e.g. le conflit convergence-accommodation). L’acquisition en imagerie intégrale est basée sur l’utilisation d’un dispositif lenticulaire, constitué d’un ensemble de micro-lentilles, placé devant le périphérique de capture (i.e. la caméra ou son capteur). Chaque micro-lentille produit une micro-image (MI), et chaque MI contient l’information provenant de plusieurs angles de vue. L’image intégrale résultante correspond à un ensemble de MIs, comme illustré en Figure 1.

Des caméras plénoptiques sont déjà disponibles sur le marché, avec notamment *Lytro* [4] et *Raytrix* [5]. De plus, plusieurs institutions ont déjà montré de l’intérêt pour les technologies liées au *light-field* en travaillant sur des systèmes d’affichage [6] pour une visualisation sans lunette, visant à offrir une expérience réaliste et immersive. La téléprésence est un cas d’application cible intéressant, ainsi que la réalité virtuelle avec la navigation dans une scène en relief [7]. Pour gérer efficacement la résolution élevée des images intégrales et les caractéristiques spécifiques de leur structure, de nouvelles technologies de codage sont nécessaires. On propose dans cet article un schéma de compression efficace qui exploite les techniques d’extraction de vues afin de créer une image intégrale résiduelle qui est encodée. La performance de ce schéma dépend de plusieurs paramètres pouvant prendre un nombre étendu de valeurs. On propose donc un ensemble de méthodes itératives pour sélectionner la configuration la plus efficace en jouant sur le compromis débit-distorsion et complexité.

La suite de cet article est organisée comme suit. En Section 2, les méthodes d’extractions de vues existantes sont décrites, et l’état de l’art des méthodes de codage d’images intégrales est présenté. Le schéma de compression proposé est décrit puis les résultats expérimentaux sont montrés en Section 3. Les conclusions sont tirées dans la Section 4.

## 2 État de l’art

### 2.1 Extraction de vues

Plusieurs méthodes pour extraire des vues d’une image intégrale sont décrites dans [9]. La méthode basique extrait un *patch* (une zone carrée de pixels) de chaque MI,

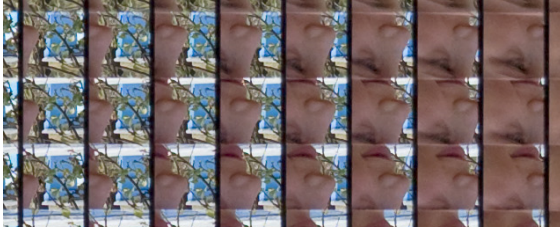


Figure 1 – *Micro-Images (MIs) - Laura [8]*

comme illustré en Figure 2 (*gauche*). Cette méthode est basée sur les caractéristiques de la *focused plenoptic camera* [10] pour laquelle chaque MI contient de l'information spatiale et de l'information angulaire. Une méthode plus basique consiste à utiliser un patch de taille  $1 \times 1$ , i.e. un pixel par MI. La taille du patch définit la profondeur (distance à la caméra) dans la scène du plan sur lequel la mise au point sera faite dans la vue extraite : plus le patch est large, moins cette distance est élevée. L'angle de vue dépend de la position relative du patch dans la MI. Une méthode plus avancée permet d'atténuer les effets de blocs en lissant les transitions entre les patches adjacents. Les pixels entourant chaque patch sont moyennés avec une pondération (les pixels les plus éloignés du centre ont un poids plus faible). Une méthode d'estimation de disparité (basée sur *block-matching*) est proposée dans [10] pour obtenir la profondeur relative des objets dans chaque MI. Une valeur de disparité par MI est estimée, correspondant à la taille du patch à utiliser. Dans les vues extraites par cette méthode dite *disparity-assisted patch blending extraction* (DAPBE [9]), tous les objets sont nets car les tailles de patch sont adaptées à la profondeur.

## 2.2 Codage d'images intégrales

Les images intégrales doivent avoir une grande résolution pour obtenir un nombre élevé de vues différentes avec une résolution suffisante. De plus, la structure en micro-images (MIs) donne un aspect de grille qui rend l'image difficile à encoder (voir Fig. 1). Une première approche consiste à appliquer la Transformée en Cosinus Discrète (*Discrete Cosine Transform - DCT*) sur les MIs, suivie d'une quantification et d'un codage sans perte [11]. Les corrélations

inter-MIs peuvent également être exploitées en utilisant une *3D-DCT* sur les MIs superposées en un volume [12]. Dans [13], une transformée en ondelettes (*Discrete Wavelet Transform - DWT*) est appliquée aux MIs, suivie d'une DCT sur les blocs de coefficients transformés (transformée hybride à 4 dimensions). Les approches basées sur les transformées correspondent bien à la structure en MIs mais donnent des gains en compression limités en comparaison aux encodeurs standards actuels (H.264/AVC [14] et HEVC [15]). Dans [16] et [17], les images intégrales sont décomposées en vues, qui sont encodées avec l'encodeur MVC [18]. Cette approche est efficace sur les images générées par ordinateur (i.e. dont les MIs sont parfaitement alignées sur les pixels) mais reste limitée pour les contenus naturels. Le mode *Self-similarity* [19] est une autre approche basée sur le même principe que le mode *Intra Block Copy* [20], qui exploite les corrélations spatiales non-locales entre les MIs. Un algorithme de *block-matching* est utilisé, comme pour le mode inter de H.264/AVC et HEVC, mais au sein de l'image courante (zone causale déjà encodée). Cet outil apporte de bons gains pour les images fixes mais reste limité pour les séquences quand la prédiction temporelle est activée.

Un schéma de codage scalable est proposé dans [21]. Ce schéma offre une fonctionnalité intéressante de scalabilité côté écran (i.e. un flux adapté aux systèmes d'affichage 2D, multi-vues, et holoscopiques). La couche 0 correspond à la vue centrale, la couche 1 correspond à un ensemble de vues additionnelles, et la couche 2 est l'image intégrale. Cette scalabilité a donc un coût, car des vues supplémentaires sont encodées. Un schéma de prédiction inter-couches est proposé pour réduire le débit dédié à la couche 2, dans lequel une image intégrale est partiellement reconstruite à partir des vues (couche 1) et ajoutée aux images de référence.

Dans la Section 3, on propose un schéma de codage original pour les images intégrales. Bien qu'il soit basé sur l'extraction de vue, et permette une certaine forme de scalabilité, il diffère des méthodes scalables existantes : son but principal est l'efficacité de codage. Il utilise avantageusement le processus d'extraction pour reconstruire un prédicteur fiable et créer une image intégrale résiduelle qui est encodée.

## 3 Schéma de codage proposé

Le schéma de compression proposé (Fig. 3) est décrit dans cette section. Dans ce schéma, une image résiduelle  $II_R$  est encodée avec HEVC (flux *résiduel*).  $II_R$  est la différence entre l'image originale  $II$  et une image reconstruite  $II^*$ .  $II^*$  est reconstruite à partir de vues extraites de l'image originale  $II$ . Le nombre de vues n'est pas limité. Les valeurs de disparité utilisées pour l'extraction et la reconstruction sont encodées sans compression, et les vues extraites sont encodées avec 3D-HEVC (flux *vues*). De par leur résolution réduite, les vues représentent un nombre de bits réduit à encoder, en comparaison à  $II$ . De plus, elles ont un as-

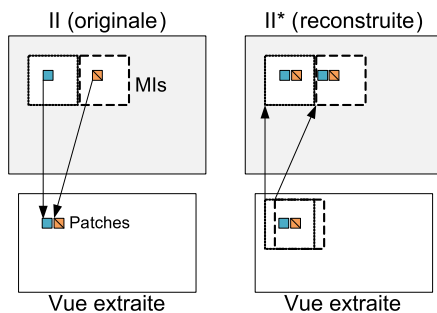


Figure 2 – *Extraction (gauche), Reconstruction (droite)*

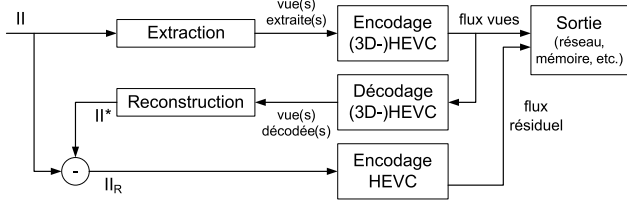


Figure 3 – Schéma de codage proposé - encodeur

pect d’images naturelles qui est moins couteux à encoder, en opposition à l’aspect donné par les MIs. Pour obtenir des vues avec un tel aspect, des méthodes d’extractions avancées sont utilisées (voir Sec. 2.1), basée sur l’extraction de patch et avec une moyenne pondérée sur les bords, ce qui empêche une reconstruction parfaite pour  $II^*$ . L’information manquante, correspondant à la différence entre  $II$  et  $II^*$ , est récupérée dans  $II_R$ . Par définition, pour une image reconstruite  $II^*$  proche de l’originale  $II$ , la différence donne des valeurs proches de zéro (un exemple de l’aspect de l’image résiduelle est illustré en Fig. 4).  $II_R$  a donc un aspect *lisse* avec peu de variations, qui est plus facile à encoder avec HEVC que  $II$ .

Comme l’illustre la Figure 2 (droite), lors de la reconstruction de  $II^*$  à partir de vues, certains pixels provenant d’angles de vue différents sont remplacés par des pixels adjacent provenant de la même vue. Cependant la transformation d’un objet en changeant d’angle de vue n’est pas limitée à une simple translation (disparité) mais implique également un mouvement angulaire. Des erreurs sont donc introduites. Un filtrage passe-bas (e.g. moyenneur) est appliqué aux vues décodées avant la reconstruction pour atténuer ces erreurs. Les hautes fréquences dans la vue (qui correspondent par exemple au bruit et aux détails) sont filtrées tout en préservant la forme des objets.

Côté décodeur, les vues sont décodées et utilisées pour reconstruire  $II^*$ , et  $II_R$  est décodée puis ajoutée (somme) à  $II^*$  pour obtenir l’image de sortie.

Il existe un compromis entre le débit et la qualité des vues, et le débit de  $II_R$ .  $II^*$  doit être la plus proche possible de  $II$  afin de minimiser le coût de  $II_R$ , sans augmenter trop le coût des vues. Plusieurs combinaisons sont possibles pour les paramètres suivants : le QP utilisé pour encoder les vues ( $QP_V$ ), le QP utilisé pour encoder l’image résiduelle ( $QP_R$ ), et la taille (en pixels) du filtre passe-bas appliqué

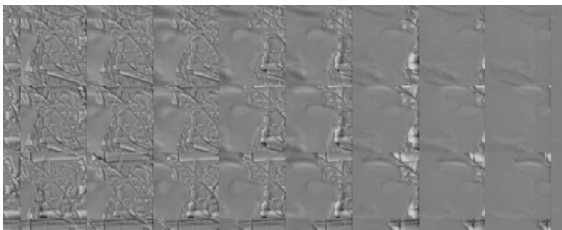
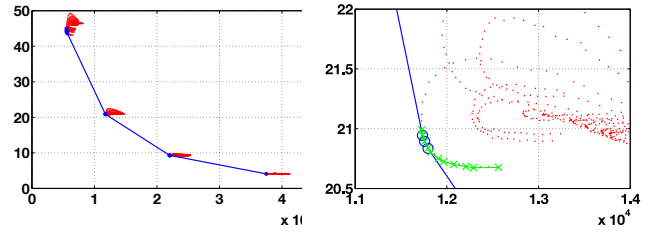


Figure 4 – Micro-Images (MIs) in a residual image - Laura [8]



(a) Toutes les configurations

(b) Focus sur  $QP_R$  20

Figure 5 – Points débit-distorsion pour toutes les configurations (Fountain) - horizontal : débit (kbits/image), vertical : EQM

aux vues décodées ( $B$ ). Dans la suite, on propose différents compromis entre débit-distorsion et complexité, en explorant plusieurs méthodes pour paramétrer ces valeurs.

### 3.1 Recherche exhaustive de la configuration optimale

Nos expérimentations incluent sept images fixes [8], listées dans le Tableau 1. Ces images sont rognées pour retirer les MIs incomplètes, et débarrassées des pixels de la grille correspondant aux limites entre les micro-lentilles [22]. La méthode dite *disparity-assisted patch blending extraction* (DAPBe [9]) est utilisée pour extraire une vue unique. Les encodages de la vue et de l’image résiduelle sont effectués avec HEVC (HM14.0) dans la configuration *Intra main* [23], et les valeurs de disparité sont codées sur 4 bits par MI. Les résultats de compression sont donnés avec la métrique Bjøntegaard Delta (BD) rate [24]. L’ancrage de référence est  $II$  encodée avec HEVC sur l’intervalle de QP  $\{25,30,35,40\}$ , et les valeurs négatives représentent un gain sur la référence.

Comme en pratique, une très large partie du débit est dédiée à  $II_R$ , la valeur de  $QP_R$  est déterminée en fonction du débit (ou de la qualité) cible, et  $QP_V$  et  $B$  sont considérés comme des paramètres à optimiser pour un  $QP_R$  donné. Pour chaque  $QP_R$  dans l’intervalle  $\{10,15,20,25\}$ , des combinaisons de valeurs pour les paramètres  $QP_V$  et  $B$  (respectivement dans les intervalles  $\{10,11,\dots,50\}$  et  $\{1,2,\dots,11\}$ ) sont testées par itérations, fournissant 1804 ( $4 \times 41 \times 11$ ) points débit-distorsion (*RD-points*, illustrés par les points rouges dans la Figure 5). Pour chaque image, le Tableau 1 montre la configuration qui donne les meilleurs résultats en BD-rate.

Un gain moyen de 15.7% (jusqu’à 31.3% pour *Fredo*) est rapporté en utilisant les combinaisons optimales de paramètres. La valeur de  $QP_V$  augmente en rapport à  $QP_R$ , offrant un compromis entre le débit de la vue et de  $II_R$ . Approximativement 97% du débit total est dédié à  $II_R$  en moyenne, principalement à cause de sa large résolution par rapport à la vue (e.g. pour *Fountain*  $6512 \times 4880$  contre  $960 \times 720$ ), qui représente les 3% restants (les valeurs de disparités utilisées pour l’extraction et la reconstruction représentent seulement 0.3%).

Les valeurs optimales pour  $QP_V$  et  $B$  sont ici sélection-

Image	BD-rate (%)	Pour chaque $QP_R$ dans {10,15,20,25}							
		$QP_V$		$B$					
Fountain	-17.0	19	21	23	29	3	3	3	3
Fredo	-31.3	18	21	25	32	3	3	3	3
Jeff	-5.9	25	30	30	32	9	9	9	7
Laura	-11.2	22	25	27	31	4	4	4	4
Seagull	-13.7	20	21	25	29	3	3	3	3
Sergio	-23.6	19	19	24	32	4	2	2	2
Zenhgyun1	-7.5	25	26	30	32	9	9	9	7
Moyenne	-15.7								

Tableau 1 – Gains BD-Rate avec les configurations optimales de  $QP_V$  et  $B$  pour chaque  $QP_R$ . Les valeurs négatives représentent un gain sur la référence

nées exhaustivement parmi les 1804 points, et dépendent de l'image testée. En Section 3.2, les résultats fournis par cette étude préliminaire sont utilisés pour déterminer un processus d'optimisation débit-distorsion (*Rate Distortion Optimisation - RDO*) qui sélectionne automatiquement les meilleures valeurs pour un  $QP_R$  donné.

### 3.2 Processus local d'optimisation débit-distorsion

La Figure 5 illustre les valeurs RD obtenues pendant la recherche exhaustive de la meilleure combinaison pour  $QP_V$ ,  $QP_R$  et  $B$ . On définit l'enveloppe convexe globale (GCH, en bleu) comme l'enveloppe convexe de tous les points, et l'enveloppe convexe locale (LCH, en vert) comme l'enveloppe convexe d'un ensemble de points ayant la même valeur de  $QP_R$ . Les points optimaux pour un  $QP_R$  donné sont situés à l'intersection  $S$  entre LCH et GCH. D'après nos données expérimentales, on peut observer que pour chaque  $QP_R$ , il existe au moins un point qui appartient à GCH (i.e. cette intersection n'est pas vide). Utiliser uniquement LCH peut donner des configurations sous-optimales, comme le montre sur la Figure 5 les points marqués d'une croix verte et qui sont situés sur la partie droite de GCH. GCH ne peut cependant être connue qu'en encodant l'image avec plusieurs  $QP_R$ , ce qui multiplie le nombre de combinaisons testées.

L'objectif dans cette section est de pouvoir sélectionner la configuration, pour un  $QP_R$  donné, qui fournit des valeurs de débit et distorsion (respectivement R et D) qui minimisent le coût  $D + \lambda R$ , où  $\lambda$  est la pente de LCH dans  $S$  (donc de GCH). Dans la Figure 5, cela revient à trouver parmi les points marqués d'une croix, ceux qui sont également marqués d'un cercle.

La pente de GCH entre deux valeurs consécutives de  $QP_R$  évolue exponentiellement par rapport à  $QP_R$ . Une estimation de  $\lambda = f(QP_R)$  est donc possible. La fonction définie par  $\lambda = 2^{aQP_R+b}$  (avec  $a = 0.34$  et  $b = -15.8$ ) a une excellente correspondance avec les données.

Une première méthode  $All_{RDO}$  est proposée, où les combinaisons de  $QP_V$  et  $B$  sont successivement testées pour un

Image	BD-rate (%)	Temps (%)	Pour chaque $QP_R$ dans {10,15,20,25}							
			$QP_V$		$B$					
Fountain	-17.0	48284	19	21	23	27	3	3	3	3
Fredo	-31.1	47067	18	21	25	28	3	3	3	3
Jeff	-5.9	48729	25	30	30	32	9	9	9	7
Laura	-11.2	49065	22	25	27	30	4	4	4	4
Seagull	-13.7	48836	19	21	25	29	3	3	3	3
Sergio	-23.5	48036	20	21	24	28	4	2	2	2
Zenhgyun1	-7.5	48554	25	26	31	30	9	9	9	7
Moyenne	-15.7	48367								

Tableau 2 – BD-Rate, variations du temps de codage et configurations associées pour la méthode  $All_{RDO}$

$QP_R$  donné. La combinaison qui donne les valeurs de débit et distorsion minimisant le coût  $D + \lambda R$  (avec  $\lambda$  déterminé comme ci-dessus) est sélectionnée. Les conditions de test sont telles que décrites en Section 3.1. Le Tableau 2 donne le BD-rate et la variation de temps de codage (pour chaque image et en moyenne) en référence à l'ancrage HEVC.

Les combinaisons sélectionnées par  $All_{RDO}$  sont très proches des configurations optimales déterminées dans la Section 3.1 (même valeurs pour  $B$  et seulement de légères différences pour quelques valeurs de  $QP_V$ ), et les gains de 15.7% sont préservés, ce qui montre la robustesse de l'estimation de  $\lambda = f(QP_R)$ . Le temps total d'encodage pour toutes les itérations est très large, avec une multiplication du temps de référence par 484 en moyenne. On note que les intervalles de valeurs testées pour  $QP_V$  et  $B$  ne sont pas utilisés entièrement et peuvent donc être réduits pour diminuer le nombre d'itérations. Deux variantes sont proposées dans la suite afin de réduire encore ce nombre.

Avec  $B_{RDO}$  et  $B_{MSE}$ , les itérations sur  $B$  sont faites uniquement pour une valeur de  $QP_V$  (e.g. pour  $QP_V = 10$  dans nos expérimentations), et la meilleure valeur de  $B$  est conservée pour les autres itérations sur  $QP_V$ . La meilleure valeur pour  $B$  est celle qui minimise le coût  $D + \lambda R$  dans  $B_{RDO}$  (même processus d'optimisation que pour  $All_{RDO}$ ), et celle qui minimise l'Erreur Quadratique Moyenne (EQM) entre  $\Pi^*$  et  $\Pi$  dans  $B_{MSE}$ . Le Tableau 3 montre les résultats BD-rate moyens et les variations en temps de codage associées pour ces méthodes.

Pour  $B_{RDO}$ , le temps d'encodage total est réduit significativement (de 484 fois le temps de référence à 55 fois) car le nombre d'itérations est fortement réduit. Les gains de 15.7% sont préservés car  $B$  ne varie pas significativement par rapport à  $QP_V$ . Les résultats de  $B_{MSE}$  montrent que le temps d'encodage peut encore être réduit (à 44 fois la référence) en sélectionnant  $B$  sans encoder l'image résiduelle à chaque itération, avec un gain presque aussi bon (15.3% en moyenne, e.g. avec une diminution de 1.7% pour *Seagull*, et 0.8% pour *Sergio*). On note que le nombre d'itérations sur  $QP_V$  peut être réduit encore en évitant de chercher sur tout l'intervalle {10,11,...,50}. Par exemple, on peut observer que le coût  $D + \lambda R$  a un minimum local par rapport à

	All <sub>RDO</sub>	BR <sub>D</sub> O	BMSE	QP <sub>V</sub> <sub>fixed</sub>	All <sub>fixed</sub>
BD-rate	-15.7	-15.7	-15.3	-15.5	-8.5
Temps	48367	5526	4443	136	120

Tableau 3 – Gains BD-Rate (%) et variations du temps de codage associée(%), en moyenne

QP<sub>V</sub> pour  $B$  et QP<sub>R</sub> donnés. Les itérations peuvent donc être stoppées quand le coût commence à augmenter.

### 3.3 Détermination empirique de la configuration

Deux variantes supplémentaires sont définies, QP<sub>V</sub><sub>fixed</sub> et All<sub>fixed</sub>, dans lesquelles la valeur de QP<sub>V</sub> est fixée empiriquement pour chaque QP<sub>R</sub>. Avec QP<sub>V</sub><sub>fixed</sub>,  $B$  est sélectionnée par rapport à l'EQM entre II\* et II (comme pour B<sub>MSE</sub> dans la Sec. 3.2), tandis que pour All<sub>fixed</sub>, la valeur de  $B$  est également fixée.

Les résultats de QP<sub>V</sub><sub>fixed</sub> dans le Tableau 3 montrent qu'assigner un QP<sub>V</sub> à un QP<sub>R</sub> réduit largement le temps d'encodage (seulement 1.4 fois la référence) et donne quand même un gain moyen proche de l'optimal avec 15.5%. Bien que le nombre d'images de test disponibles soit limité, les paramètres diffèrent peu d'une image à l'autre. Cette robustesse suggère des gains similaires sur d'autres images intégrales. Le temps de codage est de seulement 1.2 fois la référence pour All<sub>fixed</sub>. Par contre, le gain moyen descend à 8.5%, avec des pertes pour *Jeff* et *Zenhyun1*. La valeur optimale de  $B$  dépend fortement du contenu de l'image et les itérations sur ce paramètre peuvent améliorer de manière significative la performance de codage, avec seulement une légère augmentation du temps de codage.

Des résultats préliminaires montrent des performances similaires entre le schéma proposé et le mode *Intra Block Copy* [20] (même principe que *Self-Similarity* [19]), et que combiner les deux méthodes améliore encore l'efficacité.

Le Tableau 4 montre les variations du temps d'encodage et de décodage du schéma proposé par rapport à la référence, ainsi que le pourcentage du temps total dédié à chaque tâche, pour l'image *Fountain* avec la méthode QP<sub>V</sub><sub>fixed</sub>. Le schéma proposé a un temps d'encodage total de 1.3 fois le temps d'encodage de II avec HEVC (référence). 79% du

Temps (%)	Variation ancrage	HEVC				
		Extr.	Rec.	Vue	II <sub>R</sub>	Autres
Encodage	130	7	8	2	79	4
Décodage	240	.	31	1	46	22

Tableau 4 – Fountain - Variation du temps de codage par rapport à l'ancrage, avec QP<sub>V</sub><sub>fixed</sub>, et pourcentage du temps total pour chaque tâche, incluant : extraction, reconstruction, codage/décodage de la vue et de l'image résiduelle, et filtrage, différence et somme (autres).

temps est dédié à l'encodage de II<sub>R</sub>. L'extraction de la vue prend 7% du temps total, principalement à cause du temps dédié à l'estimation de disparité. Les onze itérations des étapes de filtrage, reconstruction, et différence représentent 12%. Le temps de décodage est de 2.4 fois le temps de référence, avec 46% dédié au décodage de II<sub>R</sub>. Il ne dépend pas du nombre d'itérations côté encodeur. La reconstruction (31%) et la somme de II<sub>R</sub> et II\* (22%) représentent un pourcentage plus élevé au décodeur car le processus de décodage avec HEVC est bien plus rapide que l'encodage. Cette augmentation est plus élevée en bas débit, où le temps de décodage d'HEVC est encore réduit davantage, tandis que le temps de reconstruction et de la somme ne varie pas.

## 4 Conclusions

Un schéma efficace de compression d'images intégrales est proposé dans cet article. Une image intégrale résiduelle et une vue extraite  $y$  sont encodées. L'image résiduelle est issue de la différence entre l'image originale et une image reconstruite à partir de la vue. Un gain en BD-rate de 15.7% en moyenne, et jusqu'à 31.3%, est obtenu par rapport à la référence HEVC. La configuration du QP utilisé pour encoder la vue, et de la taille du filtre passe-bas appliqué à cette vue, ont un large impact sur l'efficacité du codage. Une méthode robuste d'optimisation débit-distorsion est d'abord modélisée pour sélectionner la meilleure configuration en préservant les gains optimaux. On limite ensuite le nombre d'itérations pour réduire le temps d'encodage, tout en conservant les gains. On montre finalement qu'il est possible d'assigner une valeur unique de QP pour la vue à une valeur de QP donnée pour l'image résiduelle, avec une perte d'efficacité minimale, et que la taille du filtre passe-bas peut être sélectionnée avec un nombre réduit d'itérations. De cette étude résulte un codec réaliste du point de vue du compromis entre performances de codage et complexité. Les perspectives pour les futurs travaux incluent des tests du schéma avec plusieurs vues extraites, différents filtres pour la (les) vue(s), et l'utilisation de méthodes d'extractions plus avancées.

## Références

- [1] F. Dufaux, B. Pesquet-Popescu, et M. Cagnazzo. *Emerging technologies for 3D video : content creation, coding, transmission and rendering*. Wiley Eds, 2013.
- [2] G. Lippmann. Epreuves reversibles donnant la sensation du relief. *J. Phys. Theor. Appl.*, 7(1) :821–825, 1908.
- [3] A. Lumsdaine et T. Georgiev. Full resolution light-field rendering. *Indiana University and Adobe Systems, Tech. Rep*, 2008.
- [4] <https://www.lytro.com/>.
- [5] <https://raytrix.de/>.
- [6] M. P. Tehrani, T. Senoh, M. Okui, K. Yamamoto, N. Inoue, et T. Fujii. [m31261][FTV AHG] Multiple

- aspects. Dans *ISO/IEC JTC1/SC29/WG11*, October 2013.
- [7] M. P. Tehrani, T. Senoh, M. Okui, K. Yamamoto, N. Inoue, et T. Fujii. [m31103][FTV AHG] Introduction of super multiview video systems for requirement discussion. Dans *ISO/IEC JTC1/SC29/WG11*, October 2013.
- [8] <http://www.tgeorgiev.net/>.
- [9] J. F. O. Lino. 2D image rendering for 3D holoscopic content using disparity-assisted patch blending. *Thesis to obtain the Master of Science Degree*, October 2013.
- [10] T. Georgiev et A. Lumsdaine. Focused plenoptic camera and rendering. *Journal of Electronic Imaging*, 19(2):021106, 2010.
- [11] M. Forman, A. Aggoun, et M. McCormick. Compression of integral 3D TV pictures. Dans *Fifth International Conference on Image Processing and its Applications*, pages 584–588, Edinburgh, UK, July 1995. IET.
- [12] M. C. Forman, A. Aggoun, et M. McCormick. A novel coding scheme for full parallax 3D-TV pictures. Dans *ICASSP Proceedings*, volume 4, pages 2945–2947, Nagoya, Japan, August 1997. IEEE.
- [13] E. Elharar, A. Stern, O. Hadar, et B. Javidi. A hybrid compression method for integral images using discrete wavelet transform and discrete cosine transform. *Journal of display technology*, 3(3):321–325, September 2007.
- [14] D. Marpe, T. Wiegand, et G. J. Sullivan. The H. 264/MPEG4 advanced video coding standard and its applications. *Communications Magazine*, 44(8):134–143, 2006.
- [15] G. J. Sullivan, J. Ohm, W.-J. Han, et T. Wiegand. Overview of the high efficiency video coding (HEVC) standard. *TCSVT*, 22(12):1649–1668, 2012.
- [16] J. Dick, H. Almeida, L. D. Soares, et P. Nunes. 3D holoscopic video coding using MVC. Dans *EUROCON*, pages 1–4, Lisbon, Portugal, April 2011. IEEE.
- [17] S. Shi, P. Gioia, et G. Madec. Efficient compression method for integral images using multi-view video coding. Dans *18th ICIP*, pages 137–140, Brussels, Belgium, September 2011. IEEE.
- [18] J.-R. Ohm. Overview of 3D video coding standardization. Dans *3DSA*, Osaka, Japan, June 2013.
- [19] C. Conti, P. Nunes, et L. D. Soares. New HEVC prediction modes for 3D holoscopic video coding. Dans *19th ICIP*, pages 1325–1328, Orlando, FL, September 2012. IEEE.
- [20] D.-K. Kwon et M. Budagavi. Fast intra block copy (intrabc) search for hevc screen content coding. Dans *ISCAS*, pages 9–12. IEEE, 2014.
- [21] C. Conti, P. Nunes, et L. D. Soares. Inter-layer prediction scheme for scalable 3-d holoscopic video coding. *Signal Processing Letters*, 20(8):819–822, 2013.
- [22] T. Georgiev, K. C. Zheng, B. Curless, D. Salesin, S. Nayar, et C. Intwala. Spatio-angular resolution tradeoffs in integral photography. *Rendering Techniques*, 2006:263–272, 2006.
- [23] F. Bossen. Test conditions and software reference configurations. *JCT-VC L1100*, January 2013.
- [24] G. Bjøntegaard. Calculation of average PSNR differences between RD-curves. Dans *VCEG Meeting*, Austin, USA, April 2001.