Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
وزارة التعليم العالي و البحث العلمي

BADJI MOKHTAR-ANNABA UNIVERSITY
UNIVERSITE BADJI MOKHTAR-ANNABA

جامعة باجي مختار- عنابة

Faculté des Sciences de l'Ingéniorat
Département d'Informatique

Année : 2013/2014

THESE

Présentée en vue de l'obtention du diplôme de ***DOCTORAT***

Analyse et compression de la vidéo multi-vues

Option
Informatique

Par
BEKHOUCH Amara

Directeur de Thèse: Pr. Noureddine DOGHMANE **Univ. Badji Mokhtar, Annaba**

Devant le Jury

Président:	**Pr. Med Tayeb LASKRI**	**Univ. Badji Mokhtar, Annaba**
Examinateurs:	**Pr. Lynda DIB**	**Univ. Badji Mokhtar, Annaba**
	Pr. Abdelmalik TALEB-AHMED	**Univ. Valenciennes, France**
	Pr. Mouldi BEDDA	**Univ. Aljouf, Arabie Saoudite**
	Pr. Mohammed KHAMADJA	**Univ. Constantine**

Année Universitaire : **2013/2014**

Remerciements

Je remercie en premier Le Bon Dieu, Le Tout Puissant,

Je tiens aussi à exprimer mes vifs remerciements à mon directeur de thèse, Monsieur Noureddine DOGHMANE, Professeur à l'Université de Annaba. Il m'a constamment soutenu, encouragé et stimulé pendant ce travail de doctorat. Ses nombreuses remarques ont montré une très vaste connaissance des sujets abordés et qui m'ont permis d'améliorer mon travail.

Je voudrais ensuite remercier sincèrement Monsieur Med Tayeb LASKRI, Professeur à l'Université de Annaba, d'avoir accepté de présider cette thèse.

Je tiens également à remercier vivement Melle Lynda DIB, Professeur à l'Université de Annaba et Monsieur Mouldi BEDDA, Professeur à l'Université de Aljouf (Arabie Saoudite), pour avoir accepté de juger ces travaux en tant qu'examinateurs. Mes vifs remerciements s'adressent aussi à Monsieur Abdelmalik TALEB-AHMED, Professeur à l'Université de Valenciennes (France) et Monsieur KHAMADJA Mohammed, Professeur à l'Université de Constantine, d'avoir bien voulu juger ce travail et pour l'intérêt qu'ils ont porté à ce travail.

Bien évidemment je remercie mes parents et ma famille pour leurs confiances et leurs soutiens sans failles pendant ces années de thèse, et pour m'avoir supporté pendant ces longues années d'études.

J'ai sans doute oublié plein de monde, toutes mes excuses, je sais que vous ne m'en tiendrez pas rigueur.

ملخص

مثل أي نموذج فيديو موجود، ضغط الفيديو المتعددة العرض يتطلب كفاءة عالية للضغط، فضلا عن وصول عشوائي أسرع إلى أية وجهة عرض معينة. ويمكن تحقيق هذه الكفاءة للضغط من خلال استغلال المعلومة المتكررة سواء كانت زمنية أو الموجودة في وجهات العرض المختلفة. إن الوصول العشوائي الزمني و العرضي يتم تحسينه من خلال إضافة ما يسمى بالصور المضغوطة «I » ، و هو الأمر الذي يتطلب معدل بث أكثر بقليل. في هذه الأطروحة، يتم اقتراح وتقييم طريقة تنبؤ جديدة للفيديو المتعددة العرض و التي تقوم أساسا على الحد من تعقيد مشاهدة الوصول العشوائي في أي وقت معين و تحسين استغلال المعلومة المتكررة في ما يسمى بالصور P. يمكن ضمان هذه التحسينات عن طريق الاختيار المناسب لموقع العرض I ، الاختيار المناسب كذلك لأنواع العروض المختلفة ، وعدد من كل نوع من الصور التي يتم استخدامها. بالمقارنة مع مخطط التنبؤ ، والذي يقوم على ما يسمى بالبنية الهرمية ذات الصور B و هو الذي يستخدم كنموذج ما يسمي ب النموذج المشترك للفيديو المتعدد العروض، وقد أظهرت النتائج التجريبية أن هيكل التنبؤ المقترح يوفر أداء أفضل للضغط حيث يصل تحسين معدل البث إلى 6.81 ٪ مع جودة ضغط مماثلة محسوبة بما يسمى ذروة نسبة الإشارة إلى الضوضاء و بشكل ملحوظ يحسن كذلك الوصول العشوائي بين العروض بنسبة تصل إلى 25 ٪ .

الكلمات المفتاحية: الوصول العشوائي بين العروض، كفاءة الضغط، المعلومة المتكررة الزمنية، المعلومة المتكررة بين وجهات العرض المختلفة.

Abstract

As any existing video codec, multi-view video coding requires a highly efficient compression as well as a faster random access to any given view. Such efficiency for the coding approach can be achieved by exploiting both the temporal and the inter-view correlation. The temporal and view random access are improved by the increase of Intra-coded pictures (I pictures), which requires more bit rate. In this thesis, we propose and evaluate an improved inter-view prediction structure mainly based on reducing the complexity of view random access at any instance Tn and improving the inter-view correlation for P-view. These improvements can be guaranteed by the appropriate choice of the position of I-view, the type of the different views, and the number of each frame type to use. Compared to the prediction scheme, which is based on the hierarchical B pictures structure and used for joint multi-view video model (JMVM), experimental results have shown that the proposed prediction structure provides better bit-rate performance of up to 6.81% with a similar compression quality measured in peak signal-to-noise ratio (PSNR) and significantly improves view random access by up to 25%.

Keywords: View random access, efficient compression, temporal correlation, inter-view correlation.

Résumé

Comme tout autre codec vidéo existant, le codage de la vidéo multi-vues nécessite une compression très efficace. Cependant, il a besoin en plus d'un accès aléatoire rapide à n'importe quelle vue donnée. Cette efficacité de la compression peut être satisfaite en exploitant simultanément la corrélation temporelle et inter-vues. Les accès aléatoires temporel et inter-vues sont améliorés par l'augmentation d'images de type *Intra*-codées (les images I). Ceci produit un débit binaire plus élevé. Dans cette thèse, nous proposons et nous évaluons une nouvelle structure de prédiction inter-vues basée principalement sur la réduction de la complexité de l'accès aléatoire inter-vues à tout instant Tn ainsi que l'amélioration de la corrélation inter-vues des vues de type P. Ces améliorations peuvent être garanties par le choix approprié de la position de la vue I, le type des différentes vues et le nombre de chaque type d'image à utiliser. Par comparaison avec le schéma de prédiction, basé sur la structure d'images B hiérarchique et utilisé par le modèle de référence « Joint Multi-view Video Model (JMVM) », les résultats expérimentaux ont montré que la structure de prédiction proposée améliore les performance en termes de débit binaire allant jusqu'à 6,81% avec une qualité de compression similaire mesurée en PSNR (Peak Signal-to-Noise Ratio). Une amélioration significativement de l'accès aléatoire inter-vues pouvant atteindre 25% est également obtenue.

Mots clés : Accès aléatoire inter-vues, efficacité de la compression, corrélation temporelle, corrélation inter-vues.

Liste des Tableaux

Liste des Figures

Liste des acronymes

AVC	Advanced Video Coding
CABAC	Context-based Adaptive Binary Arithmetic Coding
CATV	CAble TV
CAVLC	Context-based Adaptive Variable Length Coding
CIF	Common Intermediate Format
DCT	Discrete Cosine Transform
DPB	Decoded Picture Buffer
DVB	Digital Video Broadcasting
DVB-C	Digital Video Broadcasting-Cable
DVB-S	Digital Video Broadcasting - Satellite
DVB-T	Digital Video Broadcasting - Terrestrial
DVD	Digital Video Disc
FIMD	Fast Inter Mode Decision
FMD	Fast Mode Decision
FPS	Frame Per Second
FVV	Free View point Video
GGOP	Group of Group Of Pictures
GOP	Group Of Pictures
GOV	Group Of Views
HBP	Hierarchical B Pictures
HiP	High Profile
ICT	Integer Cosine Transform
IDR	Instantaneous Decoder Refresh
IEC	International Electrotechnical Commission
ISO	International Organisation for Standardization

ITU-T	International Telecommunication Union - Telecommunications Standard Sector
JMVM	Joint Multi-view Video model
JTC	Joint Technical Committee
JVT	Joint Video Team
MB	Macro-Block
MCMD	Mode Correlation-based Mode Decision
MPEG	Moving Pictures Experts Group
MSE	Mean Squared Error
MVC	Multi-view Video Coding
MVD	Multi-view Video-plus Depth
MVV	Multi-View Video
NAL	Network Abstraction Layer
NTSC	National Television System Committee
NUT	NAL Unit Type
PAL	Phase Alternating Line
PMV	Predicted Motion Vector
PSNR	Peak Signal-to-Noise Ratio
QCIF	Quart de CIF
QP	Quantization Parameter
RD	Rate-Distortion
RGB	Red, Green, Blue
RNIS	Réseau Numérique à Intégration de Services
RTC	Réseau Téléphonique Commuté
RTP	Real-time Transport Protocol
RVB	Rouge, Vert, Bleu
SAD	Sum of Absolute Differences
SD	Standard Definition
SI	Switching I pictures
SIF	Standard Interchange Format

SP	Switching P pictures
SPS	Sequence Parameter Set
SSD	Sum of Squared Differences
SVC	Scalable Video Coding
SVH	Système Visuel Humain
TVHD	TV High Definition
VCEG	Video Coding Experts Group
VCL	Video Coding Layer
VLC	Variable Length Coding

Table des Matières

2- Deuxième Chapitre
Compression de la vidéo multi-vues extension du H.264/AVC.

3- Troisième Chapitre
Les structures de prédiction du MVC.

Introduction générale

Avec le développement très rapide des technologies d'affichage, du traitement du signal, de la transmission et d'acquisition, les vidéos 3D avec leurs applications sont devenues l'alternative et l'extension légitimes de la vidéo 2D. Ces vidéos sont aujourd'hui utilisables par le grand public à travers plusieurs canaux tels que, l'internet (streaming et téléchargement), la transmission par câble et également par satellite ainsi que la radiodiffusion terrestre. La technologie clé pour ces différentes applications est la vidéo multi-vues (MVV; *Multi-View Video*).

La vidéo multi-vues

En effet, la vidéo multi-vues peut être considérée comme l'assemblage de plusieurs séquences vidéo conventionnelles capturées simultanément par plusieurs caméras, à partir de différents angles de vue. Plusieurs modes d'arrangement des différentes caméras d'acquisitions sont possibles. L'application envisagée ainsi que la qualité de la vidéo multi-vues requise (selon aussi le type des dispositifs d'affichage) influent directement sur la disposition de l'ensemble des caméras utilisées. La capture de la profondeur en même temps que la texture est aussi possible. Dans ce cas on parle de vidéo *Multi-view Video-plus Depth*, (MVD) [1][2]. Ces profondeurs sont utilisées et transmises comme des vidéos classiques utiles pour la connaissance de la géométrie de la scène capturée. La figure 01 illustre un exemple de système de capture de vidéo multi-vues.

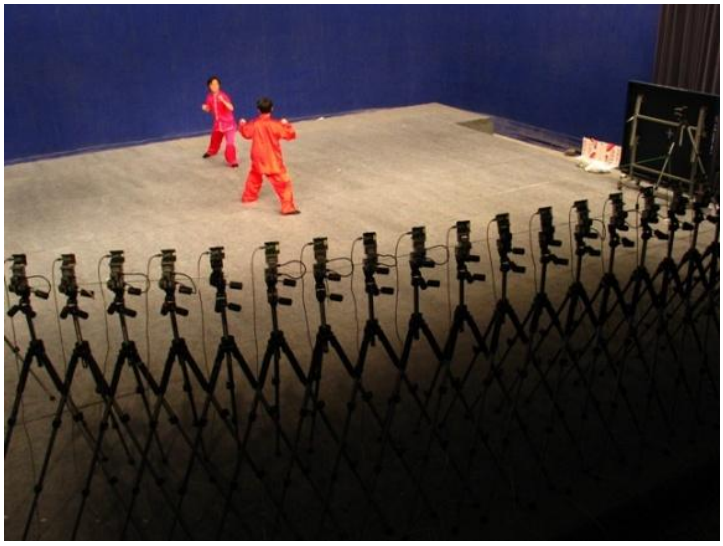


Figure 1. *Système de capture de la vidéo multi-vues [3].*

La vidéo capturée par chaque caméra peut être encodée par n'importe quel encodeur conventionnel tel que le H263, H264/AVC,…etc. Les séquences constituant la vidéo multi-vues doivent être codées par le même encodeur. Ces séquences sont indispensables pour les nouvelles applications multimédias telles que, la télévision 3D (3DTV) [4], la vidéo à point de vue libre (FVV, *Free View point Vidéo*), ce qui donne des impressions particulières plus réalistes de la scène visualisée [5]. Ceci permet, entre autres, de fournir aux téléspectateurs une vision stéréoscopique complètement nouvelle.

Applications pour la vidéo multi-vues

Contrairement à la vidéo classique (mono-vue), la vidéo multi-vues permet à l'utilisateur non seulement de regarder une scène à partir de l'angle de vue souhaité et ainsi de naviguer librement entre les vues, mais aussi de profiter des effets visuels spéciaux assurés par des dispositifs divergents. Ces applications sont faisables chacune par une architecture distincte.

Vidéo à point de vue libre

La vidéo à point de vue libre (FVV) est une application qui permet à l'utilisateur de contrôler de manière interactive un point de vue dans un espace 3D pour visionner une scène dynamique. Chaque utilisateur peut observer le point de vue désiré (un point de vue unique). La spécificité de la FVV est que les points de vue sélectionnés par les téléspectateurs ne correspondent pas seulement aux points de vue des caméras existantes. Dans la FVV, il y'a la possibilité de générer une nouvelle vue de la scène à partir de n'importe quelle position 3D ; c'est-à-dire de naviguer librement dans la scène de la vidéo 3D [4][5]. En effet, la génération d'une vue à partir de n'importe quelle position, nécessite l'utilisation de plusieurs caméras autour de la scène. L'alignement ainsi que la configuration des caméras utilisées, jouent un rôle primordial dans l'obtention d'une vidéo de bonne qualité.

Les applications qui se basent sur la vidéo à point de vue libre sont multiples. Un système de vidéo multi-vues permet par exemple à un entraineur d'évaluer mieux ses joueurs à travers plusieurs angles de vue.

La télévision 3D (3DTV)

La vidéo 3D avec ses applications peut être considérée parmi les plus importants usages da la vidéo multi-vues. Ces applications nécessitent la capture de la profondeur en même temps que la texture de la scène à visualiser [2]. C'est-à-dire qu'à partir de chaque angle de vue, une séquence de texture et également de profondeur doivent être obtenues. Plusieurs

techniques d'affichage 3D, telles que les systèmes stéréoscopiques [6] ou auto-stéréoscopiques [7] sont mises en œuvre. Les systèmes stéréoscopiques classiques, assurent une impression de profondeur d'une scène par des dispositifs spéciaux qui sont les lunettes. Dans les systèmes stéréoscopiques, seulement deux vues sont nécessaires ; l'une pour l'œil droit et l'autre pour l'œil gauche. Une autre technique de représentation d'image en relief mais sans l'intervention des dispositifs spéciaux pour la restitution de l'effet tridimensionnel, est le système auto-stéréoscopique. Dans ces systèmes de multiples techniques sont concevables, l'une des méthodes les plus utilisées actuellement par les grands fabricants, est celle basée sur un réseau lenticulaire [8]. Le principe de cette technique se base sur le placement d'un réseau de microlentilles sur la surface d'une image. Cette image est constituée elle-même d'images imbriquées obtenues chacune à partir d'un angle de vue différent. L'impression de relief est assurée par le réseau lenticulaire où chaque œil reçoit une image différente. Une autre méthode plus ancienne que le réseau de microlentilles est le système Auto-stéréoscopie à barrière de parallaxe [7]. Le principe est essentiellement le même que le réseau lenticulaire, la différence est que l'impression de relief est obtenue par une barrière qui distribue en alternance les points de vue destinés à l'un ou l'autre des deux yeux.

Compression de la vidéo multi-vues

En raison de la redondance très élevée de l'information contenue dans les différentes vues capturées simultanément ainsi que la taille considérable des informations traitées, la compression efficace de la vidéo multi-vues est alors indispensable. L'importance de la compression dans ce cas apparaît dans l'amélioration et l'optimisation du stockage et de la transmission de la vidéo sur les réseaux informatiques tels que l'internet. En fait, dans la compression de la vidéo de nombreuses contraintes sont exigées [9], à savoir :

- L'efficacité de la compression définie par un compromis débit binaire/qualité,
- L'accessibilité aléatoire temporelle sollicitée par tous les algorithmes de codage vidéo ainsi que l'accessibilité aléatoire aux différentes vues de la vidéo,
- La réduction possible de la complexité de l'algorithme de codage.

La compression de la vidéo multi-vues (MVC ; *Multi-view Video Coding*) consiste en premier lieu à générer par l'utilisation de N vues un seul flux binaire (*bitstream*). La deuxième étape est l'exploitation des différentes vues après avoir décodé le flux binaire envoyé par l'encodeur. La figure 2 illustre l'architecture générale de la compression de la vidéo multi-vues.

En effet, la compression de la vidéo multi-vues a été normalisée comme extension de la norme H.264/AVC par « The ISO/IEC JTC1/SC29/WG11 *Moving Pictures Experts Group* (MPEG) » [10][11]. Toutes les caractéristiques et les techniques de cette norme de compression vidéo sont utilisées par le MVC.

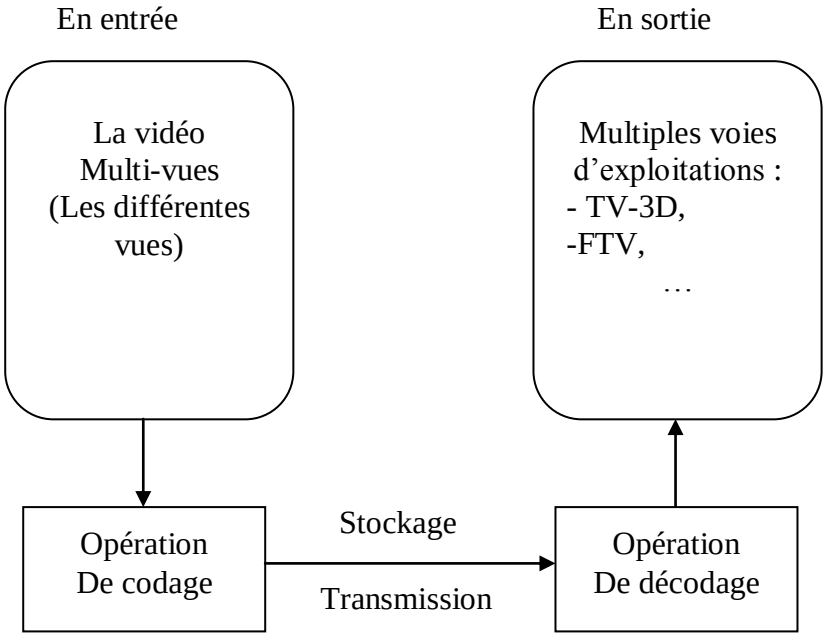


Figure 2. *Architecture générale du système MVC.*

Contribution de cette thèse

Ce travail de thèse est essentiellement orienté vers les problèmes liés généralement à la compression de la vidéo multi-vues et particulièrement à l'amélioration du débit binaire et de l'accès aléatoire inter-vues qui peuvent être considérés parmi les plus importantes exigences de la compression de la vidéo. Plus précisément, les principales contributions portent sur les points suivants :

- La principale contribution de cette thèse est d'avoir fourni une nouvelle structure de compression de la vidéo multi-vues afin d'améliorer d'une part le débit binaire tout en assurant une meilleur qualité de la vidéo et d'autre part d'accélérer l'accès aléatoire inter-vues avec une efficacité de compression meilleure que celle obtenue par le schéma implémenté dans *Joint Multi-view Video model* (JMVM) [12] développé par Joint Video Team (JVT) de ISO/IEC, *Moving Picture Experts Group* et ITU-T, *Video Coding Experts Group* (VCEG).

- La proposition d'une nouvelle méthode d'évaluation de l'accès aléatoire inter-vues pour le schéma implémenté dans le modèle de référence JMVM et également pour les schémas proposés afin de pouvoir comparer les résultats obtenus. Nous nous sommes basés dans notre évaluation sur deux métriques à savoir :
    - Le nombre d'images de référence à décoder Nbr_{img}, nécessaires à l'accès à une image donnée dans les différentes vues,
    - Le nombre maximal N_{max} pour accéder à une image donnée.

Organisation de la thèse

Outre cette introduction qui fait office à la fois de motivation et de présentation générale du problème, le manuscrit se compose de Cinq chapitres organisés comme suit :

Dans le chapitre 1, après avoir introduit quelques généralités sur la compression de la vidéo et plus particulièrement la norme H.264/AVC sur lesquelles se base le MVC, nous présenterons un rapide panorama des aspects techniques du standard afin d'éclaircir les spécificités de la norme. Dans le but de désigner la position de le MVC par rapport au H264/AVC, nous étudierons ensuite les profils et niveaux du H.264/AVC. Enfin, ce chapitre se termine par une évaluation de l'efficacité de la norme H.264/AVC plus particulièrement en termes d'efficacité de la compression (mesurée par un compromis débit binaire/qualité de la vidéo).

Nous dresserons dans le chapitre 02, les concepts de base de la compression de la vidéo multi-vues à savoir les contraintes à satisfaire, les conditions et les données de test à vérifier. La plus parts des algorithmes de compression visent à améliorer soit le débit binaire ou la diminution de la complexité de l'encodage afin de réduire la vitesse d'exécution ou d'optimiser la consommation de l'énergie soit d'accélérer l'accès aléatoire inter-vues. Nous introduisons également dans ce chapitre quelques notions distinctives du MVC. En plus de la prédiction temporelle ainsi que la compensation de mouvement le MVC profite aussi de la prédiction inter-vues et de la compensation de disparité. Nous parlerons ensuite du modèle de référence JMVM développé conjointement par JVT de ISO/IEC, MPEG et VCEG.

Le chapitre 03 est consacré à L'étude des structures de prédiction inter-vues. Toutes les structures de prédiction utilisées dans le MVC, se basent au niveau temporel sur le modèle B hiérarchique. Une analyse de l'architecture du flux binaire est présentée. Nous exposons également dans ce chapitre les méthodes d'évaluation des schémas de compression de la

vidéo multi-vues existantes en termes d'efficacité de compression et d'accès aléatoire inter-vues. Enfin, nous terminerons ce chapitre par une comparaison entre les différentes structures de prédiction abordées.

Dans le chapitre 04, nous aborderons l'approche proposée pour améliorer l'accès aléatoire inter-vues et également le débit binaire. La plupart des structures de prédiction utilisent un modèle à huit caméras de capture (huit vues) car la structure par défaut utilisée et implémentée dans le modèle de référence JMVM utilise huit vues. Après avoir détaillé le modèle à 8 caméras proposé, nous présenterons le modèle choisi dans le cas général (plus de huit caméras). Les méthodes d'évaluation concernent les structures de prédiction du modèle JMVM et aussi la structure proposée qui seront bien évidemment présentées. Enfin, pour compléter l'étude, nous examinerons les résultats sur les données de test présentées dans le chapitre 02.

Le chapitre 05 revient sur l'optimisation du modèle proposé dans le but d'améliorer précisément l'accès aléatoire inter-vues tout en fournissant un compromis débit binaire/accès aléatoire inter-vues. L'accès aléatoire inter-vues est assuré dans cette structure par l'élimination de prédiction inter-vues pour *non-anchors* image (les images non clef dans le modèle) dans quelques cas. La généralisation et l'évaluation de cette méthode sont ensuite détaillées. Enfin, ce chapitre se termine par la validation de cette méthode à travers l'exposition des résultats obtenus.

Chapitre 01 : Compression de la vidéo et le H.264/AVC

1.1. Introduction

L'augmentation de la quantité des informations (particulièrement l'image et la vidéo) transmises à travers la diversité des services de communications ne cesse de croître. Cette augmentation est due principalement aux progrès remarquables de l'informatique et de l'électronique en termes de transmission, stockage et traitement des données numériques (audio, image, vidéo,…etc). L'amélioration intéressante des bandes passantes des réseaux de télécommunications actuels ainsi que l'augmentation incessante des supports de stockage, restent limitées vis-à-vis de la grande masse des données à transmettre ou à stocker. Plusieurs standards et algorithmes de compression da la vidéo numérique sont proposés dans la littérature afin de pallier ce problème, chacun est destiné à une ou plusieurs applications ou services multimédias.

Plusieurs systèmes de compression vidéo tel que MPEG-1, MPEG-2, H.261, H.262 et H.264/AVC ont été normalisés par les deux organismes de normalisation les plus connus qui sont, l'ISO/IEC (*International Organisation for Standardization - International Electrotechnical Commission*) et l'ITU-T (*International Telecommunication Union - Telecommunications Standard Sector*). Le H.264/AVC est le résultat d'un projet collaboratif entre le Groupe d'experts en codage vidéo (VCEG) de l'ITU-T et le Groupe MPEG de l'ISO/IEC. Cette norme est conçue essentiellement pour améliorer les standards de compression vidéo antérieurs. En effet, elle permet de réduire jusqu'à 50% du débit par rapport aux autres standards.

Nous allons présenter tout au long de ce chapitre, certaines notions de base en compression vidéo. Ensuite, nous aborderons les principaux standards de compression vidéo. Nous mettons surtout le point sur les différentes évolutions que ces standards ont connus. Nous nous intéresserons plus particulièrement à la norme H264/AVC et aux techniques qu'elle emploie.

1.2. Notions de base en compression de la vidéo

1.2.1. Quelques généralités

Il est évident que la vidéo numérique, se caractérise et se distingue par sa résolution spatiale (ou *Intra*-image) et sa résolution temporelle (ou *Inter*-image). La résolution temporelle est caractérisée par la cadence de l'animation, cette cadence est exprimée elle-même par le nombre d'images par seconde (fps ; Frame Per Second). La résolution temporelle

doit être toujours supérieure à 25 fps du fait que l'œil humain, pour un observateur moyen, est capable de distinguer jusqu'à 20 images par seconde.

La résolution spatiale est définie par le nombre de lignes (nombre de pixels sur l'axe vertical) multiplié par le nombre de colonnes (nombre de pixels sur l'axe horizontal) dans chaque image. En fonction de l'application ou du service visé, plusieurs résolutions et formats sont utilisés. La grande majorité des formats est dérivée soit de standards vieillissants NTSC (*National Television System Committee* ; 720 x 480), PAL (*Phase Alternating Line* ; 720 x 576) soit de la télévision numérique haute définition (TVHD), les formats les plus importants sont regroupés comme suit:

- SIF (*Standard Interchange Format* ; 352x240) dérivés du NTSC avec une résolution temporelle de 30 fps.
- CIF (*Common Intermediate Format* ;352x288), dérivés du PAL avec une résolution temporelle de 25/30 fps.
- QCIF (*Quart de CIF* ;176x144), dérivés du PAL avec une résolution temporelle de 25/30 fps.
- Formats DVD dérivés du PAL et du NTSC (NTSC : 720 × 480, PAL: 720×576), ces formats sont appelés également SD (*Standard Definition*).
- Formats Full HD (1920×1080) et HD (1280×720) dérivés de la TVHD.

Une autre notion importante dans la représentation, le codage et la diffusion de la vidéo est la représentation de la couleur dans l'image. Cette représentation est exprimée généralement par le nombre de bits par pixel. Le mode de représentation le plus connu est le RVB (RGB en anglais) où chaque pixel est codé sur 3 octets (24 bits) dont chaque composante de couleur (R : *Red,* G *:green* et B: *blue*) est codée sur un octet soient 16777216 possibilités de couleurs différentes. Néanmoins, cet espace de couleur présente l'inconvénient d'être moins adapté au système visuel humain (SVH) par rapport aux espaces de couleurs de type luminance/chrominance. La redondance de l'information contenue dans les trois composantes du système RGB est l'un des inconvénients majeurs de cet espace de représentation de la couleur.

La majorité des standards existants utilisent un autre espace de couleur plus approprié, de type luminance/chrominance, qui est YCbCr (trois composantes, chacune est codées sur 8 bits). L'espace chromatique YCbCr utilise une composante de luminance, contient la

luminance de l'image (les niveaux de gris) et deux composantes de chrominance Cb et Cr comportent des informations sur les couleurs de l'image comme le montre la figure 1.1. Les canaux Y, Cb et Cr sont obtenus à partir des composantes R, G et B à l'aide des formules suivantes :

$$Y = 0.299 * R + 0.587 * G + 0.114 * B \quad (1.1)$$

$$Y = -0.1687\text{xR} - 0.3313\text{xG} + 0.5\text{B} + 128 \quad (1.2)$$

$$Y = \text{Cr} = 0.5\text{xR} - 0.4187\text{xG} - 0.0813\text{xB} + 128 \quad (1.3)$$

(a)

(b) **(c)** **(d)**

(e) **(f)** **(g)**

Figure 1.1. *Modes de représentation de la couleur, (a)l'image originale, (b) la composante R, (c) la composante G, (d) la composante B, (e) la composante de luminance Y, (f) la composante de chrominance Cb, (g) la composante de chrominance Cr.*

La compression d'images et de la vidéo, doit appliquer une opération de sous-échantillonnage avant de procéder au codage de l'information. Cette opération consiste à atténuer la résolution spatiale de l'image en diminuant le nombre de lignes et/ou de colonnes et ainsi, la quantité d'informations à coder. Tous les standards de compression qui utilisent un espace chromatique, favorisent le sous-échantillonnage des composantes de chrominance. Ce choix est motivé par la sensibilité réduite de l'œil humain aux variations de chrominances (Cb, Cr) qu'aux variations de luminance (Y).

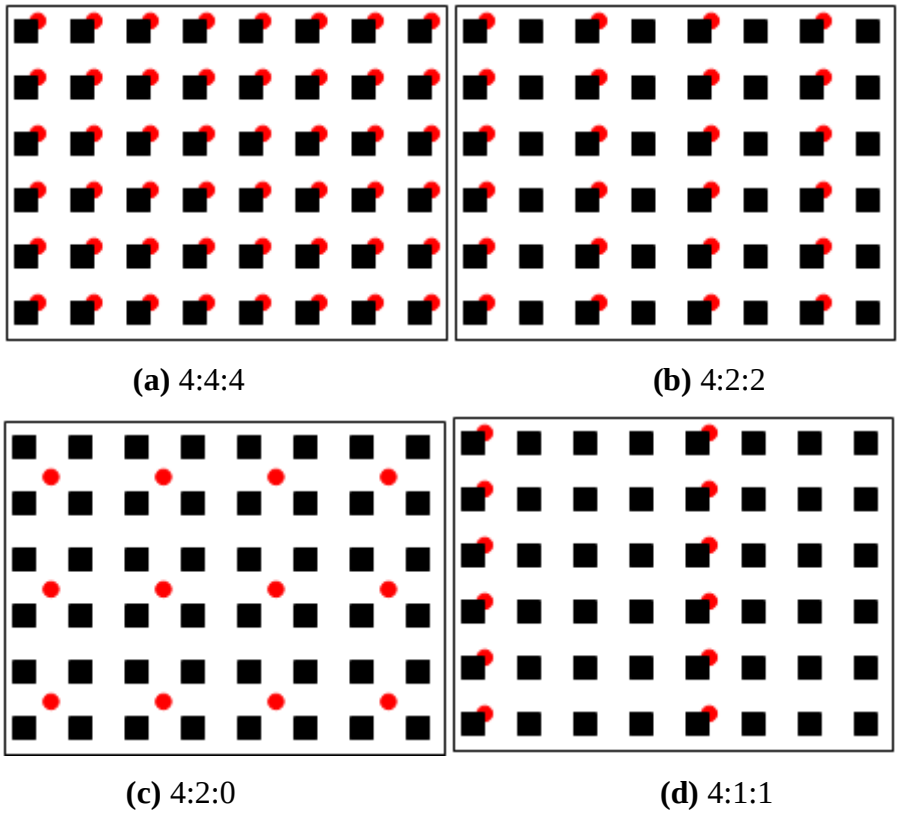
Plusieurs modes de sous-échantillonnage sont utilisés dans la littérature dont les plus importants sont exposés sur la figure 1.2. Le cas du sous- échantillonnage représenté par 4 : 4 : 4, il garde l'information entière de l'image (le nombre de lignes et de colonnes et le même pour les trois composantes). Dans le mode 4 :2 :2 une étape de décimation est nécessaire, le nombre de colonnes est réduit d'un facteur de ½ dans les deux composantes de chrominance, cette diminution est accomplie par le calcul de la valeur moyenne de deux pixels voisins horizontalement. Les deux modes de sous-échantillonnage 4 : 1 : 1 et 4 : 2 : 0 permettent d'amoindrir considérablement la teinte où un pixel sur quatre est maintenu. Dans ce cas, chaque pixel est obtenu par la moyenne des quatre pixels voisins (verticalement et/ou horizontalement en fonction du mode utilisé). Les modes les plus utilisés en compression de la vidéo sont les deux derniers.

Les deux composantes de chrominance sont sur-échantillonnées lors du décodage des images afin de restituer la résolution spatiale originale.

L'exploitation des corrélations d'informations d'une manière efficace, et l'une des techniques clés dans un système de compression de vidéos. L'élimination des redondances des informations dans la vidéo source autant que possible avant leur transmission, permet la réduction opérante du débit binaire dans le flux. Dans la compression de la vidéo, on distingue plusieurs types de redondances:

- La redondance psycho-visuelle : Ce type de codage se base principalement sur la sensibilité de l'œil humain aux différentes variations des signaux. Il s'agit d'éliminer des informations non pertinentes par rapport au système visuel humain.
- La redondance spatiale : Le codage spatial est basé sur l'utilisation de la similarité entre les pixels voisins dans chaque image prise indépendamment des autres. Les redondances spatiales peuvent être diminuées en codant seulement les différences de valeurs entre les pixels consécutifs.

- La redondance temporelle : Les valeurs des pixels de l'image t+1 sont similaires aux valeurs des pixels de l'image t. Dans ce cas, dans l'image t+1, seulement les différences de valeurs de pixels corrélés ainsi que ses positions sont transmises.
- La redondance entropique : C'est le codage sans perte dans le processus de compression d'image/vidéo. L'une des approches les plus simples est d'attribuer aux valeurs les plus fréquentes dans un signal numérique des codes plus courts.



(a) 4:4:4 **(b)** 4:2:2

(c) 4:2:0 **(d)** 4:1:1

Figure 1.2. *Les modes de sous-échantillonnage dans un espace chromatique, la luminance est représentée par des rectangles et la chrominance par des cercles.*

1.2.2. La normalisation des standards de compression vidéo

Les organismes internationaux de normalisation ISO/IEC [10] et ITU-T [9], ont partagé au cours de ces dernières décennies, la normalisation des standards de compression vidéo. Le principe de base utilisé est le même pour tous les codeurs normalisés par les deux organismes. Néanmoins, l'ITU-T vise les applications à faible débit, à titre d'exemple les applications de visioconférence et de visiophonie tandis que l'ISO/IEC s'oriente principalement vers les applications de diffusion et d'archivage qui nécessitent un débit plus élevé. Dans cette section nous décrivons les principaux standards normalisés à l'ITU-T et à l'ISO/IEC.

Les normes de l'ITU-T

La première plateforme commode de codage et de décodage des services audiovisuels numériques était la norme **H.261** [13] (de 1988 à 1993). La norme H.261 est conçue principalement pour permettre la transmission des applications de vidéophonie et vidéoconférence pour le réseau RNIS enrobant les débits de 64 Kbit/s à 2 Mbit/s. Cette norme supporte les formats d'images CIF (288x352 pixels) pour la luminance et QCIF (144x176 pixels) pour les canaux chromatiques. L'espace de couleur utilisé dans cette norme est le YCbCr (un espace de type luminance chrominance) avec un sous-échantillonnage de 4 :2 :0. Le codage utilisé est de type hybride où la redondance spatiale est diminuée par l'intervention, entre autres, d'une transformation en cosinus discrète (DCT) de l'image. Quand à la redondance temporelle, elle est minimisée par les techniques de prédiction, estimation et compensation du mouvement. Dans les deux cas, le traitement s'accomplit par la décomposition d'images en macro-block (MB) comme le montre la figure 1.3. Autrement dit, la DCT et les techniques de prédiction et d'estimation de mouvement sont appliquées toutes les deux sur des MB.

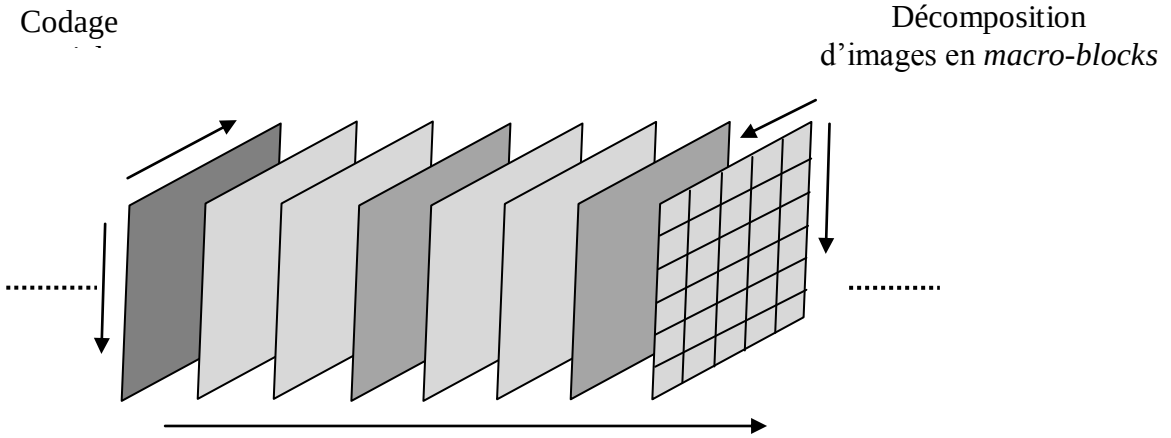


Figure 1.3. *Les modes de traitement possibles (analyse spatiale et temporelle) et décomposition d'images en macro-blocks.*

Le standard H.261 utilise des MB de 16x16 pixels pour la luminance et de 8x8 pixels pour les composantes de chrominance. La qualité de la vidéo est toujours contrôlée par la quantification des coefficients fréquentiels obtenus après DCT (compromis débit binaire/qualité vidéo). La perte d'information lors de l'opération de compression est due particulièrement à cette étape. Une amélioration du débit binaire sans perte d'informations

(élimination de la redondance entropique) est assurée également par le codage entropique appliqué après un balayage en zig-zag des MB.

L'opération de décodage intégrée dans cette norme consiste à appliquer les traitements inverses. La majorité des standards de compression vidéo (MPEG-1, MPEG-2/H.262, H.263, …) s'appuie étroitement sur cette norme. Les techniques décrites dans cette partie sont détaillées tout au long de ce chapitre.

H.263 [14] (normalisée en 1995) est une norme de codage vidéo qui se base sur les mêmes principes utilisés par la norme H.261. Les applications envisagées par ce standard sont la visiophonie et la visioconférence sur les réseaux RTC (Réseau Téléphonique Commuté) et RNIS. Cette norme peut être appliquée aussi sur Internet dans des applications de *streaming video*. Les principales améliorations et modifications apportées à la norme H.263 par rapport à H.261 sont :

- Outre, des formats d'images CIF et QCIF utilisés dans le standard H.261, le 4CIF (576x704 pixels) et le 16CIF (1152x1408 pixels) sont également adoptés dans le H.263.
- Minimisation de l'erreur de prédiction par une compensation de mouvement à base de prédiction au demi-pixel.
- Intégration d'un codage arithmétique (pour le codage entropique) avec le codage de Huffman utilisé par le H.261.
- Réduction de la taille des MB utilisées pour la prédiction à 8x8 pixels afin d'améliore l'estimation et la compensation de mouvement.
- Intégration d'un nouveau type d'image PB qui réunit les deux types d'images P et B.

Le standard H.261 a été encore amélioré dans les projets connus sous le nom de **H.263+** [15] (1998). L'organisation internationale de normalisation ITU-T, cherche toujours l'amélioration des standards développés en respectant plusieurs contraintes. L'objectif principal est constamment la diminution du débit binaire avec une meilleure qualité de la vidéo en sortie. Plusieurs options de codage ont été ajoutées dans cette norme, dont les plus importantes, celles qui améliorent appréciablement la qualité avec un débit binaire important ainsi que la robustesse face aux erreurs.

La spécificité majeure da la norme **H.263++** [16] (2000) extension de la norme H.263+, est la possibilité de la transmission vidéo en temps réel sur des réseaux à qualité de service non garantie, qui exigent un très faible débit.

Les normes de l'ISO/IEC

Le premier standard permettant le traitement et la compression de la vidéo numérique du groupe MPEG, a été normalisé par l'ISO/IEC en 1993. Le standard **MPEG-1** [17] a été fondé sur les principes de base du H.261 dont il améliore quelques parties comme par exemple la compensation de mouvement, la quantification et le type d'images utilisées. La qualité visée par cette norme est équivalente au VHS pour un débit typique de l'ordre de 1.5 Mbps en résolution SIF ou CIF. L'espace de couleur utilisé est de type luminance/chrominance (YUV) avec un sous- échantillonnage de type 4 :2 :0. Néanmoins, il est possible d'adapter le débit ainsi que la résolution. La fréquence d'image est limitée à 25 images par seconde pour PAL et à 30 images par seconde dans le cas de NTSC.

Plusieurs parties ont été définies pour cette norme, comme suit :

1. Synchronisation et multiplexage de la vidéo et de l'acoustique (1993).
2. Codage des flux vidéo (1993).
3. Codage des flux audio (le format audio MP3 est défini dans la troisième couche de cette partie) (1993).
4. Vérification de la conformité et des flux (1995).
5. produit de référence (1998).

Avec l'évolution très rapide des technologies numériques, les limites de la norme MPEG-1 sont apparues. Ainsi, une nouvelle norme qui se base sur les principes de MPEG-1 a été rapidement développée sous le nom **MPEG-2** (Proposée en 1994) [18]. Les améliorations principales adoptées dans ce standard visent les applications de télévision numérique. On peut citer à titre d'exemple, le traitement des formats entrelacés et la compatibilité entre la TV et la TVHD. Deux modes de sous échantillonnage sont utilisés 4 :2 :2 et 4 :2 :0 avec une fréquence d'images qui peut aller jusqu'à 60Hz (format entrelacé) et un débit de l'ordre de 15Mbps. Ces options couvrent les besoins de la télévision haute définition (HDTV), la télévision par câble (CATV), la vidéo numérique de qualité supérieure (DVD) et les systèmes de diffusion de télévision numérique du consortium DVB (par câble DVB-C par satellite DVB-S et diffusion hertzienne ou terrestre DVB-T). Neuf parties ont été

définies dans cette norme. La norme MPEG-2 a été supportée et intégrée par l'ITU-T sous le nom H.262.

MPEG-4 (standardisée en 1998) [19] est une évolution consistante vis-à-vis de la norme MPEG-2. Cette norme de codage vidéo peut être considérée comme la version de codage générique la plus complète dès sa conception. Le traitement et la représentation des contenus audio-visuels sous forme d'objet média est un concept complètement nouveau ajouté dans ce standard. Une grande variété de formats en termes de taille, résolution et fréquence sont possibles dans cette norme. Néanmoins, le plus important est la malléabilité de la manipulation des objets médias primitifs hiérarchiquement regroupés, tels que les vidéos, les images fixes, les sons, le son synthétique et d'autres objets. Le nombre d'applications possible par MPEG-4 est donc plus grand que dans MPEG-2 et H.263 de l'ITU-T (audiovisuelle, visioconférence, streaming sur Internet,…). D'autres modes sont également efficacement améliorés comme l'évolutivité (*scalability*) qui permet d'adapter le flux vidéo aux limitations de débit du canal de transmission et aussi la robustesse aux erreurs.

1.2.3. La norme H.264/AVC

Après le succès précieux de la norme MPEG-4 Part 2 créée par le group MPEG de l'ISO/IEC, l'ITU-T a lancé le projet H.26L (1998) [20] principalement pour doubler l'efficacité de compression par rapport aux normes existantes et de l'adapter à un nombre plus grand d'applications, services et protocoles de transport. Après avoir atteint les objectifs soulignés par H.26L, le Groupe VCEG (*Video Coding Experts Group*) de l'ITU-T et le groupe MPEG ont créé conjointement le groupe JVT (*Joint Video Coding*) en 2001 afin de réunir les expériences et de développer un standard plus avancé. Ainsi, un système de codage appelé AVC (*Advanced Video Coding*) a été développé et finalisé en 2003. L'AVC est intégré comme dernière partie au standard MPEG-4 (MPEG-4 Part 10) tandis que l'ITU-T a choisi l'appellation H.264 pour cette norme [21]. En effet, H.264, MPEG-4 Part 10/AVC, H.264/AVC (ou AVC/H.264), H.264/MPEG-4 AVC (ou MPEG-4/H.264 AVC) désignent pareillement la même norme.

Les avantages les plus importants offerts par H.264/AVC par rapport aux différents standards existants (MPEG-2, H.263, MPEG-4 Part 2,…) sont les suivants :

- L'adaptabilité du flux vidéo à une très grande variété de systèmes et de réseaux (stockage HD DVD et Blu-ray, le streaming RTP/IP, les systèmes de téléphonie,…etc),
- L'augmentation de la robustesse face aux différentes erreurs de transmission sur les divers types de réseaux utilisés,
- La réduction moyenne de 50% du débit binaire par rapport à MPEG-4 Part 2 avec une qualité similaire de la vidéo,
- L'affaiblissement du retard d'encodage, nécessaire aux applications en temps réel (le streaming RTP/IP, vidéoconférence,…).
- L'élimination du maximum d'artéfacts visuels par les techniques du «*deblockings filtres*».

Ces caractéristiques avantageuses sont assurées par l'intégration et l'amélioration de plusieurs techniques telles que par exemple :

- L'utilisation d'un pas de quantification hiérarchiquement variable selon le type et le contenu de l'image à encoder.
- La prédiction *Intra* (prédiction spatiale) plus performante à travers l'augmentation du mode de prédiction (13 mode pour la luminance et 4 pour la chrominance).
- L'atténuation de la précision de la recherche pour l'estimation de mouvement (1/4 de pixel).

Bien que cette liste ne soit exhaustive, plusieurs autres techniques sont également adoptées pour améliorer la norme H.264/AVC.

1.3. Étude approfondie du standard H.264/AVC

Comme c'est le cas pour tous les standards de codage vidéo, la norme H.264/AVC [22] ne désigne que la syntaxe du décodeur en imposant des restrictions sur le flux binaire; elle ne définit pas donc le processus de codage. Cette contrainte permet un maximum de liberté d'optimiser et d'améliorer les implémentations en termes d'efficacité de compression, robustesse aux erreurs, réduction de la complexité calculatoire,... Nous détaillons dans cette section les techniques d'encodages les plus importantes.

1.3.1. Architecture générale de la norme H.264/AVC

L'évolution des normes de codage vidéo s'articule principalement autour de l'efficacité maximale du codage ainsi que la flexibilité du flux vidéo qui permet d'adapter la syntaxe à un nombre maximum d'applications et d'environnements réseau. La cohérence des deux contraintes est soutenue dans le standard H.264/AVC (abordé également dans la plus parts des normes antérieures) par l'utilisation de deux couches conceptuelles. La figure 1.4 illustre la liaison entre les différentes couches de la norme H.264/AVC.

La représentation efficace du contenu vidéo est synthétisée dans une couche appelée VCL (*Video Coding Layer*). Les informations compressées fournies par la couche VCL sont représentées et regroupées par une couche appelée NAL (Network Abstraction Layer) d'une manière appropriée aux systèmes de transmission qui sont assurés par la couche de transport (voir la figure 1.4). La couche NAL fournit aussi des informations non-VCL comme par exemple, les paramètres d'images et de séquence vidéo, les paramètres d'affichages, …etc.

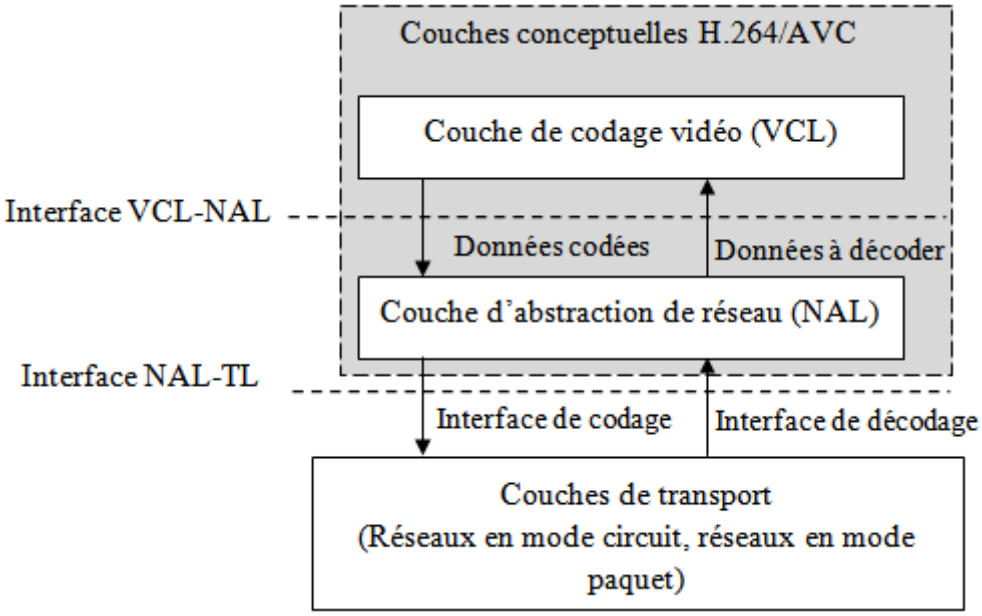


Figure 1.4. *Les deux couches VCL et NAL de la norme H.264/AVC.*

Couche d'abstraction de réseau (NAL)

Afin de répondre aux besoins des différentes applications et environnement réseau, le contenu vidéo encodé au niveau de la couche VCL est organisé dans un format spécifique par la couche NAL. Cette personnalisation des données encodées est accomplie à l'aide de plusieurs concepts dont les plus importants sont (ces concepts sont détaillés dans [23][24]) :

Les unités NAL : Les données encodées sont regroupées en plusieurs unités d'octets (un nombre entier d'octets), chacune d'elle est divisée en trois parties comme suit (voir la figure 1.5.):

- Un octet d'en-tête (header) qui contient une indication du type de données dans cette unité NAL.
- Un champ pour les informations encodées (*Payload*), représente l'information utile transmise par la couche VCL.
- Un champ de queue (*trailing*) de quelques bits (de 1 à 8 bits) permettant d'arranger la taille de l'unité à un nombre entier d'octets.

Les unités NAL Non-VCL : Ces unités contiennent généralement des informations supplémentaires utiles pour l'organisation des unités de type VCL tels que par exemple les paramètres d'images, les informations de synchronisation, …etc.

Les unités d'accès : Chaque unité d'accès contient plusieurs unités NAL de type VLC, les informations organisées dans une unité d'accès peuvent servir au décodage d'une seule image. Ces unités sont délimitées par les informations contenues dans les unités NAL non-VCL.

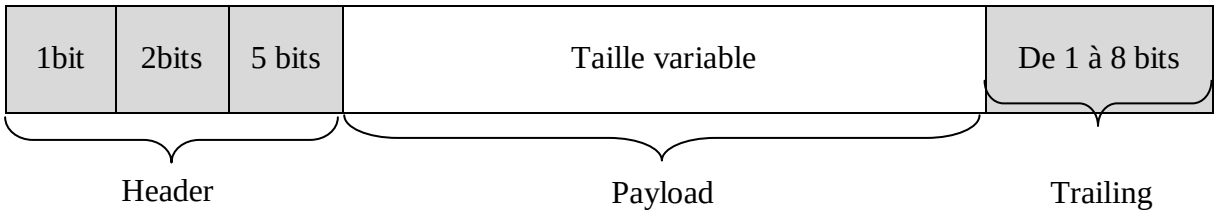


Figure 1.5. *Structure d'une unité NAL (unité non-VCL).*

Couche de codage vidéo (VCL)

Les différentes techniques de codage du signal vidéo en entrée sont appliquées au niveau de la couche VCL. La séquence vidéo est divisée en plusieurs groupes d'images appelés GOP (*group of pictures*), où chaque groupe et divisé en plusieurs images. Le choix de la taille du GOP utilisé dépend de certains facteurs tels que la fréquence trame de la vidéo à encoder. La norme H.264/AVC utilise trois types d'images pour la constitution de chaque GOP qui sont les images I codées indépendamment des autres images du GOP, les images P et B codées par une prédiction temporelle. En effet, la norme H.264/AVC utilise entre autres deux autres types d'images qui sont SI et SP exploitées généralement pour communiquer entre deux flux

différents. La figure 1.6 présente un exemple d'un GOP composé des trois types d'images I, P et B.

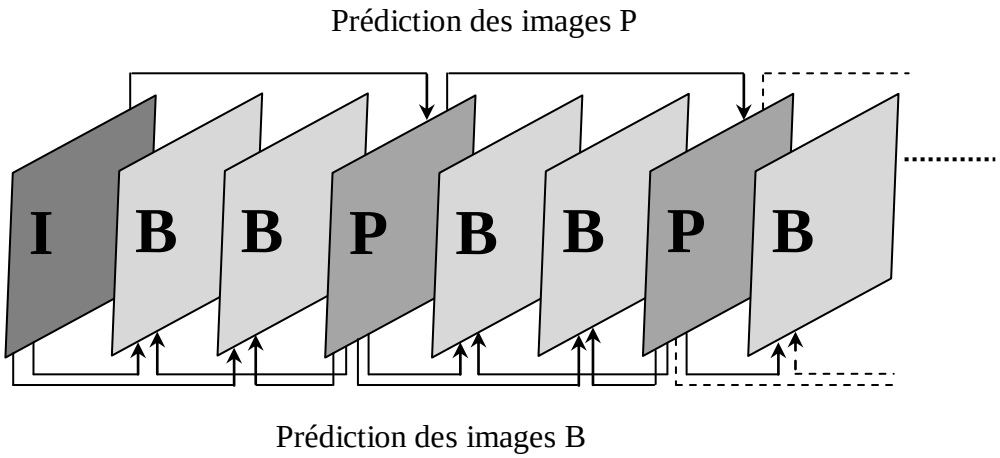


Figure 1.6. *Structure d'un GOP et illustration des différentes dépendances possibles.*

Chaque image est divisée en plusieurs tranches (*slices*) où chaque slice est divisé lui-même en macro-blocs de 16x16 pour la composante de luminance et de 8x8 pour les deux composantes de chrominances. La figure 1.7 illustre un exemple de décomposition de l'image en slices et macro-blocs. Le codage des macro-blocs dépend du type de l'image utilisée chaque fois (I, P ou B) à travers deux modes de prédiction : mode *Intra* ou mode *Inter*. Il s'agit dans le mode *Intra* d'éliminer la redondance spatiale dans l'image à encoder seulement par l'utilisation des informations de l'image elle-même (sans faire intervenir les autres images du GOP). Le mode *Inter* est utilisé pour les deux types P et B comme le montre la figure 1.6, la prédiction *Inter* consiste à éliminer la redondance temporelle entre les images du même GOP.

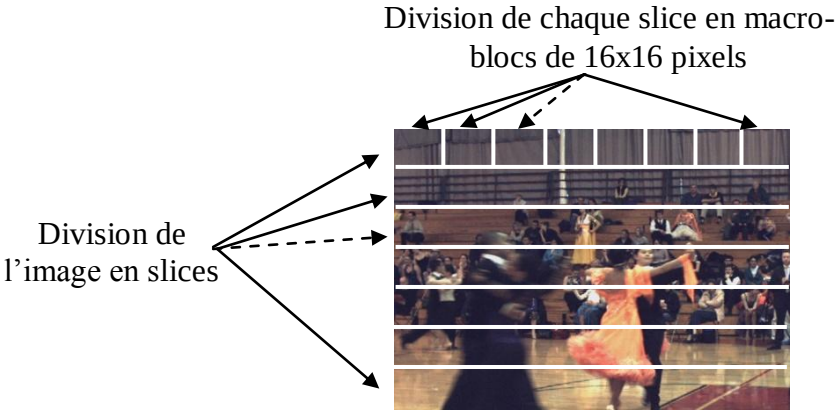


Figure 1.7. *Décomposition de l'image en plusieurs slices.*

Les deux modes de prédiction *Intra* et *Inter* s'appliquent sur les macro-blocs constituant l'image courante. Le principe de base de l'élimination de la redondance temporelle repose en particulier sur la prédiction de chaque macro-bloc à partir des images de références précédemment codées en utilisant les techniques d'estimation et de compensation de mouvement. L'étape de l'estimation de mouvement associant à chaque macro-bloc un vecteur de mouvement qui contient la position du bloc utilisée dans l'image de référence. Généralement les images P utilisent une seule image de référence pour la prédiction tandis que les images B utilisent chaque fois deux images de références (voir la figure 1.6).

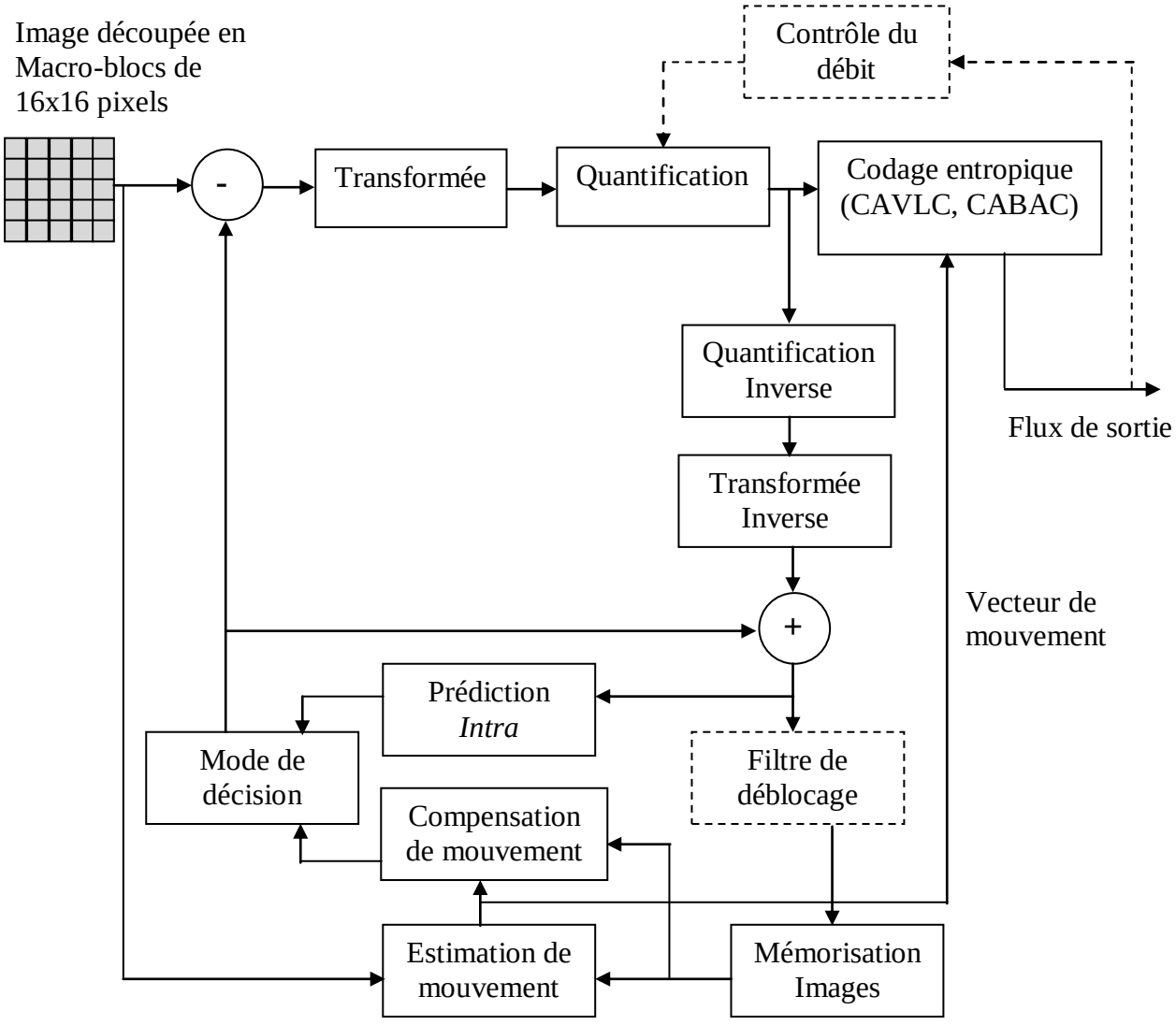


Figure 1.8. *Principe du codage d'un Macro-bloc dans la norme H.264/AVC.*

Il est plus intéressant d'utiliser la différence entre l'image à coder et les images de référence, au lieu d'utiliser les données de l'image ce qui permet d'assurer un taux de

compression élevé. Les données résiduelles provenant de cette différence sont ensuite transformées (par une DCT), quantifiées et encodées par le biais d'un codage entropique. La figure 1.8 illustre le processus de codage de la norme H.264/AVC pour un macro-bloc. Dans la suite, nous détaillerons les caractéristiques techniques les plus importantes de la norme H.264/AVC.

1.3.2 Prédiction Intra

Dans ce mode de prédiction [25] seuls les blocs adjacents (dans l'image courante) du macro-bloc à encoder sont utilisés pour la prédiction. Un macro-bloc est prédit à partir des blocs déjà encodés, décodés est mémorisés (voir la figure 1.9). Le mode *Intra* de la norme H.264/AVC se distingue de tel autre par la prédiction des macro-blocs dans le domaine spatial (les autres normes utilisent le domaine transformé pour la prédiction). Les macro-blocs de la composante de luminance dans ce mode sont prédits à travers deux classes de codage :

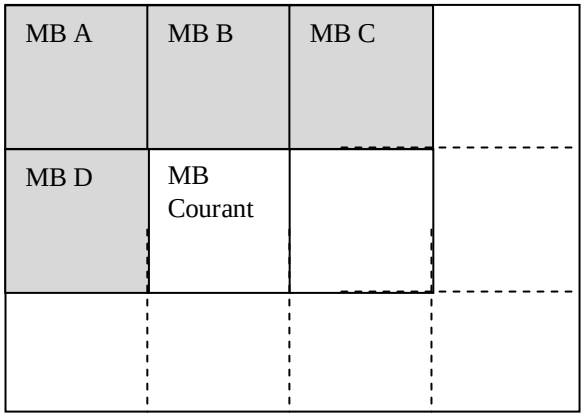


Figure 1.9. *Exemple de prédiction d'un Macro-bloc (MB) dans le mode Intra.*

Intra-4x4 : où chaque macro-bloc est divisé en seize blocs de 4x4 pixels. En effet, ce type de codage est plus utile dans les régions de détails. Chaque bloc de 4x4 pixels est prédit par neuf modes définis par l'orientation de prédiction utilisée. La figure 1.10 illustre les différents modes utilisés pour le codage *Intra*-4x4. L'encodeur de la norme H.264/AVC choisit après cette étape un seul mode ; celui le plus adéquat.

Intra-16x16 : Afin de diminuer la complexité de calcul, la norme H.264/AVC utilise dans le cas des zones d'images régulières des macro-blocs de 16x16 pixels. Dans cette classe, quatre modes de prédiction sont appliqués sur chaque macro-bloc, ensuite le codeur détermine le

mode le plus approprié au macro-bloc courant. La figure1.11 montre les quatre modes pour un seul macro-bloc.

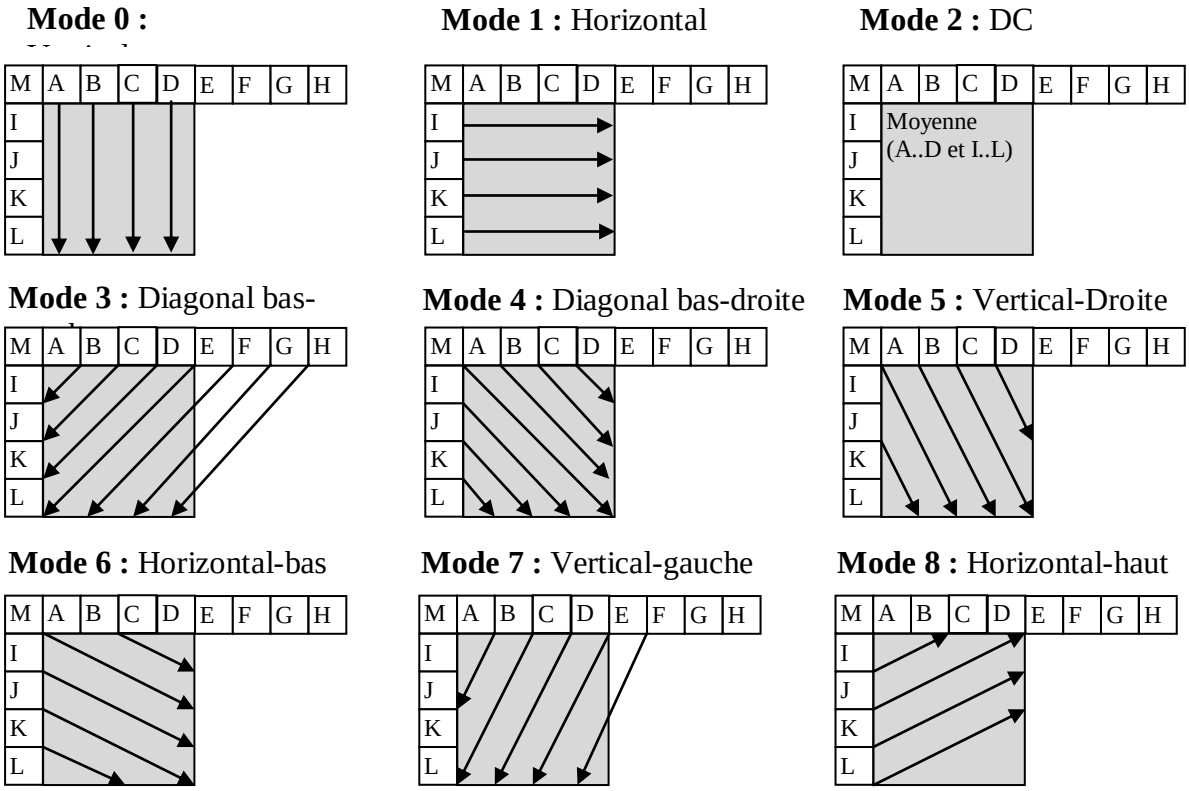


Figure 1.10. *Les modes de prédiction 4x4 de la composante de luminance.*

Les deux composantes de chrominance utilisent une seule classe de codage ; *Intra*-8x8. Les macro-blocs de 8x8 pixels de chaque composante sont estimés en utilisant quatre modes similaires à ceux de la classe *Intra*-16x16 (avec une taille différente des macro-blocs).

Le processus de codage à base de prédiction cherche toujours à fournir la plus faible erreur résiduelle afin de minimiser le débit autant que possible. Parmi les classes de codages décrites ci-dessus, un seul mode est sélectionné par le codeur de la norme H.264/AVC pour la prédiction de chaque bloc. La sélection du meilleur mode se base sur l'optimisation maximale débit-distorsion en utilisant le coût suivant:

$$J(s,c,mode/QP,\lambda) = D(s,c,mode/QP) + \lambda * R(s,c,mode/QP) \quad (1.4)$$

Où le QP désigne le paramètre de quantification (voir la section 1.3.5 pour plus de détails), λ est le multiplicateur de Lagrange qui contrôle le compromis entre D et R. Ce facteur dépend du paramètre de quantification (QP), il est défini par la formule 1.5. s et c sont

successivement le bloc original et prédit, R et D représentent le taux et la distorsion obtenus entre s et c. Le mode utilisé chaque fois est défini dans cette formule par *mode.*

$$\lambda_{mode} = 0.85.2^{(\frac{QP}{3}-4)} \quad (1.5)$$

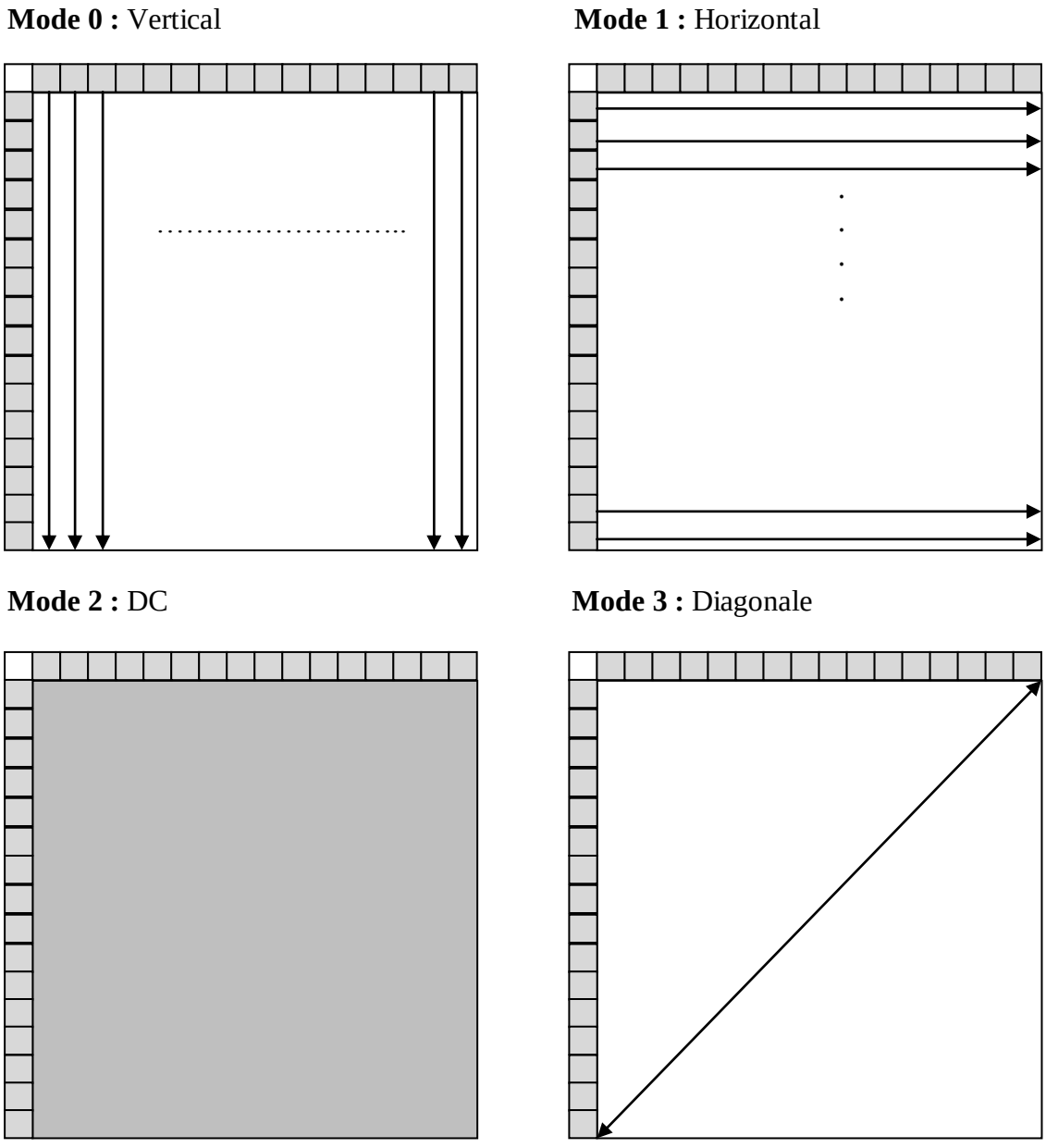


Figure 1.11. *Les modes de prédiction des blocs 16x16.*

Le taux R est le nombre de bits nécessaire pour transmettre toutes les composantes du vecteur de mouvement du bloc courant. Plusieurs mesures objectives de la distorsion D entre le bloc orignal et le bloc prédit ont été employées par le standard H.264/AVC. On peut citer à titre d'exemple :

- SAD (*Sum of Absolute Differences*) ou somme de la valeur absolue des différences définie par la formule 1.6 pour un macro-block.
- SSD (*Sum of Squared Differences*) ou somme des carrés des différences donnée par la formule 1.7 pour un macro-block.

$$SAD = \sum_{m=1}^{16} \sum_{n=1}^{16} \left| MB_{Org\,(m,n)} - MB_{ref\,(m,n)} \right| \quad (1.6)$$

$$SSD = \sum_{m=1}^{16} \sum_{n=1}^{16} (MB_{Org\,(m,n)} - MB_{ref\,(m,n)}{}^{2} \quad (1.7)$$

Avec ; MB_{ori} est le macro-bloc à encoder et MB_{ref} représente le macro-bloc de référence.

1.3.3. Prédiction Inter

La prédiction *Inter* peut être considérée comme l'étape la plus efficace en termes de taux de compression dans le processus de codage de la norme H.264/AVC [26]. La fiabilité de cette technique parvient de l'exploitation de la redondance des informations contenues dans les images successives du même GOP. Deux types d'images sont utilisés par la prédiction *Inter*, à savoir les images P utilisant une image de référence et les images B prédites à partir de deux images de référence. Les macro-blocs courants (soit des images P ou B) dépendent des macro-blocs d'une image de référence précédemment encodée, décodée puis mémorisée. Ceci est accompli au moyen d'un vecteur de mouvement qui contient le déplacement de chaque macro-bloc dans les images de référence ainsi que sa position dans l'image courante (voir la figure 1.12).

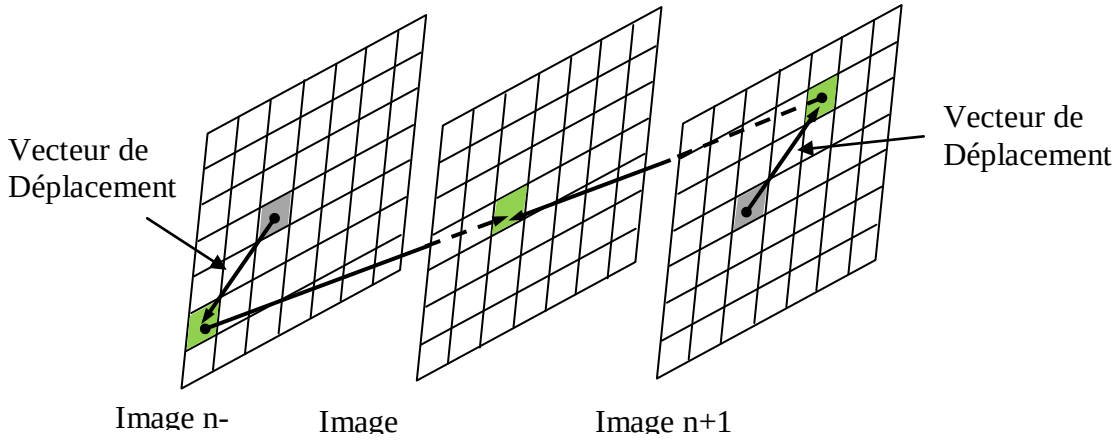


Figure 1.12. *Prédiction bidirectionnelle à compensation du mouvement.*

L'une des particularités soutenues par la norme H.264/AVC, c'est qu'elle admet un vaste éventail de combinaisons au sein d'un macro-bloc.

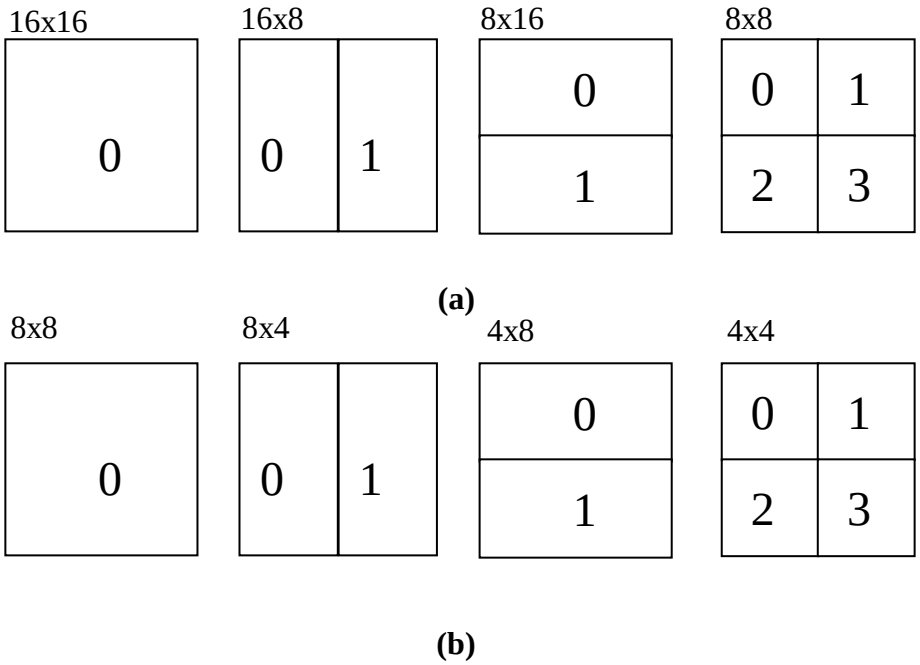


Figure 1.13. *Les possibilités de partitionnement offertes, (a) Partitions de Macro-bloc 16x16, 8x16, 16x8, 8x8, (b) Sous-partitions de macro-bloc : 8x8, 4x8, 8x4, 4x4.*

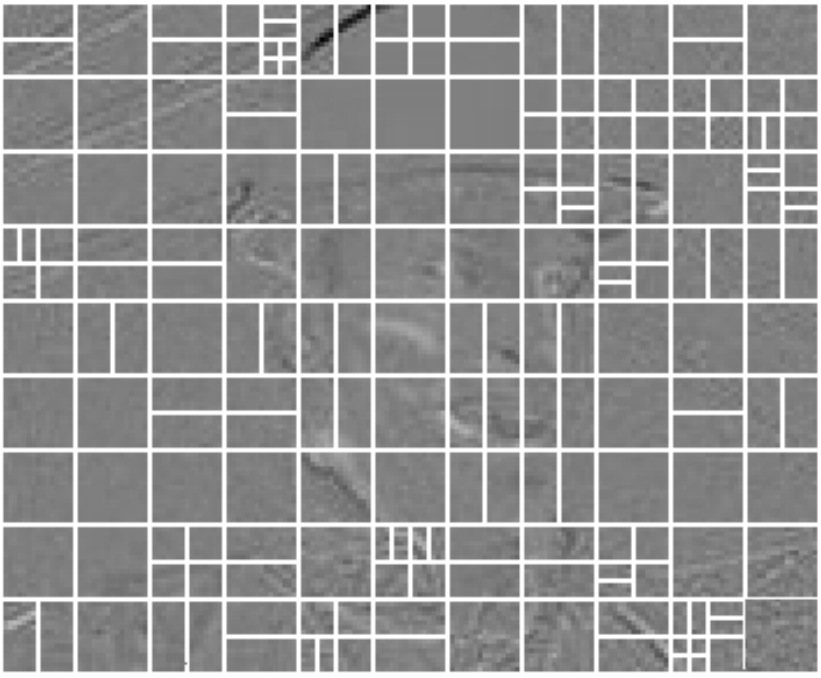


Figure 1.14. *Exemple de partitions de Macro-blocs pour l'estimation de mouvement appliqué sur une image de type P.*

Les possibilités de partitionnement offertes pour chaque macro-bloc de la composante de luminance sont illustrées sur la figure 1.13. Ainsi, les cas de figures envisageables sont :

16x16, 16x8, 8x16 et 8x8 pixels. Les quatre blocs de 8x8 pixels constituant un macro-bloc de 16x16 pixels peuvent à leur tour être partitionnés en blocs de : 8x4, 4x8 ou 4x4 pixels, l'utilisation de ces partitions est utile dans les zones d'images très détaillées.

L'étape de recherche du vecteur de mouvement est appelée estimation de mouvement. Ensuite, une étape de compensation de mouvement est nécessaire. Cette étape consiste à trouver l'information résiduelle (appelée aussi bloc résiduel). Seule l'information résiduelle transformée et quantifiée est transmise. Le mode de prédiction le plus adéquat est celui produisant l'erreur résiduelle la plus faible. Un exemple de partitions de Macro-blocs et de calcul du résidu d'une image P est illustré sur la figure 1.14.

1.3.4. Le mode de décision en H.264/AVC

A la différence des standards antérieurs, la norme H.264/AVC inclue un ensemble très riche de modes de prédiction. Pour une image P par exemple, plusieurs modes sont utilisés pour coder un macro-bloc: SKIP, 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4, *Intra*-4x4 et *Intra*-16x16. Le mode SKIP indique le cas ou la taille du bloc est de 16x16 pixels mais aucun vecteur de mouvement ou information résiduelle ne sont codés. Afin de trouver le mode optimal, le codeur applique tout un processus de test sur chaque mode (la sélection par exemple du mode minimisant le coût de Lagrange et ainsi, l'optimisation maximale débit-distorsion RD). Le cas d'une image P par exemple, la décision du mode d'un macro-bloc peut être résumée comme suit :

- Tester le mode SKIP en premier lieu : C'est le mode le plus rapide par rapport aux autres en raison de la faible complexité calculatoire qu'il possède.
- Tester les modes *Inter*/*Intra* : Il s'agit ici de déterminer le mode *Inter* (parmi les combinaisons 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4) ayant le coût minimal en appliquant le processus de calcul du débit-distorsion (RD). Ensuite, la même méthode est appliquée pour le mode *Intra*-4x4 (9 modes) et également *Intra*-16x16 (4 modes), le mode *Intra* minimisant le coût de Lagrange est choisi.
- L'étape finale consiste à choisir le mode optimal (celui ayant le faible coût) parmi les trois modes SKIP, *Intra* et *Inter*.

Cependant, le processus de choix du mode optimal appliqué à chaque bloc soutient une complexité calculatoire accrue.

1.3.5. Transformée fréquentielle et quantification

Le signal d'erreur de prédiction (l'information résiduelle) obtenu après le processus de prédiction *Intra* ou *Inter* est transformé sous une forme fréquentielle puis quantifié afin d'atténuer la redondance spatiale. A la différence des normes de compression antécédentes (MPEG-2, MPEG-4 part 2,…), qui appliquent la Transformée en Cosinus Discrète DCT [27] de taille 8x8 sur chaque bloc de l'image, le standard H.264/AVC utilise une transformée entière [24][28] sur des blocs de 4x4 pixels, voire 2x2 pixels, pour les blocs des deux composantes de chrominances. Cette transformée fréquentielle est appelée ICT (*Integer Cosine Transform*). En effet, la transformée ICT optée par la norme H.264/AVC a pour but de réduire la complexité calculatoire de l'encodeur (la DCT nécessite une complexité accrue vis-à-vis la ICT). L'utilisation des blocs de 4x4 pixels est mieux adaptée aux erreurs de prédiction locales que les blocs de 8x8 pixels. Cette transformée est calculée par l'équation :

$$Y = H.X.H^T \tag{1.8}$$

Où X représente le signal d'entrée et H est la matrice de transformation utilisée. La matrice H peut être obtenue par trois formes selon la composante et le type de prédiction utilisés. Si le type de prédiction pour la composante de luminance est *Inter* ou *Intra*, la matrice est donc de taille 4x4 et est donnée par :

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \tag{1.9}$$

Le mode *Intra*16x16 est traité d'une manière spécifique où les 16 coefficients DC (composante continue qui représente les basses fréquences) des blocs constituant chaque macro-bloc sont regroupés par la transformée de Hadamard en utilisant (voir la figure 1.15) la matrice:

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \tag{1.10}$$

Le troisième type concerne les deux composantes chromatiques Cr et Cb (utilisent des macro-blocs de taille 8x8) où la matrice H est donnée par :

$$H = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{1.11}$$

Si le mode de prédiction sélectionné est *Intra*16x16, le bloc composé des coefficients DC da la composante de luminance regroupés précédemment par la transformée de Hadamard est transmis en premier. Les informations résiduelles des différents blocs de 4x4 (16 blocs) dans le macro-bloc courant sont ensuite transmises. Finalement le bloc des coefficients DC des deux composantes de chrominance ainsi que les blocs résiduels sont respectivement envoyés.

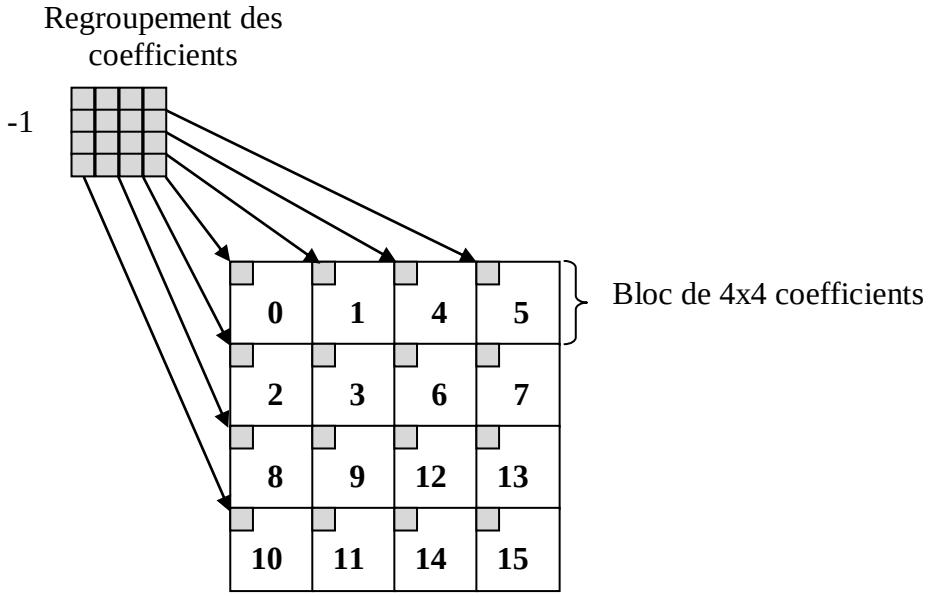


Figure 1.15. *Un exemple de regroupement des coefficients DC d'un Macro-bloc (16x16 coefficients), après transformée entière.*

Quantification

Cette étape s'applique sur les macro-blocs transformés afin d'éliminer les hautes fréquences (sont les coefficients souvent les moins importants dans la représentation fréquentielle). Ceci peut permettre d'augmenter le nombre de coefficients nuls et ainsi, d'assurer une grande fiabilité au niveau de l'encodage entropique.

A la différence des normes de codages vidéo précédentes, où le pas de quantification augmente constamment, le standard H.264/AVC utilise une quantification scalaire uniforme [28]. La qualité de la vidéo est contrôlée par le pas de quantification Q_{step} qui est également contrôlé par le paramètre de quantification QP. En effet, l'évolution du pas de quantification est logarithmique ; sa valeur est multipliée par 2 à chaque incrémentation de *6* du QP. Le

tableau 1.1 récapitule les valeurs possibles du QP et du pas de quantification. La valeur maximale possible de QP est de 52 où le taux de compression est maximisé avec une piètre qualité visuelle de la vidéo encodée. La meilleure qualité de la vidéo est obtenue pour un QP égal à 1 avec un taux de compression très faible. Ainsi, un compromis entre le taux de compression (mesuré par le débit binaire) et la qualité visuelle est contrôlé par le QP.

QP	0	1	2	3	4	5	6	7	8	9	10	…
Q_{step}	0.625	0.6875	0.8125	0.875	1	1.125	1.25	1.375	1.625	1.75	2	…
QP	…	18	…	24	…	30	…	36	…	42	…	51
Q_{step}	…	5	…	10	…	20	…	40	…	80	…	224

Tableau 1.1. *Tableau d'équivalence entre le paramètre QP et le pas de quantification.*

L'opération de quantification peut être accomplie par :

$$Cq_{ij} = round\left(\frac{C_{ij}}{Q_{step}}\right) \tag{1.12}$$

Où Cq_{ij} sont les coéfficients quantifiés, C_{ij} réprésente les coefficients obtenus après la ICT.

1.3.6. Filtrage anti-blocs

L'inconvénient inévitable des techniques de codage à base de blocs est la visibilité de la structure en blocs (les effets de blocs). Ces effets apparaissent plus particulièrement après l'opération inverse de la transformée entière appliquée sur chaque bloc ou macro-bloc et également après l'opération inverse de la quantification. Pour remédier à ce problème, le standard H.264/AVC définit un outil de filtrage du bord des blocs (un filtre anti-blocs) afin d'améliorer la qualité visuelle des séquences vidéo. En effet, le filtre anti-blocs se caractérise dans la norme H.264/AVC de ses prédécesseurs principalement par l'intégration de ce dernier dans les deux processus de codage et de décodage. Dans les normes antérieures, le filtre anti-blocs est utilisé seulement au décodage. Toutefois, l'intégration d'un filtre anti-blocs dans la phase de codage permet d'améliorer la prédiction des macro-blocs sur les images de références précédemment traitées et mémorisées. En outre, un gain de 5 à 10% en débit

binaire est obtenu avec l'utilisation de cette étape de filtrage. La figure 1.16 illustre les performances d'un tel filtre.

Figure 1.16. *La performance du filtre anti-blocs sur des images très compressées, sans et avec le filtre.*

1.3.7. Codage entropique

Le codage entropique est l'étape de compression sans perte de l'encodeur H.264/AVC, dont le but est de réduire la quantité d'informations produites par les étapes précédentes (quantification, prédiction,…) d'une façon exactement réversible. En effet, cette étape de codage est la dernière dans le processus de compression. Autrement dit, le flux binaire (*bitstream*) à transmettre ou à stocker est généré lors de l'encodage entropique.

L'idée de base derrière cette technique est que, des mots-codes de longueurs variables sont associés aux occurrences ou séquences de symboles. En effet, la longueur du mots-codes sera d'autant plus courte que l'occurrence du symbole est augmentée, ce qui permet une amélioration du taux de compression. Le H.264/AVC prend en charge deux types de codage entropique. La première méthode est fondée sur l'utilisation des mots-codes VLC, appelée CAVLC (*Context-based Adaptive Variable Length Coding*) [29]. Le deuxième type est le CABAC (*Context-based Adaptive Binary Arithmetic Coding*) [30] qui utilise un code arithmétique s'appuyant sur des tables évolutives. Toutefois, le choix entre les deux méthodes, repose sur un compromis entre la complexité calculatoire et l'efficacité de codage.

Le CAVLC fait partie de tous les profils prévus par la norme H.264/AVC (voir section 1.4). Dans ce cas un unique ensemble illimité de mots-codes est utilisé, ces mots-codes sont définis pour tous les éléments syntaxiques (les vecteurs de compensation de mouvement, type de macro-bloc, paramètre de quantification, ...), à l'exception des informations résiduelles. La

table de mots-codes unique, est définie par plusieurs codes VLC dont essentiellement celui de type *Exponentiel-Golomb* [31]. Le tableau 1.2 montre les premiers mots-codes utilisés. Les modes-codes sont organisés comme suit :

$$[N\ z\acute{e}ros][1][Code] \tag{1.13}$$

Où la taille de chaque mot-code est donnée par (2N+1) bits et la longueur du champ « Code » est de N bits. Chaque mot-code peut être obtenu en utilisant le Numéro_code de la façon suivante :

$$N = \lfloor \log_2(Num\acute{e}ro\_code + 1) \rfloor \tag{1.14}$$

$$Code = Num\acute{e}ro\_code + 1 - 2^N \tag{1.15}$$

Les informations résiduelles issues de la transformée entière sont codées par un code VLC plus compliqué offrant des résultats plus optimaux. Cette méthode, appelée adaptatif-CAVLC, est appliquée sur des blocs 4x4 ou 2x2 (le cas des composantes de chrominances) parcourus en *zigzag* (figure 1.17).

Numéro_code	**Mot-code**
0	1
1	010
2	011
3	00100
4	00101
5	00110
6	00111
7	0001000
8	0001001
9	0001010
…	…

Tableau 1.2. *Tableau de correspondance du codage Golomb exponentiel.*

Le débit binaire peut encore être diminué par l'usage du codage arithmétique CABAC (réduction de l'ordre de 10 à 15% par rapport au CAVLC). La qualité de la vidéo reste la même que celle obtenue avec l'utilisation de CAVLC (le codage entropique est exactement réversible ; sans perte). Néanmoins, ce type de codage est plus couteux en termes de complexité (nécessite plus de ressources pour coder et décoder les données) en comparaison avec l'algorithme CAVLC. Pour cette raison, la norme H.264/AVC a intégrée le CABAC seulement dans le profil *Main* et des profils supérieurs. En effet, le CABAC est fondé sur l'utilisation de plusieurs modes de probabilités, chacun est adapté à un contexte différent.

Après avoir converti tous les symboles en binaire, cette méthode choisit le modèle de probabilité le plus ajusté. L'estimation de la probabilité est ensuite optimisée par l'usage des informations des éléments voisins.

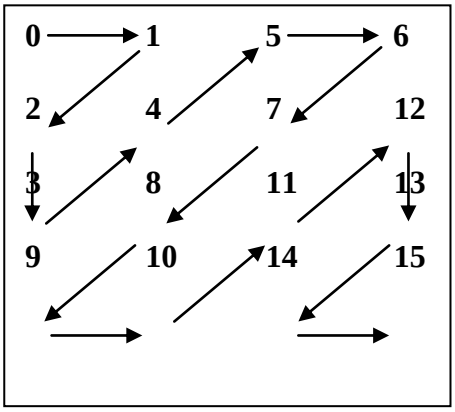


Figure 1.17. *Parcours en zigzag des blocs 4x4 de luminance.*

1.4. Profils et niveaux du H.264/AVC

Un aspect important de la norme H.264/AVC au même titre que les autres normes vidéo, est qu'elle facilite l'interopérabilité entre plusieurs applications fondées sur cette norme qui exigent une configuration semblable. Ceci est assuré par l'utilisation et la configuration de différents profils et niveaux. Un profil représente un ensemble de caractéristiques algorithmiques qui peuvent être utilisés pour générer un flux compatible. Les niveaux peuvent être considérés comme des classes d'administration des besoins de performance, de bande passante et de mémoire (classes de performance).

En effet, le choix du profil dépend de l'application. Ainsi, plusieurs profils ont été définis, chacun couvrant une catégorie d'applications précise [32][33]. Le tableau 1.3 résume les différentes fonctionnalités intégrées par chaque profil. La complexité d'implémentation est limitée par la définition de l'ensemble des fonctions que l'encodeur peut utiliser par chaque profil. Les profils possibles sont:

- **Le profil de base (*baseline profile*)** : Essentiellement recommandé pour les applications à faible débit (applications mobiles et de visioconférence) qui utilisent peu de ressources. Ce profil prend en charge la majorité des fonctionnalités de la norme H.264/AVC à l'exception de quelques-unes [23].
- **Le profil principal (*Main profile*) :** Conçu pour les applications grand public de diffusion et de stockage. A la différence du profil de base, ce profil intègre quelques

options supplémentaires tels que par exemple le codage entropique CABAC et le codage entrelacé. Toutefois, ce profil n'utilise pas l'ordonnancement flexible des macro-blocs intégré dans le profil de base (voir tableau 1.3).

- **Le profil étendu (*Extended profile*) :** La caractéristique principale du profil étendu est la robustesse face aux erreurs lors de la transmission sur des canaux bruités [34]. Ce profil est prévu pour la diffusion en flux (streaming) des vidéos et pour la transmission sans fil.

Quatre nouveaux profils, destinés aux applications haute définition (HD) appelés collectivement profils supérieurs ont été ajoutés récemment par la norme H.264/AVC :

- **Profil supérieur (*High Profile ; HiP*) :** C'est le profil de base pour les applications haute définition. Il a été adopté pour le stockage sur les disques HD DVD et Blu-ray ainsi que pour la télévision numérique haute définition (HDTV). Ce profil utilise toutes les options complexes du standard. Ainsi, il nécessite plus de ressources par rapport aux profils antécédents.

- **Profil supérieur 10 (*High 10 Profile ; Hi10P*) :** Soutient le mode de sous-échantillonnage 4 :2 :0 jusqu'à 10 bits par échantillon.

- **Profil supérieur 422 (*High 4:2:2 Profile ; Hi422P*) :** c'est le profil principal pour les applications professionnelles, il se base sur le profil Hi10P avec un sous-échantillonnage de 4 :2 :2 à 10 bits par échantillon.

- **Profil supérieur 444 (*High 4:4:4 Profile ; Hi444P*) :** Se base sur le profil Hi422P avec un sous-échantillonnage de la chrominance de 4 :4 :4 à plus de 12 bits par échantillon. Ce profil inclus une transformée résistante aux erreurs de transformation de l'espace couleur ainsi qu'un codage zonal sans pertes.

	profil de base	**profil étendu**	**profil principal**	**HiP**	**Hi10P**	**Hi422P**	**Hi444P**
tranches I et P	Oui	Oui	Oui	Oui	Oui	Oui	Oui
tranches B	Non	Oui	Oui	Oui	Oui	Oui	Oui
tranches SI et SP	Non	Oui	Non	Non	Non	Non	Non
Image de Références Multiples	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Filtre anti-blocs	Oui	Oui	Oui	Oui	Oui	Oui	Oui
codage CAVLC	Oui	Oui	Oui	Oui	Oui	Oui	Oui
codage CABAC	Non	Non	Oui	Oui	Oui	Oui	Oui
ordonnancement flexible des macro-blocs	Oui	Oui	Non	Non	Non	Non	Non
ordonnancement arbitraire des tranches	Oui	Oui	Non	Non	Non	Non	Non
tranches redondantes	Oui	Oui	Non	Non	Non	Non	Non
partitionnement des données	Non	Oui	Non	Non	Non	Non	Non
codage entrelacé	Non	Oui	Oui	Oui	Oui	Oui	Oui
format 4:2:0	Oui	Oui	Oui	Oui	Oui	Oui	Oui
format monochrome (4:0:0)	Non	Non	Non	Oui	Oui	Oui	Oui
format 4:2:2	Non	Non	Non	Non	Non	Oui	Oui
format 4:4:4	Non	Non	Non	Non	Non	Non	Oui
pixel 8 Bit	Oui	Oui	Oui	Oui	Oui	Oui	Oui
pixel 9 et 10 Bit	Non	Non	Non	Non	Oui	Oui	Oui
pixel 11 et 12 Bit	Non	Non	Non	Non	Non	Non	Oui
transformée 8×8	Non	Non	Non	Oui	Oui	Oui	Oui
matrices de quantification	Non	Non	Non	Oui	Oui	Oui	Oui
quantification Cb et Cr séparée	Non	Non	Non	Oui	Oui	Oui	Oui
codage sans-perte	Non	Non	Non	Non	Non	Non	Oui

Tableau 1.3. *Les profils inclus par la norme H.264/AVC.*

Les ressources mémoires et calculatoires nécessaires pour décoder une vidéo ainsi que les performances de codage en termes de bande passante sont délimitées dans le H.264/AVC par onze niveaux en s'appuyant sur un certain nombre de paramètres (la vitesse de traitement du décodeur, le débit binaire pour chaque profil,…). Le tableau 1.4 récapitule les niveaux définis et les paramètres utilisés.

Num-Niveau	**Macro-blocs par seconde (max)**	**taille de l'image en macro-blocs (max)**	**débit en Kbit pour les profils de base, étendu et principal (max)**	**débit en Kbit pour le profil HiP (max)**	**exemple de résolution et cadence en (Fps).**
1	1485	99	64 Kbit/s	80 Kbit/s	128×96/30.9 176×144/15.0
1b	1485	99	128 kbit/s	160 kbit/s	128×96/30.9 176×144/15.0
1.1	3000	396	192 kbit/s	240 kbit/s	176×144/30.3 320×240/10.0
1.2	6000	396	384 kbit/s	480 kbit/s	176×144/60.6 320×240/20.0 352×288/15.2
1.3	11880	396	768 kbit/s	960 kbit/s	352×288/30.0
2	11880	396	2 Mbit/s	2.5 Mbit/s	352×288/30.0
2.1	19800	792	4 Mbit/s	5 Mbit/s	352×480/30.0 352×576/25.0
2.2	20250	1620	4 Mbit/s	5 Mbit/s	720×480/15.0 352×576/25.6
3	40500	1620	10 Mbit/s	12.5 Mbit/s	720×480/30.0 720×576/25.0
3.1	108000	3600	14 Mbit/s	17.5 Mbit/s	1280×720/30.0 720×576/66.7
3.2	216000	5120	20 Mbit/s	25 Mbit/s	1280×720/60.0
4	245760	8192	20 Mbit/s	25 Mbit/s	1920×1080/30.1 2048×1024/30.0
4.1	245760	8192	50 Mbit/s	62.5 Mbit/s	1920×1080/30.1 2048×1024/30.0
4.2	522240	8704	50 Mbit/s	62.5 Mbit/s	1920×1080/64.0 2048×1088/60.0
5	589824	22080	135 Mbit/s	168.75 Mbit/s	1920×1080/72.3 2560×1920/30.7
5.1	983040	36864	240 Mbit/s	300 Mbit/s	1920×1080/120.5 4096×2048/30.0

Tableau 1.4. *Les niveaux définis par le standard H.264/AVC.*

1.5. Conclusion

Le domaine de la compression de la vidéo numérique a connu une évolution spectaculaire dès la naissance de la norme H.264/AVC. L'efficacité de compression est améliorée énormément par cette norme grâce à l'intégration de nouvelles options et le perfectionnement de plusieurs autres, en particulier la résistance face aux erreurs ainsi que la précision des fonctions prédictives. Parmi les points forts également de l'encodeur du standard H.264/AVC est la production des vidéos hautes définitions à débit similaire aux normes antécédentes (MPEG-2, MPEG-4 part 2, H.263,…). Néanmoins, cette amélioration d'efficacité de compression de vidéo est accompagnée d'une augmentation de la complexité calculatoire. Grâce à sa flexibilité, le H.264/AVC a été appliqué dans différents domaines tels que le stockage HD DVD et Blu-ray, la télévision numérique haute définition (TVHD), la téléphonie mobile.

Dans ce chapitre, nous avons présenté quelques principes de base sur la compression vidéo. Puis nous avons détaillé les différences majeures présentes entre les diverses normes de compression vidéo développées par les deux organismes de normalisation les plus connus (ISO/IEC et l'ITU-T). Finalement, nous avons présenté plus particulièrement les techniques adoptées dans l'encodeur de la norme H.264/AVC.

Chapitre 02 : Compression de la vidéo multi-vues extension du H.264/AVC

2.1. Introduction

La compression de la vidéo multi-vues est l'extension majeure de la norme H.264/AVC [10][11]. Les travaux de recherche sur le MVC sont effectués par le groupe *Joint Video Team* (JVT) créé conjointement par le groupe MPEG de l'organisme de normalisation ISO/IEC et le groupe VCEG de l'ITU-T. L'objectif de la compression de la vidéo multi-vues est d'optimiser le stockage, l'utilisation et la transmission sur les réseaux informatiques tels que l'internet tout en conservant une qualité acceptable (suivant l'application et le service utilisé) da la vidéo multi-vues. Cette optimisation s'accomplie essentiellement par l'exploitation de la corrélation entre les flux vidéo utilisés.

Pour le codage indépendant des différentes vues (élimination de la redondance spatiale et temporelle), la compression de la vidéo multi-vues utilise les techniques de codage implémentées par la norme H.264/AVC (la notion de GOP composé de plusieurs images de type I, P et B, la décomposition de chaque image en plusieurs slices, les types de macro-blocks utilisés : 16x16, 16x8,…, les techniques de prédiction et de compensation de mouvement, … etc). Pour l'élimination de la redondance inter-vues, plusieurs techniques spécifiques du MVC doivent intervenir, les plus importantes sont celles qui concernent la prédiction et la compensation de disparité.

Nous allons aborder dans ce chapitre, en premier lieu, les exigences et les données de test nécessaires au développement de la compression de la vidéo. Puis, nous dresserons un état de l'art des schémas du MVC qui se basent surtout sur la satisfaction et l'amélioration des exigences du MVC. Ensuite, nous présenterons quelques notions clés de la compression de la vidéo multi-vues telles que la prédiction et compensation de disparité ainsi que la structure générale du flux binaire de cette norme. Finalement, nous présenterons le modèle de référence appelé JMVM proposé conjointement par les deux groupe MPEG et VCEG.

2.2. Les exigences générales

Lors du développement ou de la création d'une norme de compression de vidéo, un ensemble d'exigences et de recommandations doit être respecté. Ces exigences et recommandations varient dans le cas de la compression de la vidéo multi-vues selon le service multimédia visé [4]. Parmi plusieurs recommandations, les exigences les plus importantes (définies dans [9] en détails) sont :

- Efficacité de compression : L'exigence primordiale pour la plupart des modèles de codage vidéo est l'efficacité de compression élevée. Cette efficacité est définie par un compromis entre débit binaire et qualité de la vidéo (un débit binaire à une certaine qualité ou l'inverse). Dans le cas spécifique du MVC, l'efficacité de la compression désigne un gain significatif par rapport à la compression indépendante de chaque vue. En plus du codage efficace des différentes vues, une bonne compression du MVC nécessite une configuration adéquate des caméras de capture.
- L'accès aléatoire inter-vues : L'accès à une image donnée à une instance T_n dans l'axe temporel avec un minimum d'images à décoder est l'une des exigences les plus souhaitées par la majorité des standards de compression. Ceci est appelé l'accès aléatoire temporel. Cette recommandation est toujours assurée par les images de type *Intra* (image I) dans chaque GOP. L'accès aléatoire, par exemple, à l'image B numéro cinq dans la figure 2.1, nécessite le décodage de plusieurs images qui sont, l'image I numéro zéro, l'image P numéro trois et l'image P numéro six. Le cas spécifique du MVC, nécessite en plus de l'accès aléatoire temporel, défini dans chaque vue indépendamment des autres, un accès aléatoire inter-vues qui permet d'accéder et ainsi, décoder n'importe quelle image dans les différentes vues à une instance T_n avec un minimum d'images à décoder.
- L'évolutivité (*scalability*): Ici le décodeur doit pouvoir accéder à une partie d'un flux binaire afin de générer une vidéo effective de qualité variable. L'évolutivité peut être définie au niveau spatial dans le but d'offrir plusieurs niveaux de résolution ou au niveau temporel pour fournir plusieurs fréquences temporelles du signal. En plus de l'évolutivité spatiale et temporelle, le MVC doit prendre en charge une structure de flux binaire (*bitstream*) permettant d'accéder aux vues sélectionnées avec un effort minimum de décodage. Cela, permet à la vidéo d'être affichée sur une multitude de terminaux et de l'adapter aux limitations de débit du canal de transmission.
- Le faible retard: un faible retard d'encodage et de décodage doit aussi être assuré par la compression de la vidéo multi-vues. Le Faible retard est très important pour les applications en temps réel qui utilisent la vidéo multi-vues (MVV) telles que le streaming et la vidéo conférence.
- Consommation des ressources : Le MVC devrait être efficace en termes de consommation de ressources (par exemple, la minimisation de la complexité calculatoire), telles que la taille de la mémoire utilisée, la bande passante et la

puissance du traitement. Cette recommandation est très importante également en termes de consommation énergétique dans le cas de l'utilisation de la norme dans les systèmes embarqués.

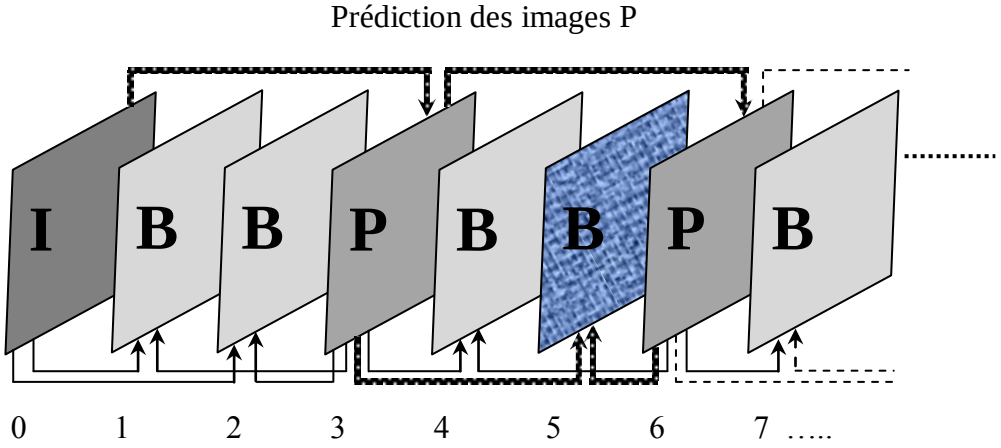


Figure 2.1. *Un exemple d'accès aléatoire temporel, l'image à consulter est l'image 5 de type B.*

A noter que seules les recommandations les plus importantes sont présentées ci-dessus (liste non exhaustive). Plusieurs autres facteurs sont aussi exigés pour la compression de la vidéo multi-vues tels que la robustesse face aux erreurs, principalement provoquées lors de la diffusion sur les différents réseaux de télécommunication. Le MVC doit également supporter une gamme d'images de résolutions différentes chacune destinée à une application ou service différent. La rétrocompatibilité doit aussi être respectée, où à tout instant, le flux binaire correspondant à une vue donnée doit être conforme aux H.264/AVC.

2.3. Le choix des données et des conditions de test

Le choix approprié des données et des conditions de test est irréfutable pour l'analyse et le développement d'une norme de codage vidéo. L'ensemble de données de test doit être représentatif des applications envisagées. En effet, les séquences utilisées pour le test et l'analyse de la compression de la vidéo multi-vues varient selon plusieurs paramètres tels que:

- Le nombre de caméras ou points de vue utilisés pour l'obtention de la même scène. Ce nombre peut varier en fonction du service envisagé.
- La distance entre les caméras dont l'effet est important sur la qualité de la compression. Une bonne similarité entre les vues est obtenue par un espacement moins distant entre les caméras.

- L'arrangement des caméras utilisées. A titre d'exemples, cet arrangement peut être 1D parallèle, 2D parallèle, 1D arc et ainsi de suite,
- Les propriétés des images constituant chaque vue en termes de résolution généralement 640x480 pixels ou 1024x 768 pixels, et de fréquence 15, 25 ou 30 images par seconde.

La configuration initiale des caméras utilisées doit être aussi envoyée au décodeur via le flux binaire. Les différentes séquences fréquemment utilisées pour le test des divers algorithmes sont synthétisées dans le tableau 2.1. Afin de pouvoir comparer notre approche avec d'autres algorithmes, nous avons utilisé également ces mêmes vidéos.

La figure 2.2 présente un exemple d'un système de capture multi-vues proposé par [35]. 32 caméras sont utilisées dans ce système avec un arrangement de type 1D arc, la distance entre deux caméras successives est de 3 degrés (la distance est calculée en degré puisque les caméras utilisées sont arrangées sous forme d'arc).

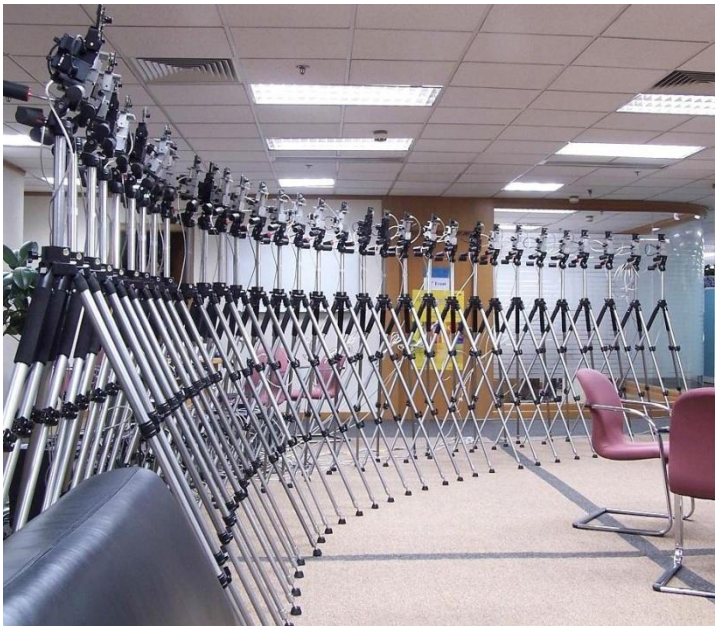


Figure 2.2. *Un système de capture multi-vues composé de 32 caméras avec un arrangement 1D arc et un espacement de 3 degrés.*

L'ensemble de données de test du MVC couvre un large éventail de différents types de contenu. Divers facteurs peuvent influer sur ce contenu comme par exemple l'endroit où la scène est capturée (intérieur ou extérieur), le système de caméras (fixes ou mobiles), et les différentes complexités du mouvement et de précision spatiale. La figure 2.3 illustre un

exemple des 8 vues pour la vidéo « *Ballroom* » tandis que la figure 2.4 montre quelques exemples de vidéo multi-vues dans une seule instance de temps Tn pour une seule vue.

Dataset	**Séquences**	**Définition image**	**Fréquence**	**Arrangement caméra**
MERL	Ballroom	640x480 (250 images)	25 fps	8 caméras, 20 cm espacement, 1D parallèle
	Exit	640x480 (250 images)	25 fps	8 caméras, 20 cm espacement, 1D parallèle
	Vassar	640x480 (250 images)	25 fps	8 caméras, 20 cm espacement, 1D parallèle
KDDI	Race1	640x480 (250 images)	30 fps	8 caméras, 20 cm espacement, 1D parallèle
Nagoya University / Tanimoto Lab	Rena	640x480	30 fps	100 cameras avec 5cm espacement; 1D parallèle

Tableau 2.1. *Les séquences utilisées pour l'évaluation.*

Les paramètres et les conditions d'essai doivent être unifiés pour une étude comparative vraisemblable entre plusieurs approches MVC. Pour chaque séquence d'essai, au moins trois débits binaires doivent être choisis, en fonction des propriétés et du contenu de la séquence à encoder. Les trois niveaux de débits binaire sont représentés par la qualité de la vidéo compressée, les qualités possibles selon le débit binaire sont faibles (mais acceptable), moyenne et haute qualité. Le débit binaire est contrôlé directement par le paramètre de quantification QP. Nous avons opté dans notre travail pour cinq niveaux du débit définis par cinq valeurs du QP.

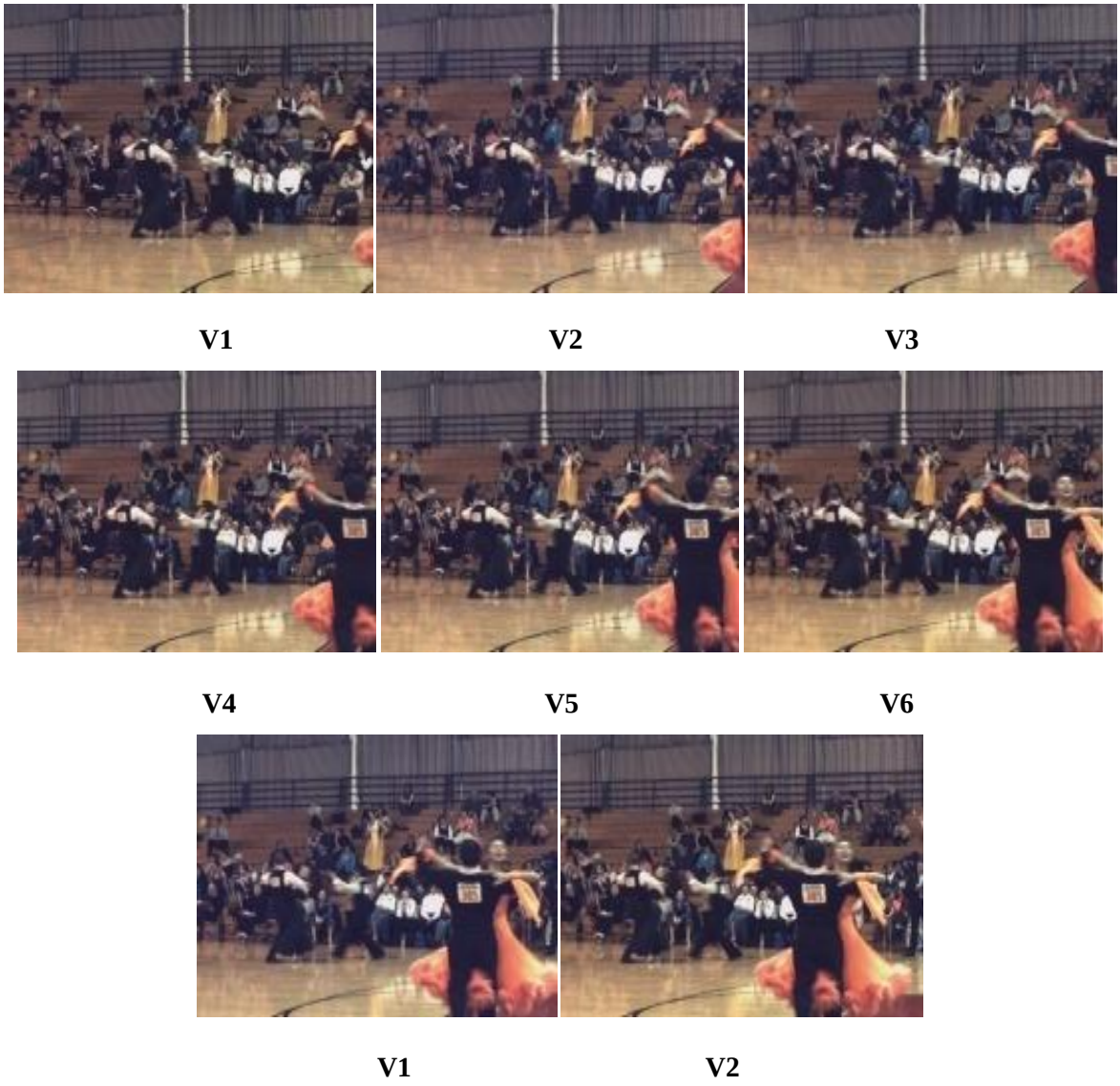


Figure 2.3. *Les huit vues (de V1 à V8) de la vidéo BallRoom capturées par un arrangement linaire des caméras.*

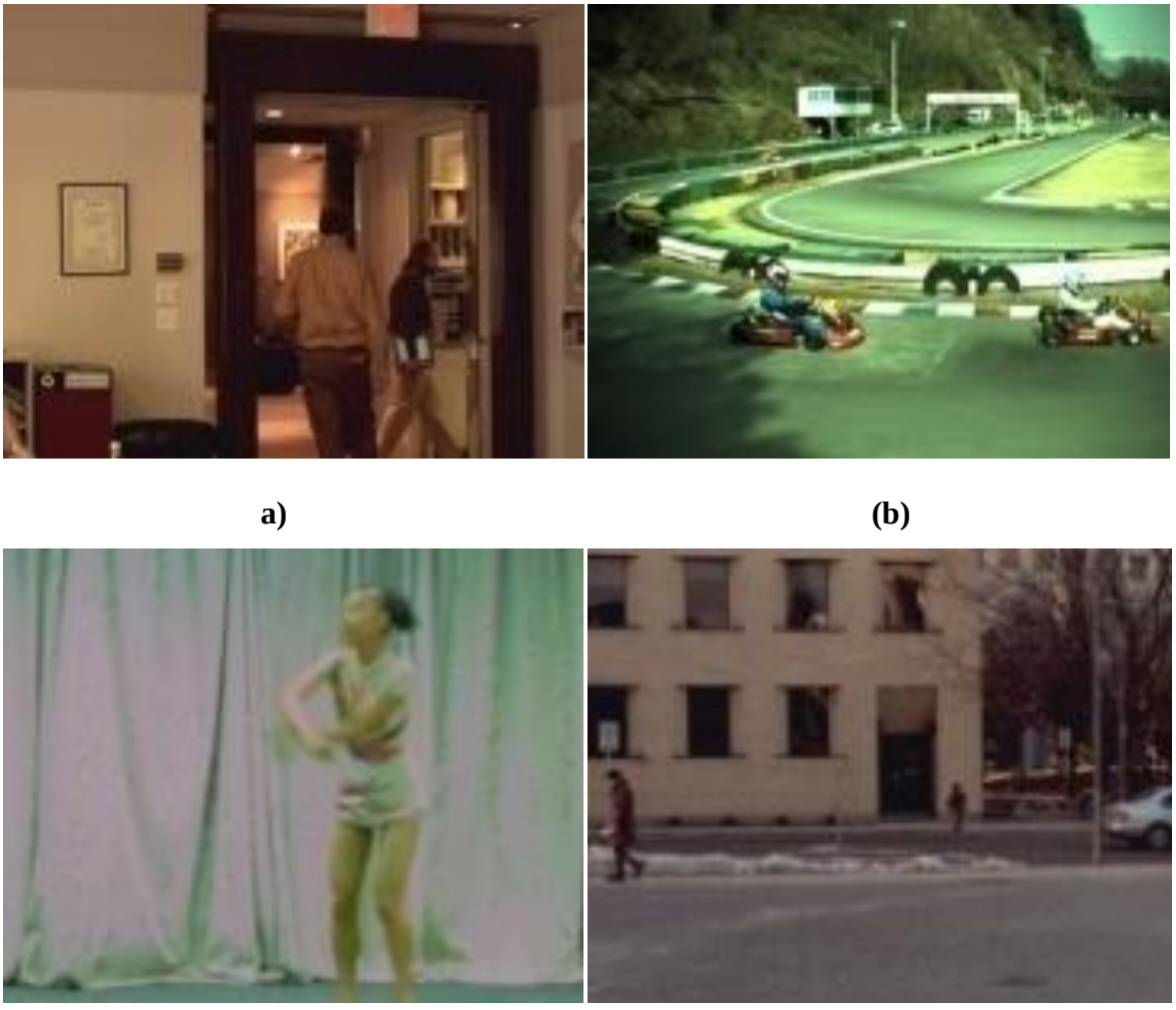
2.4. Algorithmes de compressions de la vidéo multi-vues

2.4.1. Amélioration du débit binaire

L'exigence la plus importante dans le MVC est l'efficacité de compression déterminée par un compromis débit binaire et la qualité de la vidéo. Plusieurs travaux on été proposés afin d'améliorer le débit binaire de la compression de la vidéo multi-vues avec une qualité acceptable, dont les plus importants sont les travaux de Merkle *et al* [36] et Yang *et al* [37].

Merkle et son groupe de travail ont proposé un schéma MVC à base de la structure de prédiction appelée « images B hiérarchique » (HBP ; *Hierarchical B Pictures*) détaillée dans

[38][39]. La structure HBP est implémentée par [36] dans l'axe temporel (dans chaque vue) et également dans le niveau inter-vues. La structure proposée dans [36] (détaillée dans le chapitre 03) appelée « IBP » est adoptée dans le projet commun de normalisation MVC et le modèle de référence *Joint Multi-view Video model* (JMVM) [12]. Ceci, en raison de son efficacité de compression supérieure et d'évolutivité temporelle. L'idée de base de la structure IBP est d'organiser la MVV sou forme de plusieurs GGOP (*Group of Group Of Pictures*). Chaque GGOP est composé de plusieurs GOP selon le nombre de vues utilisées. Le gain en débit binaire est dû dans cette structure aux images de type B hiérarchiquement organisées dans les deux niveaux temporel et inter-vues. Toutefois, cette structure de prédiction utilise une seule image de type « I » (codée sans référence à d'autres) par GGOP ce qui produit un accès aléatoire très lourd.



a) **(b)**

(c) **(d)**

Figure 2.4. *Séquences MVC avec un arrangement linaire de caméras. (a) Exit: 8 vues , (b) Race1 : 8 vues, (c) Rena : 100 vues , (d)Vassar :8 vues.*

Dans [37], une nouvelle structure de prédiction avec un gain significatif de débit binaire est proposée. Ceci est principalement dû à l'amélioration de l'ordre de codage des images B dans chaque vue de la vidéo multi-vues par l'utilisation d'un arbre binaire, qui est également étendu à d'autres points de vue.

L'utilisation de la technique d'arbre binaire permet de formuler une équation récursive pour déterminer un ordre de codage sous-optimal, mais efficace, et d'utiliser la programmation dynamique pour résoudre l'équation récursive efficace. La figure 2.5 (a) présente un exemple de cette technique appliquée sur un seul GOP. Ce dernier est décomposé en plusieurs sous-GOP. Ensuite, l'ordre de codage des images B dans chaque sous-GOP est désigné récursivement par un arbre binaire. L'arbre binaire présenté par exemple dans la figure 2.5 (c) génère l'ordre de codage du sous-GOP de la figure 2.5 (b). Ensuite une généralisation de la méthode est appliquée sur les différentes vues de la MVV. L'encodage récursif des images B dans les deux niveaux temporel et inter-vues, appliqué après la décomposition de chaque GOP en plusieurs sous-GOP provoque une complexité accrue de l'encodage. Ceci, conduit à un accès aléatoire inter-vues plus long.

2.4.2. Accélération de l'encodage

La complexité de la compression de la vidéo multi-vues est considérablement augmentée. Ceci est dû à l'utilisation d'un mode de décision basé sur des blocs de tailles variables. Ces blocs sont adoptés dans le MVC pour l'amélioration de l'efficacité de la compression en termes de compromis débit binaire et qualité. En raison de la complexité du codage, des efforts [40][41] ont été consacrés à développer des algorithmes afin d'accélérer le mode de décision (FMD ; *Fast Mode Decision*).

Zeng et al [40] proposent une approche plus efficace pour le mode de décision appliqué sur la structure IBP du modèle de référence JMVM (la même structure proposée par [36] est utilisée). Cette approche est appelée *Mode Correlation-based Mode Decision* (MCMD). Les points clés de cette méthode sont décrits comme suit :

- Tous les modes de prédiction utilisés dans le JMVM sont regroupés en cinq classes d'activité de mouvement. Ces modes organisés de manière hiérarchique.
- Ensuite, pour le Macro-Block (MB) actuel, une comparaison du coût du taux de distorsion (RD) du mode SKIP (voir le chapitre 01) par rapport au seuil adaptatif est

effectuée. Cette comparaison permet de sauter éventuellement le processus de vérification des modes restants.

- Si une telle résiliation anticipée n'est pas accordée, une seule des quatre classes d'activité de mouvement sera sélectionnée. Ceci permet d'identifier le mode optimal selon une analyse plus approfondie du vecteur de mouvement prédit (PMV ; *predicted motion vector*) du MB actuel.
- Le seuil d'adaptation précité et le PMV sont calculés en utilisant le *mode corrélation*.

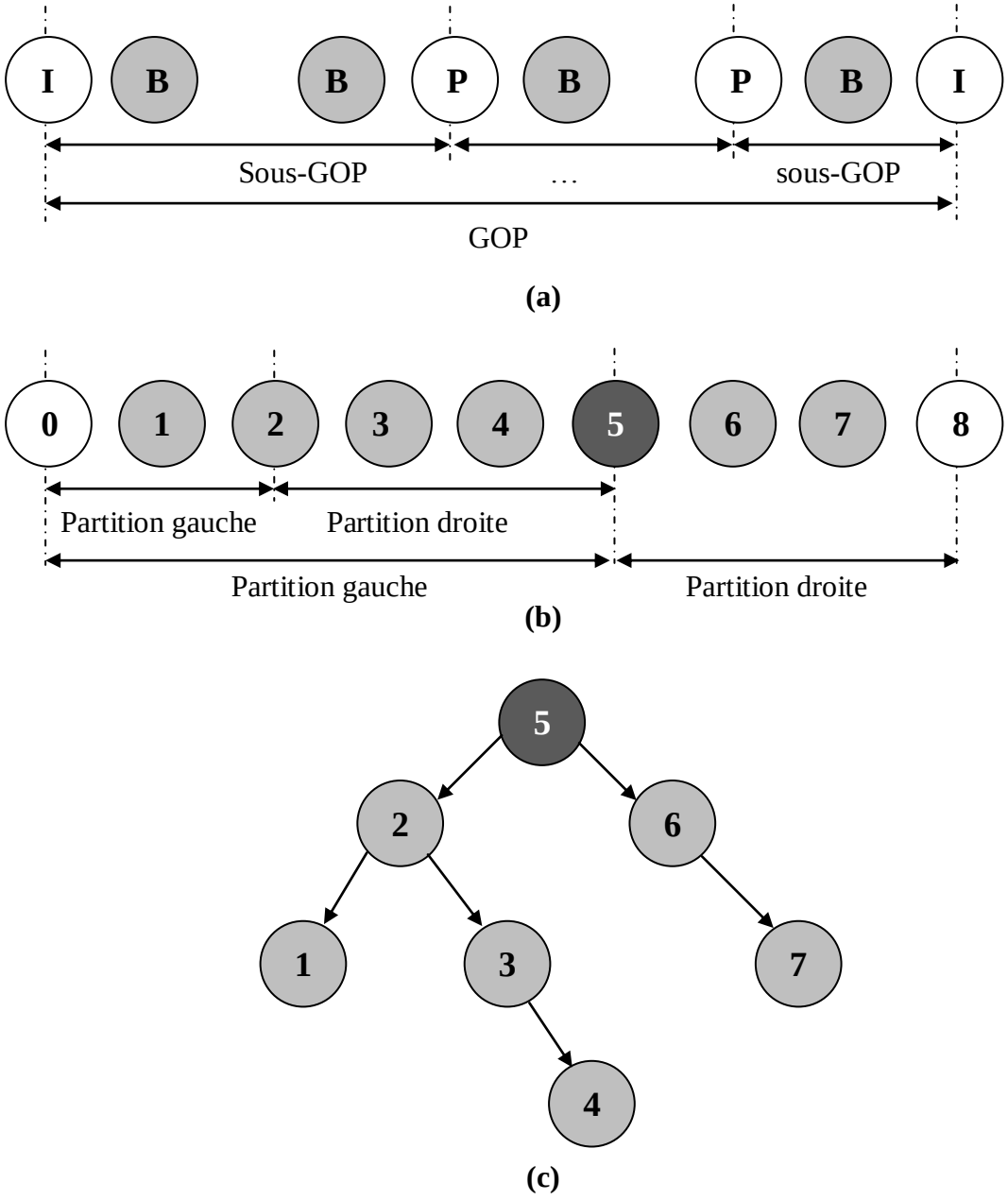


Figure 2.5. *Structure de prédiction proposée par [37], (a) structure d'un GOP, (b) partitionnement d'un sous-GOP, (a) arbre binaire de l'ordre de codage.*

L'approche MCMD proposée par [40] réduit considérablement la complexité de calcul tout en conservant pratiquement la même qualité de codage vidéo et le débit total ciblé que celle du mode de décision exhaustive. Cette méthode par défaut mise en œuvre dans le modèle de référence JMVM [12]. Néanmoins, cette approche présente l'inconvénient d'être très lourde en accès aléatoire inter-vues. En effet, elle utilise la même structure de prédiction IBP du modèle de référence JMVM, mais avec une perte négligeable en qualité et une augmentation légère du débit binaire.

Dans [41], une autre méthode qui vise l'amélioration du mode de décision *Inter* basée sur la segmentation et les corrélations de texture est proposée, cette approche est appelée *Fast Inter Mode Decision* (FIMD). Chaque image est segmentée en trois régions de texture en utilisant les coûts RD du mode *Intra* ainsi que les seuils de segmentation. Ensuite, les résultats de segmentation sont utilisés dans la prise du mode *Inter* ultérieur pour enregistrer le temps de calcul tout en conservant la qualité de l'image. Les coûts RD du mode SKIP classés par le mode *Intra* sont utilisés pour exploiter la corrélation du coût RD entre les vues à la place de l'utilisation de vecteurs de mouvement (obtenus après la prédiction et l'estimation inter-vues). Afin d'accélérer la décision du mode SKIP, le coût RD de ce dernier est comparé avec celui calculé selon les différentes régions de texture. La direction de prédiction inter-vues est également optimisée. Elle a été basée sur la réutilisation de l'estimation de mouvement pour la sélection du meilleur mode pour l'estimation de disparité. En outre, la corrélation entre le mode *Inter*8 × 8 et les régions de texture est associée à l'activité de mouvement visant à réduire l'estimation de mode *Inter*8×8. Cette méthode est aussi lourde en accès aléatoire inter-vues. En effet, la même structure de prédiction (IBP) est utilisée avec une perte négligeable en efficacité de la compression (en termes de débit binaire et qualité de la vidéo).

2.4.3. Amélioration de l'accès aléatoire inter-vues

Plusieurs approches de la compression de la vidéo multi-vues, basées sur une prédiction mixte inter-vues et temporelle, ont été proposées au cours de la dernière décennie [42][43]. Dans [42], la structure de prédiction proposée a montré une amélioration significative de l'accès aléatoire inter-vues. Dans ce cas les vues de type B (les vues qui commencent par une image B, voir le chapitre 03 pour plus de détails) utilisées dans la structure par default du modèle JMVM (structure IBP), sont remplacées par des vues codées par une prédiction à échelle réduite. Les points de vue codés par prédiction à échelle réduite sont obtenus en exploitant un compromis entre la distorsion due à la quantification et la distorsion due au

sous-échantillonnage. Dans la structure de prédiction [42], l'amélioration de l'accès aléatoire inter-vues est principalement due à la décomposition de toutes les vues en plusieurs groupes de vues (GOV ; *Group Of Views*). Chaque GOV doit contenir une vue de base "I" (où la première image est de type I) ou une vue de type "P" où les autres vues sont codées par une prédiction à échelle réduite. La figure 2.6 illustre un exemple de cette structure de prédiction où seulement huit vues sont utilisées avec deux GOV. Cependant, cette méthode de codage de la vidéo multi-vues conduit à une complexité accrue. Cette complexité est encore plus prononcée lors de l'utilisation de séquences vidéo avec trop de textures. Ceci peut avoir comme conséquence, une dégradation considérable de la qualité subjective.

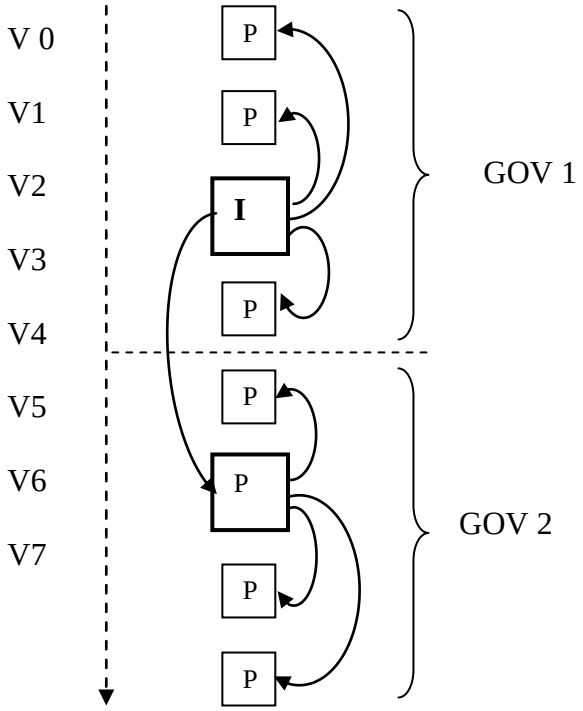


Figure 2.6. *La structure de prédiction proposée par [42], un exemple de 8 vues avec deux GOV.*

L'approche proposée dans [43] est basée sur une méthode adaptative de la compression de la vidéo multi-vues. La mise en œuvre d'un tel algorithme se fait à travers l'analyse et la sélection d'une meilleure structure de prédiction à partir de trois types qui ont été désignés comme ‘’modes’’. L'analyse et la sélection sont effectuées par l'utilisation d'un algorithme de commutation de mode pour le codage. En outre, la sélection est basée principalement sur l'analyse de corrélation spatio-temporelle. Cette dernière peut être obtenue selon la fonction de coût de Lagrange calculée pour les images codées au cours du processus de compression.

L'utilisation de l'algorithme de commutation de mode dans le codage de la vidéo multi-vues a amélioré l'accès aléatoire inter-vues et également la flexibilité des vues évolutives.

Cependant, l'utilisation de trois modes de prédiction produit un retard significatif lors du décodage ainsi qu'une consommation élevée de la mémoire, qui représentent des exigences très importantes pour le MVC [9]. Ces inconvénients sont dus aux modes de choix de la meilleure structure nécessitant à chaque fois la comparaison des trois structures de prédiction en utilisant la fonction du coût de Lagrange. Cette fonction doit être calculée à chaque fois pour l'image codée pendant le processus de codage.

2.5. Description de la compression de la vidéo multi-vues

La compression de la vidéo multi-vues a été normalisée comme extension de la norme H.264/AVC. Ainsi, toutes les caractéristiques et les techniques de cette norme de compression vidéo sont utilisées par le MVC. A titre d'exemple nous rappelons l'idée de la division de la séquence en groupe de plusieurs images (GOP) où chaque groupe est composé de trois types d'images ou moins qui sont les images I (image *Intra*), P (image de prédiction) et B (image de prédiction bidirectionnelle). De même, la décomposition de chaque image en plusieurs slices (tranches) composés de plusieurs macro-blocks (un carré de 16 pixels par 16 pixels) [10]. Enfin, l'utilisation des techniques d'estimation et de compensation de mouvement ainsi que le choix du meilleur macro-block sont également maintenus dans le MVC. Nous examinons dans cette section certaines fonctionnalités clés du MVC.

2.5.1. Prédiction inter-vues

L'une des solutions les plus simples est de coder indépendamment les différentes séquences par l'intervention d'un codec vidéo tel que H.264/AVC [10] [23] [44]. Cette méthode est appelée compression « *Simulcast* ». Le principe de fonctionnement ainsi que la conception de cette structure de prédiction sont détaillés dans le troisième chapitre. Une autre solution possible pour le codage de la MVV est d'utiliser seulement la corrélation inter-vues (rarement utilisée) [45]. Dans cette solution, les images de chaque vue sont prédites à partir des images des vues adjacentes. Cette technique est appelée l'estimation de disparité. Les informations concernant le déplacement des différents blocs sont stockées dans un vecteur de disparité. Cependant, une solution plus optimale pour l'élimination de l'information redondante est d'exploiter la corrélation entre les différentes vues en plus de la corrélation temporelle. Ainsi, les informations non pertinentes peuvent être éliminées, entre autres, par la

combinaison entre prédiction temporelle et inter-vues. Autrement dit, l'utilisation de l'estimation de disparité en plus de l'estimation de mouvement dans les différentes vidéos de la même scène est très efficace en termes d'élimination de la redondance inter-vues.

Par conséquent, cette section présente une analyse des propriétés de corrélation temporelle et inter-vues par l'étude des dépendances statistiques qui peuvent être exploitées pour la prédiction. La figure 2.7 montre les différentes possibilités de prédiction (temporelle et inter-vues) à partir des images voisines, où S indique les vues constituant la MVV et T représente l'axe temporel.

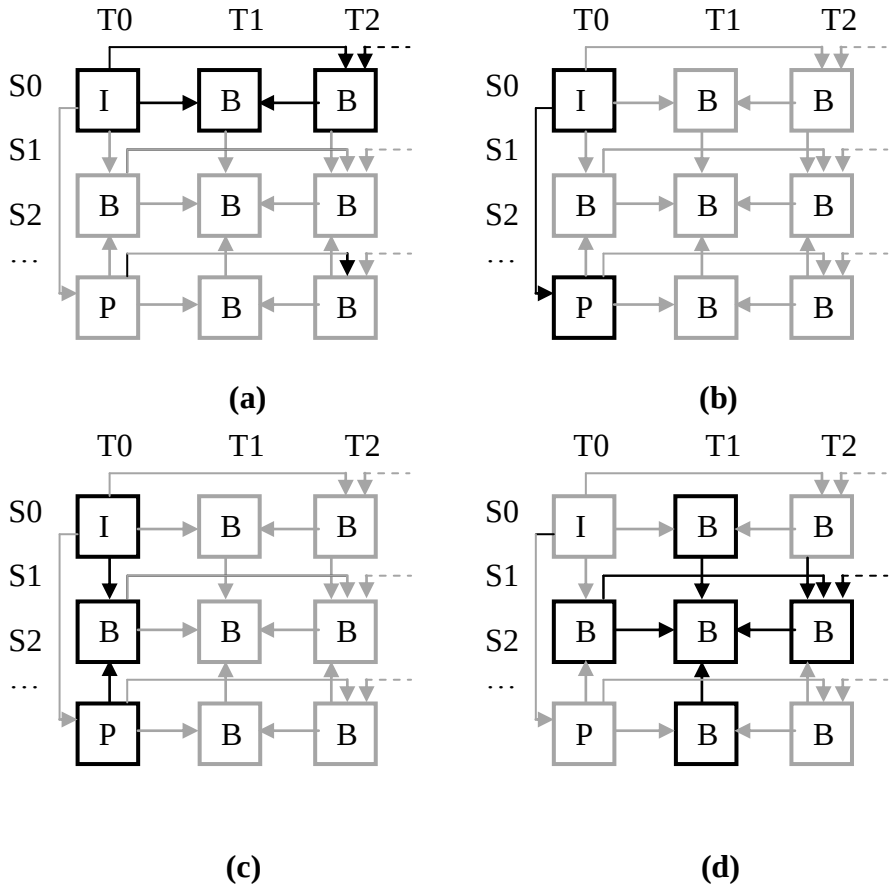


Figure 2.7. *Les différents types de prédiction dans le MVC, (a) prédiction temporelle, (b) prédiction inter-vues appliquée sur une image P, (c) prédiction inter-vues appliquée sur une image B, (d) prédiction mixte appliquée sur une image B.*

Le but de cette analyse est de déterminer le type de prédiction le plus performant en terme d'optimisation du taux de distorsion. La figure 2.7 (a) illustre un exemple d'une structure où seulement la prédiction temporelle est appliquée, cette solution nécessite toujours l'utilisation d'une image de référence codée indépendamment des autres dans chaque vue de

la MVV. Si la solution choisie par le codeur est la deuxième, où seulement la prédiction inter-vues est utilisée, chaque vue doit disposer également à chaque instant T_n d'une image de référence. La figure 2.7 (b) montre un exemple de cette solution appliquée sur une image P (nécessite une seule image de référence). Tout comme le niveau temporel, les images B nécessitent deux images de référence dans la prédiction inter-vues. A la différence de la prédiction temporelle et inter-vues, la prédiction mixte (temporelle/inter-vues) utilise quatre images de référence, deux dans le niveau temporel et le même nombre dans le niveau inter-vues, pour le codage de chaque image B (voir la figure 2.7 (d)).

Paramètres	**Valeurs**
Paramètre de quantification(QP)	22
Taille GOP	12,15
Nombre GOP	02
Plage de recherche	32 pixels
Mode *Intra*	désactivé
Codage entropique	CABAC
Vues utilisées (ordre)	I, B, P

Tableau 2.2. *Les paramètres utilisés pour l'encodage.*

Le codage prédictif mixte (temporel/inter-vues) est fondé principalement sur les techniques d'estimation et de compensation de mouvement et également de disparité. Ces techniques sont basées sur la minimisation de la fonction du coût de Lagrange calculée par la formule 1.4. Pour chaque bloc *Si* d'une image à encoder, une estimation de mouvement et/ou de disparité est effectuée. Le résultat de cette estimation est un vecteur de mouvement ou de disparité noté V_i. Le vecteur V_i sélectionné est celui qui minimise la fonction de Lagrange J. Ce vecteur est choisi dans un intervalle de recherche M dans l'image de référence du niveau temporel ou inter-vues. La figure 2.8 illustre un exemple de recherche d'un bloc à partir de deux niveaux temporel et inter-vues. Le vecteur V_i peut être estimé en utilisant la formule suivante :

$$V_i = arg\ min\{D(S_i,m) + \lambda.R(S_i,m)\} \quad (2.1)$$

Où D et R représentent successivement la distorsion calculée entre le bloc original et le bloc prédit (voir les formules 1.5 et 1.6) et le taux qui définit le nombre de bits du vecteur de mouvement du bloc courant.

Afin d'analyser les dépendances statistiques temporelles et inter-vues de la MVV, nous avons testé une structure de prédiction mixte sur plusieurs séquences (celles décrites dans le tableau 2.1) multi-vues. Dans cette analyse seulement les images B utilisant quatre images de référence ont été utilisées. Les mêmes paramètres de codage sont utilisés pour toutes les séquences. Le tableau 2.2 montre quelques exemples de ces paramètres. Les résultats obtenus en termes de pourcentage pour chaque type de prédiction appliqué lors de l'encodage des images B sont présentés dans le tableau 2.3 et la figure 2.9. Le codeur choisit le mode (parmi plusieurs) qui minimise la fonction de Lagrange J pour chaque bloc à partir de deux modes de prédiction ; temporel et inter-vues.

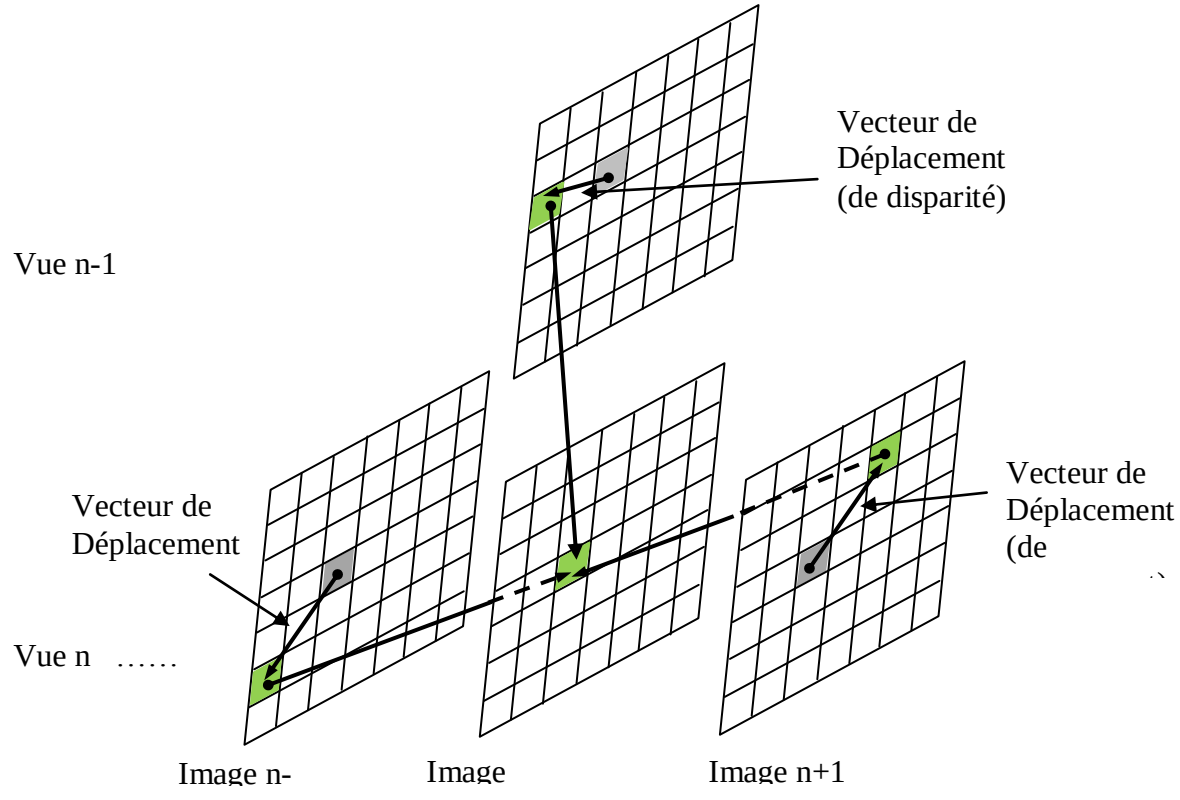


Figure 2.8. *Prédiction bidirectionnelle inter-vues.*

Séquences utilisées	**Temporel (%)**	**Inter-vues(%)**
Ballroom	82.19	17.81
Exit	85.34	14.66
Vassar	88.07	11.93
Race1	97.16	2.84
Rena	80.63	19.37

Tableau 2.3. *Analyse de la prédiction temporelle et inter-vues.*

Les résultats obtenus montrent clairement que le mode de prédiction le plus efficace pour toutes les séquences est le mode temporel. Néanmoins, la prédiction inter-vues est également utile. Le fait d'utiliser une portion de blocs par l'encodeur, ceci permet une amélioration considérable de l'efficacité de compression. Après cette analyse, il est claire que la prédiction mixte peut être considérée comme un facteur important, voire indispensable, pour la compression de la vidéo multi-vues.

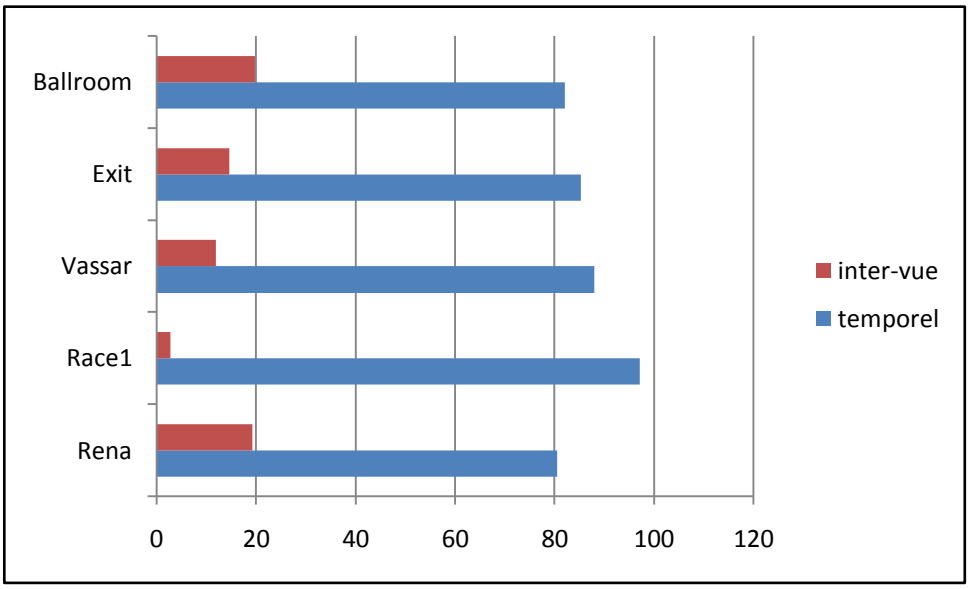


Figure 2.9. *Efficacité de la prédiction temporelle par rapport à la prédiction inter-vues pour toutes les séquences.*

En général, la relation entre la prédiction temporelle et inter-vues dépend fortement des propriétés de la vidéo multi-vues utilisée. En particulier la densité temporelle et spatiale d'une

part, et la complexité de la scène d'autre part. Le mode de capture de la scène (intérieur, extérieur,…), les paramètres des caméras utilisées, le type d'arrangement ainsi que la distance entre les caméras, sont tous des facteurs importants qui influent directement sur la qualité d'encodage de la MVV.

2.5.2. Mémorisation des imagés décodées

La complexité d'encodage du MVC dépend pleinement de la complexité de la norme H.264/AVC. A noter qu'il peut s'agir d'une autre norme utilisée pour l'encodage des différentes vues, MPEG-2 à titre d'exemple. Il est évident que l'usage de l'estimation de disparité parallèlement avec l'estimation de mouvement, augmentent considérablement la complexité de l'encodage du MVC. Dans l'exemple de la figure 2.7 (d), une image B nécessite quatre images de références au lieu de deux. Ainsi, le meilleur mode de prédiction d'un macro-block, doit être sélectionné après la recherche dans quatre images de référence. La compensation de disparité accroît donc le temps nécessaire au choix du meilleur mode. La mémoire tampon DPB (*Decoded Picture Buffer*) utilisée pour mémoriser les images décodées (mémorise les images à utiliser pour prédiction) est également augmentée. La taille de la DPB dépend du profil et du niveau définis. Dans le MVC la taille est calculée par l'équation suivante (pour le profil supérieur de la vidéo multi-vues et stéréo) [46].

$$Taille_{DPB} = \min\left(\frac{Facteu\,r_{MVC} * MaxMBs}{Himage * Limage}, \max(1, \lfloor log_{2.}(nbr_{vues}) \rfloor) * 16\right) \qquad (2.2)$$

Où $Facteur_{MVC}$ est toujours égal à 2, MaxMBs varie selon le niveau utilisé et représente le nombre de macro-blocks maximum par seconde (voir tableau 1.4). *Himage* et *Limage* désignent successivement la hauteur et la largeur de l'image en macro-blocks, le nbr_{vues} représente le nombre de vues utilisées. La formule 2.2 montre que la taille de la DPB doit être comprise entre 16 et Log_2 (nbr_{vues})*16. Cette exigence est assurée par le *min* qui précède la formule. La taille maximale permise dans le la norme H.264/AVC est de 16 comme le montre la formule suivante :

$$Taille_{DPB} = \min\left(\frac{Facteu\,r_{MVC} * MaxMBs}{Himage * Limage}, 16\right) \qquad (2.3)$$

La taille de la DPB peut être aussi calculée en fonction de la taille du GOP et le nombre de vues. Dans ce cas-là, l'ordre d'organisation du flux vidéo non compressé est pris en charge (pour plus de détails sur l'ordre de décodage, voir le chapitre 03). Deux arrangements

fondamentaux de l'ordre d'organisation sont définis par le JVT [47]. Le premier type s'articule autour du décodage des images des vues en respectant l'ordre de ces dernières ainsi que l'architecture en GOP. Ce type est appelé *view-first*. La deuxième méthode est appelée *time-first*. L'ordre d'organisation dans cette méthode est désigné par l'axe temporel (l'ordre des images dans chaque GOP est écarté). La taille de la DPB est calculée dans le premier type (*view-first*) en utilisant la formule suivante [48]:

$$Taille_{DPB} = (log_2(taille_{GOP}) + 2 * taille_{GOP} + nombre\_vues) \quad (2.4)$$

Par exemple dans le cas où la taille du GOP utilisée est de 16 avec une MVV de trois vues, la taille DPB est de 39. Dans le deuxième cas (time-first), la taille de la DPB est obtenue de la même façon en appliquant la formule suivante :

$$Taille_{DPB} = (log_2(taille_{GOP}) + 1) * nombre\_vues + 2 \quad (2.5)$$

Avec les mêmes conditions de l'exemple précédent (la taille du GOP est de 16 avec trois vues), la taille DPB est de 17. La taille du DPB dépend aussi des structures de prédiction temporelle est inter-vues utilisées. Cependant, la compression de la vidéo multi-vues sans faire intervenir la prédiction inter-vues, ne nécessite qu'une taille DPB d'une seule vue, ceci est quel que soit le nombre de vues.

2.5.3. Structure du flux binaire MVC

La compression de la vidéo multi-vues doit permettre un usage multiple des applications et des services nécessitant l'accès seulement à une partie (une version 2D) du contenu de la vidéo. En effet, cette partie doit être facilement extraite et décodée à partir du flux binaire produit par le MVC. Autrement dit, le flux binaire global contenant toutes les vues encodées. L'idée de base de cette propriété est l'intégration d'un flux binaire d'une vue codée indépendamment de toutes les autres vues, appelée vue de base, dans le flux binaire comprimé du MVC. Le flux binaire de la vue de base doit être compatible avec les profils du décodeur, de la norme utilisée pour le codage de chaque vue comme par exemple la norme H.264/AVC, d'une seule vue. Ensuite, si le dispositif utilisé est un récepteur 2D, la vue de base (le flux binaire) pourrait être alors extraite et ainsi, décodée. Dans le cas où le dispositif représente un récepteur 3D, ce dernier peut décoder le flux binaire MVC complet y compris les vues restantes, c'est-à-dire autres que la vue de base.

Tout comme la norme H.264/AVC, le MVC utilise deux catégories d'unité NAL. La première est réservée aux informations codées des images, appelée unité NAL de type VCL tandis que la deuxième est une unité non-VCL contenant les informations supplémentaires nécessaires aux décodages des différentes images. En effet, le MVC utilise un nouveau type d'unité NAL, appelé « slice codé de l'extension MVC » contenant les informations des images codées des différentes vues autres que la vue de base. Pour différencier l'unité NAL de base des autres unités NAL (des autres vues), une unité NAL de préfix est ajoutée à ces dernières. Ainsi, le décodeur de la norme H.264/AVC peut ignorer facilement les nouveaux types d'unité NAL lors de l'utilisation d'un récepteur 2D. Ainsi, il décode seulement l'unité de base ; conforme aux unités NAL définis dans H.264/AVC. L'unité NAL de type VCL contenant l'image indépendamment codée des autres (image *Intra*) est appelée IDR slice (*Instantaneous Decoding Refresh*). Autrement dit, c'est à partir de cette unité que les autres unités sont décodées. La figure 2.10 montre un exemple d'unité NAL de type VCL destinée aux vues non de base.

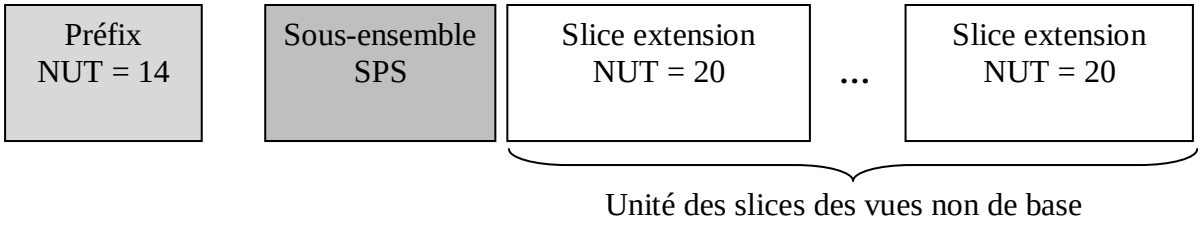


Figure 2.10. *Structure d'un flux binaire MVC contenant les unités NAL de type VCL décodées seulement par le décodeur MVC.*

Dans la norme H.264/AVC, plusieurs types d'unités sont réservés aux extensions futures. Les types utilisés par le MVC sont indiqués dans la figure 2.10 par les NUT (*NAL Unit Type*). Le type numéro 14 est utilisé pour désigner l'unité NAL de préfix tandis que le type 20 représente les slices codés des vues non de base. En effet, les informations du flux binaire MVC ne se limitent pas aux unités décrites ci-dessus. Par conséquent, d'autres unités NAL indispensables pour le décodage sont insérées entre les unités NAL des slices codés. A titre d'exemple, ces unités peuvent représenter les paramètres importants s'appliquant à une large séquence d'images (SPS ou *Sequence Parameter Set*) comme le montre la figure 2.10. Dans le MVC, le SPS peut contenir des informations de dépendance nécessaires pour la prédiction inter-vues [49].

Dans le standard H.264/AVC, une unité NAL est constituée d'une en-tête de 1 octet et un champ (pour l'information utile) transmis par la couche VCL en plus d'un champ *trailing*

(voir chapitre 01). Dans le MVC, cette structure est maintenue à l'exception des unités NAL numéro 14 et 20 qui indiquent la présence de trois octets supplémentaires dans l'en-tête, comme illustré dans la figure 2.11.

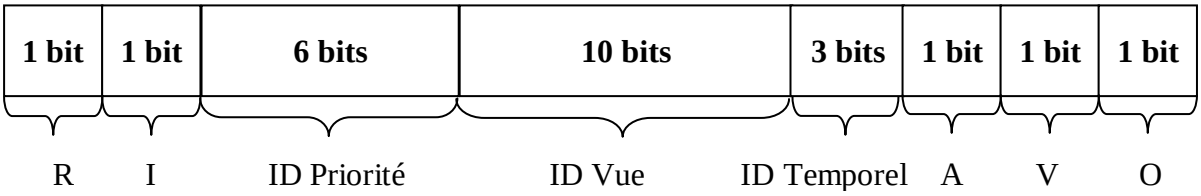


Figure 2.11. *Les trois octets supplémentaires ajoutés à l'en-tête.*

Les trois octets présentés dans la figure 2.11 sont ajoutés à l'octet d'en-tête utilisé par les unités NAL da la norme H.264/AVC (voir la figure 1.5). Les champs d'en-tête présentés dans la figure 2.11 peuvent être expliqués comme suit:

- ID Temporel: Ce champ indique le niveau hiérarchique de la couche temporelle (voir le troisième chapitre).
- ID Priorité : La priorité de l'unité NAL courante est spécifiée par ce champ. Une valeur élevée de cet indicateur représente une priorité inférieure.
- ID vue: L'identificateur de la vue encodée est stocké dans cette partie de l'en-tête.
- A: Cet indicateur spécifie si le slice encodé concerne une image de type *anchor* ou *non-anchor* (voir le chapitre 03 pour plus de détails). Quand le champ A contient un 1, il s'agit alors d'un slice *anchor,* si ce champ contient un 0, c'est slice d'une image *non-anchor*.
- I: Dans ce cas, le contenu de ce bit indique si le slice encodé concerne une image de référence (le contenu est égal à 1) ou pas (le contenu égale 0).
- V: Si l'image encodée est utilisée pour la prédiction inter-vues, ce champ doit contenir un 1, dans le cas contraire (où le bit est à 0) l'image est utilisée seulement dans le niveau temporel.
- R et O: Ce sont des bits réservés pour une extension future.

2.5.4. Profils et niveaux

Les profils dans le MVC désignent comme pour les standards de codage vidéo existants un ensemble de caractéristiques algorithmiques pouvant être utilisés pour générer un flux conforme. En effet, la compression de la vidéo multi-vues définit deux profils reposant sur le

profil supérieur (High Profile, HiP) de la norme H.264/AVC, avec certaines différences. Le premier profil est appelé profil supérieur (High Profile) de la vidéo multi-vues. Il prend en charge plusieurs points de vue, généralement plus que deux vues, et ne supporte pas le codage entrelacé. Contrairement au premier profil, le second prend en charge les outils de codage entrelacé mais il est limité seulement à deux vues.

La figure 2.12 illustre les deux profils avec quelques options supportées par chaque profil. Le profil supérieur de la norme H.264/AVC ainsi qu'un autre profil appelé profil de base limité [50] (*constrained baseline*) sont utilisés dans certains cas par les deux profils , à savoir multi-vues supérieur et stéréo supérieur, comme il est présenté dans la figure 2.12. Il est possible d'avoir un flux binaire qui est conforme à la fois au profil supérieur de la vidéo stéréo quand il y a seulement deux vues codées, et au profil supérieur de la vidéo multi-vues où les outils de codage entrelacés ne sont pas utilisés. Dans ce cas, un drapeau signalant leur compatibilité est défini.

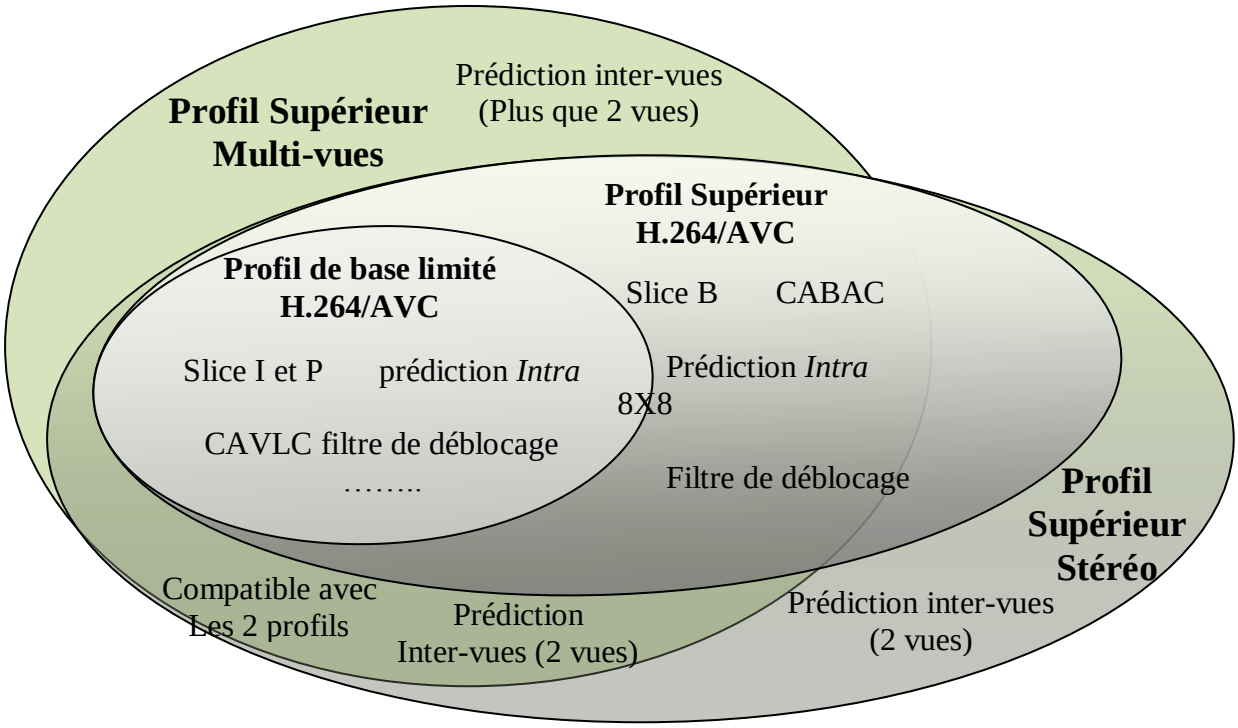


Figure 2.12. *Les différents profils et outils supportés par la compression de la vidéo multi-vues [37].*

La vue de base du MVC peut être encodée en utilisant soit le profil supérieur de la norme H.264/AVC, soit le profil de base limité. L'utilisation de ces deux profils est conditionnée par les contraintes décrites dans les deux profils supérieurs. Celui de la vidéo multi-vues et celui

de la vidéo stéréo. A titre d'exemple, lorsque le profil supérieur de la norme H.264/AVC est utilisé pour la vue de base, les outils de codage entrelacé autorisés dans ce profil (High H.264/AVC), ne peuvent pas être utilisés. En effet, ils ne sont pas pris en charge dans le profil supérieur de la vidéo multi-vues.

Comme pour les normes de codage de la vidéo existantes, les niveaux de la vidéo multi-vues imposent des contraintes sur le flux binaire produit par l'encodeur du MVC. Ainsi, les limites de la complexité et des ressources mémoires utilisées pour le décodage des différentes vues sont définies dans le MVC à travers certains niveaux. Les paramètres utilisés comprennent le débit maximal en termes de macro-blocks par seconde, la taille d'image maximale, le débit binaire global, …etc.

La majorité des limites de niveau sont conservées afin de permettre un flexible passage entre le décodage d'un flux binaire d'une seule vue et le décodage du flux binaire de la vidéo multi-vues. Toutefois, d'autres limites de niveau sont augmentées, telles que la capacité de la mémoire tampon d'image décodée et le débit maximal défini pour chaque profil.

2.6. Joint Multi-view Video Model (JMVM)

La compression de la vidéo multi-vues est développée comme extension de la norme H.264/AVC qui est le résultat d'un projet collaboratif entre le Groupe d'experts en codage vidéo (VCEG) de l'ITU-T et le Groupe MPEG de l'ISO/IEC. Afin de pouvoir comparer les résultats d'encodage des chercheurs et ainsi d'améliorer l'efficacité de codage du MVC, un modèle de référence appelé *Joint Multi-view Video model* (JMVM) [12] est fourni. Le modèle de référence JMVM est développé conjointement par le groupe JVT qui rassemble les deux groupes MPEG et VCEG.

En effet, le modèle de référence JMVM est développé à travers plusieurs versions, il se base principalement sur l'élimination de la redondance de l'information entre les différentes vues (similarité inter-vues), entre les images de chaque vue (similarité temporelle) et l'information redondante dans chaque image (similarité spatiale). Les technique clés utilisées sont les mêmes que celles utilisées dans le standard H.264/AVC, à savoir codage spatial, estimation et compensation de mouvement, en plus de l'estimation et la compensation de disparité.

La structure de prédiction sélectionnée et utilisée comme structure par défaut pour le modèle de référence JMVM est celle proposée par [36]. Cette structure de prédiction est appelée IBP (*Intra*-coded, *Bi-predicted*, *Predicted*) ou par fois IBP-HBP, les principales caractéristiques de cette structure de prédiction sont détaillées dans le chapitre 03. La structure de prédiction IBP utilise pour l'élimination de la redondance de l'information une prédiction mixte inter-vues/temporelle. Un exemple de cette structure avec un GOP de 8 images et une MVV composée de huit vues est illustré dans la figure 3.6 du troisième chapitre. Chaque vue et codée par une structure de prédiction nommée « images B hiérarchique » (un exemple de cette structure de prédiction est présenté dans la figure 3.1 dans le chapitre 03) détaillée dans [38][39].

2.7. Conclusion

La compression de la vidéo multi-vues, utilisée dans de nombreuses applications telles que la télévision 3D et la vidéo à point de vue libre, doit satisfaire un compromis débit binaire/qualité face au développement rapide des technologies d'acquisition et d'affichage 3D. Ce compromis est obtenu par l'exploitation de la similarité inter-vues (prédiction, estimation et compensation de disparité) parallèlement avec la similarité temporelle (prédiction et compensation de mouvement) et spatiale.

Actuellement, de nombreux chercheurs tentent d'améliorer la compression de la vidéo multi-vues tout en respectant les exigences de ce standard normalisé par le group JVT créé conjointement par les deux groups MPEG et VCEG. La grande majorité des travaux ont été basés soit sur l'accélération de l'encodage soit sur l'amélioration de l'efficacité de compression traduite par un compromis entre le débit binaire et la qualité de la vidéo. Toutefois, une autre contrainte du MVC est très insistante où certains chercheurs ont travaillé là-dessus. Cette contrainte représente l'accès aléatoire inter-vues. Elle est assurée généralement par le bon choix de la structure de prédiction. Pour cela, nous présenterons dans le chapitre suivant les structures de prédiction inter-vues les plus utilisées.

Nous avons présenté dans ce chapitre les concepts de base et les caractéristiques techniques les plus pertinentes de la vidéo multi-vues. Nous avons également abordé les exigences les plus importantes dans le MVC. Finalement, nous avons exposé succinctement le modèle référence utilisé pour MVC (JMVM).

Chapitre 03 : Les structures de prédiction du MVC

3.1. Introduction

Dans le but de compresser beaucoup plus efficacement les différentes séquences constituant la vidéo multi-vues, trois types de prédictions sont combinées :

- La prédiction spatiale (*Intra*-image) qui s'effectue par l'utilisation des blocs voisins dans l'image elle-même afin de réduire davantage la redondance spatiale.
- La prédiction temporelle (*Inter*-images) qui s'effectue par GOP composés de trois types d'images ; l'image I spatialement codée et sert de référence aux images suivantes du groupe et des vues, les images prédictives P et les images bidirectionnelles B. Le nombre d'images par GOP change suivant la norme utilisée. Le MPEG-2 par exemple utilise généralement 12 images pour une fréquence de 25 images par seconde tandis que ce nombre dans le cas du H.264/AVC peut dépasser 40 images. L'objectif de cette prédiction est l'élimination de la redondance temporelle due à la corrélation d'images adjacentes dans chaque GOP.
- Une prédiction inter-vues, profite généralement du fait qu'il y a peu de changement d'une vue à l'autre pour la minimisation de la redondance inter-vues. D'où, l'amélioration de l'efficacité de compression.

Nous allons aborder dans ce chapitre plusieurs points. Premièrement, la présentation de la structure de prédiction temporelle utilisée pour les différentes vues du MVC. Cette structure est appelée « images B hiérarchique » (*Hierarchical B Pictures ;* HBP) qui peut être utilisée d'une manière indépendante pour le codage de la MVV. Puis nous détaillerons les caractéristiques les plus importantes des structures de prédiction inter-vues dans le codage de la MVV. Ensuite, nous présenterons quelques spécificités pour les structures de prédiction inter-vues vis-à-vis du codage temporel. Les méthodes d'évaluation utilisées pour le MVC seront ensuite présentées. Finalement, nous renforcerons cette étude par une expérimentation afin de comparer les différentes structures de prédictions étudiées.

3.2. Prédiction temporelle

Tout standard de codage vidéo tel que par exemple le H.264/AVC, peut être utilisé pour l'encodage et le décodage indépendant de chaque vue de la vidéo multi-vues. Cependant, cette méthode permet un gain énorme en temps de calcul, mais avec une faible efficacité de compression. Ceci, est dû au fait que les similarités inter-vues ne sont pas exploitées.

Á la différence de la majorité des normes de codage vidéo classiques utilisant la structure de prédiction IBBP… [51], le standard H.264/AVC est adopté pour une nouvelle structure de prédiction dyadique nommée images B hiérarchie ou HBP (voir [38] pour une description détaillée). Cette section présente les concepts de base des deux structures de prédiction HBP et Simulcast. La flexibilité accrue du H.264/AVC, essentiellement, grâce à la disponibilité de la technique d'image de référence multiple [52], permet une meilleure utilisation de la structure HBP par rapport aux anciennes normes de codage vidéo. Cette structure peut être utilisée efficacement pour le codage indépendant des différentes vues, où chaque vue doit utiliser une image de type I par GOP.

3.2.1. La structure « images B hiérarchique »

Dans les structures de prédiction classiques, une image B n'est prédite qu'à partir des images I/P précédentes et suivantes. Cette propriété n'est plus exigée dans la structure HBP, autrement dit, les images B peuvent servir de référence pour des images B.

Comme dans les structures de prédiction classiques IBBP…, la structure HBP doit utiliser une image *Intra*-codée dans des intervalles réguliers. Parfois ces images figurent dans des intervalles irréguliers. Il est à noter que la première image de la séquence est toujours *Intra*-codée, cette image est notée comme image IDR. La structure HBP, exige l'utilisation d'au moins une image de type I (*Intra*-codée) ou P (image prédictive ou *Inter*-codée) appelée image clé par GOP. La figure 3.1 illustre un exemple typique de cette structure de prédiction. Cette exigence peut permettre un accès aléatoire plus rapide à tout image dans l'axe temporel. Si le GOP courant utilise une image P comme image clé, cette dernière elle-même doit être prédite, par l'estimation et la compensation de mouvement, à partir de l'image clé (image de référence) du GOP précédent. L'ensemble des images qui sont temporellement situées entre l'image clé du GOP courant et celle du GOP précédent sont généralement de type B (les images bidirectionnelles). D'une manière générale, les images B du même GOP précèdent leur image clé dans l'ordre d'affichage et l'utilisent comme image de référence pour leurs encodages (voir la figure 3.1). Les images bidirectionnelles (les images B) de chaque GOP sont organisées de façon hiérarchique afin de permettre une prédiction hiérarchique. La figure 3.1 montre un exemple d'une structure de prédiction HBP avec quatre niveaux hiérarchiques dyadiques.

En effet, la structure de prédiction HBP ne se limite pas au cas dyadique (cas présenté dans la figure 3.2). En outre, cette structure de prédiction peut être ajustée de manière

adaptative dans le temps. Ce cas de figure est appelé HBP non dyadique [53] (voir la figure 3.2 avec seulement trois niveaux hiérarchiques). Dans les deux cas, chaque image B est prédite à partir des images B d'un niveau hiérarchique plus haut. Les images B de plus haut niveau hiérarchique sont prédites en utilisant les images clés (niveau hiérarchique 0) du GOP courant et précédent. Dans l'arbre hiérarchique, les images B ayant le plus bas niveau ne servent pas de références pour aucune image. L'organisation des images B en plusieurs niveaux hiérarchiques a arborée une efficacité très importante pour l'évolutivité temporelle (*temporal scalability*) dans la norme de codage vidéo évolutive H.264/SVC [53] (*Scalable Video Coding*).

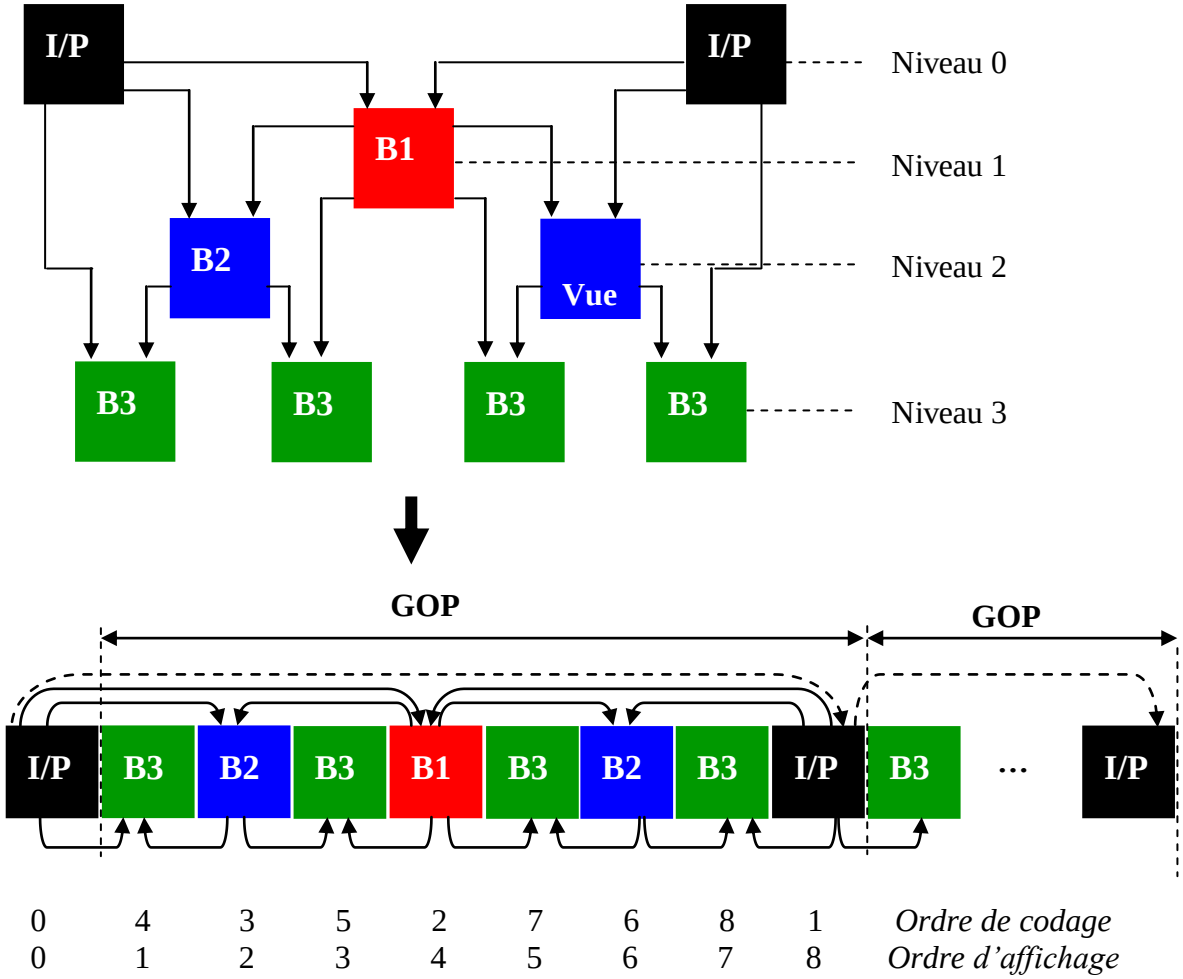


Figure 3.1. *La structure de prédiction "images B hiérarchique".*

En effet, l'ordre de codage diffère de l'ordre d'affichage. Les images de référence sont toujours codées avant qu'elles ne soient utilisées pour l'estimation et la compensation de mouvement. Tout d'abord, l'image clé du GOP courant est codée, ensuite, les images restantes sont codées selon le niveau hiérarchique. L'ordre d'encodage des images B ayant le même

niveau hiérarchique est désigné dans ce cas-là par l'ordre d'affichage. Autrement dit, la première image dans l'ordre temporel est premièrement codée à partir des images B du niveau supérieur (voir la figure 3.1). L'exemple de la Figure 3.1, montre que la première image codée est l'image clé (image n ° 8. dans l'ordre d'affichage). Puis les images B selon les niveaux hiérarchiques (image numéro 4, 2 et 1…) sont ensuite codées.

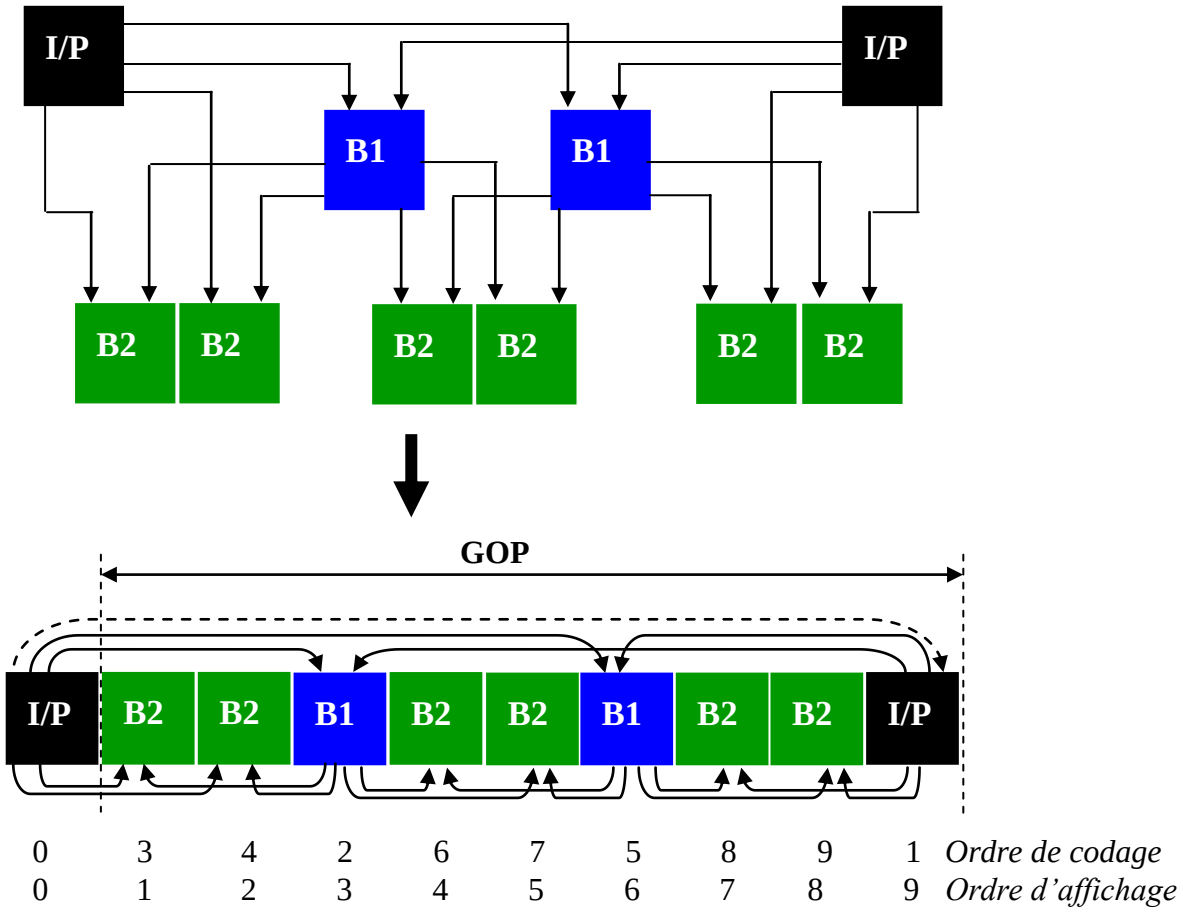


Figure 3.2. *La structure de prédiction "images B hiérarchique" non dyadique.*

Une contrainte très importante influant directement sur la conception de la structure HBP est celle de la mémoire tampon DPB maximale nécessaire (voir la section 2.5 pour plus de détails). Toutefois, l'augmentation du nombre des niveaux hiérarchiques dans chaque GOP agrandie le nombre des images de référence nécessaires. Une autre exigence sur laquelle se base le choix de la conception la plus adéquate est le faible retard discuté dans le deuxième chapitre (définies dans [9] en détails). L'utilisation des images B hiérarchique peut ne pas être fiable vis-à vis de cette contrainte. Pour cela une autre structure hiérarchique fondée sur les images P (des images P hiérarchique) est aussi définie. Dans cette structure chaque image P n'a besoin que d'une seule image de référence pour sa prédiction et estimation de mouvement.

Un exemple de cette structure est présenté dans la figure 3.3 où l'ordre d'encodage est le même que l'ordre d'affichage.

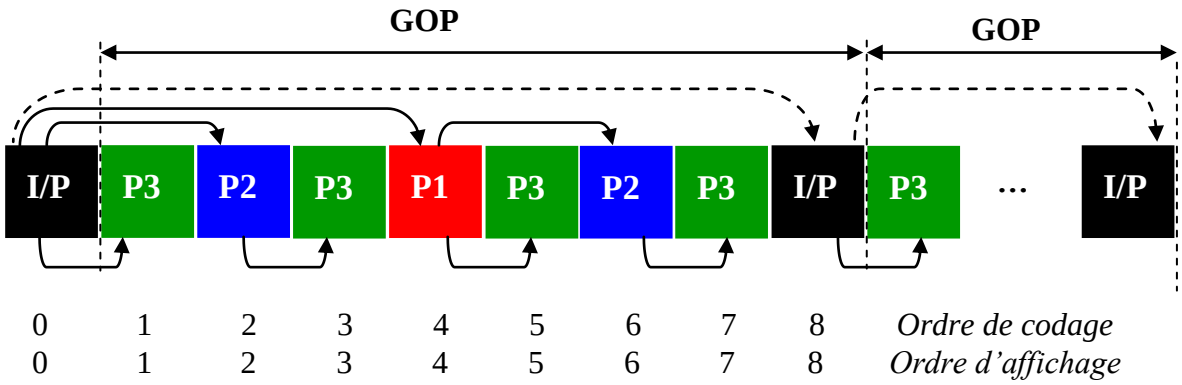


Figure 3.3. *Structure de prédiction hiérarchique fondée sur les images P.*

3.2.2. La structure de prédiction Simulcast

Le codage Indépendant des différentes vues de la vidéo multi-vues en utilisant la même structure de prédiction comme « images B hiérarchique », est l'une des méthodes les plus simples de la compression de la vidéo multi-vues. Cette méthode est appelée la compression "Simulcast". Dans ce cas, chaque groupe de groupes d'images (GGOP ; Group of Group of pictures) est composé de huit vues (dans le cas où la vidéo multi-vues est composée de huit vues) et 8 images par GOP. Dans toutes les vues, les images I sont appelées images clés *(anchor picture)*. Tandis que les images B, situées entre deux images I, sont appelées images non-clés (*non anchor pictures*). Dans la figure 3.4, les T_n représentent le temps pendant que les S_n indiquent les différents points de vue. Un exemple de la structure de prédiction Simulcast où chaque GOP est composé de 8 images est illustré dans la figure 3.4.

Cette méthode a l'avantage d'être simple à mettre en œuvre et moins coûteuse en temps de calcul. Cet avantage est dû principalement au fait que la compression Simulcast n'exploite pas de la similarité inter-vues. Autrement dit, que la prédiction et la compensation de disparité ne sont pas utilisées dans ce cas. Un autre avantage très important de cette structure de prédiction est qu'elle permet un accès aléatoire inter-vues très rapide. La rapidité en accès aléatoire inter-vues obtenue par cette structure est justifiée par le fait, que toutes les images clés des différentes vues sont *Intra*-codées. L'inconvénient majeur de cette approche est qu'elle est moins efficace en termes de compromis entre le débit binaire et la qualité de la vidéo encodée.

L'augmentation en débit binaire est due aussi aux images *Intra*-codées qui nécessitent un débit binaire élevé par rapport aux images P et B.

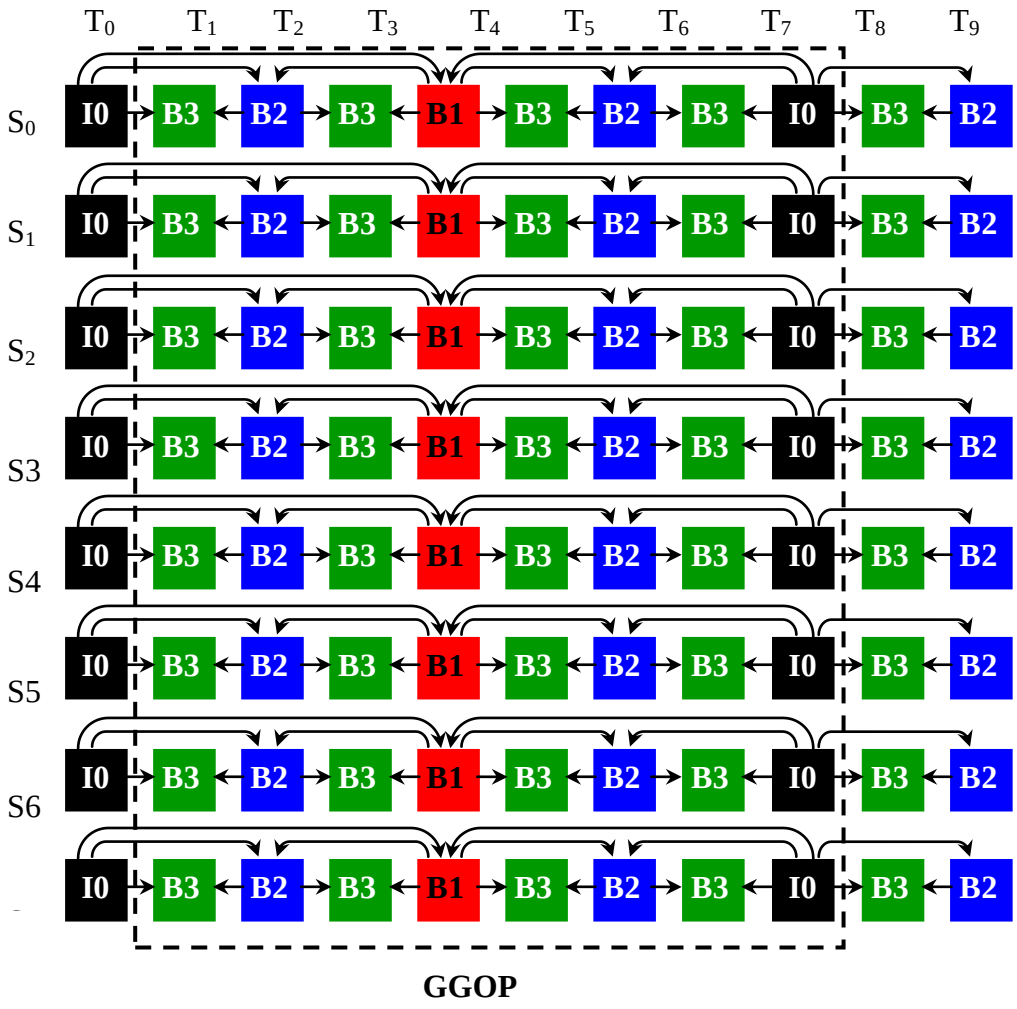


Figure 3.4. *La structure de prédiction Simulcast.*

3.3. Prédiction inter-vues

Contrairement à la méthode de compression "Simulcast" où uniquement le codage temporel est impliqué, les structures de prédiction inter-vues profitent de la corrélation contenue dans les vues voisines en appliquant une prédiction mixte (avec l'estimation et la compensation de mouvement et de disparité). Les structures de prédiction inter-vues utilisent généralement une seule image *Intra*-codée (image I; qui représente l'image clé de la vue qu'elle contient) pour chaque GGOP. La séquence qui contient les images clés de type I est codée sans l'intervention de la compensation de disparité (seulement par codage spatial et temporel). Les images P et B peuvent aussi être des images clés lorsqu'elles figurent au début des différentes séquences. La vue contenant l'image I (c'est l'image IDR du GGOP) est

appelée la vue de base. Les autres vues utilisent seulement une prédiction inter-vues pour ses images clés et combinent les deux types de prédiction (temporelle et inter-vues) pour le codage des différentes images non-clés.

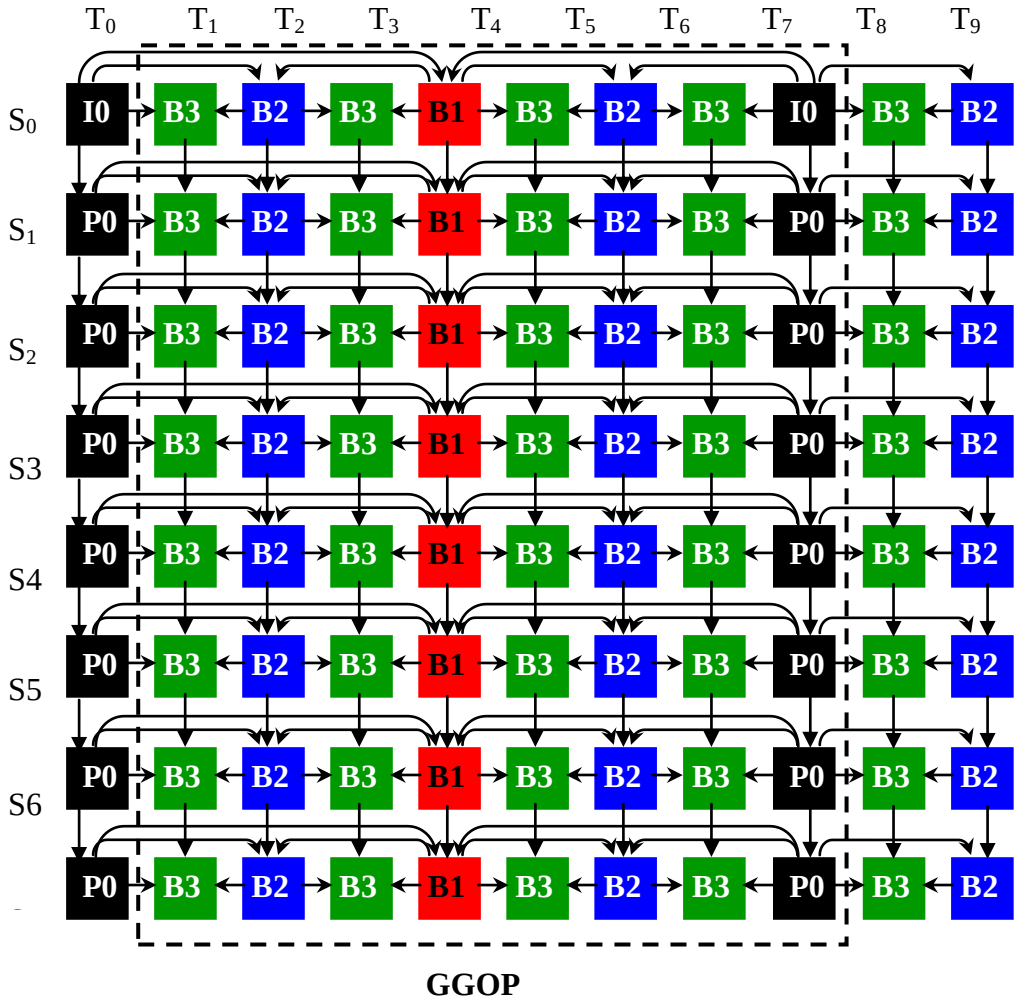


Figure 3.5. *Un exemple de la structure de prédiction IPP, la taille de chaque GGOP est égale à 8x8 (huit vues et huit images par GOP).*

3.3.1. La structure de prédiction IPP

Plusieurs structures de prédiction inter-vues sont mises en œuvre selon un certain nombre de critères tels que, le nombre des images de référence utilisées que ce soient des images clés ou non-clés. Le type des images clés utilisées par les différentes vues ou séquences, qui doivent être de type I, P ou B, est aussi un critère important pour la conception d'une structure de prédiction. Une vue est appelée généralement par le type de son image clé. Les vues possibles sont alors vue-I, vue-P ou vue-B. La structure de prédiction, appelée IPP [36], est l'une des structures pouvant assurer une efficacité de compression importante représentée par le gain significatif en débit binaire. La principale caractéristique de cette structure de

prédiction inter-vues est que les images clés des différentes vues autres que la vue de base (la vue-I), sont de type P. Ces vues sont appelées les vues-P. En d'autres termes, toutes les vues commencent par des images de type P, à l'exception de la vue de base. La prédiction de l'image clé est toujours faite à partir de l'image clé de la vue précédente soit de type vue-P ou vue-I. En effet, la prédiction des images non-clés dans les vues-P est un peu différente. En effet, elles utilisent une prédiction mixte temporelle/inter-vues. Les images non-clés dans ce cas sont prédites à partir de trois images. Deux images sont au niveau temporel et la troisième est celle de la vue précédente (prédiction inter-vues). Un exemple de cette structure est présenté dans la figure 3.5 (8 vue et 8 images par GOP).

Si on compare les résultats obtenus par cette structure de prédiction (IPP) avec la structure Simulcast en termes d'efficacité de compression, un gain considérable est assuré. Cependant, l'utilisation de plusieurs vue-P successives dans cette structure avec une seule vue-I par GGOP provoque un accès aléatoire inter-vues très lent vis-à-vis la structure de prédiction Simulcast.

3.3.2. La structure de prédiction JMVM (IBP)

Une autre structure de prédiction qui exploite la corrélation inter-vues est proposée par [36]. Cette structure est appelée IBP où chaque lettre exprime un type de vue. Elle est choisie et utilisée comme la structure par défaut dans le modèle de référence JMVM. Tout comme le cas de la structure IPP, l'IBP utilise également une seule vue de base celle qui contient les images de type I (*Intra*-coded) par GGOP. Toutefois, cette structure utilise trois types de vues au lieu de deux, comme dans le cas de la structure IPP; vue I, vue-P. Ces trois vues sont la vue de base (vue-I), vue-P et vue-B. Les images clés dans le cas des vues-B sont de type B (bidirectionnel). Dans ce modèle, chaque vue-B doit être située entre une vue-I/P et une autre de type vue-P. Les vues-B de la structure présentée dans la figure 3.6 sont S1, S3 et S5. Contrairement à la structure de prédiction IPP qui utilise au maximum trois images pour la prédiction de ses images non-clés, chaque image non-clé de chaque vue-B dans la structure IBP doit être codée en utilisant quatre images de référence. Il s'agit de deux images dans le niveau temporel et deux images voisines dans les vues de références. Néanmoins, les images clés des vues-B utilisent seulement une prédiction inter-vues (deux images de la vue précédente est suivante). Les vues de référence, pouvant être utilisées pour la prédiction par les autres vues, dans cette structure présentée dans la figure 3.6 sont S0, S2, S4 et S6. Les images clés des vues-P utilisent dans cette structure une seule image de référence. Il s'agit de

l'image clé de la vue précédente de type vue-I/P. Afin de réduire la complexité calculatoire, les images non-clés des vues-P sont prédites seulement à partir de deux images de référence du niveau temporel. Voici quelques exemples qui récapitulent ces notions de bases :

- L'image S1/T6 non-clé (de la vue-B) dans la figure 3.6 est prédite à partir des deux images de référence S1/T4 et S1/T8 du niveau temporel en plus de S0/T6 et S2/T6 au niveau inter-vues. Ces images de référence sont elles-mêmes des images non-clés.

- L'image S2/T6 non-clé (de la vue-P) dans la même figure est codée en utilisant seulement deux images de référence S2/T4 et S2/T8 dans le niveau temporel.

- L'image S1/T0 clé (une vue-B) est prédite à partir de deux images clés (dans l'axe inter-vues), S0/T0 (de la vue de base) et S2/T0 (d'une vue-P).

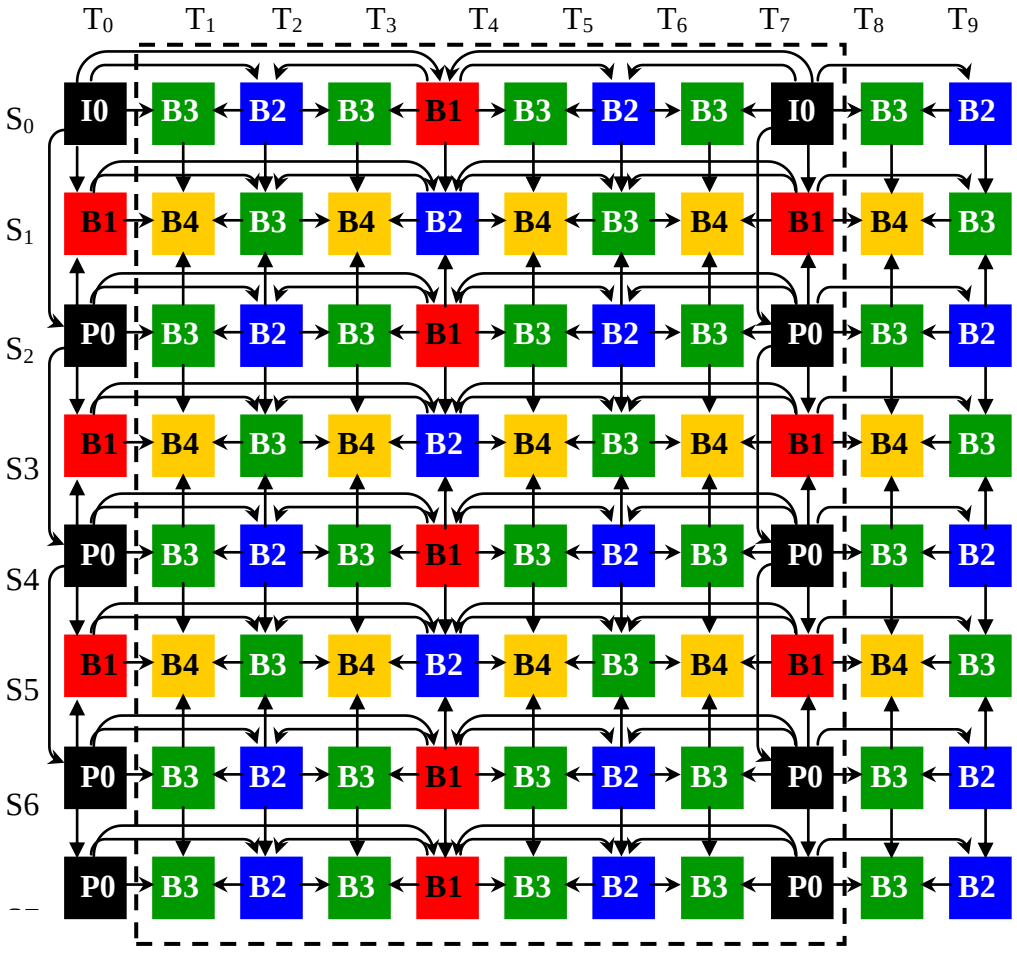


Figure 3.6. *Un exemple de la structure de prédiction IBP où chaque GGOP est composé de huit vues et 8 images par GOP.*

Un cas particulier de la structure IBP se distingue lorsque le nombre total de points de vue est pair. Dans ce cas-là, deux vues-P doivent être implémentées successivement (la vue S6 et S7 de la figure 3.6). La particularité de la dernière vue S7 est qu'elle utilise trois images de référence au lieu de deux (le cas des vues-P qu'elle précède) pour la prédiction de ses images non-clés.

L'utilisation de quatre images de référence (le cas des vue-B) au lieu de deux ou trois images pour encoder une image non-clé possède une efficacité de compression très importante par rapport à la compression Simulcast et similaire à la structure IPP. L'accès aléatoire inter-vues est aussi amélioré d'une façon considérable en comparaison avec la structure IPP. Seulement, cette caractéristique produit une complexité accrue, d'où la nécessité de décodage d'un nombre maximum pour pouvoir accéder à une image donnée Sn/Tn défini par la formule 3.2. Ceci, nécessite toujours l'amélioration de l'accès aléatoire inter-vues.

3.4. Spécificités des structures de prédiction MVC

Plusieurs caractéristiques spécifiques aux structures de prédiction inter-vues sont envisageables. Ces propriétés peuvent être classées par exemple en fonction de l'ordre d'organisation et d'encodage qui définit la conception du flux vidéo avant et lors de la compression. Ensuite, nous présentons les différentes propriétés concernant les GGOP.

3.4.1. Ordre d'organisation et d'encodage dans le MVC

Les structures de prédiction inter-vues présentées précédemment peuvent être implémentées par toute norme de codage vidéo mais avec une taille étendue de la DPB. La raison principale pour laquelle la taille de la DPB a été augmentée dans l'extension MVC, est que le nombre des images de référence pour chaque image n'est pas le même. En effet, il peut aller jusqu'a quatre images pour les images B. Le flux vidéo non compressé à utiliser comme entrée de l'encodeur doit respecter un ordre spécifique sur lequel se base la désignation de la taille DPB. L'organisation du flux vidéo non compressé permet de combiner les différentes vues à coder en un seul flux vidéo non compressé. Cette organisation est effectuée suivant deux types de parcours définis par le groupe JVT [47] où les différentes images constituant chaque GGOP sont stockées dans le flux en entrée tout en respectant le parcours utilisé.

Le premier type d'organisation appelé *view-first* [54] commence premièrement par le stockage des images clés des différentes vues. Puis comme son nom l'indique, il respecte dans chaque GOP l'ordre des images qu'il contient et dans chaque GGOP l'ordre des vues composant la MVV. La figure 3.7 illustre l'ordre de stockage en utilisant cette technique appliquée sur un GGOP de huit vues et huit images par GOP. L'ordre de stockage des images de chaque GGOP dans le flux en entrée est spécifié dans le deuxième type par l'axe temporel. Cette méthode est appelée *time-first* [55]. La figure 3.8 montre un exemple de parcours sur lequel se base cette organisation. Dans cet exemple les images des vues sont stockées les unes après les autres sans tenir compte de l'organisation du GOP. Dans la structure de prédiction IPP par exemple, le mieux est d'utiliser le deuxième type car les vues dans cette structure sont codées successivement. Ceci, n'empêche pas que l'encodeur et le décodeur choisissent n'importe quels types d'organisation pour toutes structures de prédiction inter-vues.

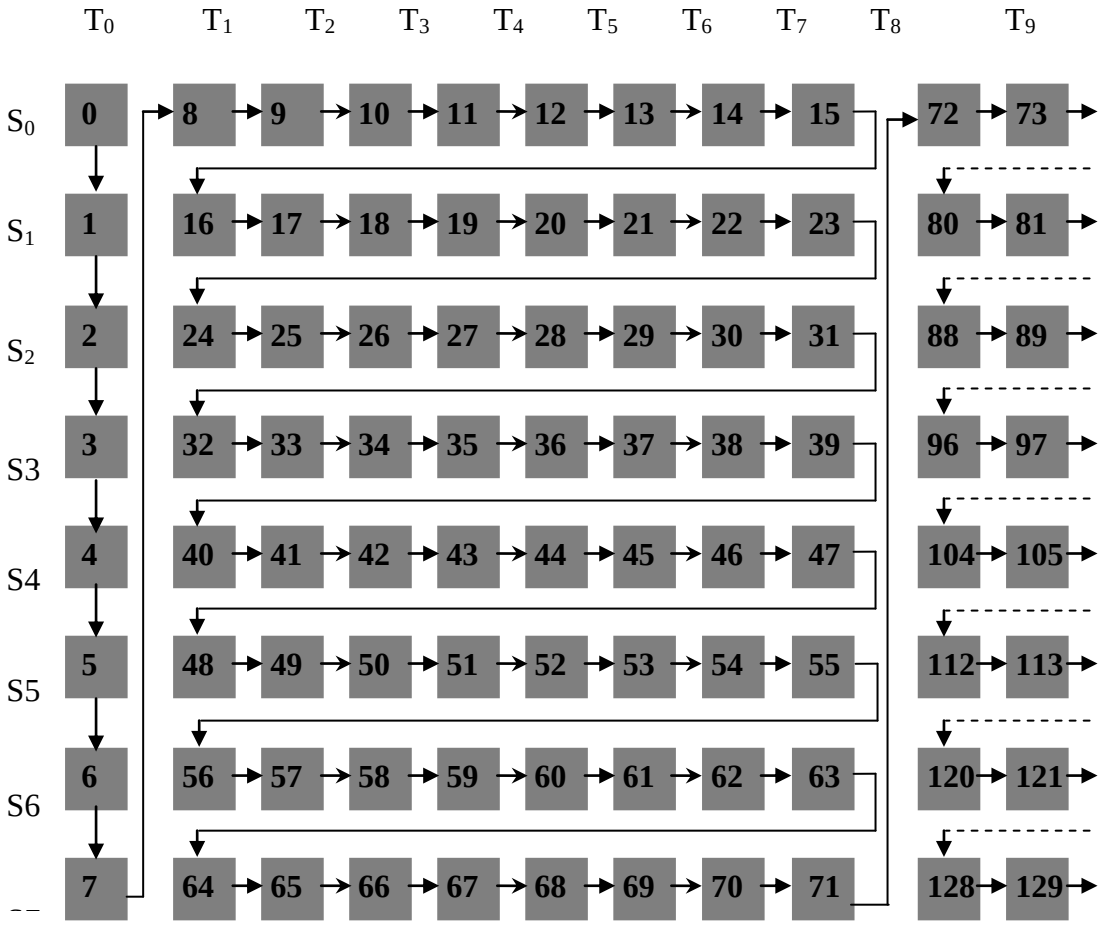


Figure 3.7. *Organisation view-first pour le stockage des images dans le flux vidéo non compressé en entrée de l'encodeur.*

La sélection des images de référence et la gestion de la mémoire de chaque structure de prédiction sont contrôlées par des paramètres appropriés de l'encodeur [37]. En outre, l'encodeur de la norme H.264/AVC utilise tout simplement une spécification de syntaxe de haut niveau pour signaler que le flux binaire représente une séquence multi-vues. Ensuite, le décodeur peut définir la taille appropriée DPB, décoder le flux binaire avec les outils existants, et arrive à inverser la réorganisation du flux vidéo présenté dans de la figure 3.7 comportant toutes les vues.

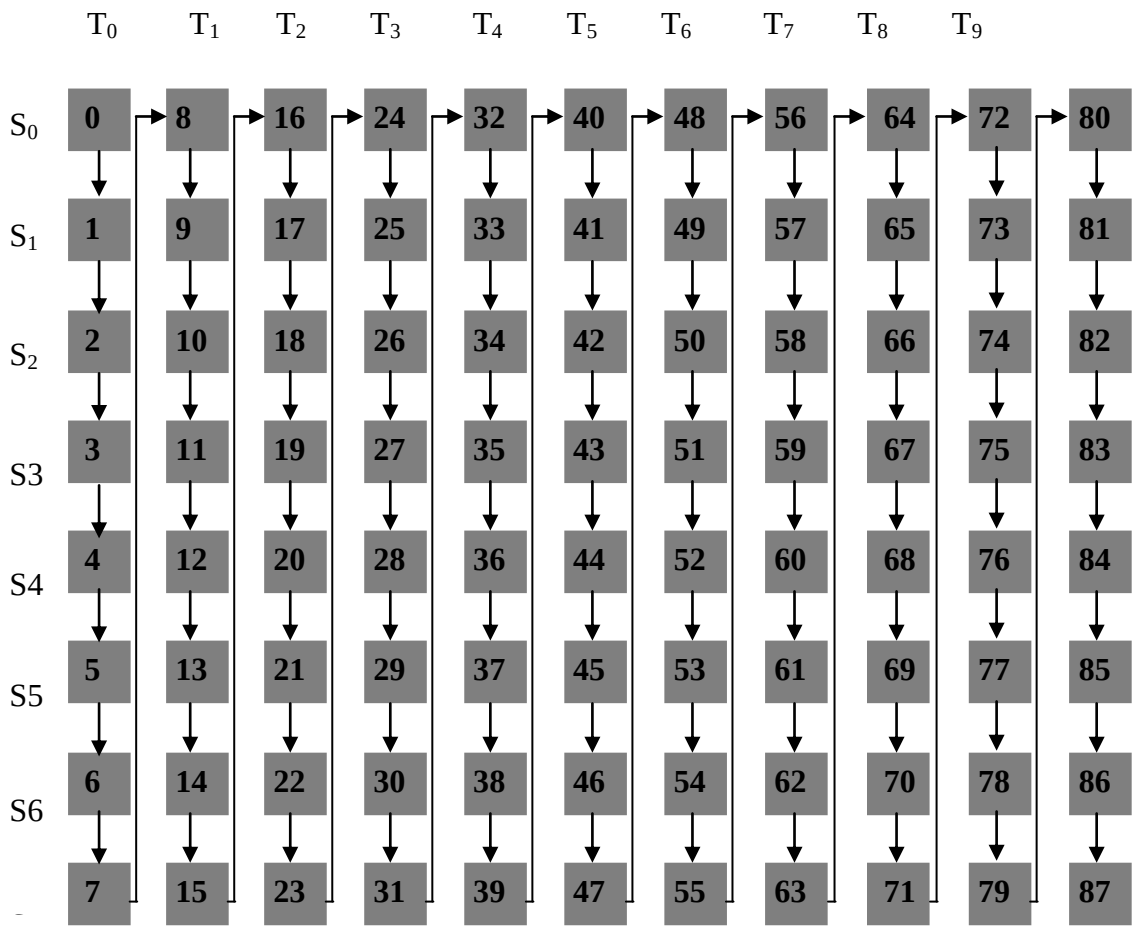


Figure 3.8. *Organisation time-first pour le stockage des images dans le flux vidéo non compressé en entrée de l'encodeur.*

En effet, l'ordre d'encodage des images de chaque vue n'est pas le même que l'ordre d'organisation. L'ordre d'encodage varie suivant la structure de prédiction utilisée. Dans le cas de la structure Simulcast et IPP par exemple les vues sont codées les unes à la suite des autres tandis que la structure IBP exige un ordre spécifique. La première vue à coder dans le cas de cette structure de prédiction (IBP) est la vue de base. Les vues-P doivent ensuite être codées pour être utilisées comme références pour les vues-B codées à la fin du processus

d'encodage. Elles ne servent de références pour aucune vue dans cette structure. C'est-à-dire l'ordre d'encodage dans le cas de la figure 3.6 est S0, S2, S1, S4, S3, S6, S5 et S7.

3.4.2. La structure du GOP et de la prédiction inter-vues

Dans la compression de la vidéo multi-vues, toutes les vues doivent avoir la même taille du GOP. En effet, la taille du GOP peut varier suivant plusieurs critères tels que la fréquence trame de la séquence à encoder. Le choix de la taille de chaque GOP est primordial. A titre d'exemple, l'agrandissement du GOP peut accroitre le niveau hiérarchique temporel. La figure 3.9 (b) illustre un exemple de GOP de taille 15. Les niveaux hiérarchiques temporels et également inter-vues sont augmentés par 1. Le cas d'un GOP de 12 images (voir la figure 3.9 (a)) le niveau hiérarchique reste le même avec un nombre plus élevé d'images B de plus haut niveau.

Séquences	**Partitionnement temporel**		
	Nombre d'images	**Nombre de GOP**	**Taille dernier GOP**
Ballroom	250	20*GOP_12	9 images
Exit	250	20*GOP_12	9 images
Vassar	250	20*GOP_12	9 images
Race1	250	16*GOP_15	9 images
Rena	300	19*GOP_15	14 images

Tableau 3.1. *Partitionnement temporel des données de test.*

Les structures de prédiction présentées ci-dessus ainsi que celles de la figure 3.9 peuvent être appliquées directement à des vidéos multi-vues capturées par des caméras avec un arrangement 1D. Cependant, le codage d'autres vidéos multi-vues capturées par des caméras avec un arrangement tableau 2D nécessite une légère modification de ces structures (voir la figure 3.10). Néanmoins, nous nous sommes intéressés dans cette thèse qu'aux vidéos multi-vues avec un arrangement 1D des caméras.

Le tableau 3.1 présente la structure de partitionnement temporel pour chaque ensemble de données. Il s'agit des séquences multi-vues présentées dans le chapitre 02. Comme le montre

le tableau 3.1, le dernier groupe est codé suivant le nombre d'images restant, c'est-à-dire avec un niveau hiérarchique différent.

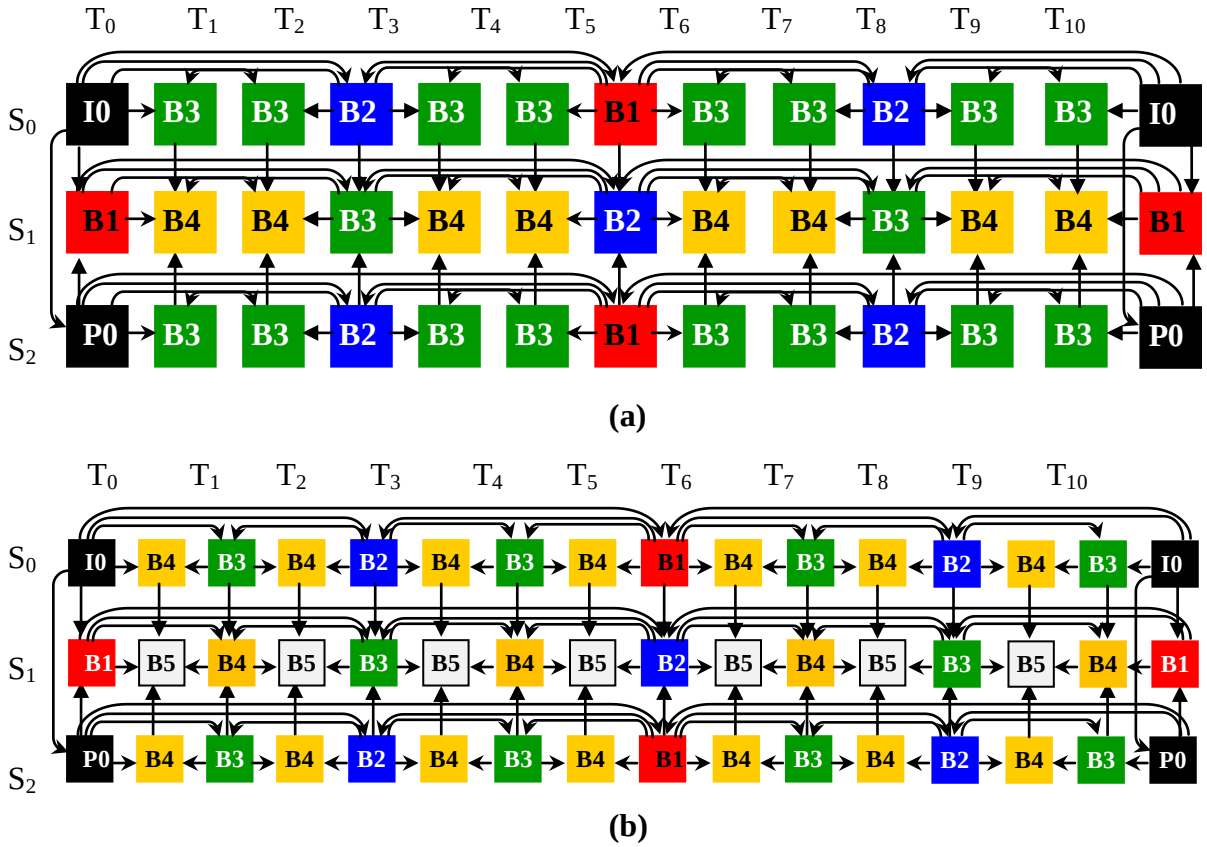


Figure 3.9. *Structure de prédiction IBP, (a) taille GOP est 12, (b) taille GOP égale 15.*

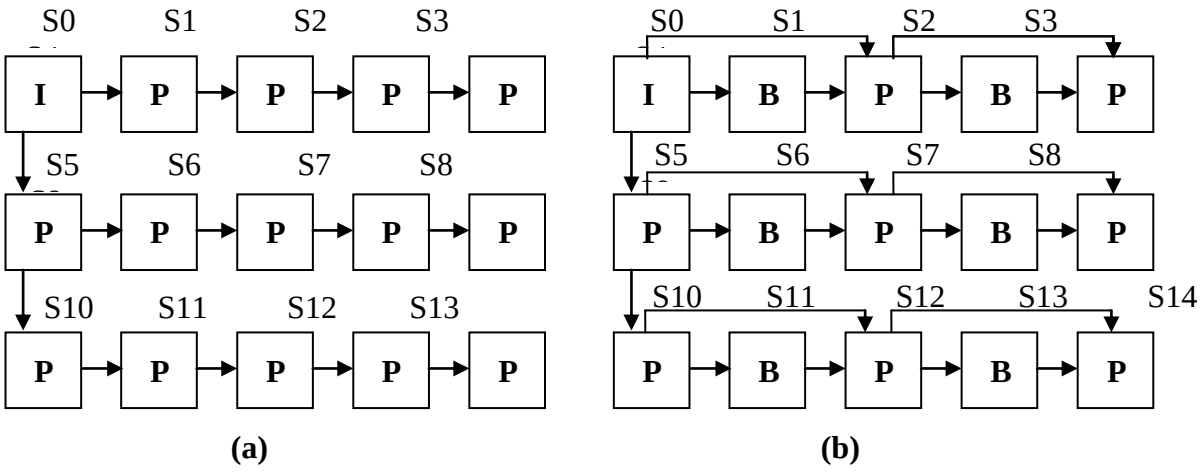


Figure 3.10. *Les structures de prédiction avec un arrangement tableau 2D des caméras de capture, (a) adaptation à la structure IPP (seulement la relation entre les vues), (b) adaptation à la structure IBB (seulement la relation entre les vues).*

3.5. Evaluation des structures de prédiction

La robustesse d'une structure de prédiction inter-vues peut être estimée par plusieurs méthodes. La plus importante contrainte à évaluer est l'efficacité de la compression représentée par un compromis débit binaire et la qualité de la vidéo. L'accès aléatoire inter-vues qui doit être le plus rapide possible est aussi une exigence très importante. Cette contrainte peut être évaluée par la rapidité d'accès à une image donnée à une instance T_n. Bien que plusieurs autres contraintes soient aussi importantes.

3.5.1. La qualité de la vidéo

L'évaluation des algorithmes de codage de la vidéo en général en termes d'efficacité de compression peut être obtenue en utilisant des mesures objectives et subjectives. La mesure objective la plus couramment utilisée en codage vidéo est le ‘*'Peak Signal-to-Noise-Ratio*''. Cependant, cette métrique ne prend pas en compte la dimension temporelle. Le PSNR est généralement calculé en utilisant le signal de luminance. Les composantes de chrominances peuvent être aussi utilisées pour l'évaluation. Le PSNR est donné par:

$$PSNR \quad = 10.log_{10}\left(\frac{255^2}{MSE}\right) \tag{3.1}$$

Avec la MSE (*Mean Squared Error*) ou erreur quadratique moyenne entre la vidéo originale et la vidéo compressée. En règle générale, le PSNR est déterminé en fonction du débit binaire, ou bien encore en fonction du pas de quantification utilisée pour l'obtention de débit binaire désiré, afin d'évaluer l'efficacité de la compression de la vidéo. Dans le cas du MVC, cette métrique est obtenue par le calcul de la moyenne des PSNR des différentes séquences ou vues. Il est à noter qu'une bonne qualité de la vidéo compressée est interprétée par une valeur élevée du PSNR. Les valeurs typiques pour le PSNR dans la compression d'images et de la vidéo se situent entre 30 et 50 dB (décibel).

Cependant, cette mesure de distorsion présente l'inconvénient majeur de n'être basée que sur le signal. Les propriétés du système visuel humain ne sont pas donc prises en compte par le PSNR. En conséquence, une évaluation subjective de la vidéo encodée est parfois nécessaire. Ce genre d'évaluation se base sur les caractéristiques subjectives de la qualité de la vidéo. Elle est préoccupée par la façon dont la vidéo est perçue par un observateur et désigne son opinion sur une séquence vidéo en particulier. Dans ce cas les vidéos à évaluer peuvent être accompagnées de leurs versions originales. En conséquence, ces tests subjectifs

exigent un effort énorme en termes de temps et de ressources humaines. Les tests subjectifs d'évaluation de la qualité peuvent être divisés en trois grandes familles : Les méthodes comparatives, les méthodes à simple stimulus et les méthodes à double stimulus. Les conditions de déroulement de ces différentes méthodes pour le MVC sont spécifiées dans la Recommandation UIT -R Rec . BT.500 -11 [56].

Nous nous sommes intéressés seulement au PSNR dans l'évaluation des méthodes proposées en termes de qualité. Nous pouvons justifier ce choix par le fait que cette mesure est utilisée pour des raisons comparatives avec d'autres approches où les mêmes paramètres et séquences sont utilisés.

3.5.2. Accès aléatoire inter-vues

L'accès aléatoire inter-vues le plus rapide possible est une autre recommandation très importante pour les schémas MVC. L'évaluation de l'accès aléatoire temporel est désignée par la capacité d'une norme à accéder et décoder une image donnée à une instance de temps T_n. Le même principe est utilisé dans le cas du MVC en ajoutant la contrainte de multiple vues ou séquences. L'utilisation d'une image *Intra* (de type I) par GGOP (la taille de chaque GGOP est égal à la *taille_GOP*nombre_vues)* peut permettre d'accélérer relativement l'accès aléatoire inter-vues. Le groupe JVT a choisi d'évaluer cette exigence en calculant le nombre maximum d'images à décoder pour accéder à une image donnée dans un GGOP. Ce nombre dépend essentiellement de plusieurs paramètres comme le nombre total de vues utilisées et le niveau hiérarchique maximum. Il s'agit du niveau hiérarchique inter-vues et non pas temporel. Dans le cas de la structure de prédiction IBP, l'image nécessitant le nombre maximum est une image B ayant le plus bas niveau hiérarchique et se trouve dans la vue-B la plus éloignée de la vue de base (voir la figure 3.11). Le nombre maximum d'images à décoder est donné pour la structure IBP comme suit :

$$N_{MAX} = 3 * Hierarchy_{MAX} + 2 * \lfloor (Nbr_{views} - 1)/2 \rfloor \quad (3.2)$$

Le niveau hiérarchique le plus bas dans la structure IBP est représenté par $Hierarchy_{MAX}$. Quand à Nbr_{views} il représente le nombre total de vues utilisées. Par exemple, afin d'accéder à l'image B4 (S5/T7) sur la figure 3.11, nous avons besoin de décoder les 18 images nécessaires pour la prédiction dans l'ordre hiérarchique. Ces images peuvent être données selon le type comme suit:

- Images de type I : S0/T0, S0/T8

- Images de type P : S2/T0, S4/T0, S6/T0, S2/T8, S4/T8, S6/T8

- Images B1 : S5/T0, S5/T8, S4/T4, S6/T4

- Images B2 : S5/T4, S4/T6, S6/T6

- Image B3 : S5/T6, S4/T7, S6/T7.

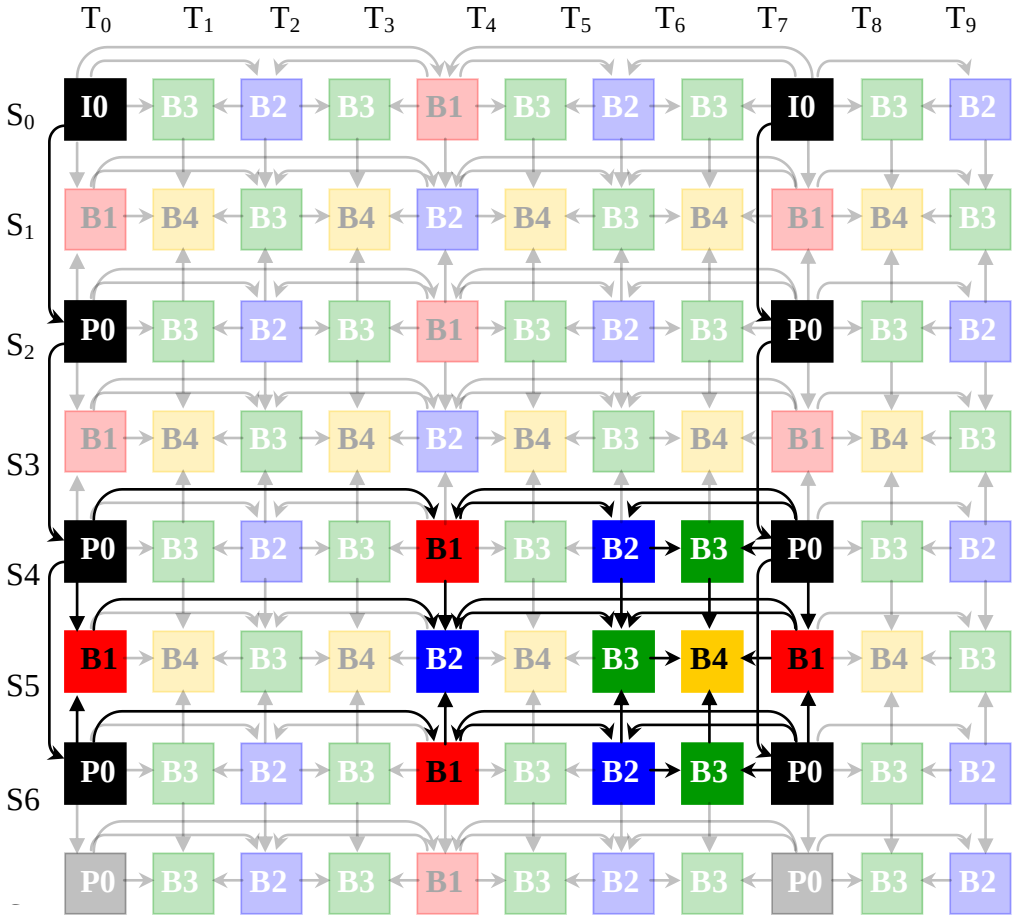


Figure 3.11. *L'accès aléatoire inter-vues pour la structure IBP.*

La figure 3.11 montre le même exemple (l'accès aléatoire inter-vues à l'image S5/T7) par ordre de décodage. Les images B ayant le même niveau hiérarchique, figurant dans la même vue (S5/T5 , S5/T3, S5/T1), nécessitent également dans cet exemple 18 images à décoder.

Si la taille du GOP est de 15 au lieu de 8, le niveau maximum de la hiérarchie augmente de 1. Dans ce cas le nombre maximum à décoder nécessaire pour accéder à une image augmente de trois. Le nombre des images clés (images I et P) dans cet exemple reste le même, les images ajoutées sont de type B. qui sont de type image-B (image-B4 selon cet exemple).

Le nombre des images clés à décoder n'augmente qu'avec l'accroissement du nombre de vues.

L'accès aléatoire inter-vues pour la structure IPP est plus lourd. Le nombre maximum N_{max} où la méthode de calcul n'est pas proposée par le JVT, est plus grand que dans la structure IBB. Afin de pouvoir comparer les approches proposées avec cette structure nous avons préconisé une méthode de calcul du N_{max} pour cette structure (voir le chapitre 04) en appliquant le même principe. La vue qui contient l'image nécessitant un nombre maximum à décoder pour son accès aléatoire inter-vues est toujours la dernière vue. La figure 3.12 illustre un exemple appliqué sur un GGOP de taille (*huit vues) * (huit images par GGOP),* le N_{MAX} est égal dans ce cas à 39 images.

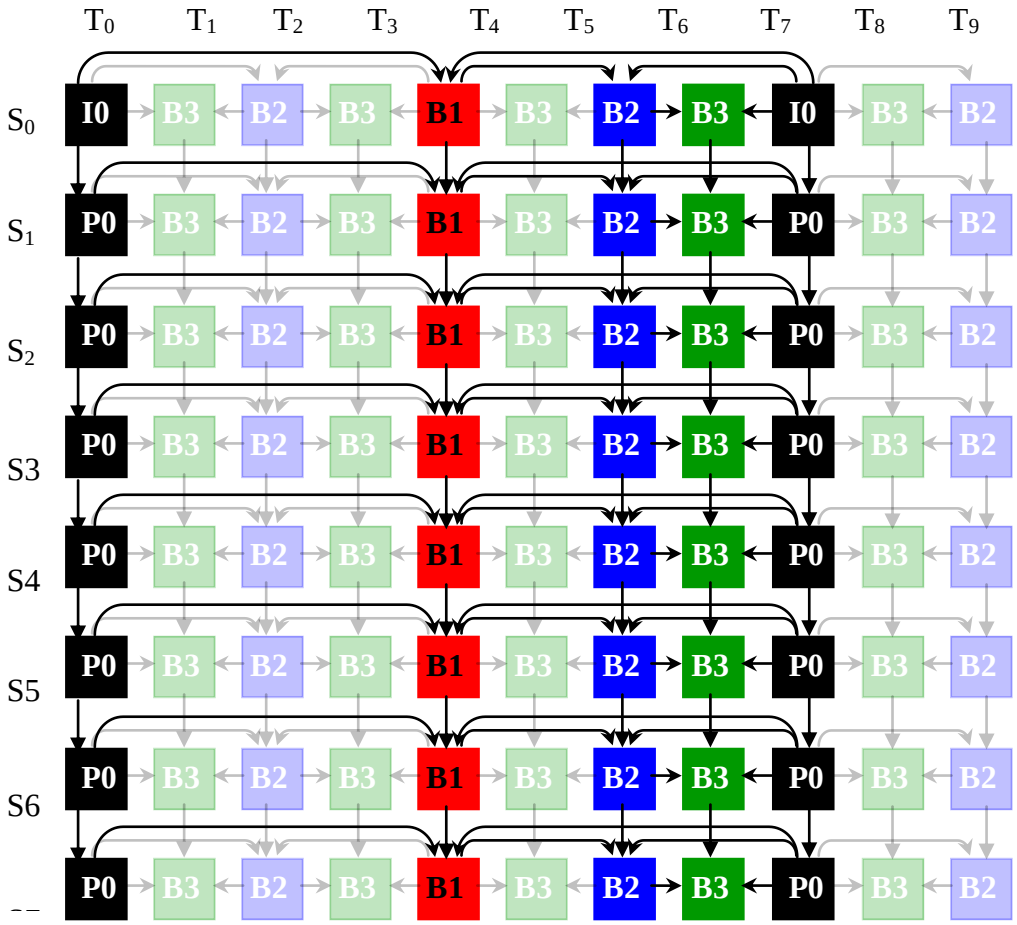


Figure 3.12. *L'accès aléatoire inter-vues pour la structure IPP.*

3.6. Expérimentation et résultats

L'ensemble des vidéos utilisées pour l'évaluation et l'analyse de la compression de la vidéo multi-vues, changent suivant plusieurs paramètres comme le nombre de caméras

utilisées lors de l'opération de capture, le type d'arrangement des différentes caméras, la résolution d'image et la fréquence d'affichage. Les propriétés des vidéos en termes de fréquence trame, résolution et options de capture utilisées pour l'expérimentation dans ce chapitre sont présentées dans le tableau 2.1 (chapitre 02). La figure 2.3 donne un exemple des 8 vues pour la vidéo « *Ballroom* » dans une seule instance de temps *Tn* tandis que les autres vidéos sont représentées dans la figure 2.4.

En effet, les structures de prédiction à base d'images B hiérarchique présentées dans la deuxième section de ce chapitre sont contrôlées par des paramètres appropriés de l'encodeur MVC. Le tableau 3.2 présente certains paramètres utilisés dans notre évaluation de la compression de la vidéo multi-vues. Le paramètre essentiel permettant le contrôle de la qualité de la vidéo compressée à travers plusieurs valeurs est celui de quantification (QP, quantization parameter), une étude comparative entre les différentes structures étudiées en utilisant ces paramètres est présentée dans [57].

Paramètres	**Valeur(s)**
Paramètre de quantification (QP)	22, 27, 32, 37, 40
Taille GOP	12, 15
Précision recherche	64
Codage entropique	CABAC
Filtrage anti-blocs	Activé
Ordre d'encodage Simulcast et IPP	S0, S1, S2, S3, S4, S5, S6, S7
Ordre d'encodage IBP	S0, S2, S1, S4, S3, S6, S5, S7

Tableau 3.2. *Paramètres d'évaluation utilisés.*

L'augmentation de ce paramètre amoindrit aussi bien la qualité de la vidéo compressée, en conduisant à une faible valeur de PSNR, et le débit binaire nécessaire. Dans cette section tous les résultats sont obtenus en fonction de la variation du QP.

Dans un souci de comparaison des résultats obtenus nous avons choisi 5 valeurs pour QP. La taille du GOP dépend de la fréquence de la vidéo. Nous avons utilisé la taille du GOP égale 12 pour les trois vidéos « *Ballroom* », « *Exit* » et « *Vassar* » où la fréquence trame est

de 25Hz. Alors que nous avons opté pour un GOP de taille 15 pour les vidéos « *Race1* » et « *Rena* » dont la fréquence trame est de 30. Le codage entropique utilisé est le CABAC (*Context-adaptive binary arithmetic coding*) car il produit d'excellents résultats en termes de compression mais avec une grande complexité. La précision de recherche spécifie la plage maximale pour la compensation de mouvement.

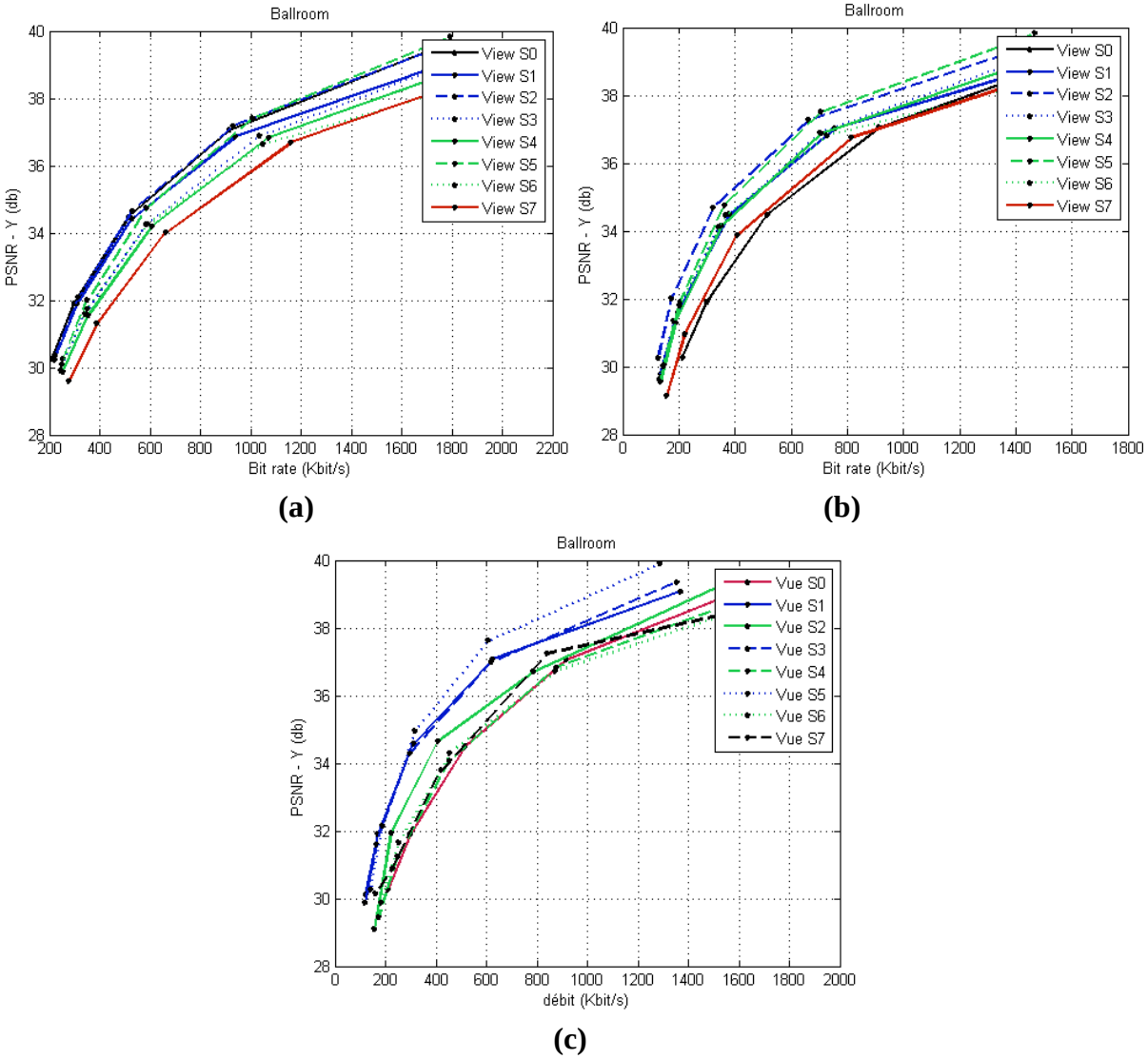


Figure 3.13. *Comparaison de l'efficacité de la compensation entre les différentes vues de la vidéo Ballroom, (a) pour la structure Simulcast, (b) pour la structure IPP, (c) pour la structure IBP.*

Les résultats obtenus montrent clairement l'avantage d'exploiter la compensation de disparité à partir de deux vues adjacentes en plus de la compensation temporelle. Les vues qui profitent de la compensation de mouvement dans la structure IBP sont S1, S3 et S5. Les tableaux présentés à la fin de ce chapitre récapitulent les résultats obtenus en termes de débit

binaire et PSNR pour les trois structures Simulcast, IPP et IBP en utilisant les cinq vidéos MVV de test. La figure 3.13 (c) illustre également le gain en débit binaire des trois vues par rapport aux autres avec une qualité de PSNR approximativement stable. Les autres vues S0, S2, S4 et S6 n'exploitent que la compensation temporelle à l'exception de l'image clef du GOP qui doit être obtenue à partir de l'image clef du GOP (de type I ou P) de la vue paire précédente. C'est ce qui explique le débit binaire élevé vis-à-vis des vues impaires (S1, S3 et S5). Le cas exceptionnel ici est le codage de la vue 7 qui n'utilise qu'une seule vue pour la prédiction inter-vues et nécessite donc un débit binaire plus élevé que celui des vues S1, S3 et S5. Le gain en débit binaire pour la structure IPP est obtenu pour les sept vues de S1 à S7, utilisant la prédiction inter-vues, par rapport à la vue de base S0 avec un PSNR qui diffère légèrement d'une vue à l'autre. La figure 3.13 (b) illustre le gain en débit des sept vues pour toutes les valeurs du QP utilisées. En effet, la différence entre les vues pour la structure Simulcast réside dans le PSNR comme le montre la figure 3.13 (a). Ceci est dû à la nature de chaque vue codée.

QP		**22**	**27**	**32**	**37**	**40**
Ballroom	Δbit-rate (%)	15.50	24.24	31.06	34.64	34.75
	ΔPSNR(dB)	0.02	0.08	0.01	-0.10	-0.19
Exit	Δbit-rate (%)	7.73	15.28	20.15	22.85	22.79
	ΔPSNR(dB)	-0.005	0.002	-0.04	-0.17	-0.23
Vassar	Δbit-rate (%)	4.34	14.30	25.75	38.27	43.40
	ΔPSNR(dB)	0.01	0.047	-0.03	-0.25	-0.38
Race1	Δbit-rate(%)	13.65	19.16	24.43	27.21	26.55
	ΔPSNR(dB)	0.05	0.09	0.06	-0.04	-0.09
Rena (8 vues)	Δbit-rate (%)	18.30	24.78	28.94	28.24	25.67
	ΔPSNR(dB)	0.05	0.05	-0.14	-0.37	-0.47

Tableau 3.3. *L'évaluation de l'efficacité de la compression de la structure IBP par rapport à la structure de prédiction Simulcast.*

Par comparaison entre la structure IBP et simulcast, utilisant seulement une prédiction temporelle, un gain d'environ 43.40 (en utilisant la vidéo « *Vassar* ») de la structure IBP est obtenu pour une qualité très dégradée de la vidéo compressée. Le tableau 3.3 récapitule les

résultats obtenus pour les 5 vidéos de test. En effet, la structure de prédiction IPP offre un gain dans le débit binaire pouvant atteindre 10.52, en utilisant la vidéo « *Race1* », par rapport à la structure de prédiction IBP. Néanmoins, la structure IPP est très lourde en accès aléatoire inter-vues (voir la figure 3.12) en comparaison avec la structure IBP. Le gain en débit binaire dans ce cas, est dû à la prédiction de sept vues par compensation de disparité dans la structure IPP avec seulement quatre vues pour la IBP. Dans les deux analyses la qualité mesurée par PSNR est approximativement similaire.

QP		**22**	**27**	**32**	**37**	**40**
Ballroom	Δbit-rate (%)	-0.77	-2.48	-4.22	-6.12	-6.58
	ΔPSNR(dB)	0.01	0.006	0.01	0.05	0.05
Exit	Δbit-rate (%)	0.29	-1.08	-2.89	-3.64	-3.41
	ΔPSNR(dB)	0.009	0.02	0.06	0.10	0.09
Vassar	Δbit-rate (%)	1.11	1.28	-0.70	-4.21	-4.70
	ΔPSNR(dB)	0.006	0.03	0.05	0.09	0.09
Race1	Δbit-rate(%)	-2.21	-3.67	-5.91	-9.33	-10.52
	ΔPSNR(dB)	-0.001	0.004	-0.0005	-0.02	0.01
Rena	Δbit-rate (%)	-4.29	-6.54	-9.02	-7.65	-5.01
	ΔPSNR(dB)	-0.02	-0.04	0.001	-0.01	0.01

Tableau 3.4. *Comparaison entre les deux structures de prédiction IBP et IPP en termes d'efficacité de compression.*

Le tableau 3.4 récapitule les résultats de la comparaison entre IPP et IBP en termes de gains en débit binaire et PSNR. Sachant que le gain en PSNR est donné par la formule 4 et en débit binaire par la formule suivante:

$$\Delta_{bit\_rate} = \frac{(bit\_rate - bit\_rate_{IBP})}{(bit\_rate)} \times 100[\%] \qquad (3.3)$$

Où *bit-rate* peut être le débit binaire de l'une des deux structures de prédiction IPP, ou Simulcast.

$$\Delta PSNR = PSNR\_IBP - PSNR \qquad (3.4)$$

Où PSNR représente celui de la structure IPP ou Simulcast. Les résultats synthétisés dans le tableau 3.3 et 3.4 montrent que la différence du PSNR entre les structures de prédiction

étudiées est toujours marginale. Ceci peut être expliqué par l'intérêt de l'utilisation de l'estimation et de la compensation de disparité parallèlement avec les techniques temporelles.

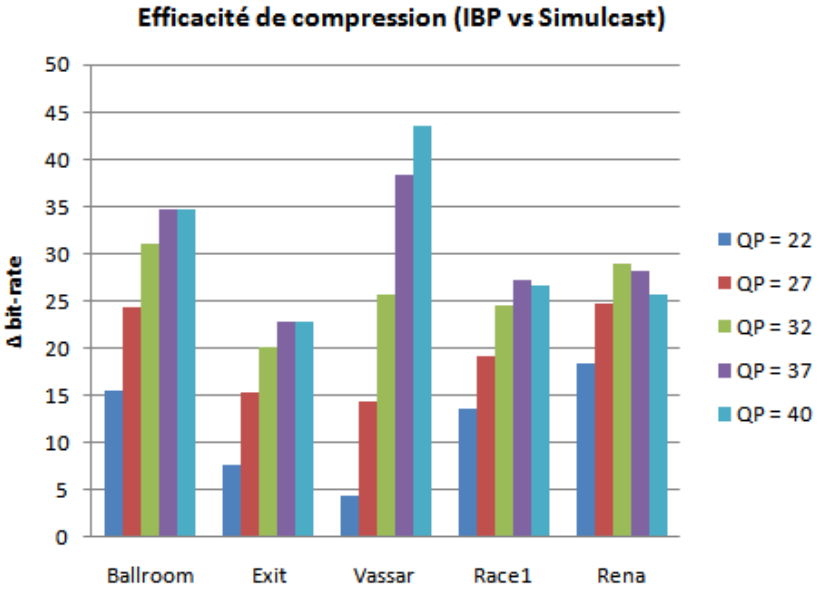


Figure 3.14. *Le gain en débit binaire apporté par une structure du MVC à base de prédiction mixte (IBP) par rapport à une structure à base de prédiction temporelle (Simulcast).*

Les résultats obtenus pour la comparaison entre les deux structures IBP par rapport à Simulcast et IBP par rapport à IPP sont représentés successivement par les graphs des deux figures 3.14 et 3.15. Les résultats illustrés par ces deux figures montrent clairement l'intérêt de cette analyse. Cette étude a été aussi démontrée dans plusieurs autres travaux [36].

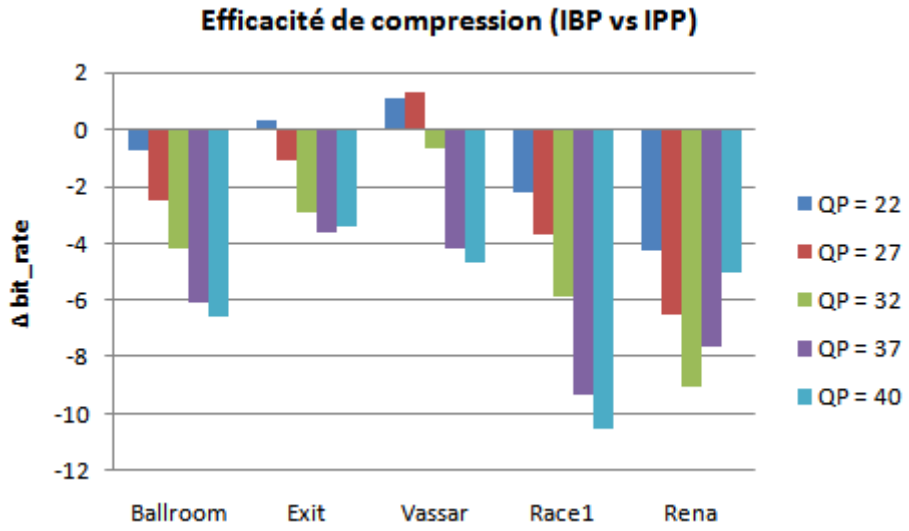


Figure 3.15. *Comparaison entre deux structures à base de prédiction mixte (IBP et IPP) en utilisant les 5 vidéos de test.*

	Vue	QP = 22	QP = 27	QP = 32	QP = 37	QP = 40
Débit binaire (Kbit/s)	S0	1692,5459	914,1892	513,7622	298,2216	211,2216
	S1	1767,0108	940,0378	526,4919	308,2486	219,827
	S2	1709,0054	927,9405	527,9189	312,6649	222,373
	S3	1894,2649	1032,4649	583,8162	339,6216	242,0703
	S4	1964,5243	1069,5459	605,7568	352,9297	252,3297
	S5	1792,1243	1006,7838	586,0054	349,1351	252,7189
	S6	1997,1568	1046,8595	590,3351	348	250,0432
	S7	2105,6162	1158,8054	660,6865	386,3892	278,1027
Moyenne		1865,2810	1012,0783	574,3466	336,90133	241,0858
PSNR (Db)	S0	39,3778	37,0531	34,4901	31,906	30,2651
	S1	39,0101	36,8649	34,4301	31,8994	30,2408
	S2	39,3997	37,1793	34,6639	32,1003	30,3681
	S3	39,3173	36,8894	34,262	31,607	29,9297
	S4	39,226	36,8354	34,2037	31,5477	29,8691
	S5	39,8269	37,4187	34,7471	32,0053	30,2612
	S6	38,7045	36,6417	34,2665	31,747	30,0958
	S7	39,1633	36,6944	34,0187	31,3111	29,6089
Moyenne		39,2532	36,9471	34,3852	31,7654	30,0798

Tableau 3.5. *Les résultats obtenus en termes de débit binaire et PSNR pour la structure de prédiction Simulcast en utilisant la vidéo Ballroom (en détails pour chaque vue).*

	Vue	QP = 22	QP = 27	QP = 32	QP = 37	QP = 40
Débit binaire (Kbit/s)	S0	1692,6	914,2432	513,8162	298,2757	211,2757
	S1	1562,6919	756,2486	380,5135	195,5135	133,5892
	S2	1428,773	660,6216	322,2	173,1838	124,3514
	S3	1507,827	701,8973	339,4865	181,3135	129,4703
	S4	1534,4378	706,8811	349,9514	187,6162	134,1243
	S5	1467,827	705,0595	360,6162	203,4757	146,2216
	S6	1621,8973	726,2378	366,0054	200,7838	144,3405
	S7	1695,8541	813,9243	406,4757	219,6703	157,173
Moyenne		1563,9885	748,1391	379,8831	207,4790	147,5682
PSNR (Db)	S0	39,3778	37,0531	34,4901	31,906	30,2651

	S1	39,0469	37,0272	34,4941	31,6023	29,7893
	S2	39,4134	37,2772	34,6982	32,0088	30,256
	S3	39,3094	36,9004	34,1189	31,3469	29,6315
	S4	39,2383	36,8874	34,1551	31,2853	29,5611
	S5	39,8389	37,5098	34,7643	31,898	30,0433
	S6	38,7467	36,8172	34,4706	31,795	30,0081
	S7	39,1533	36,7582	33,8861	30,9581	29,1521
Moyenne		39,2655	37,0288	34,3846	31,6	29,8383

Tableau 3.6. *Les résultats obtenus en termes de débit binaire et PSNR pour la structure de prédiction IPP en utilisant la vidéo Ballroom (en détails pour chaque vue).*

	Vue	QP = 22	QP = 27	QP = 32	QP = 37	QP = 40
Débit binaire (Kbit/s)	S0	1692,6108	914,2541	513,827	298,2865	211,2865
	S1	1367,8216	625,0865	310,173	170,0378	121,3081
	S2	1580,5459	785,427	406,6378	220,773	154,2324
	S3	1353,2595	616,3622	294,5784	161,1946	118,7081
	S4	1759,4919	874,1946	451,9027	247,4432	171,7622
	S5	1286,4324	606,3459	316,4378	186,4324	139,6108
	S6	1838,3459	871,7027	453,4541	251,6541	180,8811
	S7	1730,2703	840,3027	420,4054	225,7081	160,5135
Moyenne		1576,0972	766,7094	395,9270	220,1912	157,2878
PSNR (Db)	S0	39,3778	37,0531	34,4901	31,906	30,2651
	S1	39,0849	37,074	34,5801	31,8996	30,1211
	S2	39,4274	36,7093	34,657	31,9304	29,0972
	S3	39,3613	36,9891	34,2953	31,61	29,877
	S4	39,234	36,8421	34,0706	31,2374	29,4498
	S5	39,9006	37,6302	34,9535	32,1413	30,2803
	S6	39,1372	36,7258	34,3153	31,6494	29,8775
	S7	38,726	37,2584	33,8059	30,8933	30,1454
Moyenne		39,2811	37,0352	34,3959	31,6584	29,8891

Tableau 3.7. *Les résultats obtenus en termes de débit binaire et PSNR pour la structure de prédiction IBP en utilisant la vidéo Ballroom (en détails pour chaque vue).*

	Structure	MVV	QP = 22	QP = 27	QP = 32	QP = 37	QP = 40
Débit binaire (Kbit/s)	Simulcast	Ballroom	1865,2810	1012,0783	574,3466	336,9013	241,0858
		Exit	1028,9642	467,0398	251,6168	149,7250	107,6567
		Vassar	1218,2777	486,4810	231,7702	125,7337	85,4527
		Race1	2439,4082	1279,9239	699,3124	416,6634	306,4552
		Rena	910,4210	477,9475	259,6063	149,4137	108,4623
	IPP	Ballroom	1563,9885	748,1391	379,8831	207,4790	147,5682
		Exit	952,1952	391,4175	195,2344	111,4392	80,3702
		Vassar	1178,4182	422,3256	170,8668	74,4682	46,1898
		Race1	2060,5402	997,9317	498,9495	277,3989	203,6419
		Rena	713,1111	337,3969	169,1974	99,5830	76,7576
	IBP	Ballroom	1576,0972	766,7094	395,9270	220,1912	157,2878
		Exit	949,3439	395,6682	200,8932	115,4979	83,1189
		Vassar	1165,3020	416,8993	172,0716	77,6047	48,3628
		Race1	2106,2080	1034,5663	528,4486	303,2823	225,0802
		Rena	743,7515	359,4873	184,4687	107,2056	80,6106
PSNR (Db)	Simulcast	Ballroom	39,2532	36,9471125	34,3852	31,7654	30,0798
		Exit	40,0016	38,4777	36,6160	34,4013	32,7932
		Vassar	38,8594	36,9976	35,1565	33,1772	31,7573
		Race1	39,7215	36,9374	34,2293	31,5139	29,7681
		Rena	44,8534	42,0893	39,2146	36,4908	34,8662
	IPP	Ballroom	39,2655	37,0288	34,3846	31,6	29,8383
		Exit	39,9875	38,4534	36,5004	34,1196	32,4581
		Vassar	38,8729	37,0133	35,0648	32,8226	31,2718
		Race1	39,7795	37,02975	34,2978	31,4870	29,6634
		Rena	44,9342	42,1924	39,0682	36,1274	34,3704
	IBP	Ballroom	39,2811	37,0352	34,3959	31,6584	29,8891
		Exit	39,9966	38,4804	36,5690	34,2245	32,5569
		Vassar	38,8791	37,0449	35,1178	32,9184	31,3695
		Race1	39,7780	37,0339	34,2972	31,4653	29,6767
		Rena	44,908	42,1443	39,0699	36,1121	34,3890

Tableau 3.8. *Les résultats obtenus en termes de débit binaire et PSNR pour les trois structures de prédiction inter-vues étudiées en utilisant les 5 vidéos de tests (résultat globale de chaque MVV).*

3.7. Conclusion

En plus du compromis débit binaire et qualité de la vidéo, un compromis efficacité de compression et accès aléatoire inter-vues doit aussi être satisfait. La structure IBP représente la meilleure structure de prédiction inter-vues offrant un débit binaire acceptable avec un accès aléatoire inter-vues relativement rapide. Cette efficacité est due à l'exploitation de la corrélation inter-vues, contrairement à la structure de prédiction Simulcast où le codage des vue est indépendant, et également à l'utilisation des vues-B entre vues-I/P et vue-P. Ceci, peut permettre une accélération importante de l'accès aléatoire inter-vues contrairement à la structure de prédiction IPP. En effet, cette dernière n'utilise que des vues de type P successives ce qui permet d'alourdir l'accès aléatoire inter-vues.

Nous avons passé en revue dans ce chapitre une présentation des différentes structures de prédiction possibles pour la compression de la vidéo multi-vues. Ensuite, nous avons dressé une comparaison entre les structures avec et sans l'exploitation de la corrélation inter-vues. Nous avons également détaillé certaines spécificités des structures de prédiction MVC. Enfin, nous avons renforcé cette étude par une analyse expérimentale. Ces résultats seront exploités pour une étude comparative lors de la présentation des chapitres suivants qui seront entièrement consacrés à nos contributions.

Chapitre 04 : Amélioration de l’accès aléatoire inter-vues

4.1. Introduction

En raison de l'intérêt spectaculaire de la compression de la vidéo multi-vues, ainsi que les limites et les inconvénients des approches existantes, nous proposons dans notre travail une nouvelle structure de prédiction inter-vues basée sur le modèle d'images B hiérarchique. Afin de pouvoir comparer nos résultats avec les structures existantes, en particulier celle utilisée dans le modèle de référence JMVM, une nouvelle méthode d'évaluation de l'accès aléatoire inter-vues a été mise en œuvre au cours de ces travaux de recherche. Ce modèle d'évaluation ne concerne pas seulement la structure proposée, mais aussi la structure IPP et IBP. Notre méthode est principalement utilisée pour calculer le nombre d'images nécessaires à la construction d'une image donnée, ainsi que sa capacité d'estimer le nombre maximum d'images à décoder pour atteindre une image particulière. Les résultats expérimentaux ont montré l'efficacité de notre méthode appliquée à la compression de la vidéo multi-vues. Par conséquent, cette technique permet une amélioration considérable de l'accès aléatoire inter-vues. De même, elle permet d'améliorer le débit pour une qualité vidéo similaire.

Nous allons présenter tout au long de ce chapitre, la fiabilité de l'approche proposée appliquée sur huit vues en termes d'amélioration de l'efficacité de la compression et particulièrement en termes d'accélération de l'accès aléatoire inter-vues. Ensuite, nous aborderons la généralisation de la méthode proposée. Autrement dit, comment fonctionne la structure pour un nombre de vues supérieur à huit. L'étude du cas général de la méthode proposée est aussi effectuée vis-à-vis de l'amélioration du débit et de l'accès aléatoire inter-vues. La troisième section de ce chapitre présente le modèle d'évaluation proposé pour notre structure et également pour les deux structures de prédiction IPP et IBP. Finalement, nous présenterons en détail une étude expérimentale afin de montrer le gain significatif obtenu par l'approche proposée en termes d'efficacité de compression et essentiellement en termes d'accès aléatoire inter-vues.

4.2. Présentation de l'approche proposée

Le choix approprié de la position de la vue de base est irréfutable pour la rapidité d'accès aléatoire inter-vues. Ceci est dû au fait que la vue de base serve de référence dans le GGOP pour les autres vues. La séquence utilisée dans la structure proposée comme vue de base est la vue S3. Ce choix est justifié par l'amélioration de l'accès aléatoire inter-vues des vues S4 et S5 utilisant la vue de base S3 directement pour la prédiction. Les images des deux vues S6 et S7 profitent aussi de cette propriété à travers la vue S5 sans passer par la première vue S0. La

figure 4.1 illustre un exemple d'organisation des vues pour la structure IBP et celle proposée. Les trois vues S0, S1 et S2 bénéficient également d'un accès aléatoire inter-vues rapide par la prédiction directement depuis la vue S3.

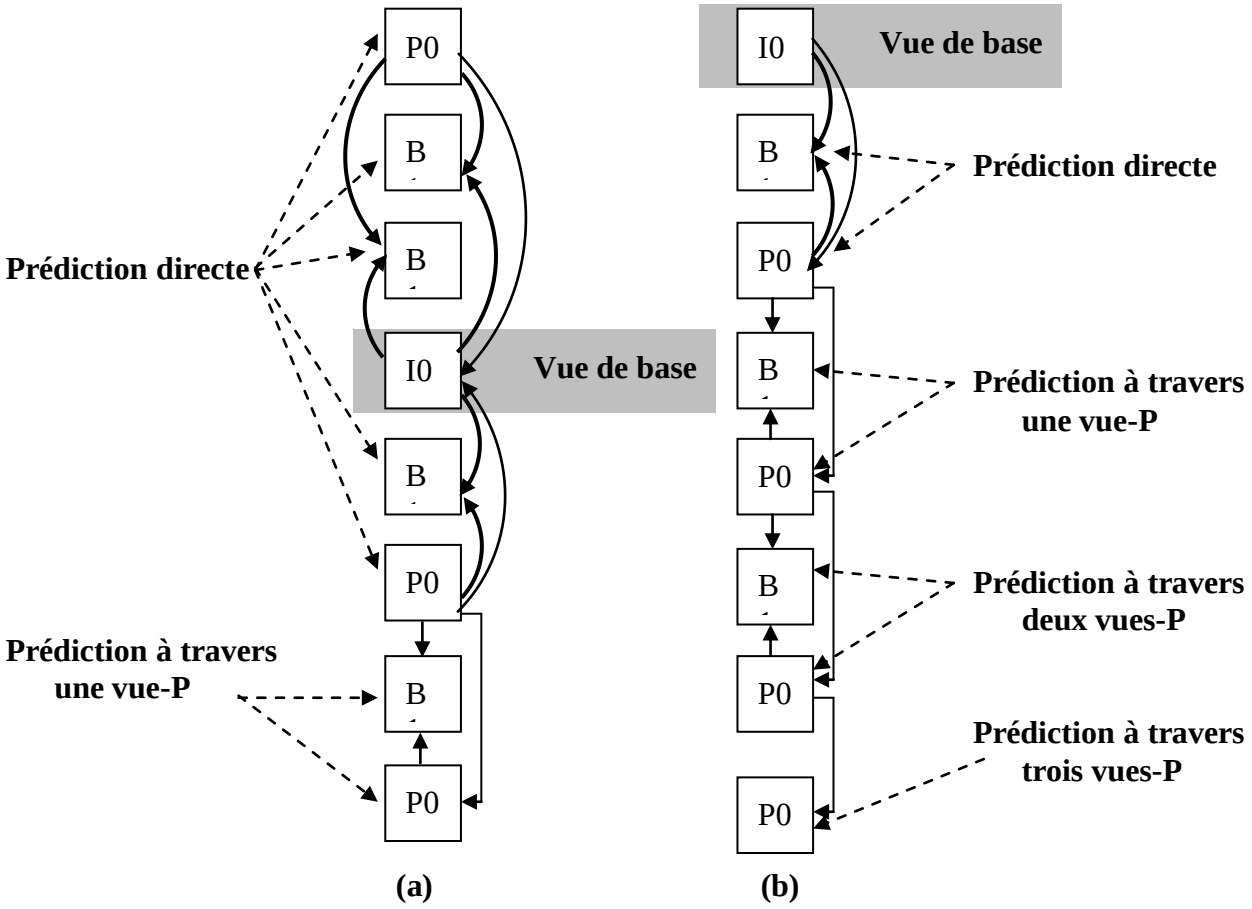


Figure 4.1. *Le type des différentes vues, (a) pour la structure proposée, (b) pour la structure IBP.*

La position de la vue de base est choisie de façon qu'elle permette l'utilisation de deux vues-B successives, avant ou après la séquence ou la vue de base. Un exemple de vues-B précédant la vue de base I est présenté dans la figure 4.1 (a). La prédiction des images non-clés pour les vues-B successives à partir des vues-B est prohibée. Cela signifie que seules les vues de référence autorisées sont les vues-I/P. La méthode favorite de vues-B successives après la vue de base est expliquée dans la section suivante. L'utilisation de deux vues-B successives a un double objectif. Le premier est d'atteindre un meilleur débit binaire avec la même qualité visuelle de la vidéo mesurée par PSNR. Le gain en débit est obtenu par le remplacement des vues-P par des vues-B (voir la figure 4.2) offrant ainsi un meilleur compromis débit binaire/qualité de la vidéo que les vues-P et vues-I. Le deuxième objectif est de réduire le nombre maximal d'accès à une image donnée Sn/Tn. Ceci permet ainsi,

d'accélérer l'accès aléatoire inter-vues pour toute image Sn/Tn dans un GGOP. La figure 4.2 illustre un exemple de prédiction par l'utilisation de deux vues-B successives.

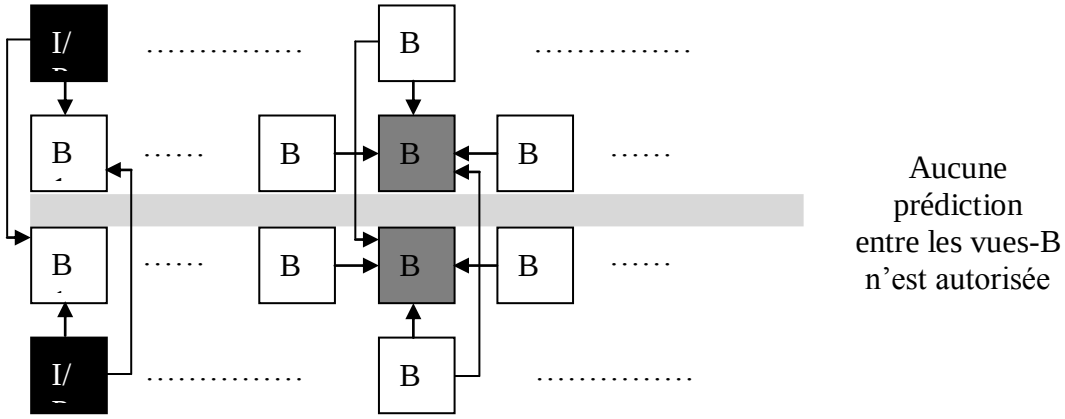


Figure 4.2. *Prédiction des images non-clés des vues-B.*

Nous avons essayé dans notre travail d'améliorer le débit binaire tout en offrant un accès aléatoire inter-vues le plus rapide possible. Pour cette raison, la prédiction inter-vues des images non-clés de la première et la dernière vue-P est également ajoutée. Le numéro de la dernière vue dépend du nombre de vues utilisées. Une étude détaillée du cas général est présentée dans la section suivante. Cette contrainte permet de prédire chaque image B à partir de trois images de référence et non pas de deux comme le cas de la structure IBP (voir la figure 4.3). En effet, la prédiction de ces images dans les autres cas a un effet suspensif sur l'accès aléatoire inter-vues comme l'expliquent les deux exemples suivants :

Premier cas (sans l'utilisation de la prédiction): Dans la figure 4.4 présentant le schéma général de la structure proposée, l'accès à l'image S6/T2 nécessite le décodage de treize images. Ces images peuvent être regroupées suivant la vue qu'elles contiennent comme suit :

- Vue S7 : S7/T0, S7/T2, S7/T4, S7/T8.
- Vue S6 : S6/T0, S6/T4, S6/T8.
- Vue S5 : S5/T0, S5/T2, S5/T4, S5/T8.
- Vue S3 : S3/T0, S3/T8.

Deuxième cas (avec l'utilisation de la prédiction) : Ce cas signifie que la prédiction des images non-clés de la vue-P doit être ajoutée. Dans ce cas-là, les images non-clés de la vue S5 dans la figure 4.4, doivent utiliser les images non-clés de la vues de base pour la prédiction. Ainsi, l'accès à l'image S6/T2 nécessite le décodage de deux images S3/T2 et

S3/T4 de la vue de base en plus des treize images du premier cas. L'accès à une image de niveau hiérarchique plus bas comme l'image S6/T1 de la figure 4.4, impose un nombre plus élevé d'images à décoder. Ces notions sont détaillées dans la section 4.4.

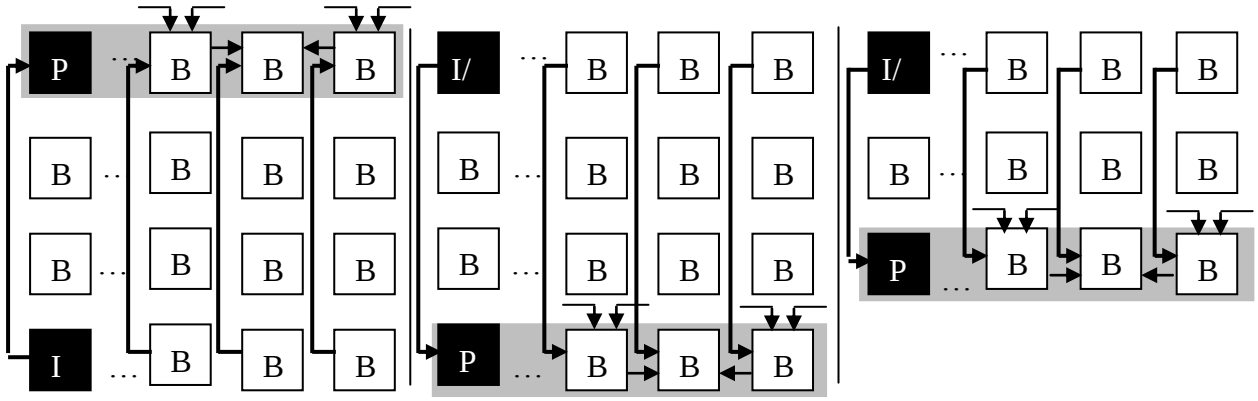


Figure 4.3. *Les cas de figures pour la prédiction des images non-clés des vues-P, (a) pour la première vue-P, (b) Pour la dernière vue-P avec deux vues-B successives, (c) pour la dernière vue avec une seule vue-B.*

Le gain souhaité dans ce cas-là, est moins important que dans les autres cas de prédiction à base de compensation de disparité. Ceci est dû à l'éloignement de l'image de référence inter-vues de l'image à encoder (voir figure 4.3 (a) et (b)). Néanmoins, cette modification améliore le débit binaire total des vues-P. La figure 4.3 illustre les possibilités envisageables pour cette amélioration. Par exemple, le cas (b) de cette figure n'est réalisable que lorsque la MVV est composée de sept ou plus que neuf vues. Les différents cas de cette amélioration sont présentés dans la section suivante. Le nombre minimum autorisé dans notre étude est huit vues. Le schéma général de la structure de prédiction proposée avec huit vues et huit images par GOP est présenté dans la figure 4.4. Cette figure montre que malgré l'augmentation du nombre de vues entre les vues-I/P, le niveau hiérarchique reste le même que la structure IBP.

4.3. Généralisation de l'approche proposée

En effet, huit vues ne peuvent pas mettre en évidence l'utilité de la structure proposée. A cet effet, l'étude de cette structure pour plus de huit caméras semble intéressante. Nous nous limitons dans cette étude à 16 vues de la vidéo "*Rena*" composée de 100 vues. Les autres vidéos multi-vues sont composées généralement de huit vues. L'approche proposée peut être appliquée quel que soit le nombre de vues. Nous avons opté pour 16 caméras pour bien détailler les cas de figures possibles et en même temps économiser le temps d'exécution lors

des expérimentations. En effet, l'utilisation de plusieurs vues nécessite un temps d'exécution prohibitif.

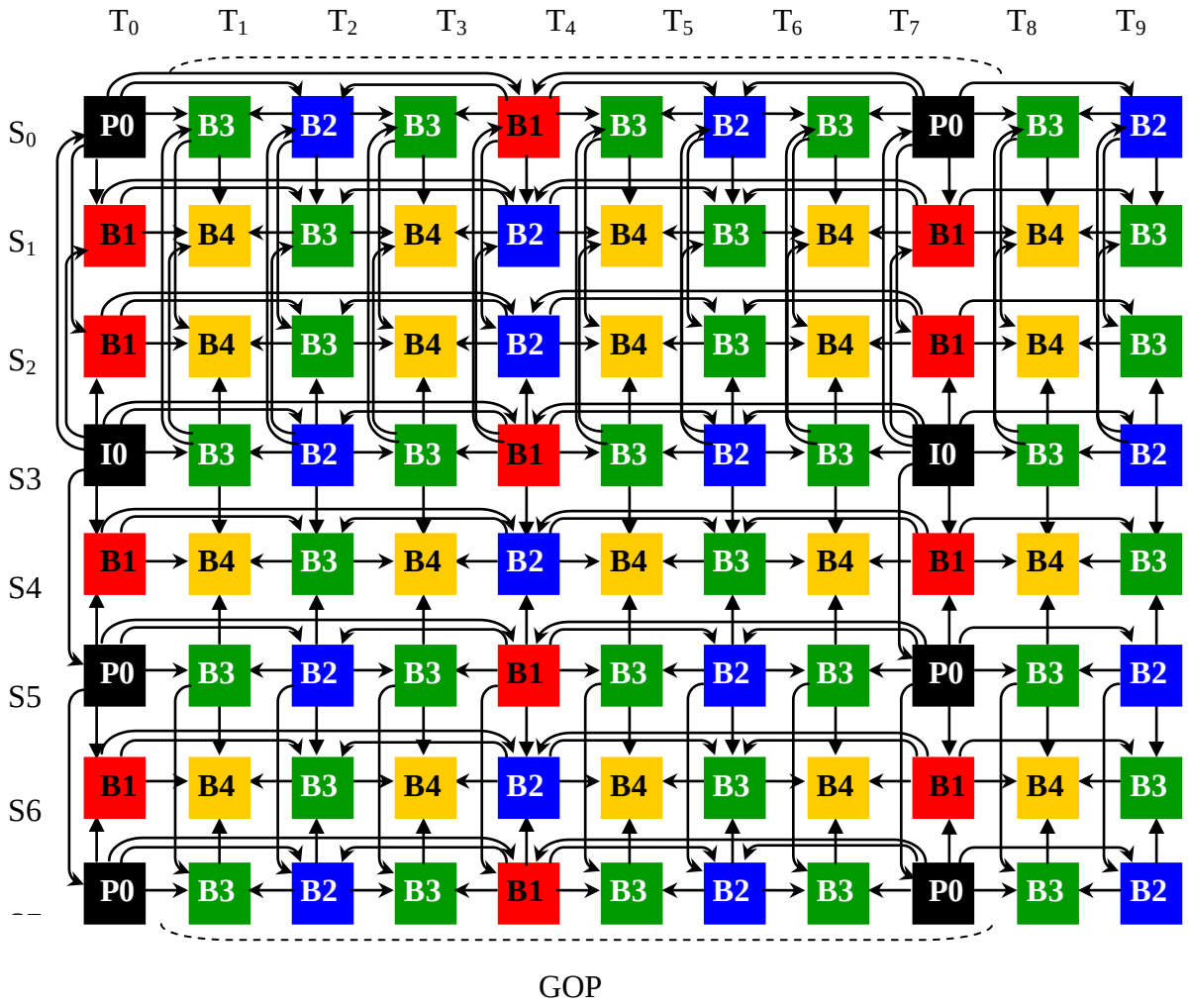


Figure 4.4. *La structure de prédiction proposée en utilisant huit vues et huit images par GOP.*

Pour assurer une meilleure efficacité de la compression de la vidéo multi-vues en termes de débit binaire et de qualité de la vidéo, le type ainsi que l'arrangement des différentes vues situées après la vue de base, sont mis en œuvre en fonction du nombre de vues utilisées. L'idée de base est d'utiliser un nombre maximum de vues-B successives groupées en paires de deux vues comme le montre la figure 4.5. Cette méthode ne doit jamais utiliser des vues-P successives afin d'éviter de ralentir l'accès aléatoire inter-vues, comme dans le cas de la structure IPP. Dans le cas, par exemple, d'une vidéo multi-vues composée de huit vues, où le nombre de vues après la vue de base (la vue S3) est de quatre; l'ordre de vues utilisé est IBPBP. L'utilisation de deux vues-B successives nécessite également l'utilisation de deux

vues-P successives. À cet effet, l'ordre des vues après la vue de base sera I, B, B, P, P qui n'est pas souhaitable. En effet, l'utilisation de deux vues successives de type P est strictement interdite dans l'approche proposée. L'ordre général ainsi que le type, varient en fonction du nombre de caméras utilisées. Il peut être établi suivant le nombre de vues utilisé comme suit:

$$Ordre = \begin{cases} "I/P, B, P, B, P" & if\ (Nbr_{views}\ MOD\ 3) = 2 \\ "I/P, B, B, P, B, P" & if\ (Nbr_{views}\ MOD\ 3) = 0 \\ "I/P, B, B, P, B, B, P" & if\ (Nbr_{views}\ MOD\ 3) = 1 \end{cases} \quad (4.1)$$

Où chaque vue est représentée par une lettre I, P ou B et *Ordre* représente toujours les dernières vues de chaque choix quel que soit le nombre de caméras utilisées. En effet, les premières vues précédant la vue de base restent inchangées (S0, S1 et S2). Les nombres de vues possibles qui doivent être modifiés sont, 5 ("I/P, B, P, B, P"), 6 ("I/P, B, B, P, B, P") ou 7 ("I/P, B, B, P, B, B, P "). Ainsi, quel que soit le nombre de caméras utilisées, l'une des trois options doit être mise en œuvre. Avec l'augmentation du nombre de caméra, les vues qui précèdent l'un des trois choix, présentés ci-dessus, doivent être toujours de la forme « I/P, B, B, P ». Les trois choix sont conçus de telle sorte qu'ils permettent une ou plusieurs améliorations de l'architecture générale comme suit :

- Le premier choix permet d'éviter le cas de vues-P successives. Ce cas est autorisé dans la structure IBP dans les deux dernières vues lorsque le nombre total des vues est pair. Toutefois, la structure IPP utilise seulement les vues-P successive. En effet, l'élimination de ce cas, permet d'éviter l'alourdissement de l'accès aléatoire inter-vues tout en réduisant le débit binaire,

- Dans le second choix, nous bénéficions de l'utilisation de vues-B successives ce qui permet d'améliorer l'efficacité de compression. En effet, les vues-B nécessitent moins de débit binaire par rapport aux autres vues. Ce choix permet également l'amélioration de l'accès aléatoire inter-vues où le nombre de vues-P est moindre que dans le cas « I/P, B, P, B, P.

- Le dernier choix, profite encore de l'utilisation des paires de vues-B successives. Ceci a un effet important sur l'amélioration du débit et également sur l'accès aléatoire inter-vues.

La figure 4.5 illustre les divers cas en fonction du nombre de caméras. Dans cette figure, seulement les images clés des différentes vues sont utilisées.

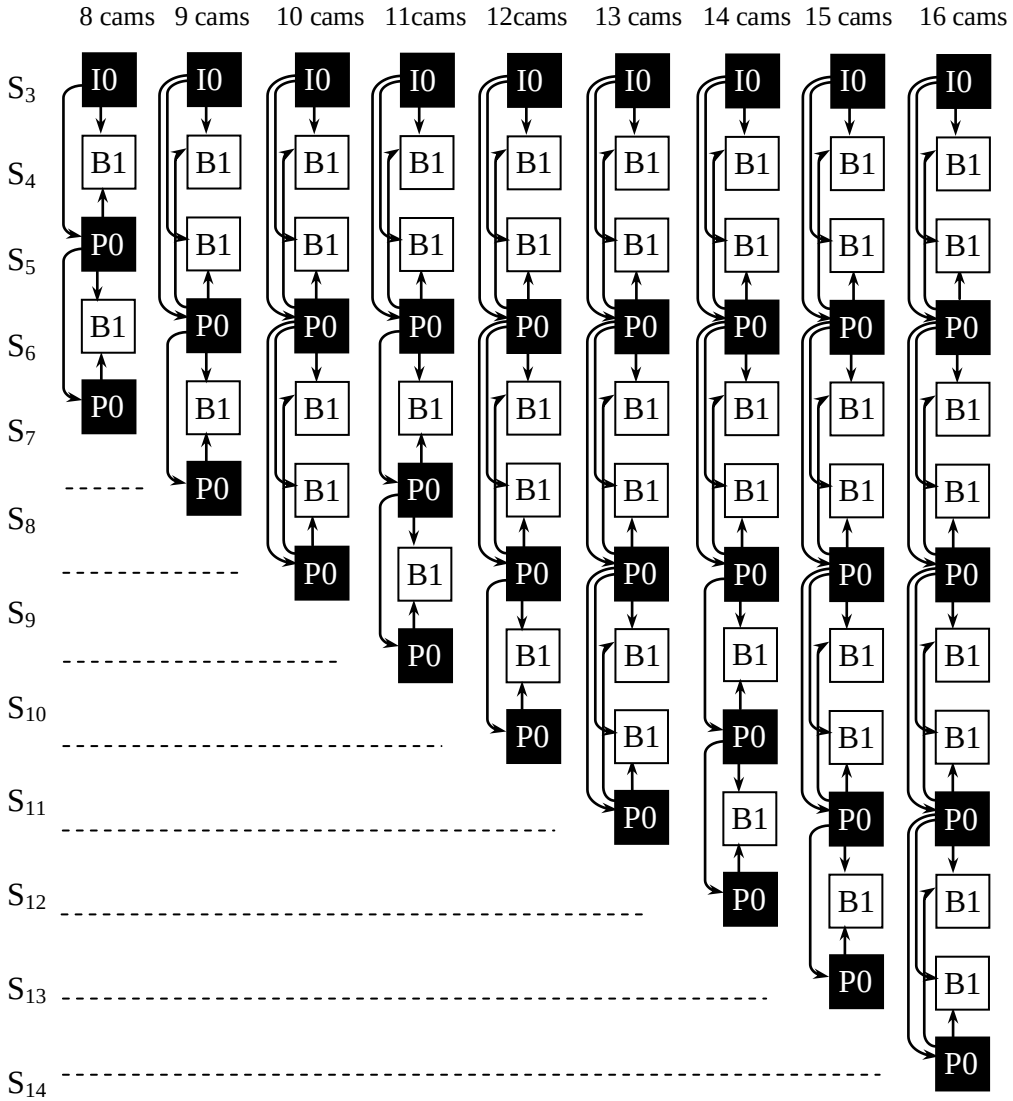


Figure 4.5. *Généralisation de l'approche proposée.*

Parmi les trois choix présentés dans l'expression 4.1, le troisième donne les meilleurs résultats en termes de débit et d'accès aléatoire inter-vues. Ce choix permet d'utiliser un maximum de vues-B successives groupées par paires.

4.4. Méthodes d'évaluation proposées

Dans le travail proposé, l'évaluation de l'accès aléatoire inter-vues est effectuée en utilisant deux métriques N_{MAX} et Nbr_{img}. Le N_{MAX} est utilisé aussi dans l'architecture IBP, ce nombre est détaillé dans le chapitre 03. Le Nbr_{img} proposé représente le nombre d'images de référence nécessaires à décoder à chaque fois pour pouvoir accéder à une image donnée

Sn/Tn. Ce nombre est contrôlé directement par le nombre de vues-P dans un GGOP. Autrement dit, lorsque nous augmentons le nombre de vues-P, le nombre Nbr_{img} augmente aussi. Le Nbr_{img} est différent de la métrique N_{MAX} présentée dans le chapitre 03 pour la structure IBP. Dans ce cas, nous essayons de trouver le nombre requis pour chaque image dans S1 à S7. Nous n'avons pas pris en compte les images de la vue de base S0. En effet, le Nbr_{img} pour ces images est toujours le même pour toutes les structures. Nous présentons dans cette section les deux métriques pour l'approche proposée et également pour les autres structures étudiées IPP et IBP. Nous n'avons pas proposé ces métriques pour la structure Simulcast parce que toutes les vues sont de type I.

4.4.1. Évaluation des structures étudiées

Le nombre N_{MAX} peut être calculé pour une seule image par GGOP. Il s'agit de l'image qui nécessite le décodage du plus grand nombre d'images de référence pour son accès aléatoire inter-vues. Nous avons essayé dans ce travail de proposer le N_{MAX} pour la structure de prédiction IPP afin de pouvoir comparer les résultats obtenus par notre structure avec cette approche. Ainsi, le nombre N_{MAX} pour la structure IPP, peut être obtenu en utilisant la hiérarchie maximale et le nombre de vues par l'équation proposée ci-dessous:

$$N_{MAX} = (Hierarchy_{MAX} + 2) * Nbr_{views} - 1 \qquad (4.2)$$

Où $Hierarchy_{MAX}$ est le niveau hiérarchique maximal de la structure de prédiction HBP (*Hierarchical B Pictures*) utilisée dans le niveau temporel de la structure IPP. Ce nombre est égal à 3 dans le cas de la structure IPP de la figure 3.5. Le Nbr_{views} représente le nombre total de vues utilisées. L'image qui nécessite un nombre maximum d'images à décoder lors de son accès aléatoire dans la structure IPP, se trouve toujours dans la dernière vue. Le nombre d'images à décoder dans la vue courante *Sn*, celle qui contient l'image à consulter, est égal à $Hierarchy_{MAX} + 1$. Dans les vues de référence *Sn-1*, *Sn-2*, …etc, ce nombre augmente de 1, il doit avoir la valeur $Hierarchy_{MAX} + 2$.

Pour évaluer l'accès aléatoire inter-vues nous proposons une méthode générale pour le calcul du nombre d'images Nbr_{img} pour les trois structures IPP, IBP et proposée. Ce nombre peut être estimé de manière différente selon le type de l'image à consulter, image clé ou non-clé. Mais aussi selon le type de vue contenant cette image, vue-P ou vue-B. Ce nombre peut donc être estimé en fonction du niveau hiérarchique de l'image à consulter indépendamment de la taille de chaque GOP. D'une manière générale, les images ayant le plus haut niveau

hiérarchique, aurons un accès aléatoire inter-vues plus rapide. Dans le cas de la structure IPP, le nombre Nbr_{img} peut être obtenu pour les images clés qui sont toutes de type P en soustrayant 1 du numéro d'ordre (2, 3, …8) de la vue courante. Le Nbr_{img} dépend dans le cas des images non-clés du niveau hiérarchique de chaque image B. Il dépend également du numéro d'ordre de la vue où cette image se trouve. Pour calculer ce nombre, l'expression 4.2 peut être modifiée de la façon suivante:

$$Nbr_{img} = (Hierarchy + 2) * Num_{views} - 1 \tag{4.3}$$

Où *Hierarchy* désigne le niveau hiérarchique de l'image à consulter et Num_{View} représente le numéro d'ordre de la vue courante. Le nombre Nbr_{img} peut être calculé de la même façon pour la structure de prédiction IBP en prenant en considération un troisième type de vue à savoir la vue-B. Les images clés et non-clés des vues-B de cette structure nécessitent chaque fois un nombre Nbr_{img} plus grand que dans le cas de la structure IPP. Le nombre Nbr_{img} peut être calculé dans la structure IBP pour les images clés qu'elles soient de type B, cas des vues-B, ou de type P, cas des vues-P, en utilisant l'expression:

$$Nbr_{img} = \alpha + \left\lfloor \frac{Num_{view}}{2} \right\rfloor where \;\; \alpha = \begin{cases} 0 & for\ anchor\ \ frames\ of\ \ P\_views \\ 1 & for\ anchor\ \ frames\ of\ \ B\_views \end{cases} \tag{4.4}$$

Les images clés sont les plus rapides en termes d'accès aléatoire inter-vues. En effet, chaque image clé utilise pour la prédiction soit une ou deux images de référence successivement pour les vues-P et vues-B. Contrairement aux images clés, les images non-clés nécessitent à chaque fois pour leur accès aléatoire inter-vues, le décodage d'un nombre d'images plus élevé. Tout comme les images clés, les images non-clés des vues-B sont plus lourdes en accès aléatoire inter-vues que dans le cas des vues-P. Le nombre Nbr_{img} proposé pour les images non-clés de la structure de prédiction IBP, peut être estimé par:

$$\begin{aligned} &Nbr_{img} = \beta * (Hierarchy + \alpha) + 2 * \left\lfloor \frac{Num_{view}}{2} \right\rfloor \\ &where \begin{cases} \beta = 1, \alpha = 1 \;\; for\ non-anchor\ \ frames\ of\ \ P\_views \\ \beta = 3, \alpha = 0 \;\; for\ non-anchor\ \ frames\ of\ \ B\_views \end{cases} \end{aligned} \tag{4.5}$$

Lors de l'accès aléatoire inter-vues à une image Sn/Tn, l'ordre de décodage des images nécessaires dans un GGOP est le suivant :

- Le décodeur commence par les images clés des différentes vues utilisées comme vues de référence. Pour chaque vue, deux images sont décodées, celles du GOP courant et celle du précédent.
- La même opération se répète jusqu'à la vue contenant l'image à consulter. Ceci est interprété par $2*(Num_{view}/2)$ dans la formule 4.5 quel que soit le type de la vue.
- Si la vue contenant l'image non-clé à consulter est une vue-B, le nombre d'images à décoder dans la vue elle-même est identique au niveau hiérarchique de cette image. Dans ce cas-là, la vue précédente et la vue suivante, doivent contenir plus que deux images de références. Autrement dit, elles ne se limitent pas aux images clés. A cet effet, nous avons interprété ceci, dans l'expression 5, par 3* *Hierarchy*.
- Lorsque l'image à consulter figure dans une vue-P, le nombre des images nécessaires dans cette dernière est égal à *Hierarchy + 1.*

Il est important de noter que l'usage des GOP avec une taille plus grande mais avec le même niveau hiérarchique n'augmente pas le nombre Nbr_{img}. À titre d'exemple, avec des GOP de taille 12, le niveau hiérarchique est le même que dans le cas des GOP de taille 8. Si le niveau hiérarchique augmente avec la taille des GOP, ceci accroît particulièrement le N_{MAX} et également le nombre Nbr_{img} pour les images non-clés de plus bas niveau de la structure hiérarchique. Toutefois, le Nbr_{img} pour les images clés des deux vues-B/P reste le même quel que soit le niveau utilisé.

Un cas exceptionnel se distingue lorsque les images non-clés de la dernière vue-P de la structure IBP utilisent trois images de référence pour la prédiction. Ainsi, deux images de référence sont utilisées dans le niveau temporel et la troisième pour l'inter-vues. Ce cas n'est possible que lorsque deux vues-P successives sont utilisées où le nombre de vues est pair. La vue S7 de la figure 3.6 montre un exemple de ce cas de figure. Ainsi, le Nbr_{img} proposé pour l'estimation du nombre d'images à décoder pour pouvoir accéder à une image non-clés est donné comme suit:

$$Nbr_{img} = 1 + 2*(Hierarchy + \left\lfloor \frac{Num_{view}}{2} \right\rfloor) \tag{4.6}$$

4.4.2. Évaluation de l'approche proposée

L'approche proposée permet un gain considérable en accès aléatoire à l'image Sn/Tn, qui nécessite le décodage d'un nombre maximum d'images de référence et qui se calcule à travers la métrique N_{MAX}. La figure 4.6 illustre un exemple d'accès aléatoire à l'image S6/T7 qui nécessite le nombre maximum d'images à décoder. Les images S6/T5, S6/T3 et S6/T1 imposent le même nombre d'images pour leur accès aléatoire inter-vues. Tout comme la structure IBP, l'image Sn/Tn se trouve dans la vue-B la plus distante de la vue de base. Le nombre N_{MAX} peut être obtenu de la même façon que dans la structure IPP et IBP en tenant compte de la position différente de la vue de base ainsi que l'utilisation des vues-B successives. L'utilisation d'une prédiction inter-vues pour les images non-clés de la dernière vue-P n'a aucun effet sur l'accès aléatoire inter-vues à l'image Sn/Tn. La vue-B contenant cette image représente l'avant dernière vue de la MVV. À cet effet, la vue qui suit et celle qui précède la vue courante doivent contenir le même nombre d'images de référence utilisées par la vue courante. Le N_{MAX} peut être obtenu pour la structure proposée avec seulement huit vues en utilisant l'expression 4.7. Tandis que le N_{MAX} pour le cas général sera présenté dans la section suivante:

$$N_{MAX} = 3 * Hierarchy_{MAX} + 2 * \left\lfloor \frac{Nbr_{views} - 2}{3} \right\rfloor \tag{4.7}$$

La $Hierarchy_{MAX}$ est donnée par l'indice des images qui ne peuvent servir de référence à d'autres images. Il s'agit des images ayant le plus bas niveau hiérarchique. Ce niveau est égal à quatre dans la structure proposée avec des GOP de 12 images. Par contre il est égal à 5 avec un GOP de 15 images. Dans l'expression 4.7, $3*Hierarchy_{MAX}$ représente les images de référence utilisées pour une prédiction directe de l'image Sn/Tn à partir de la vue courante et les deux vues adjacentes. Les images utilisées comme référence pour une prédiction indirecte (voir la figure 4.1) sont toujours au nombre de deux. Il s'agit des images clés des vues précédentes jusqu'à la vue de base. Le nombre de ces images est obtenu par $2*\llcorner(Nbr_{Views}-2)/3\square$ comme il est présenté dans l'expression 4.7 et la figure 4.6.

En effet, le nombre N_{MAX} ne peut montrer la fiabilité de l'approche proposée que par l'utilisation d'un nombre élevé de vues. Pour cela, nous avons proposé d'évaluer toutes les images du GGOP, à l'exception de la vue de base, à travers le calcul du nombre Nbr_{img}. La modification de la position de la vue de base exige une évaluation différente pour les vues

existant avant et après la vue de base. La première et la dernière vue-P sont également évaluées en tenant compte de l'utilisation d'une prédiction inter-vues pour leurs images non-clés. Dans la structure proposée, ce nombre est obtenu pour les images clés des vues qui se trouvent après la vue de base par :

$$Nbr_{img} = \beta + \left\lfloor \frac{(Num_{view} - 3) + \alpha}{3} \right\rfloor \; where \begin{cases} \beta = 0, \alpha = 1 \;\; for \; P \; anchor \;\; frames \\ \beta = 1, \alpha = 2 \;\; for \; B \; anchor \;\; frames \end{cases} \tag{4.8}$$

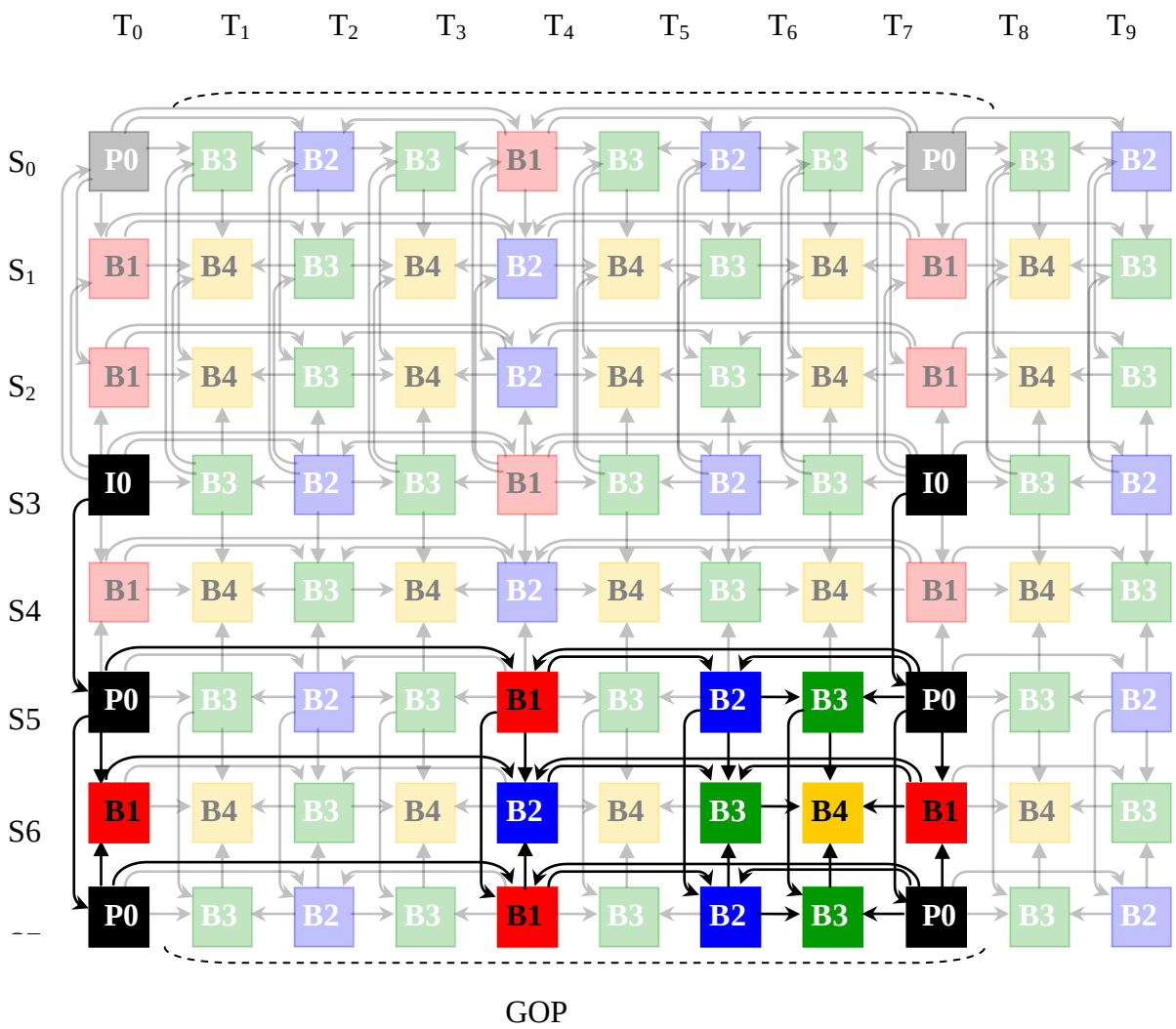


Figure 4.6. *L'accès aléatoire inter-vues pour le schéma proposé.*

Par rapport à la structure IPP et IBP le nombre Nbr_{img} pour les images clés est significativement amélioré dans la structure proposée. À titre d'exemple, le Nbr_{img} maximum devant être utilisé par une image clé de type P est de sept dans la structure IPP et quatre pour la structure IBP. Alors que le Nbr_{img} ne dépasse pas trois dans l'approche proposée. Cette

amélioration est plus importante pour les images non-clés. Dans ce cas le calcul du Nbr_{img} pour les vues qui suivent la vue de base est généré par:

$$Nbr_{img} = \delta * (Hierarchy + \beta) + 2 * \left\lfloor \frac{(Num_{view} - 3) + \alpha}{3} \right\rfloor$$
$$where \begin{cases} \delta = 1,\ \beta = 1 \ \ and \ \ \alpha = 1 \ for\ P\ non-anchor\ \ frames \\ \delta = 3,\ \beta = 0 \ \ and\ \alpha = 2 \ for\ B\ non-anchor\ \ frames \end{cases} \quad (4.9)$$

Le Nbr_{img} de l'expression 4.9 est calculé par le même principe que N_{max} de l'expression 4.7. Dans le cas des vues-B, l'obtention du nombre d'images de référence à partir de la vue courante et les vues adjacentes est effectuée en replaçant le $Hierarchy_{MAX}$ par *Hierarchy*. Dans le cas des vues-P : $3*Hierarchy_{MAX}$ est remplacé par *Hierarchy +1*. La valeur maximale pouvant être affectée à la variable *Hierarchy* est de quatre pour les vues-B et trois pour les vues-P. Les images B3 des vues-P nécessitent quatre images de référence dans la vue elle-même ce qui justifie l'utilisation de *Hierarchy +1* pour ces vues.

En effet, l'utilisation de trois images de référence par les images non-clés de la première et la dernière vue-P permet un gain en débit binaire avec une légère augmentation du Nbr_{img}. Dans ce cas particulier, le Nbr_{img} peut être estimé pour ces images non-clés par l'expression 4.10. Les autres vues-P se trouvant entre la première et la dernière vue, ne peuvent tirer profit de cette propriété. En effet, ces vues peuvent alourdir l'accès aléatoire inter-vues non seulement pour les vues-P mais aussi pour les vues-B, comme il est présenté dans la section 4.2.

$$Nbr_{img} = 2*(Hierarchy + \alpha) + 1 \ where \ \alpha = \begin{cases} 1 & for\ \ first\ P\_view \\ \left\lfloor \frac{(Num_{view} - 2)}{3} \right\rfloor & for\ last\ P\_view \end{cases} \quad (4.10)$$

Le Nbr_{View} des images clés est toujours égal à 1 pour la première vue-P et peut être calculé par l'équation 4.8 pour la dernière vue-P. Les vues-B précédant la vue de base peuvent être également évaluées différemment. Les images clés dans ces vues sont prédites en utilisant deux images clés de type I/P de la vue suivante et la vue précédente. En outre, les images non-clés de B2 à B4 dans les deux vues, peuvent être estimées par l'équation:

$$Nbr_{img} = 3 * Hierarchy + 2 \quad (4.11)$$

4.4.3. Évaluation dans le cas général

De la même façon on peut généraliser la méthode d'évaluation pour les deux métriques N_{MAX} et Nbr_{img}. La méthode de généralisation doit tenir compte des trois choix utilisés pour la désignation du type et de l'ordre des diverses vues. D'autres paramètres tels que le niveau hiérarchique, le nombre total de vues ainsi que le numéro d'ordre de la vue contenant l'image à évaluer sont également utilisés. Le nombre maximum d'images de référence N_{MAX} requis pour accéder à une image donnée Sn/Tn quel que soit le nombre de vues, peut alors être estimé comme suit:

$$N_{MAX} = 3 * Hierarchy_{MAX} + 2 * \left\lfloor \frac{(Nbr_{views} - 3) + \alpha}{3} \right\rfloor \quad where\ \alpha = \begin{cases} 1 & if\ (Nbr_{views}\ MOD\ 3) = 2 \\ 0 & if\ (Nbr_{views}\ MOD\ 3) = 0 \\ -1 & if\ (Nbr_{views}\ MOD\ 3) = 1 \end{cases} \tag{4.12}$$

L'augmentation du nombre de vues dans la structure proposée ne signifie pas un niveau hiérarchique plus haut. Ainsi, le $Hierarchy_{MAX}$ est toujours égal à quatre. Le N_{MAX} peut être d'autant plus amélioré selon les vues-B successives utilisées. A cet effet, le meilleur choix permettant la réduction maximale du nombre N_{MAX} est le "I/P, B, B, P, B, B, P". Ce cas est envisageable lorsque le nombre de vues égal à 10, 13, 16, …etc, où toutes les vues-B après la vue de base sont organisées en paires de vues-B successives. Le nombre Nbr_{img} reste le même pour les images clés et non-clés des vues qui précédent la vue de base. Les images clés après la vue de base peuvent être évaluées en utilisant l'équation suivante quel que soit le type de la vue.

$$Nbr_{img} = \beta + \left\lfloor \frac{(Num_{view} - 3) + \alpha}{3} \right\rfloor$$

$$where \begin{cases} \beta = 0, \quad \alpha = \begin{cases} 1 & if\ (Nbr_{views}\ MOD\ 3) = 2 \\ 0 & if\ (Nbr_{views}\ MOD\ 3) = 0 \\ -1 & if\ (Nbr_{views}\ MOD\ 3) = 1 \end{cases} & for\ P\ anchor\ \ frames \\ \beta = 1, \quad \alpha = 2 & for\ B\ anchor\ \ frames \end{cases} \tag{4.13}$$

Les images pouvant faire la différence en termes d'accès aléatoire inter-vues entre la structure de prédiction proposée et les autres structures étudiées dans cette thèse (voir chapitre 03) sont les images non-clés, et particulièrement des vues-B. La généralisation de la méthode permettant d'obtenir le Nbr_{img} pour les images non-clés des vues-P et vues-B est obtenue dans ce cas par:

$$Nbr_{img} = \delta * (Hierarchy + \beta) + 2 * \left\lfloor \frac{(Num_{view} - 3) + \alpha}{3} \right\rfloor$$

$$where \begin{cases} \delta = 1, \beta = 1 \; and \; \alpha = \begin{cases} 1 \quad if \; (Nbr_{views} \; MOD \; 3) = 2 \\ 0 \quad if \; (Nbr_{views} \; MOD \; 3) = 0 \\ -1 \; if \; (Nbr_{views} \; MOD \; 3) = 1 \end{cases} & for \; P \; non-anchor \; frames \\ \delta = 3, \beta = 0 \; and \; \alpha = 2 & for \; B \; non-anchor \; frames \end{cases} \quad (4.14)$$

Le cas particulier de la dernière vue-P peut être évalué au-delà de huit vues, en utilisant l'expression 4.10 pour les trois choix.

4.5. Configuration requise

Toutes les séquences MVV utilisées pour l'analyse et l'évaluation de la compression de la vidéo multi-vues, sont celles exploitées pour l'évaluation des trois structures de prédiction présentées dans le chapitre 03. Nous avons utilisé également les mêmes paramètres d'encodages employés durant l'implémentation des trois structures Simulcast, IPP et IBP. Certains paramètres sont présentés dans le tableau 3.2 dans le chapitre 03. Seulement seize vues de la vidéo "*Rena*" sont utilisées pour démontrer et évaluer la généralisation de l'approche proposée. En effet, cette vidéo est composée de 100 vues. Cependant, l'explication de la généralisation de l'approche proposée nécessite seulement dix séquences ou vues. Néanmoins, dans le but de confirmer la fiabilité de cette approche, nous avons opté pour seize points de vue. Ceci, permet d'utiliser plusieurs paires de vues-B successives.

Nombre de vues	**Ordre d'encodage**
8 vues	S3, S0, S1, S2, S5, S4, S7, S6
9 vues	S3, S0, S1, S2, S6, S4, S5, S8, S7
10 vues	S3, S0, S1, S2, S6, S4, S5, S9, S7, S8
11 vues	S3, S0, S1, S2, S6, S4, S5, S8, S7, S10, S9
12 vues	S3, S0, S1, S2, S6, S4, S5, S9, S7, S8, S11, S10
13 vues	S3, S0, S1, S2, S6, S4, S5, S9, S7, S8, S12, S10, S11
14 vues	S3, S0, S1, S2, S6, S4, S5, S9, S7, S8, S11, S10, S13, S12
15 vues	S3, S0, S1, S2, S6, S4, S5, S9, S7, S8, S12, S10, S11, S14, S13
16 vues	S3, S0, S1, S2, S6, S4, S5, S9, S7, S8, S12, S10, S11, S15, S13, S14

Tableau 4.1. *L'ordre d'encodage selon le nombre de vues utilisées.*

Tout comme la structure de prédiction inter-vues IBP, l'ordre d'encodage des différentes vues doit être défini au préalable afin de garantir une estimation et compensation de mouvement fiables. Le tableau 4.1 illustre l'ordre d'encodage des vues pour les divers cas présentés précédemment. L'ordre d'encodage des trois structures Simulcast, IPP et IBP a été expliqué dans le chapitre précédent. La première vue à encoder est toujours la vue de base. En effet, elle ne nécessite aucune vue de référence pour le codage de ses images. Ensuite, cette vue serve de référence pour les vues-P exigeant une seule vue de référence. Finalement, les vues-B utilisent pour l'encodage des leurs images clés et non-clés les deux vues précédemment encodées de type I/P. Pour l'ordre d'encodage des vues-P entre-elles, nous avons choisi d'encoder en premier lieu celle qui précède la vue de base. La vue-P qui figure juste après la vue de base doit être ensuite encodée. Les autres vues-P sont encodées auparavant, les unes après les autres. L'ordre des vues-B est désigné par l'ordre des vues-P. Il est à noter que les vues précédant la vues de base peuvent être encodées à la fin du processus de compression. Ceci est justifié par le fait que la première vue-P ne serve de référence à aucune autre vue de type P. En effet, les valeurs du paramètre de quantification QP sont choisies pour des raisons comparatives. Néanmoins, les deux valeurs 37 et 40 produisent des vidéos de qualité très dégradée.

4.6. Expérimentation et résultats

Nous essayons dans cette section d'évaluer le travail exposé dans les sections précédentes de ce chapitre en comparaison avec les trois structures de prédiction Simulcast, IPP et IBP. L'évaluation de l'approche proposée s'articule principalement sur les deux exigences les plus importantes décrites dans [09]. La première recommandation est l'efficacité de la compression représentée par un compromis débit binaire/ qualité de la vidéo. L'exigence à satisfaire sur laquelle se base notre proposition est l'amélioration de l'accès aléatoire inter-vues.

4.6.1. Evaluation de l'efficacité de la compression

Dans cette section, nous évaluons l'efficacité de la compression en utilisant le débit binaire (en Kbit/s), le gain en débit binaire Δbit-rate (taux) et la qualité de la vidéo compressée mesurée par ΔPSNR (en dB). Le débit binaire et le PSNR de chaque MVV sont obtenus par la moyenne des résultats des vues qui composent la MVV. Les résultats obtenus pour la vidéo "*Ballroom*" pour chaque vue sont exposés dans le tableau 4.2. Le débit total est

le PSNR pour les autres vidéos de test sont présentés dans le tableau 4.3. Les résultats obtenus montrent l'intérêt d'exploiter la prédiction et la compensation de disparité en plus de la prédiction et la compensation de mouvement. Dans la structure proposée les vues exploitant la prédiction et la compensation de disparités dans un modèle de 8 caméras sont S1, S2, S4 et S6 qui sont toutes des vues-B. Ces types de vues permettent de réduire le débit binaire requis tout en offrant une qualité acceptable de la vidéo compressée en comparaison avec les autres vues à savoir la vue de base et les vues-P. La figure 4.7 illustre le gain en débit binaire des vues-B par rapport aux vues-P, de même pour les vues-P en comparaison avec la vues de base vue-I.

	Vue	QP = 22	QP = 27	QP = 32	QP = 37	QP = 40
Débit binaire (Kbit/s)	S0	1547,6649	780,1243	399,8216	210,8757	142,7351
	S1	1417,4703	646,1514	317,8973	173,8865	122,9568
	S2	1273,2054	575,9622	284,9676	160,3405	118,0486
	S3	1895,2811	1033,4811	584,8324	340,6378	243,0865
	S4	1384,3297	633,0595	306,9189	166,4811	121,0378
	S5	1637,7351	838,7243	443,6108	245,5189	176,2595
	S6	1440,7405	623,9676	313,5784	182,027	136,4
	S7	1760,2595	879,1405	458,8054	252,0054	179,0486
Moyenne		1544,5858	751,3263	388,804	216,4716	154,9466
PSNR (Db)	S0	39,3756	37,0641	34,4094	31,6106	29,8277
	S1	39,11	37,1016	34,6514	32,0267	30,2774
	S2	39,4807	37,3998	34,9198	32,3234	30,6049
	S3	39,3173	36,8894	34,262	31,607	29,9297
	S4	39,272	36,9473	34,2578	31,5676	29,8857
	S5	39,8491	37,4121	34,6556	31,8319	30,0345
	S6	38,7765	36,8765	34,6081	32,1086	30,3981
	S7	39,1853	36,6931	33,873	31,0312	29,2417
Moyenne		39,2958	37,0479	34,4546	31,7633	30,0249

Tableau 4.2. *Les résultats obtenus en termes de débit binaire et PSNR pour la structure de prédiction proposée en utilisant la vidéo Ballroom (en détails pour chaque vue).*

Les vues-P, utilisent uniquement la prédiction et la compensation temporelle, à l'exception de la première et la dernière vue qui sont successivement S0 et S7 dans le cas de 8

vues. Ceci, explique le haut débit de ces vues par rapport aux vues S1, S2, S4 et S6. Le numéro de la dernière vue change en fonction du nombre de vues utilisées. Cette vue peut être S7 ou S8, ...etc. La vue de base S3 donne le débit le plus élevé par rapport aux autres vues-P/B. Ceci est dû au fait qu'elle utilise seulement la prédiction et la compensation de mouvement et contient en plus les images spatialement codées I. Un exemple de seize vues est présenté dans la figure 4.7 (e) où le gain en débit binaire des vues-B, par rapport aux autres de type I et P est plus remarquable par rapport aux autres figures.

	MVV	QP = 22	QP = 27	QP = 32	QP = 37	QP = 40
Débit binaire (Kbit/s)	Ballroom	1544,5858	751,3263	388,804	216,4716	154,9466
	Exit	946,8087	396,777	202,4675	117,4533	84,4270
	Vassar	1160,1108	413,9223	174,2385	81,9229	52,2412
	Race1	2071,091	1009,5711	515,1926	294,2817	217,1752
	Rena	722,3823	349,2861	179,5624	105,6273	80,4554
PSNR (Db)	Ballroom	39,2958	37,0479	34,4546	31,7633	30,0249
	Exit	39,9885	38,4458	36,5128	34,1849	32,5098
	Vassar	38,8592	36,9705	35,0088	32,8141	31,2263
	Race1	39,7454	37,0140	34,2959	31,4702	29,6671
	Rena	44,8593	42,0433	39,0099	36,1470	34,4410

Tableau 4.3. *Les résultats obtenus en termes de débit binaire et PSNR pour la structure de prédiction proposée en utilisant les 5 vidéos de test (résultat globale de chaque MVV).*

La Figure 4.8 présente la variation du débit binaire suivant les 5 valeurs de QP utilisées dans le codage de toutes les vidéos de test. Les mêmes valeurs de QP sont utilisées pour les quatre structures de prédiction : Simulcast , IPP, IBP et celle que nous avons proposée. Dans cette étude, nous constatons une amélioration considérable dans le débit binaire avec une qualité similaire de la vidéo par rapport aux autres structures étudiées. L'amélioration du débit binaire doit respecter l'intérêt principal pour lequel la structure a été proposée, à savoir l'amélioration de l'accès aléatoire inter-vues. La figure 4.8 montre que le débit binaire obtenu par IPP, IBP et la structure proposée est plus significatif par rapport au codage «simulcast » qui fournit l'accès aléatoire le plus rapide. Cependant, les images des vues de la structure Simulcast profitent d'un accès aléatoire direct à partir de la vue courante sans faire intervenir les autres vues.

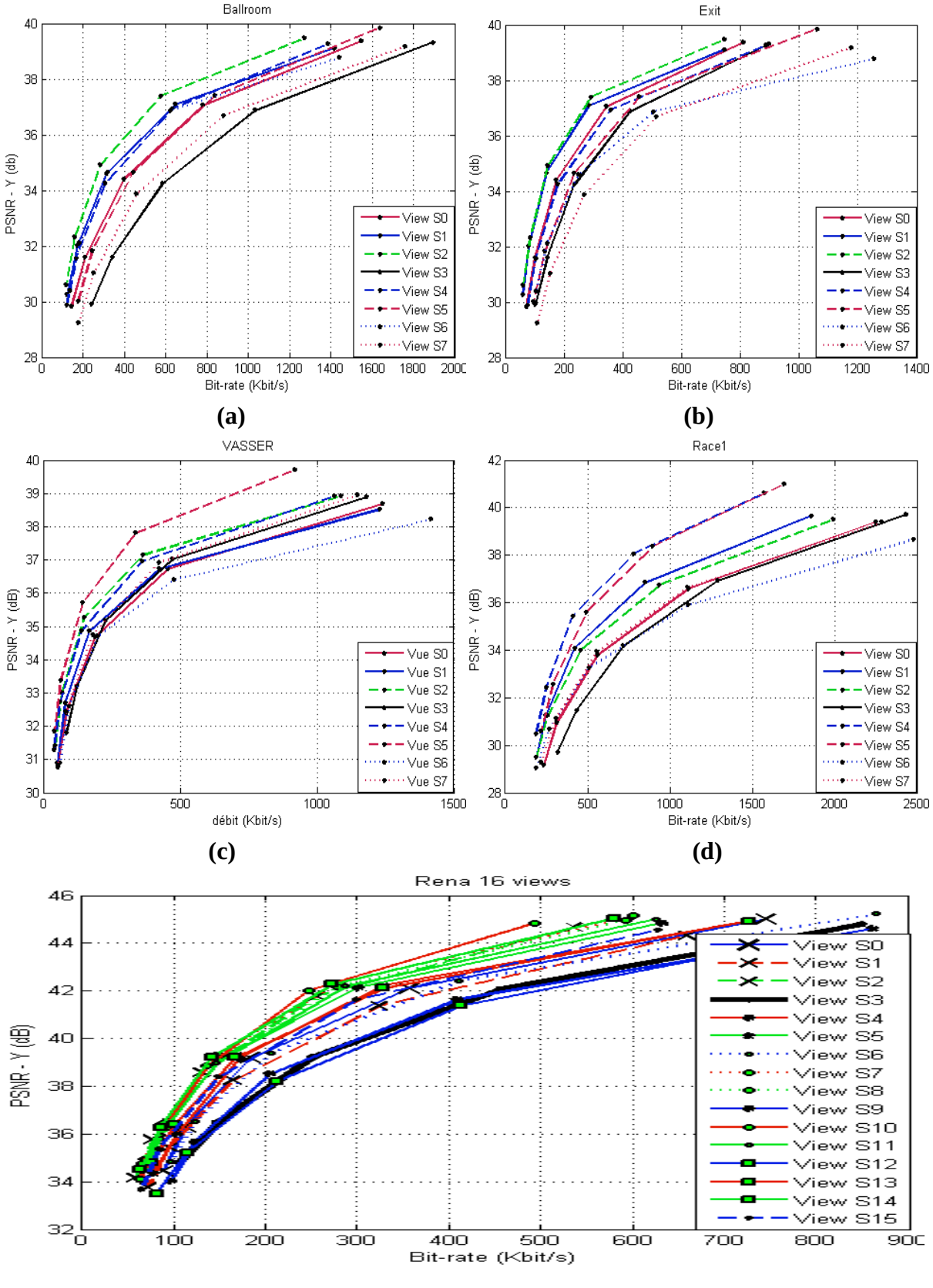
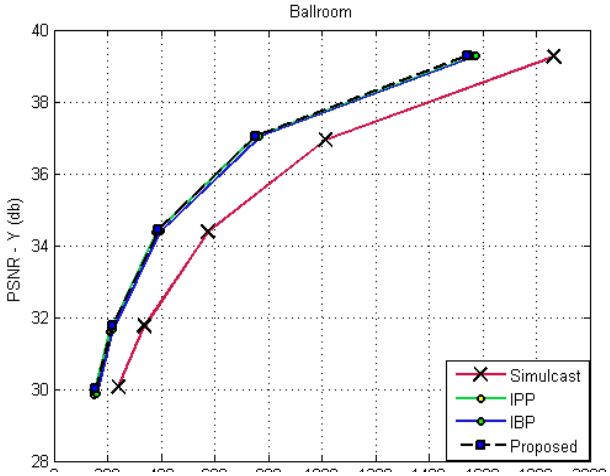
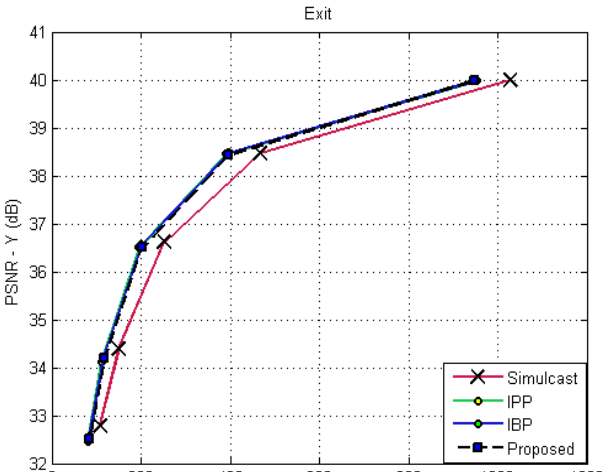


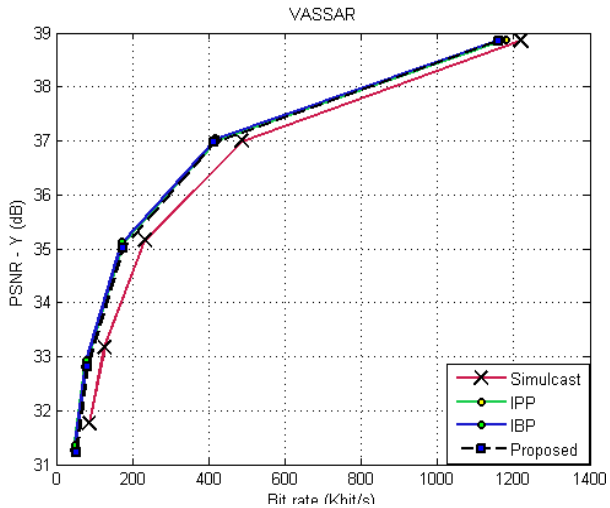
Figure 4.7. *L'efficacité de la compression entre les différentes vues pour les 5 vidéos de test.*



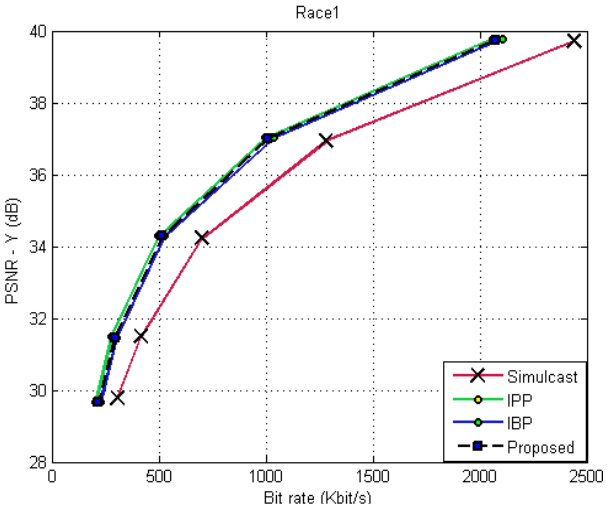
(a) Ballroom



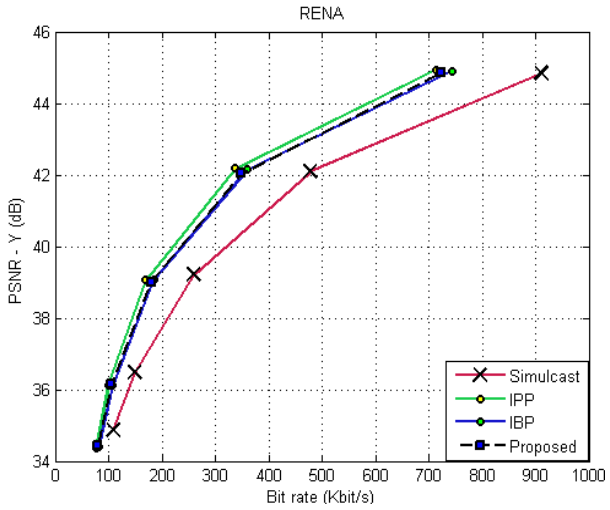
(b) Exit



(c) Vassar



(d) Race1



(e) Rena

Figure 4.8. *La variation du débit binaire en fonction des 5 valeurs du QP.*

En effet, la structure de prédiction IPP produit un gain obsolète dans le débit binaire par rapport à notre approche qui peut atteindre 2%. Toutefois, la structure IPP est extrêmement lourde en accès aléatoire inter-vues ce qui mettra en évidence le gain en débit binaire. D'autre part, par l'utilisation de 8 caméras, le système proposé montre un gain d'environ 3% en débit binaire par rapport à la structure de prédiction IBP (voir le tableau 4.4). Le gain en débit binaire est obtenu par:

$$\Delta_{bit\_rate} = \frac{(bit\_rate - bit\_rate_{proposed})}{(bit\_rate)} \times 100[\%] \tag{4.15}$$

Où *bit-rate* peut être le débit binaire de l'une des trois structures de prédiction IPP, IBP de JMVM ou Simulcast.

	MVV	QP = 22	QP = 27	QP = 32	QP = 37	QP = 40
Débit binaire (Kbit/s)	Ballroom	1,99	2,00	1,78	1,69	1,49
	Exit	0,27	-0,28	-0,78	-1,69	-1,57
	Vassar	0,44	0,71	-1,26	-5,56	-8,02
	Race1	1,67	2,42	2,51	2,97	3,51
	Rena	2,87	2,84	2,66	1,47	0,19
PSNR (Db)	Ballroom	0,01	0,01	0,06	0,10	0,13
	Exit	-0,008	-0,03	-0,05	-0,04	-0,05
	Vassar	-0,01	-0,07	-0,11	-0,10	-0,14
	Race1	-0,03	-0,012	-0,001	0,005	-0,009
	Rena	-0,0487	-0,101	-0,06	0,0349	0,052

Tableau 4.4. *L'évaluation de l'efficacité de la compression de l'approche proposée par rapport à la structure de prédiction IBP.*

La consistance de l'approche proposée peut apparaître encore mieux dans le cas de plus de 8 caméras. Le tableau 4.5 montre les résultats obtenus en comparaison avec la structure de prédiction IBP. Nous avons utilisé seulement la structure de prédiction IBP. En effet, son accès aléatoire inter-vues est plus rapide que celui de l'approche IPP avec une efficacité de compression pratiquement similaire. L'efficacité de compression de la structure IBP est également plus importante par rapport à la compression Simulcast. La vidéo multi-vues utilisée dans ce cas est "*Rena*", où nous avons utilisé seulement 16 vues. L'évaluation est

effectuée en utilisant le gain en débit binaire (Δ bit-rate) et en PSNR de la composante de luminance Y (Δ PSNR) qui est égal à :

$$\Delta PSNR = PSNR_{proposed} - PSNR_{IBP} \tag{4.16}$$

La variation de Δbit-rate dépend du nombre de vues et du choix utilisé parmi les trois choix proposés à savoir "I/P , B, P , B, P" , "I/P , B , B, P , B , P "ou" I/P , B, B , P, B , B, P". D'autre part, le Δ bit-rate dépend aussi de la structure IBP. Par exemple, si la structure IBP contient deux vues-P successives avec l'implémentation du premier choix pour la structure proposée, le gain dans ce cas-là est faible. Ce cas est réalisable avec un modèle de huit caméras où le gain en débit binaire est de 2,87 %. En revanche, le cas d'un modèle à 13 caméras, la structure proposée utilise le troisième choix tandis que la structure IBP ne contient pas de P-vues successives. Le gain en débit peut ainsi atteindre 6,81% pour QP égal à 27. La figure 4.9 illustre le gain obtenu dans les différents cas avec les diverses valeurs de QP.

QP		**8 Vu**	**9Vu**	**10Vu**	**11Vu**	**12Vu**	**13Vu**	**14Vu**	**15Vu**	**16Vu**
22	Δbit-rate (%)	2,87	5,49	5,40	4,39	5,31	6,13	4,87	5,50	5,08
	ΔPSNR(dB)	-0,05	-0,07	-0,06	-0,06	-0,05	-0,05	-0,05	-0,03	-0,02
27	Δbit-rate (%)	2,84	6,05	5,59	5,10	6,06	6,81	5,62	6,08	5,46
	ΔPSNR(dB)	-0,10	-0,12	-0,10	-0,10	-0,12	-0,09	-0,11	-0,06	-0,06
32	Δbit-rate (%)	2,66	6,00	4,97	5,09	5,72	6,13	5,37	5,42	4,35
	ΔPSNR(dB)	-0,06	-0,06	-0,05	-0,06	-0,08	-0,05	-0,08	-0,02	-0,01
37	Δbit-rate(%)	1,47	4,30	2,96	3,15	3,31	4,30	2,93	3,17	2,31
	ΔPSNR(dB)	0,03	0,01	0,03	0,01	0,0002	0,02	-0,009	0,04	0,05
40	Δbit-rate (%)	0,19	3,39	2,35	1,90	2,00	3,45	1,43	2,26	1,85
	ΔPSNR(dB)	0,05	0,06	0,10	0,08	0,07	0,08	0,05	0,08	0,11

Tableau 4.5. *L'évaluation de l'efficacité de la compression de l'approche proposée par rapport à la structure de prédiction IBP en utilisant 16 vues de la vidéo Rena.*

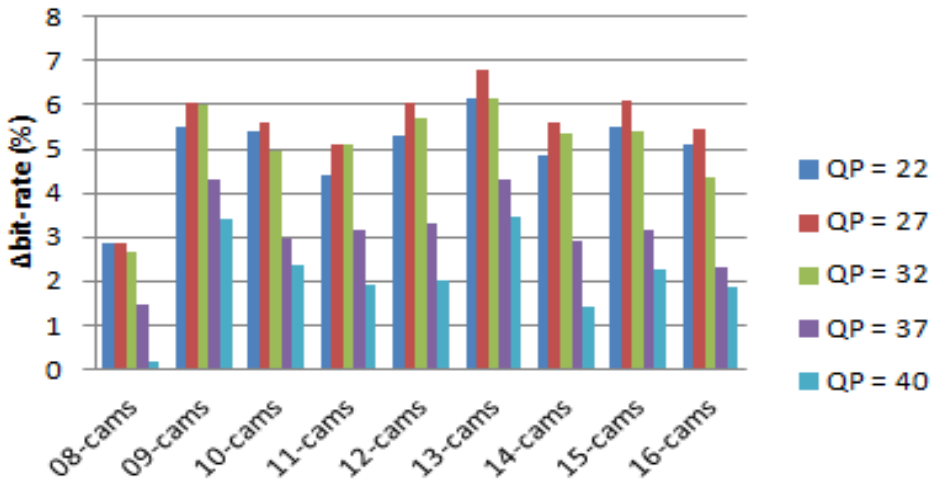


Figure 4.9. *Le gain en débit binaire pour 16 caméras.*

4.6.2. L'accès aléatoire inter-vues

Cette section expose les résultats relatifs à une analyse comparative de la performance en termes d'accès aléatoire inter-vues du modèle proposé avec les deux structures IPP et IBP. Cette analyse est achevée à travers le taux ΔN_{MAX} et le T_{img} qui représente le taux du Nbr_{img} de chaque image en fonction de la taille du GGOP. Le T_{img} peut être obtenu différemment, en fonction du type d'ordre et du type de l'image dans chaque vue, tel que présenté dans les sections précédentes.

La modification de la position de la vue de base peut diminuer le nombre maximal N_{MAX}, à décoder pour accéder à une image donnée. Par exemple, l'accès à l'image B4 située dans S6/T7 nécessite le décodage d'un nombre maximum d'images de référence jusqu'à la vue de base S3. L'utilisation de deux vues-B successives, peut également permettre un gain significatif en N_{MAX} , ce gain est déterminé par:

$$\Delta N_{MAX} = \frac{(IBP\_N_{MAX} - proposed\_N_{MAX})}{IBP\_N_{MAX}} \times 100[\%] \tag{4.17}$$

Où IBP_N_{MAX} et Proposed_N_{MAX}, indiquent respectivement, le nombre maximum d'images de référence à décoder pour accéder à une image dans la structure de prédiction IBP et dans la structure proposée. Le travail présenté dans ce chapitre avec ces métriques, cette configuration et ces données de test a été publié dans [58].

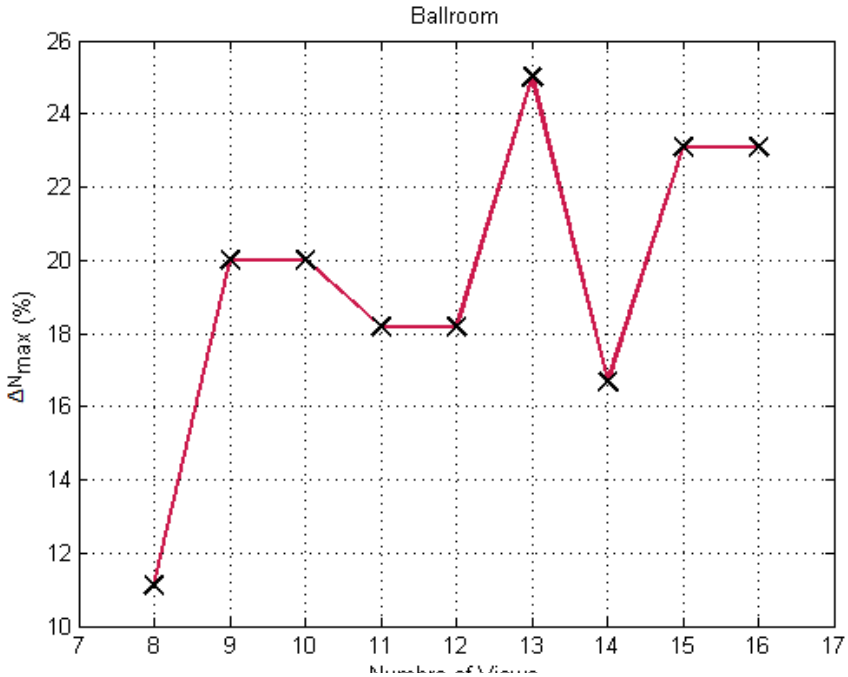


Figure 4.10. *Le gain dans le nombre maximum (ΔN_{MAX}) à décoder, en fonction du nombre de caméras.*

L'augmentation du nombre de vues à décoder conduit à l'utilisation d'un nombre maximum de paires vues-B successive. Ce nombre permet l'utilisation du troisième choix ("I / P , B , B, P , B , B, P") parmi les trois choix étudiés ci-dessus. Ce choix peut fournir un gain important dans Δ N_{MAX} pouvant atteindre les 25% pour 13 caméras, comme le montre la Figure 4.10. Le plus grand gain est obtenu pour 13 vues, où IBP_N_{MAX} est égal à 24 et Proposed_N_{MAX} équivaut à 18. Le gain est moins important dans le cas de 16 caméras (23,07%). Ceci est justifié par le fait, que la valeur IBP_N_{MAX} - Proposed_N_{MAX} (26-20) reste la même pour 13 caméras (24-18) avec une augmentation de la IBP_N_{MAX}. Ces résultats sont obtenus pour un GOP de taille 12 images.

Toutefois, dans le présent travail, nous nous sommes particulièrement intéressés à l'accélération de l'accès aléatoire inter-vues à toutes les images dans les différentes vues. Ces images peuvent être calculées au moyen de la métrique Nbr_{img} et évaluées en utilisant le taux T_img estimé comme suit :

$$T\_img = \frac{(Nbr_{img})}{Size\_GGOP} \times 100[\%] \tag{4.18}$$

La taille du GGOP est obtenu par :

$$Size\_GGOP = (Nbr_{views} * Size\_GOP) \tag{4.19}$$

Afin de démontrer la fiabilité de l'approche proposée, nous avons utilisé un GOP de taille 8. Ainsi, le Nbr_{img} à évaluer est de (8-1)*8 dans le cas de huit vues et de (16-1)*8 pour le cas de seize vues. Les images de la vues de base ne sont pas prises en charge dans les deux cas. En effet, le Nbr_{img} reste le même quel que soit la structure de prédiction utilisée. Les données résumées dans le tableau 4.6 sont obtenues pour un modèle de huit caméras. Les résultats obtenus en utilisant seize vues sont exposés dans le tableau 4.7.

		IPP	**IBP**	**Proposée**
< 5%	Nbr_{img}	3	5	7
	T_{img}	4.6875	7.8125	10.9375
>=5%&<10%	nbr_{img}	4	10	8
	T_{img}	6.25	15.625	12.5
>=10%&<15%	Nbr_{img}	8	10	12
	T_{img}	12.5	15.625	18.75
>=15%&<20%	Nbr_{img}	3	9	11
	T_{img}	4.6875	14.0625	17.1875
>=20%	Nbr_{img}	38	22	18
	T_{img}	59.375	34.375	28.125

Tableau 4.6. *Le gain en accès aléatoire inter-vues.*

Par comparaison des résultats obtenus avec ceux de la structure IPP en utilisant le taux *T_img*, la méthode proposée permet de réduire ce dernier de plus de 50 % comme le montrent la figure 4.11 et le tableau 4.7. Si nous prenons l'exemple de la structure IBP, il ya seulement deux vues qui utilisent directement la vue de base S0 pour la prédiction, qui sont S1 de type vue-B et S2 de type vue-P. Toutefois, dans la méthode proposée cinq vues sont directement prédites à partir de la vue de base S3 , où trois d'entre elles sont des vues-B et deux vues-P. Dans le cas de plus de huit caméras, ce nombre doit être de six (quatre vues-B et deux vues-P). En outre, 6,25% d'images d'un GGOP qui nécessitent le décodage de plus de 28% des images du GGOP pour son accès aléatoire inter-vues dans la structure IBP, nécessitent seulement 25% dans le cas du schéma proposé.

Le gain distinct dans le taux *T_img* de l'approche proposée par rapport à la méthode IBP et la structure IPP est plus intéressant dans le cas de seize caméras. Comme il est présenté sur la figure 4.12. Par comparaison avec la structure IPP, le taux *T_img* dépasse les 61%, soit environ quatre fois de plus par rapport à notre structure où le T_img ne dépasse pas 16 %. Avec une analyse des résultats saisis, nous estimons que 20,31% des images d'un seul GGOP qui exigent un T_img entre 15,62% et 20,31% dans la structure IBP, ont besoin, dans notre structure, d'un T_img moins de 15,62%.

		IPP	**IBP**	**Proposée**
< 5%	Nbr_{img}	07	19	24
	T_{img}	5.469	14.844	18.75
>=5%&<10%	Nbr_{img}	16	30	36
	T_{img}	12.5	23.437	28.125
>=10%&<15%	Nbr_{img}	17	44	52
	T_{img}	13.281	34.375	40.625
>=15%&<20%	Nbr_{img}	8	23	08
	T_{img}	6.25	17.969	6.25
>=20%	Nbr_{img}	72	4	0
	T_{img}	56.25	3.125	0

Tableau 4.7. *Le gain en accès aléatoire inter-vues.*

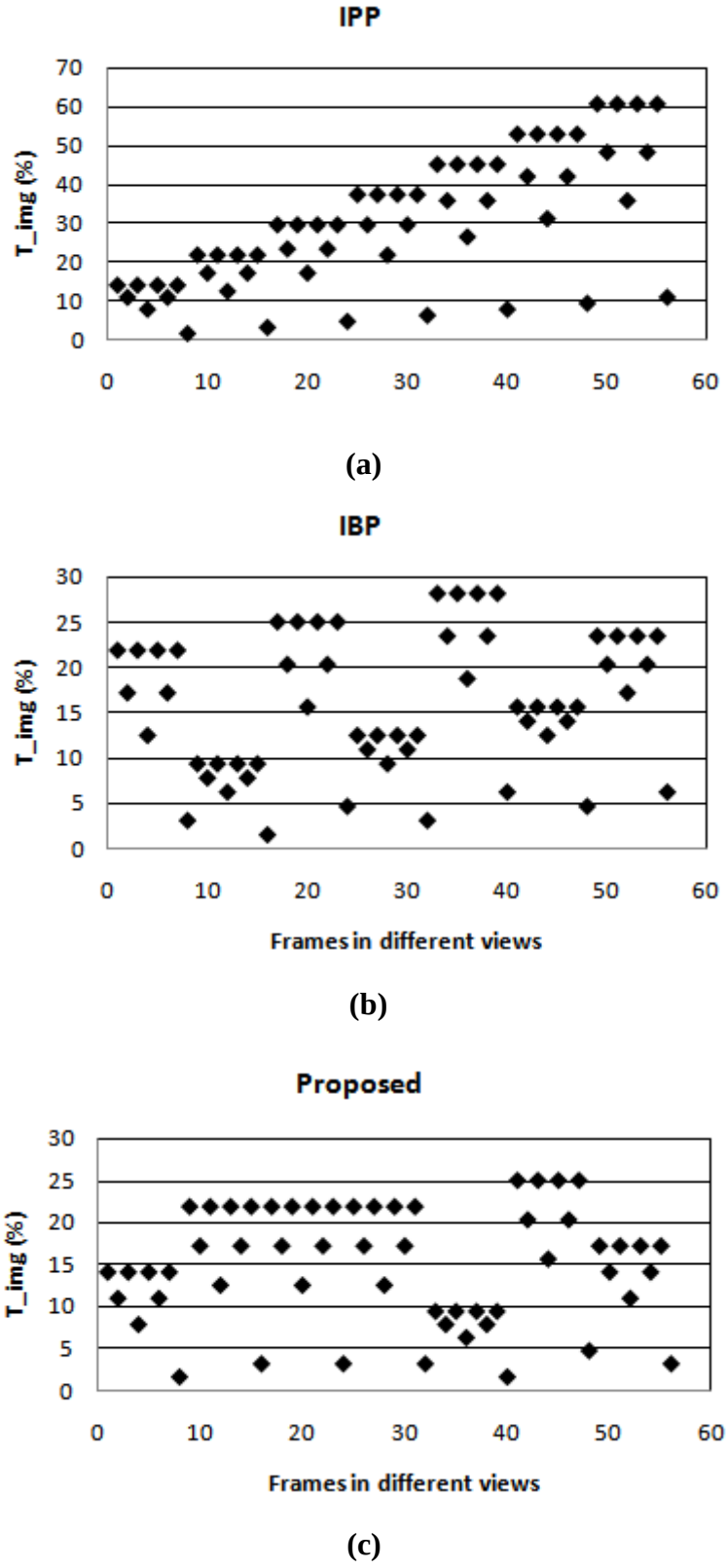
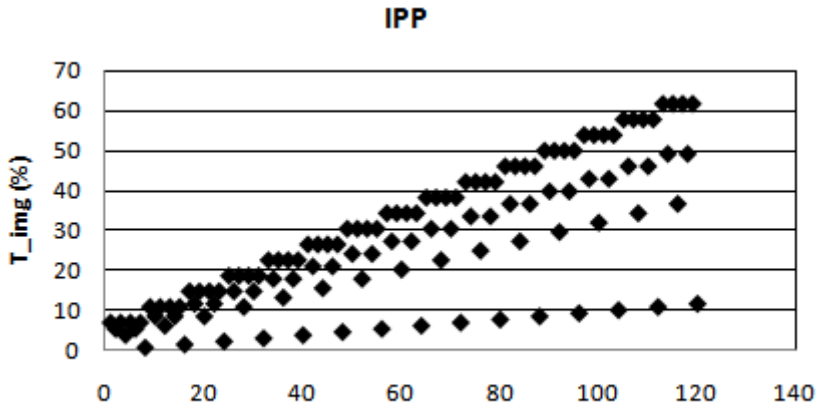
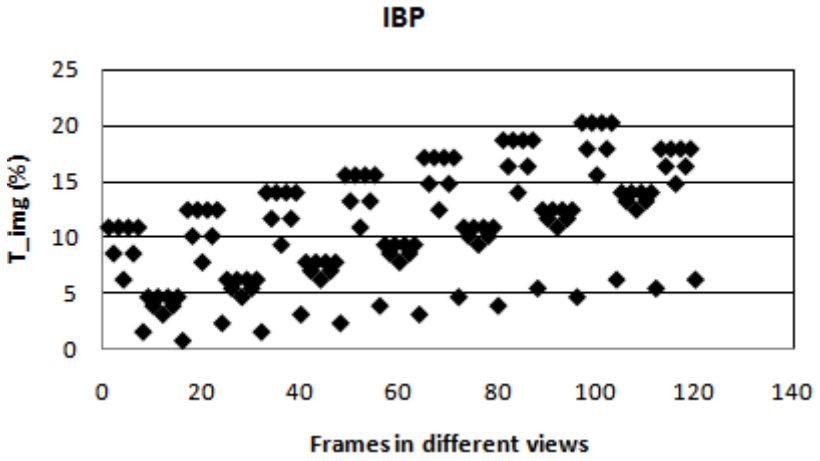


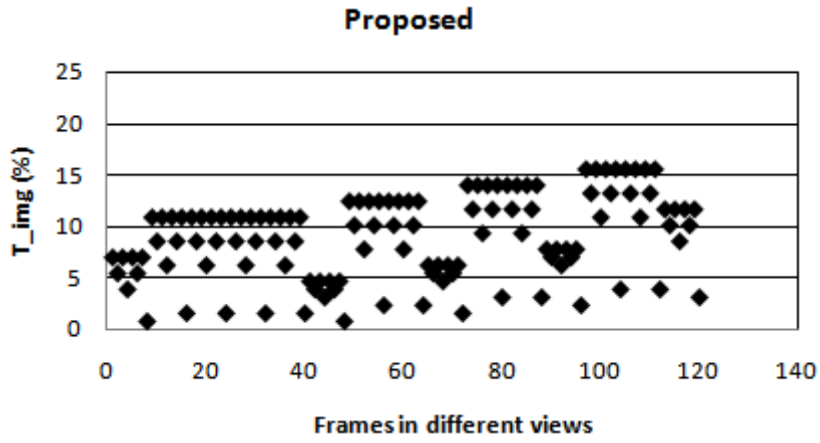
Figure 4.11. *Le gain dans l'accès aléatoire inter-vues pour les différentes images. (a) la structure IPP, (b) la structure IBP, (c) la structure proposée.*



(a)



(b)



(c)

Figure 4.12. *Le gain dans l'accès aléatoire inter-vues pour les différentes images dans le cas de 16 caméras. (a) la structure IPP, (b) la structure IBP, (c) la structure proposée.*

4.7. Conclusion

Afin d'évaluer et d'améliorer l'accès aléatoire inter-vues et également le débit binaire requis pour la compression de la vidéo multi-vues, une nouvelle structure de prédiction inter - vues, ainsi qu'une méthode d'évaluation adéquate ont été proposées. Pour accélérer l'accès aléatoire inter-vues lors de l'opération de décodage, nous recommandons l'utilisation d'un maximum de paires de vues-B successives. Ceci peut varier en fonction de la possibilité donnée par le nombre de vues utilisées. L'utilisation de deux vues-B successives peut aussi compenser le débit requis. La position de la vue de base "I" a également été modifiée afin d'accélérer l'accès aléatoire inter-vues à toute image dans les différentes vues. L'approche proposée a été évaluée et comparée avec les méthodes existantes en utilisant l'accès aléatoire inter-vues au moyen du gain ΔN_{MAX} et aussi le taux d'images T_img. Le ΔN_{MAX} exprime le nombre maximum d'images de référence à décoder tandis que le T_img décrit l'évaluation de l'accès aléatoire à toutes les images dans un GGOP.

L'efficacité de la compression qui est représentée par le compromis entre le débit binaire et la qualité de la vidéo compressée a également été utilisée dans l'évaluation du système proposé. Les résultats expérimentaux ont montré un gain significatif en ΔN_{MAX} qui atteint 25 % par rapport à la structure de prédiction IBP et également une réduction très importante du pourcentage T_img par rapport aux deux structures IBP et IPP. Les résultats expérimentaux ont également montré l'important gain en débit binaire obtenu par la structure proposée qui est d'environ 6,81% avec une qualité de la vidéo multi-vues similaire mesurée par le PSNR vis-à-vis de la structure de prédiction IBP.

Chapitre 05 : Compromis débit binaire et accès aléatoire inter-vues

5.1. Introduction

L'utilisation des paires de vues-B successives a permis un gain considérable en débit binaire. Ceci, sans alourdir l'accès aléatoire inter-vues. Bien au contraire, elle a garanti un gain qui peut atteindre 25% en accès aléatoire inter-vues. Ceci est dû à l'utilisation du même niveau hiérarchique que la structure IBP tout en diminuant les vues-P. L'utilisation de plus que deux vues-B successives sans augmenter le niveau hiérarchique améliore davantage l'accès aléatoire inter-vues. Néanmoins, ceci augmente également le débit binaire nécessaire. Dans le but d'améliorer encore plus l'accès aléatoire inter-vues tout en conservant la même efficacité de la compression, un nouveau modèle de prédiction inter-vues est proposé. Ce modèle vise essentiellement l'amélioration du schéma à des paires de vues-B successives présenté dans le chapitre 04. La méthode proposée se base sur l'utilisation de trois vues-B successives avec une augmentation du niveau hiérarchique afin de conserver un meilleur débit binaire. Néanmoins, en raison de l'augmentation du niveau hiérarchique, cette technique peut alourdir l'accès aléatoire inter-vues. Pour remédier à ce problème la prédiction inter-vues des images non-clés dans certaines vues-B est éliminée. L'approche proposée est détaillée dans ce chapitre en comparant les deux propositions ; avec et sans prédiction inter-vues des images non-clés de certaines vues-B.

Dans ce chapitre nous organisons le travail comme suit : Nous présentons premièrement la méthode proposée pour l'élimination de la prédiction inter-vues de certaines vues-B tout en justifiant notre choix des vues-B. Ensuite, nous détaillerons le choix préconisé sur seulement huit caméras avec et sans l'élimination de la prédiction inter-vues prévue. La présentation de la généralisation de la méthode proposée est ensuite achevée vis-à-vis de l'amélioration du débit et de l'accès aléatoire inter-vues. La troisième section présente le modèle d'évaluation proposé pour les diverses structures proposées ; sans et avec élimination de prédiction inter-vues. Enfin, nous validerons cette approche par une étude expérimentale détaillée permettant de présenter le gain significatif en accès aléatoire inter-vues obtenu par rapport à la structure présentée dans le chapitre précédent.

5.2. Elimination de la prédiction inter-vues pour les images non-clés

L'augmentation excessive du nombre des vues-B successives entre deux vues-I/P, sans l'augmentation du niveau hiérarchique des images clés et non-clés, accroit le débit binaire nécessaire. Cette contrainte est justifiée par l'éloignement des vues de référence I et P les unes des autres. Les figures 5.1 illustrent un exemple de trois vues-B successives, pour les images

clés et non-clés sans l'augmentation du niveau hiérarchique. Autrement-dit, le niveau hiérarchique maximum est de quatre.

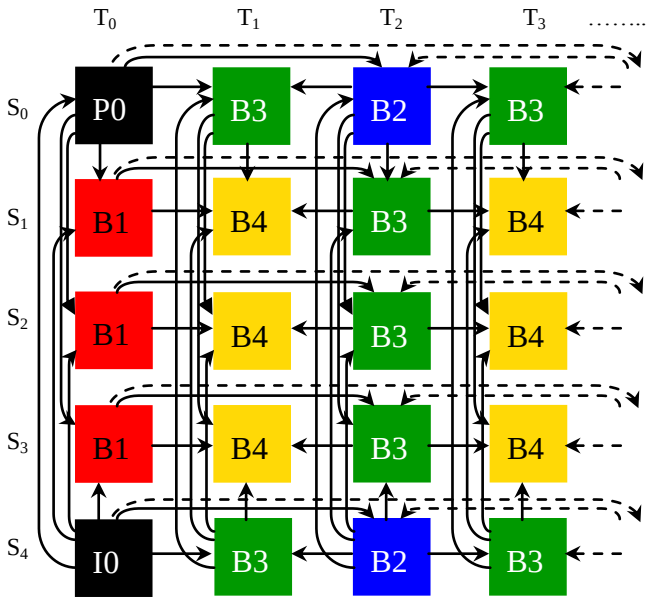


Figure 5.1. *Architectures à trois vues-B successives sans l'augmentation du niveau hiérarchique.*

Pour que la prédiction inter-vues soit efficace, les vues de référence de type I et P doivent être les proches possible les unes des autres. Á cet effet, la corrélation entre les vues-B et les vues-I/P augmente et la compensation de disparité s'améliore. Pour ce faire, nous avons utilisé dans notre proposition trois vues-B successives avec l'ajout d'un niveau hiérarchique. L'incrustation du niveau hiérarchique inter-vues permet de diminuer la distance entre les vues de référence. Dans ce cas-là, l'une des trois vues-B successives doit servir de référence pour les deux autres vues-B. Ceci est faisable par l'augmentation du niveau hiérarchique des images clés et non-clés pour deux vues-B parmi les trois vues successives. La figure 5.2 (a) montre un exemple des images clés et non-clés où le niveau hiérarchique maximum devient cinq pour les images non-clés et deux pour les images clés (B2). La vue S2 de type B1 de la figure 5.2 (a) est utilisée en plus des vues-I/P comme vue de référence pour les vues S1 et S3 de type B2. Ceci permet de résoudre le problème de la prédiction distante posée par les vues-B successives sans l'augmentation du niveau hiérarchique (voir la figure 5.1). Un nombre plus élevé de vues-B signifie la diminution obligatoire de vues-P. Ceci permet un gain considérable en débit binaire avec une qualité similaire de la vidéo encodée. Néanmoins, l'ajout d'un niveau hiérarchique a l'inconvénient d'augmenter également le nombre maximum d'images de référence nécessaires à la consultation d'une image donnée SnTn. Á

titre d'exemple, l'accès aléatoire à l'image S3/T1 de la vue S3 nécessite le décodage de 19 images réparties comme suit :

- Cinq images de la vue S0 qui sont : S0/T0, S0/T1, S0/T2, S0/T4, S0/T8.
- Cinq images de la vue S2 qui sont : S2/T0, S2/T1, S2/T2, S2/T4, S2/T8.
- Quatre images de la vue S3 qui sont : S3/T0, S3/T2, S3/T4, S3/T8.
- Cinq images de la vue S4 qui sont : S4/T0, S4/T1, S4/T2, S4/T4, S4/T8.

L'accès à la même image dans le cas de la figure 5.1 nécessite seulement le décodage de 14 images de référence réparties comme suit:

- Cinq images de la vue S0 qui sont : S0/T0, S0/T1, S0/T2, S0/T4, S0/T8.
- Quatre images de la vue S3 qui sont : S3/T0, S3/T2, S3/T4, S3/T8.
- Cinq images de la vue S4 qui sont : S4/T0, S4/T1, S4/T2, S4/T4, S4/T8.

Dans le but de respecter les exigences de la MVV détaillées dans [9] et d'améliorer le débit binaire et l'accès aléatoire inter-vues, un compromis entre l'efficacité de compression et l'accès aléatoire inter-vues doit être satisfait. Ainsi, le nombre d'images de référence utilisées pour la prédiction d'images non-clés des vues-B2 doit être diminué. Autrement dit, chaque image non-clé d'une vue-B2 doit utiliser deux images de référence dans le niveau temporel et une seule image inter-vues. La quatrième image de référence éliminée est celle de la vue-B1. Dans ce cas-là, le nombre d'images de référence, nécessaire à la consultation de l'image S3/T1 est réduit à 13 trouvant sur les vues suivantes :

- Deux images de la vue S0 qui sont : S0/T0, S0/T8.
- Deux images de la vue S2 qui sont : S2/T0, S2/T8.
- Quatre images de la vue S3 qui sont : S3/T0, S3/T2, S3/T4, S3/T8.
- Cinq images de la vue S4 qui sont : S4/T0, S4/T1, S4/T2, S4/T4, S4/T8.

Á la différence des images non-clés, les images clés des vues-B2 utilisent les vues-B1 pour la prédiction inter-vues. En effet, ces images n'altèrent pas beaucoup l'accès aléatoire inter-vues. Nous n'avons pas opté dans cette approche pour la prédiction des images non-clés de la première. En effet, les vues-P sont relativement distantes les unes des autres, ce qui empêche une compensation de disparité pertinente. Ce cas est autorisé seulement pour la

dernière vue-P si le nombre de vues-B successives est inférieur à 3. Ce nombre change selon le nombre de caméras utilisées. Une étude détaillée du cas général est présentée dans la section 5.3. Un exemple de l'approche proposée après l'élimination de la quatrième image de référence est présenté dans la figure 5.2 (b).

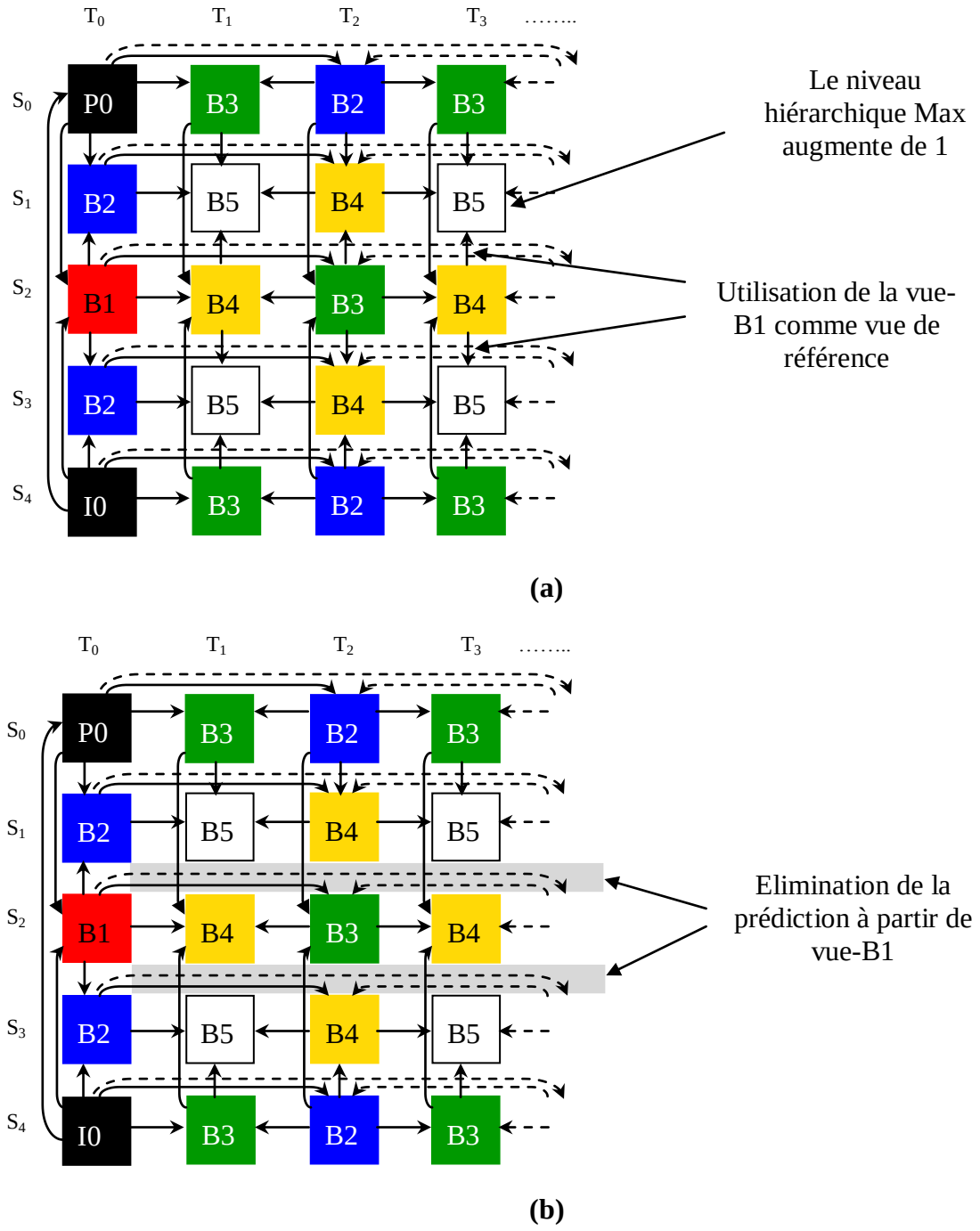


Figure 5.2. *Architectures à trois vues-B successives, (a) utilisation de la vue-B1 comme vue de référence, (b) élimination de la prédiction des images non-clés à partir de la vue-B1.*

5.3. Amélioration de la structure de prédiction inter-vues

La position de la vue de base est choisie de telle sorte qu'elle permet l'utilisation de trois vues-B successives. La vue de base est S4 dans la figure 5.2. La vérification de la fiabilité de la structure proposée est accomplie via l'implémentation de deux structures de prédiction. Ces structures sont appelées *Proposed1* et *Proposed2* (détaillée dans [59]) désignant successivement l'approche avec et sans prédiction des images non-clés des vues-B2. Par cette technique la structure *Proposed1* fournit un gain considérable en débit binaire par rapport à *Proposed2*. En effet, ce gain est dû à l'utilisation de toutes les images de référence pour la prédiction des images non-clés des vues-B2 par *Proposed1*. En contrepartie, la structure *Proposed2* n'utilise que trois images de référence. En revanche, la structure *Proposed2* offre une amélioration très importante de l'accès aléatoire inter-vues vis-à-vis de *Proposed1* pour les mêmes raisons.

La figure 5.3 présente le schéma général de la structure *Proposed2* pour huit vues où la taille de chaque GOP est de 8. La structure *Proposed1* doit être obtenue de la même façon que dans la figure 5.3 à l'exception des vues-B2 où il faut ajouter les images non-clés des vues-B1 voisines comme images de référence pour les images non-clés des vues-B2. Ceci n'est applicable que lorsque le nombre de vues-B successives est de trois. Si on compare la structure *Proposed2* avec celle détaillée dans le chapitre 04, les vues profitant d'un accès aléatoire inter-vues à partir de la vue de base sont S0, S2, S3, S5, S6, S7. Le cas de la structure du quatrième chapitre sont au nombre de cinq. En plus, seulement deux vues-P sont utilisées dans cette structure ce qui permet l'amélioration de l'efficacité de la compression et également l'accès aléatoire inter-vues.

5.4. L'approche pour plusieurs caméras

Tout comme la méthode proposée dans le chapitre 04, l'importance de l'amélioration présentée dans ce chapitre peut apparaitre plus clairement par plus de huit vues. En effet, l'augmentation du nombre de vues-B successives accroît également les cas possibles après la vue de base. Les structures de prédiction *Proposed1* et *Proposed2* doivent avoir le même ordre général de vues et les mêmes choix adoptés après la vue de base. La seule différence entre les deux structures est l'utilisation ou non des vue-B1 comme vues de référence par les vues-B2. D'une manière générale, l'approche doit vérifier dans le cas général deux points essentiels :

- Les vues utilisées ne doivent jamais contenir des vues-P successives pour des raisons d'amélioration de l'accès aléatoire inter-vues.

- L'utilisation maximum des vues-B successives est le point primordial. En effet, cette approche doit favoriser l'utilisation de trois vues-B successives puis deux vues-B, sinon une seule vue-B. Ceci est contrôlé principalement par le nombre de vues après la vue de base I.
- La prédiction inter-vues des images non-clés de la dernière vue-P n'est possible que lorsque le nombre de vues-B entre cette vue et la vue-P précédente est inférieur à trois.

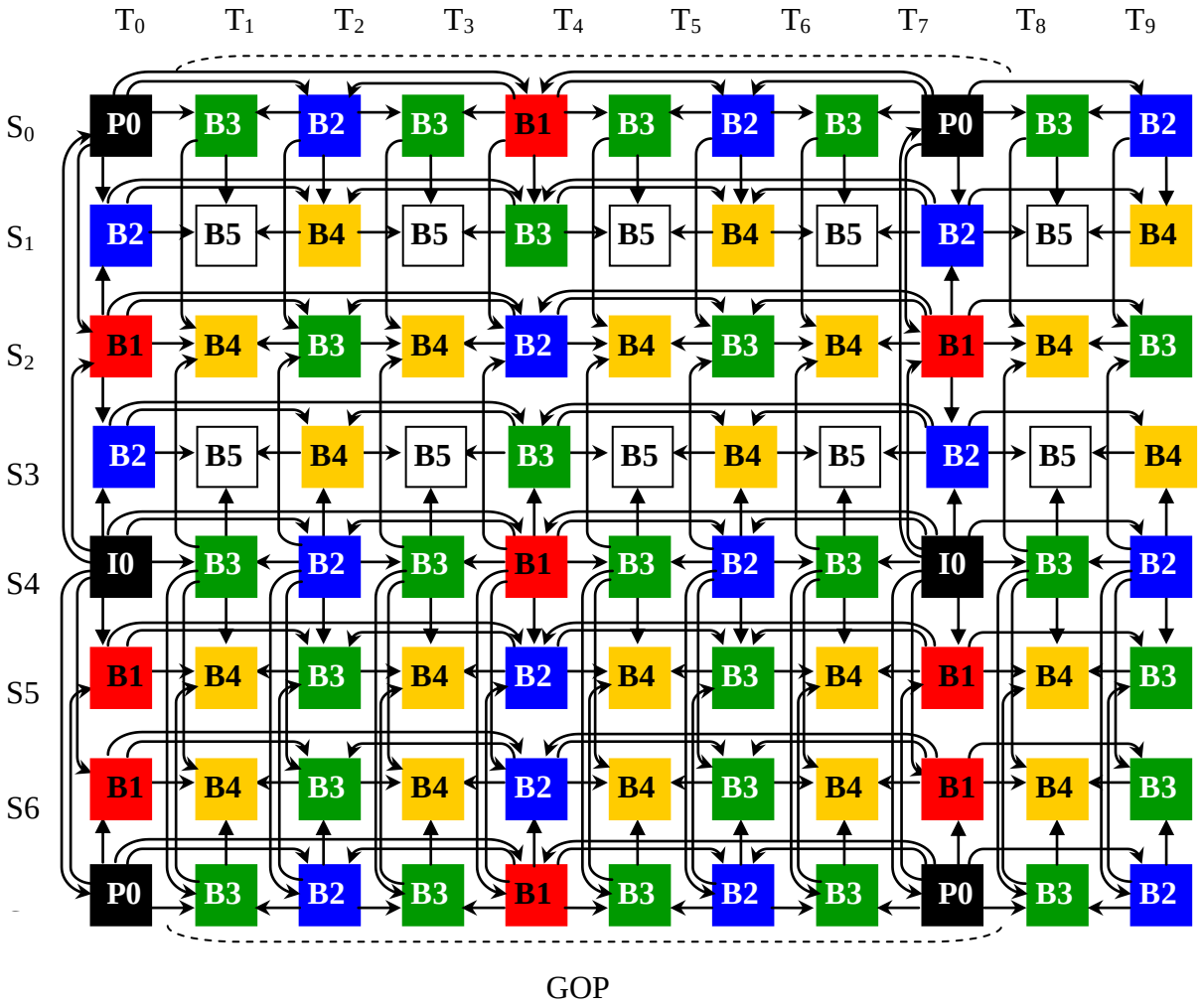


Figure 5.3. *La structure de prédiction proposée sans l'utilisation de prédiction pour les images non-clés des vues-B2.*

L'ordre des vues figurant après la vue de base change en fonction du nombre de caméras utilisées tout en respectant les trois points exigés ci-dessus. La figure 5.4 présente les cas possibles en se limitant à 16 vues. Normalement 11 vues suffisent amplement, après la 11ème caméra les cas se répètent. Dans la figure 5.4, la démonstration est effectuée seulement à l'aide des images clés. Les images non-clés suivent leurs images clés pour la désignation du niveau hiérarchique. L'interdiction de l'utilisation des vues-P consécutives et la préférence de

trois vues-B successives ont imposé quatre choix en fonction du nombre de vues utilisées. L'expression permettant chaque fois la désignation du choix adéquat, peut être donnée comme suit:

$$Ordre = \begin{cases} "I/P, B_1, B_1, P" & if\ (Nbr_{views}\ MOD\ 4) = 0 \\ "I/P, B_2, B_1, B_2, P" & if\ (Nbr_{views}\ MOD\ 4) = 1 \\ "I/P, B_1, B_1, P, B_1, P" & if\ (Nbr_{views}\ MOD\ 4) = 2 \\ "I/P, B_2, B_1, B_2, P, B_1, P" & if\ (Nbr_{views}\ MOD\ 4) = 3 \end{cases} \quad (5.1)$$

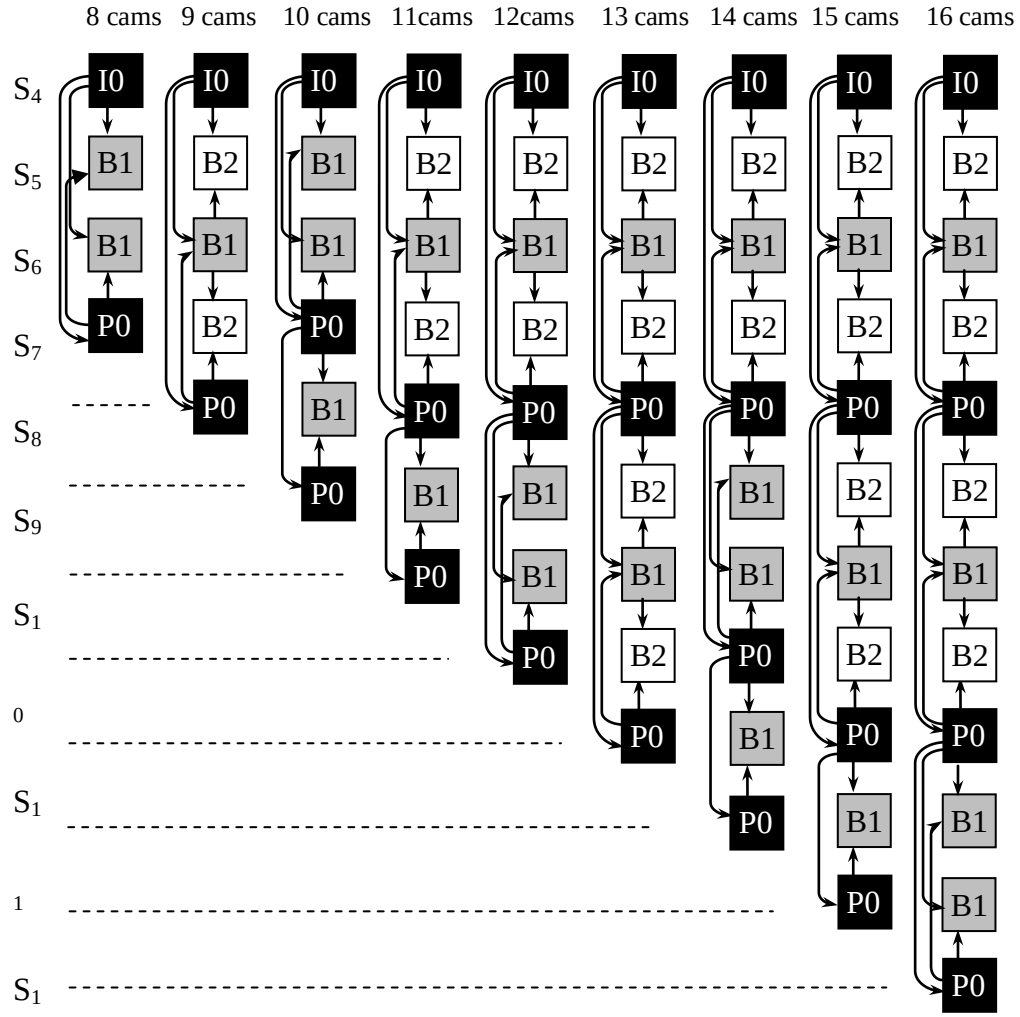


Figure 5.4. *Généralisation de l'approche proposée en utilisant 16 vues, seules les images clés sont utilisées.*

Chaque fois, le choix s'applique sur les vues situées après le cas « I/P, B_2, B_1, B_2, ». Par exemple le cas de 8 vues, le choix adéquat est «I, B_1, B_1, P » les vues précédant ce choix sont « P, B_2, B_1, B_2, ». Un autre exemple semble utile est lorsque le nombre de vues est 14, dans ce

cas le chois approprié est « P, B_1, B_1, P, B_1, P », les vues qui précèdent ce choix vérifient aussi la contrainte « I, B_2, B_1, B_2, ».

5.5. Évaluation proposée pour l'accès aléatoire inter-vues

Cependant, l'évaluation de l'accès aléatoire des deux structures *Proposed1* et *Proposed2* est accomplie à l'aide du N_{MAX} et Nbr_{img} détaillés dans le chapitre 04. Nous essayons dans cette section d'estimer les deux métriques N_{MAX} et Nbr_{img} pour les deux cas de l'approche proposée, avec et sans prédiction des images non-clés des vues-B2. L'évaluation de la structure *Proposed1* est effectuée à titre comparatif afin de démontrer l'importance d'élimination de la prédiction à partir de la vue-B1 présentée dans la section précédente. Il est évident que l'augmentation du niveau hiérarchique maximum accroît le nombre d'images maximum à décoder N_{MAX}. En effet, dans le cas de la structure *Proposed1*, ce nombre dépend fortement du nombre de blocs de trois vues-B successives utilisées. L'adoption d'une prédiction à quatre images de référence pour les images non-clés de vues-B2 produit un nombre plus grand de N_{MAX} que la structures *Proposed2*, IBP et la structure proposée dans le chapitre 04. Le N_{MAX} de la structure *Proposed1* dans le cas général peut être calculé comme suit :

$$
\begin{aligned}
&if\ (Nbr_{view}) <= 12 \\
&\quad N_{MAX} = 4 * hierarchy_{MAX} - 1 \\
&else \\
&\quad N_{MAX} = 4*(Hierarchy_{MAX}) - 1 + 2*((Nbr_{view} - 5)\ div 4 - \alpha) \\
&where \\
&\quad \alpha = \begin{cases} 1 & if\ (Nbr_{view} - 5) \mod 4 == 0,2,3 \\ 2 & if\ (Nbr_{view} - 5) \mod 4 == 1 \end{cases}
\end{aligned}
\tag{5.2}
$$

Où $hierarchy_{MAX}$ explique le niveau hiérarchique maximal qui est égale dans cette structure à 5. Le Nbr_{view} désigne le nombre de vues utilisées chaque fois. Si le nombre de vues Nbr_{view} utilisé est inferieur ou égal à 12, le N_{MAX} est égal à 19. Dans ce cas-là, les images nécessitant un N_{MAX} égal à 19 varient selon le nombre de vues comme suit :

- Pour 8 et 10 vues : Les images non-clés B5 des vues S1 et S3.
- Pour 9 et 12 vues : Les images non-clé B5 des vues S1, S3, S5 et S7.

En effet, dans le cas de 12 caméras, la vue-B la plus éloignée de la vue de base est S10. Malgré cela, les images B5 de cette vue nécessitent seulement 16 images à décoder pour un accès aléatoire. Ceci, est dû à l'utilisation de trois vues-B successives. Si le nombre de

caméras est strictement supérieur à 12, le nombre de blocs de trois vues-B successives augmente (voir la figure 5.4). Dans ce cas-là, le N_{MAX} peut être calculé différemment comme le montre le tableau 5.1.

Nbr_{view}	Nbr blocs de 3 vues-B successives après vue-I	Vues contenant les images nécessitant N_{MAX}	N_{MAX}
13, 15, 16	2	S9, S11	21
14	1	S5, S7	19
17	3	S13, S15	23

Tableau 5.1. *Le N_{MAX} pour plus de 12 caméras pour la structure Proposed1.*

La deuxième métrique Nbr_{img}, utilisée pour l'évaluation, doit être calculée pour toutes les images du GGOP à l'exception des images de la vue de base. En effet, l'estimation du Nbr_{img} varie en fonction du type de l'image à consulter et de la vue contenant cette dernière. Les images des vues précédant la vue de base doivent être estimées aussi d'une manière particulière. Nous évaluons premièrement les images des vues situées après la vues de base, puis nous reviendrons aux images des vues précédant la vue-I. Les images clés des deux structures de prédiction *Proposed1* et *Proposed2* utilisent le même nombre d'images de référence pour la prédiction. Ainsi, le Nbr_{img} est obtenu pour ces images de la même façon pour les deux structures en utilisant :

$$Nbr_{img} = \alpha + \left\lfloor \frac{Num_{view} - \beta}{4} \right\rfloor$$
$$where \begin{cases} \alpha = 0, \beta = 2 & for\ anchor\ frames\ of\ P-views \\ \alpha = 1, \beta = 1 & for\ anchor\ frames\ of\ B1-views \\ \alpha = 2, \beta = 1 & for\ anchor\ frames\ of\ B2-views \end{cases} \quad (5.3)$$

Le Num_{view} représente le numéro d'ordre de la vue courante. L'expression 5.3 est conçue de telle sorte que nous pouvons l'adapter aux quatre choix présentés ci-dessus.

Tout comme les images clés, la même méthode peut être appliquée pour l'évaluation des images non-clés des vues-B1 et vues-P pour les deux structures *Proposed1* et *Proposed2*. Néanmoins, le Nbr_{img} nécessaire aux images des vues-B2 doit être estimé différemment pour les deux structures. L'expression permettant d'obtenir le Nbr_{img} dans le cas des deux vues, P et B1 est donnée par :

$$Nbr_{img} = \alpha * (Hierarchy + \beta) + 2 * \left\lfloor \frac{Num_{view} - \delta}{4} \right\rfloor$$
$$where \quad \begin{cases} \alpha = 1, \beta = 1, \delta = 2 & for\ non-anchor\ frames\ of\ P-views \\ \alpha = 3, \beta = 0, \delta = 1 & for\ non-anchor\ frames\ of\ B1-views \end{cases} \tag{5.4}$$

Où *Hierarchy* dénote le niveau hiérarchique de l'image à consulter. Afin de bien éclaircir le fonctionnement de la formule 5.4, un exemple pour chaque cas peut être donné comme suit :

- Le cas des vues-P : Si la MVV est composée de 12 vues, l'accès aléatoire à l'image S8/T2 de type B2, nécessite le décodage de 5 images. Avec Num_{view} égalant 9 et un niveau hiérarchique de 2.
- Le cas des vues-B1 : L'accès aléatoire à l'image S10/T2 de type B3 où la MVV est composée de 12 vues est égal 13 images. Dans ce cas, le numéro d'ordre Num_{view} de cette image est de 11, le niveau hiérarchique *Hierarchy* est égal 3.

Les images non-clés des vues-B2 nécessitent dans le cas de la structure Proposed1 un nombre Nbr_{img} qui peut être donné par :

$$Nbr_{img} = 4 * Hierarchy - 3 + 2 * \left\lfloor \frac{Num_{view} - 1}{4} \right\rfloor \tag{5.5}$$

Avec l'utilisation d'une MVV de 15 vues à titre d'exemple, l'accès aléatoire à l'image S11/T3 de type B5, nécessite le décodage de 21 images. Dans ce cas, le Num_{view} est égal à 12 et *Hierarchy* est de 5.

Le cas particulier de la dernière vue-P où le nombre de vues-B successives est moins de 3 peut aussi être traité séparément. Le Nbr_{img} des images non-clés dans ce cas est le même pour les deux structures Proposed1 et Proposed2. Ce nombre peut être estimé comme suit :

$$if\ (Nbr_{view}\ MOD\ 4) \quad != 1$$
$$Nbr_{img} = 2 * (Hierarchy) + 1 + 2 * \left\lfloor \frac{(Num_{view} - 2)}{4} \right\rfloor \tag{5.6}$$

Ce cas est possible lorsque le nombre de vues est de 8, 10, 11, 12, 14, 15, 16, 18, …etc. Á titre d'exemple, dans le cas où le nombre de vues utilisées est de 16, l'accès à l'image S15/T5 de type B3 est réalisé après le décodage de 13 images.

Le deuxième cas particulier dans cette approche est celui des vues précédant la vue-I en l'occurrence S0, S1, S2 et S3. Le Nbr_{img}, dans ce cas-là, est de 1, 2 et 3 respectivement pour les vues P, B1 et B2. En outre, les images non-clés des deux vues-P et B1 sont évaluées pour les deux structures analysées dans ce chapitre en utilisant :

$$Nbr_{img} = \alpha * Hierarchy + \beta \quad where \quad \begin{cases} \alpha = 1, \beta = 3, & for\ non-anchor\ frames\ of\ P-views \\ \alpha = 3, \beta = 2 & for\ non-anchor\ frames\ of\ B1-views \end{cases} \tag{5.7}$$

Le cas de vues-B2, ce nombre peut être estimé de la même façon par :

$$Nbr_{img} = 4 * Hierarchy - 1 \tag{5.8}$$

L'évaluation de la structure *Proposed2* passe par les mêmes étapes. Ainsi, nous essayons d'estimer tout d'abord le N_{MAX}, ensuite, le Nbr_{img} pour le cas particulier des vues-B2 avant et après la vue-I sera établi. Le nombre maximum d'images à décoder N_{MAX} peut être calculé par:

$$N_{MAX} = 3*(Hierarchy_{MAX} - 1) + 2*\left\lfloor \frac{Nbr_{view} - 2}{4} \right\rfloor \tag{5.9}$$

En effet, les images nécessitant un N_{MAX} dans la structure *Proposed1* sont toujours de type B5 des vues-B2. L'élimination de la prédiction de ces images à partir de la vue-B1 dans la structure *Proposed2* a amélioré le N_{MAX}. Ceci a conduit à la diminution du nombre d'images de référence utilisées par les images B5. Ainsi, après cette modification les images nécessitant un N_{MAX} sont devenues dans la structure *Proposed2* les images B4 des vues-B1 comme le montre le tableau 5.2.

Les images non-clés des vues-B2 situées après la vue-I peuvent être calculées en utilisant :

$$Nbr_{img} = 2 * Hierarchy + 1 + 2 * \left\lfloor \frac{Num_{view} - 1}{4} \right\rfloor \tag{5.10}$$

Le cas spécifique des images non-clés des vues-B2 précédant la vue de base est traité de la façon suivante :

$$Nbr_{img} = 2 * Hierarchy + 3 \tag{5.11}$$

Nbr_{view}	Nbr blocs de 3 vues-B successives après vue-I	Vues contenant les images nécessitant N_{MAX}	N_{MAX}
8	0	S2, S5, S6	14
9	1	S2, S6	14
10	0	S8	16
11	1	S9	16
12	1	S9, S10	16
13	2	S10	16
14	1	S12	18
15	2	S13	18
16	2	S13, S14	18

Tableau 5.2. *Le N_{MAX} pour la structure de prédiction Proposed2.*

Nombre de vues	**Ordre d'encodage**
8 vues	S4, S0, S2, S1, S3, S7, S5, S6
9 vues	S4, S0, S2, S1, S3, S8, S6, S5, S7
10 vues	S4, S0, S2, S1, S3, S7, S5, S6, S9, S8
11 vues	S4, S0, S2, S1, S3, S8, S6, S5, S7, S10, S9
12 vues	S4, S0, S2, S1, S3, S8, S6, S5, S7, S11, S9, S10
13 vues	S4, S0, S2, S1, S3, S8, S6, S5, S7, S12, S10, S9, S11
14 vues	S4, S0, S2, S1, S3, S8, S6, S5, S7, S11, S9, S10, S13, S12
15 vues	S4, S0, S2, S1, S3, S8, S6, S5, S7, S12, S10, S9, S11, S14, S13
16 vues	S4, S0, S2, S1, S3, S8, S6, S5, S7, S12, S10, S9, S11, S15, S13, S14

Tableau 5.3. *L'ordre d'encodage selon le nombre de vues utilisées.*

5.6. Expérimentation et résultats

La validation du travail réalisé est toujours accomplie à travers l'étude et le test des deux exigences les plus importantes de la compression de la vidéo multi-vues, à savoir : l'efficacité de la compression et l'accès aléatoire inter-vues. Nous comparons dans cette section les

résultats des deux structures étudiées dans ce chapitre avec ceux de la structure proposée et présentée dans le chapitre précédent.

5.6.1. Evaluation de l'efficacité de la compression

Nous avons utilisé dans l'implémentation des deux structures *Proposed1* et *Proposed2* les mêmes vidéos MVV et paramètres utilisés durant l'implémentation des structures étudiées dans cette thèse à l'exception de l'ordre d'encodage. Le tableau 5.3 illustre l'ordre d'encodage en fonction du nombre de vues utilisées. Il est à noter que l'ordre des cinq premières vues reste inchangé. En effet, l'ordre d'encodage présenté dans le tableau 5.3 n'est pas forcément le seul cas possible. Á titre d'exemple, les quatre vues précédant la vue-I, peuvent être encodées après les autres vues.

	Vue	QP = 22	QP = 27	QP = 32	QP = 37	QP = 40
Débit binaire (Kbit/s)	S0	1609,8973	829,4757	435,0649	233,6486	158,9135
	S1	1350,8432	598,4108	291,0324	156,5405	111,5838
	S2	1340,2486	626,1135	315,8757	178,2432	130,7081
	S3	1348,8378	611,4162	293,373	159,6	114,4649
	S4	1965,5405	1070,5622	606,773	353,9459	253,3459
	S5	1297,0486	619,7297	323,4162	187,9514	139,2054
	S6	1500,0541	661,0216	332,9135	191,7189	141,6054
	S7	1927,2865	987,2757	522,1351	285,3892	201,6865
Moyenne		1542,46958	750,50068	390,072975	218,3797	156,4392
PSNR (Db)	S0	39,3745	37,0676	34,4024	31,658	29,8609
	S1	39,1005	37,1123	34,6769	32,0408	30,3149
	S2	39,4789	37,3904	34,9169	32,3182	30,5639
	S3	39,3707	36,9955	34,3053	31,6379	29,951
	S4	39,226	36,8354	34,2037	31,5477	29,8691
	S5	39,9233	37,6422	34,9835	32,238	30,4424
	S6	38,7667	36,8665	34,5819	32,0828	30,3728
	S7	39,1718	36,6907	33,8848	30,9991	29,2292
Moyenne		39,30155	37,075075	34,494425	31,81531	30,07553

Tableau 5.4. *Les résultats obtenus en termes de débit binaire et PSNR pour la structure Proposed1 en utilisant la vidéo Ballroom (en détails pour chaque vue).*

Les résultats obtenus en termes de débit binaire pour les deux structures de prédiction en utilisant la vidéo *Ballroom* sont présentés respectivement dans les deux tableaux 5.4 et 5.5. Les résultats synthétisés dans les deux tableaux, montrent clairement le gain en débit binaire des deux vues S1 et S3 de la structure *Proposed1* par rapport à *Proposed2*. Néanmoins, ce gain est négligeable vis-à-vis de l'inconvénient de cette structure en termes d'accès aléatoire inter-vues. Le débit binaire obtenu pour les deux structures est tracé en fonction des cinq valeurs QP utilisées comme le montre la figure 5.5. Cette figure illustre la similarité de l'efficacité de compression entre les trois structures Proposed1, Proposed2 et celle présentée dans le chapitre 04 nommée *Proposed.*

	Vue	QP = 22	QP = 27	QP = 32	QP = 37	QP = 40
Débit binaire (Kbit/s)	S0	1609,8973	829,4757	435,0649	233,6486	158,9135
	S1	1421,2919	642,8216	313,173	166,6216	117,6216
	S2	1340,2486	626,1135	315,8757	178,2432	130,7081
	S3	1408,3568	643,8703	310,2541	165,9784	118,4649
	S4	1965,5405	1070,5622	606,773	353,9459	253,3459
	S5	1297,0486	619,7297	323,4162	187,9514	139,2054
	S6	1500,0541	661,0216	332,9135	191,7189	141,6054
	S7	1927,2865	987,2757	522,1351	285,3892	201,6865
Moyenne		1558,7155	760,1087	394,9506	220,4372	157,6939
PSNR (Db)	S0	39,3745	37,0676	34,4024	31,658	29,8609
	S1	39,0951	37,0958	34,6174	31,9698	30,2433
	S2	39,4789	37,3904	34,9169	32,3182	30,5639
	S3	39,3546	36,9534	34,245	31,5632	29,8859
	S4	39,226	36,8354	34,2037	31,5477	29,8691
	S5	39,9233	37,6422	34,9835	32,238	30,4424
	S6	38,7667	36,8665	34,5819	32,0828	30,3728
	S7	39,1718	36,6907	33,8848	30,9991	29,2292
Moyenne		39,2988	37,0677	34,4794	31,7971	30,0584

Tableau 5.5. *Les résultats obtenus en termes de débit binaire et PSNR pour la structure Proposed2 en utilisant la vidéo Ballroom (en détails pour chaque vue).*

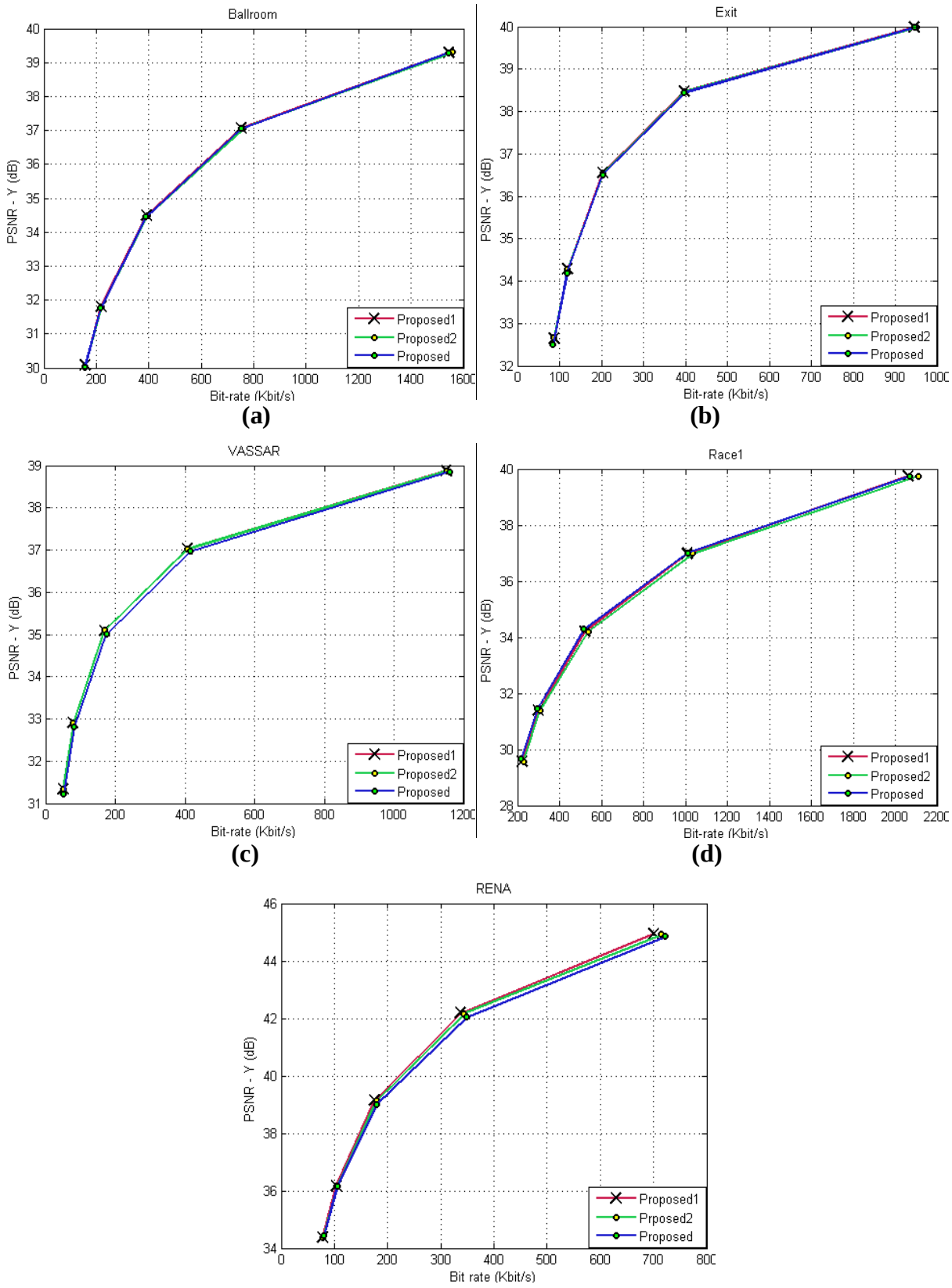
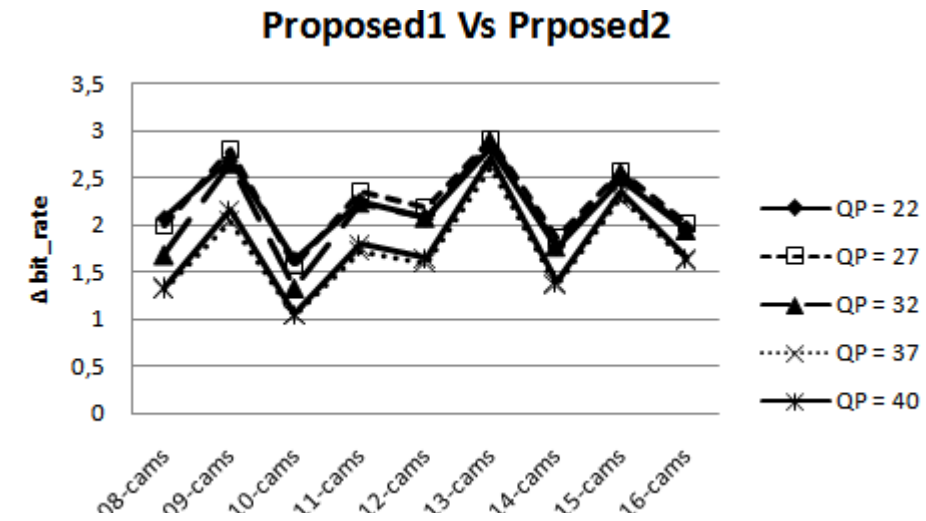


Figure 5.5. *La variation du débit binaire en fonction des 5 valeurs du QP, (a) vidéo Ballroom , (b) vidéo Exit, (c) vidéo Vassar, (d) vidéo Race1, (e) la vidéo Rena.*

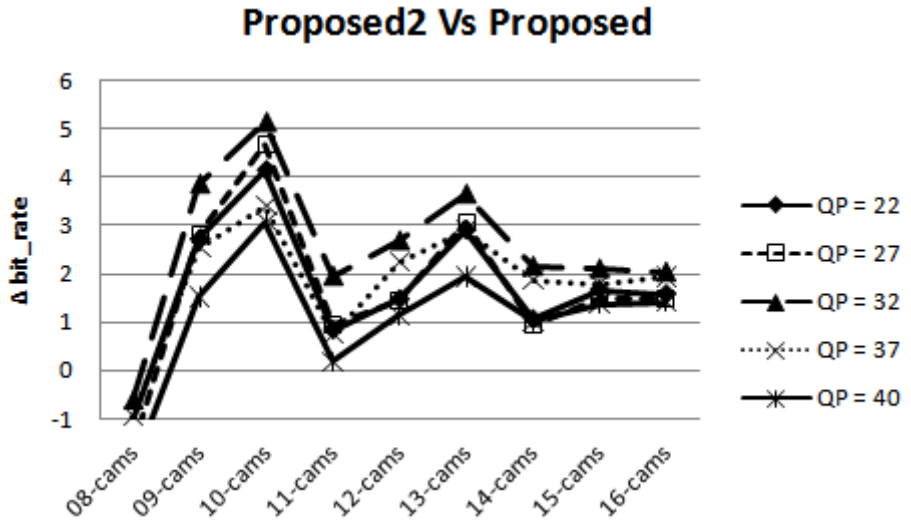
Les résultats des cinq vidéos de test utilisées dans la figure 5.5 sont présentés dans le tableau 5.6. Le gain en débit binaire de la structure *Proposed1* par rapport à la structure *Proposed2* en utilisant 16 vues de la vidéo Rena est illustré dans la figure 5.6 (a). Comme le montre cette figure, le gain maximum est obtenu en utilisant 13 vues. En effet, le cas de 13 vues, seulement des blocs de 3 vues-B successives sont utilisés. La figure 5.6 (b) présente le gain de la structure favorisant deux vues-B successives par rapport à la structure *Proposed*. Le gain maximum est obtenu dans ce cas en utilisant 13 vues. Ceci est dû au nombre élevé de paires de vues-B successives.

		MVV	QP = 22	QP = 27	QP = 32	QP = 37	QP = 40
Proposed1	Débit binaire (Kbit/s)	Ballroom	1542,4695	750,5006	390,0729	218,3797	156,4391
		Exit	945,1587	397,1189	202,6675	118,5385	86,2108
		Vassar	1149,0567	406,4162	168,8087	78,6153	50,0298
		Race1	2061,2967	1008,6104	519,9124	298,5867	222,6632
		Rena	700,0304	336,8856	175,4413	103,2365	77,7958
	PSNR (dB)	Ballroom	39,3015	37,0750	34,4944	31,8153	30,0755
		Exit	39,9959	38,4766	36,554	34,2886	32,6595
		Vassar	38,8780	37,0308	35,0944	32,9074	31,3365
		Race1	39,7578	36,9930	34,2205	31,3944	29,5798
		Rena	44,9511	42,1928	39,1440	36,1825	34,3859
Proposed2	Débit binaire (Kbit/s)	Ballroom	1558,7155	760,1087	394,9506	220,4371	157,6939
		Exit	949,7202	399,0520	203,5452	118,9628	86,3473
		Vassar	1150,7398	407,1824	169,0722	78,6520	50,0864
		Race1	2108,0934	1035,3430	535,8815	305,9595	227,7039
		Rena	714,9045	343,7080	178,4497	104,6321	78,8576
	PSNR (dB)	Ballroom	39,2988	37,0677	34,4794	31,7971	30,0584
		Exit	39,9957	38,4743	36,5512	34,2860	32,6576
		Vassar	38,8804	37,0327	35,0974	32,9062	31,3366
		Race1	39,7461	36,9856	34,2074	31,3767	29,5522
		Rena	44,9230	42,1592	39,1233	36,1698	34,3664

Tableau 5.6. *Les résultats obtenus en termes de débit binaire et PSNR pour les deux structures de prédiction Proposed1 et Proposed2 en utilisant les 5 vidéos de test (résultat global de chaque MVV).*



(a)



(b)

Figure 5.6. *Le gain en débit binaire par rapport à la structure Proposed2, (a) gain par la structure Proposed1, (b) gain par la structure Proposed à deux vues-B successives.*

5.6.2. Évaluation de l'accès aléatoire inter-vues

L'évaluation de l'accès aléatoire inter-vues est effectuée à l'aide des taux, ΔN_{MAX} et T_{img} calculés respectivement par N_{MAX} et Nbr_{img} étudiés ci-dessus. La méthode de calcul du ΔN_{MAX} et T_{img} est présentée dans le chapitre 04.

Nous avons utilisé dans cette expérimentation des GOP de taille 8 afin de simplifier les calculs sachant que la méthode offre des bon résultats quelle que soit la taille du GOP. Le tableau 5.7 illustre les résultats obtenus en termes de N_{MAX} et ΔN_{MAX} pour les quatre

structures IBP, Proposed1, Proposed2 et celle utilisant deux vues-B successives. L'amélioration en accès aléatoire inter-vues estimée par le N_{MAX} apparait plus clairement dans la figure 5.7. Le gain ΔN_{MAX} par rapport à la structure IBP, augmente progressivement avec l'augmentation des cas de trois vues-B successives comme l'exemple de 9, 13, 15 et 16 caméras. Ce taux abouti à 33.33% dans le cas de 13 caméras où seulement des blocs de trois vues-B successives sont utilisés (voir le tableau 5.7). En comparaison avec la structure à deux vues-B successives appelée *Proposed*, le ΔN_{MAX} dépend du nombre de vues-P après la vue-I et aussi le nombre total de vues utilisées. Á titre d'exemple, le cas de 8 et 9 vues le gain en N_{MAX} et maximum. Ceci est justifié par l'utilisation de deux vues-P après la vue-I par la structure *Proposed* avec une seule pour la structure *Proposed2*. Le cas de 10 vues, les deux structures de prédiction nécessitent le même nombre N_{MAX}, ceci donne un ΔN_{MAX} égale à 0.

	N_{MAX}				ΔN_{MAX} (%)		
	IBP	Proposed	Proposed1	Proposed2	IBP Vs Pro2	Pro Vs Pro2	Pro1 Vs Pro2
8 V	18	16	19	14	22.22	12.5	26.31
9 V	20	16	19	14	30	12.5	26.31
10 V	20	16	19	16	20	0	15.77
11 V	22	18	19	16	27.27	11.11	15.77
12 V	22	18	19	16	27.27	11.11	15.77
13 V	24	18	21	16	33.33	11.11	23.81
14 V	24	20	19	18	25	10	5.26
15 V	26	20	21	18	30.77	10	14.28
16 V	26	20	21	18	30.77	10	14.28

Tableau 5.7. *Le gain en nombre maximum N_{MAX} en fonction du nombre de caméras.*

L'utilisation du Nbr_{img} et le T_{img} permet une évaluation plus significative de la structure *Proposed2*. Le Nbr_{img} et ainsi, le T_{img} sont estimés pour toutes les images (clés et non-clés) des différentes vues autres que la vue de base. La figure 5.8 résume les différentes valeurs obtenues pour les deux structures *Proposed1* et *Proposed2*. La figure 5.8 (b) montre clairement le gain significatif par rapport à la première structure, basée sur deux vues-B

successives détaillée dans le chapitre 04 (voir la figure 4.12 (c)). Le tableau 5.8 résume les valeurs obtenues on utilisant T_{img} pour les trois structures *Proposed1*, *Proposed2* et celle présentée dans le chapitre 04.

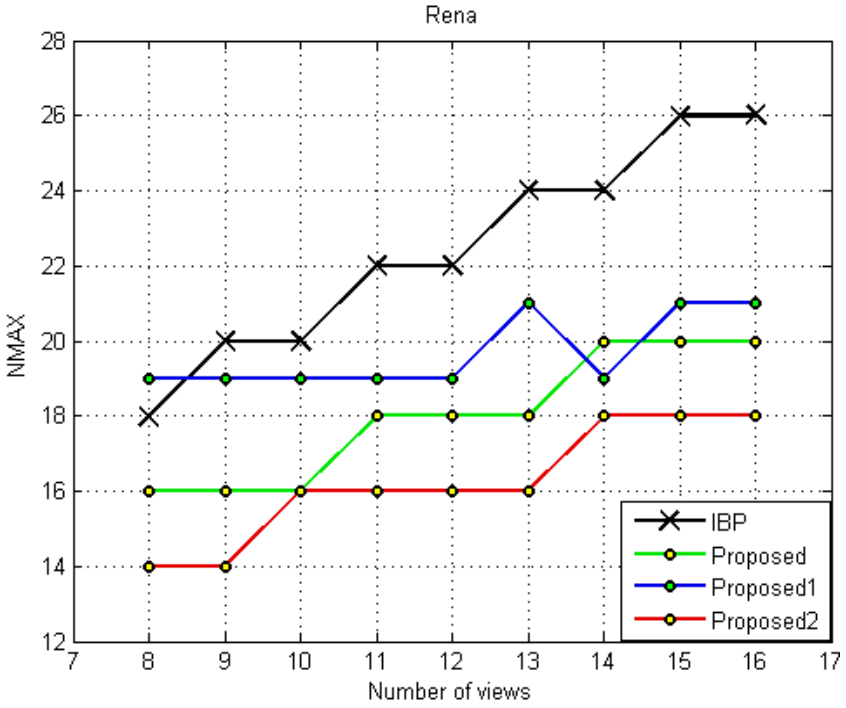


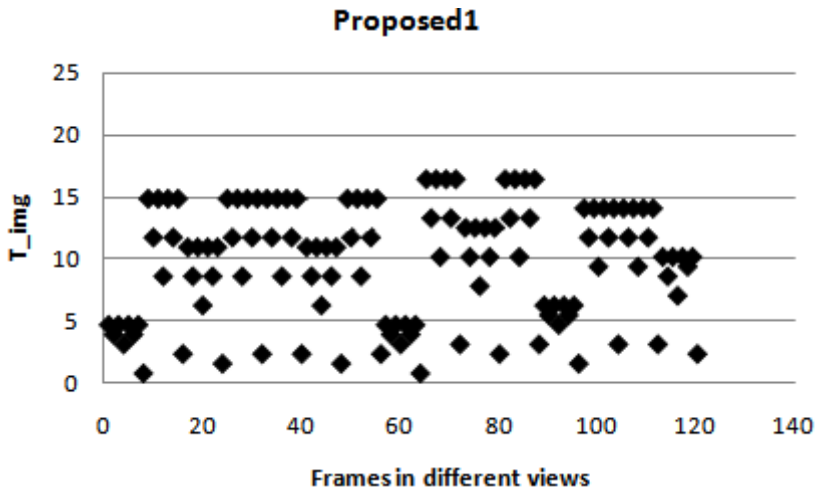
Figure 5.7. Le N_{MAX} *des structures de prédiction: IBP, Proposed, Proposed1 et Proposed2, estimé en fonction du nombre de vues.*

		Proposed	**Proposed1**	**Proposed2**
< 5%	Nbr_{img}	24	30	30
	T_{img}	18.75	23.44	23.44
>=5%&<10%	Nbr_{img}	36	22	32
	T_{img}	28.125	17.19	25.00
>=10%&<15%	Nbr_{img}	52	60	58
	T_{img}	40.625	46.875	45.31
>=15%&<20%	Nbr_{img}	08	08	0
	T_{img}	6.25	6.25	0

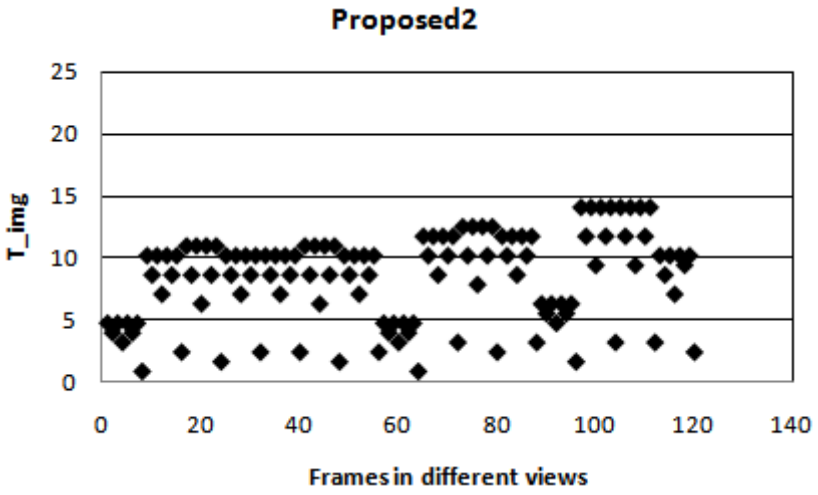
Tableau 5.8. *Le gain en accès aléatoire inter-vues estimé par le* Nbr_{img}.

Les résultats présentés dans le tableau 5.8 sont exploités par l'utilisation de 16 vues. La taille du GGOP est égal ainsi à 8x16. Ces résultats montrent un gain significatif de la structure

Proposed2 par rapport aux autres structures. Á titre d'exemple, le nombre d'images nécessitant moins que 5% de l'ensemble d'images du GGOP pour le décodage est plus élevé dans la structure *Proposed2* que dans la structure *Proposed*. En revanche, le nombre d'images sollicitant plus que 15% pour le décodage est de 08 pour les deux structures *Proposed* et *Proposed2* et de 0 dans la structure *Proposed2*. Cette amélioration est mieux présentée dans la figure 5.9. En effet, la somme totale du Nbr_{img} obtenue pour toutes les images, peut être utile pour l'évaluation de la fiabilité de la structure *Proposed2* par rapport aux autres. Le tableau 5.9 montre que le gain en Nbrimg total est très important. En effet, il dépasse 23% par comparaison avec la structure IBP.



(a)



(b)

Figure 5.8. *L'accès aléatoire inter-vues pour les différentes images. (a) la structure Proposed1, (b) la structure Proposed2.*

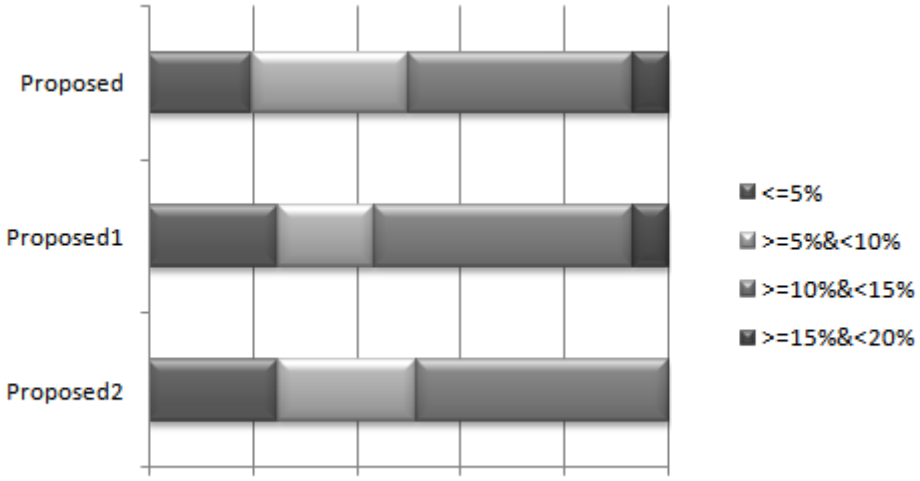


Figure 5.9. *Le gain en accès aléatoire inter-vues pour les trois structures Proposed, Proposed1 et Proposed2.*

	ΣNbr_{img}				ΔNbr_{img} (%)		
	IBP	Proposed	Proposed1	Proposed2	IBP Vs Proposed2	Proposed Vs Proposed2	Proposed1 Vs Proposed2
16 Vues	1680	1379	1484	1280	23.81	7.18	13.75

Tableau 5.9. *Évaluation en fonction du nombre total Nbr_{img} des différentes images.*

5.7. Conclusion

Dans le but d'améliorer en plus l'accès aléatoire inter-vues obtenu par l'approche à deux vues-B successives tout en conservant le débit binaire nécessaire stable, nous avons proposé dans ce chapitre une nouvelle structure de prédiction inter-vues. Nous avons créé une méthode d'évaluation de l'accès aléatoire pour cette structure de prédiction. Cette approche est améliorée par l'utilisation de trois vues-B successives au lieu de deux. Ceci a imposé l'augmentation du niveau hiérarchique avec une légère amélioration en débit binaire. L'inconvénient d'un niveau hiérarchique de plus, été le ralentissement de l'accès aléatoire inter-vues. Nous avons ainsi, solutionné ce problème par l'élimination de la prédiction inter-vues des images non-clés dans certaines vues-B. Nous avons défini également les cas de figures possibles pour les vues après la vue de base S4. Quatre choix sont présentés selon le nombre total des vues utilisées pour la MVV. Les choix préconisés sont mis en œuvre en tenant compte des cas non souhaitables des vues-P successives.

La fiabilité de cette approche en termes d'efficacité de la compression et d'accès aléatoire inter-vues est validée par une analyse expérimentale. Une amélioration de l'accès aléatoire inter-vues par rapport à la première approche basée sur deux vues-B successives est obtenue. Les résultats expérimentaux ont montré également que les deux approches conduisent à une efficacité similaire en termes de débit binaire avec un évident gain de la structure basée sur deux vues-B successives.

Conclusion générale et perspectives

La compression de la vidéo multi-vues, utilisée dans de nombreuses applications telles que la télévision 3D et la vidéo à point de vue libre, doit satisfaire un compromis débit binaire/ qualité, face au développement rapide des technologies d'acquisition et d'affichage 3D. La majorité des normes de compression vidéo existantes doivent également satisfaire l'exigence de l'accès aléatoire temporel le plus rapide possible. Dans le cas de la compression de la vidéo multi-vues, l'accès aléatoire le plus rapide à toute vue se trouvant à une instance *Tn* doit aussi être vérifié.

Dans cette thèse, nous nous sommes concentrés sur l'amélioration de l'accès aléatoire inter-vues à travers le développement d'une nouvelle approche de prédiction basée sur la prédiction et la compensation mixte, temporelle et inter-vues. Nous apportons une contribution à la fois sur les structures de prédiction inter-vues, et l'évaluation de ces structures en termes d'accès aléatoire inter-vues. En résumé, nous avons proposé ce qui suit :

- Une nouvelle structure de prédiction inter-vues. Nous préconisons dans cette structure l'utilisation des paires de vues-B successives. Ceci, afin de permettre un accès aléatoire inter-vues plus rapide lors de l'opération de décodage et d'améliorer l'efficacité de la compression. Cette efficacité est mesurée par un compromis entre le débit binaire à une certaine qualité et la qualité de la vidéo compressée à un certain débit. L'utilisation des paires de vues-B successives est assurée par un choix adéquat de la position de la vue de bas (vue-I),
- La généralisation de l'approche proposée afin de l'appliquer à tout nombre de vues, la généralisation est accomplie par l'utilisation de 16 vues dans le but de conformer la fiabilité de l'approche, tandis que 10 vues suffisent amplement,
- L'amélioration de l'approche proposée en termes d'accès aléatoire inter-vues par l'utilisation de trois vues-B successives au lieu de deux, ce qui exige d'incruster le niveau hiérarchique temporel et inter-vues. Toutefois, cette modification a provoqué une amélioration du débit binaire conservé dans la première approche avec un accès aléatoire plus lourd. Nous avons ainsi utilisé seulement trois images de référence pour chaque image non-clé dans les vues-B2. Ceci permet de régler le problème et d'obtenir un accès aléatoire plus rapide que dans la première approche proposée,

- Une nouvelle méthode d'évaluation de l'accès aléatoire inter-vues. Cette méthode est basée sur la génération et le calcul du nombre d'images Nbr_{img} de référence nécessaires pour le décodage d'une image donnée à une instance Tn. Ce nombre doit être calculé pour toutes les images, clés et non-clés, des différentes vues à l'exception de la vue de base. En effet, cette vue contienne les images spatialement codées. La deuxième métrique générée pour l'évaluation des approches proposées est le nombre maximum N_{MAX} d'images de référence qui doit être utilisé pour le décodage d'une seule image. Le N_{MAX} est également utilisé dans la structure IBP du modèle de référence JMVM. Ces nombres sont générés de la façon suivante :
    - La proposition d'une méthode de calcul du Nbr_{img} pour la structure de prédiction IBP afin de pouvoir comparer les résultats obtenus.
    - La proposition d'une méthode de calcul du Nbr_{img} et aussi N_{MAX} pour la structure de prédiction IPP pour des raisons de comparaison des résultats obtenus.
    - Finalement, la généralisation des deux nombres pour les approches proposées.

Les résultats expérimentaux ont montré un gain significatif en accès aléatoire inter-vues mesuré par le N_{MAX} qui atteint 25% par rapport à la structure de prédiction IBP. Ce gain dépasse 33% par l'approche améliorée. Également une réduction très importante du nombre Nbr_{img} d'images de référence nécessaire pour le décodage d'une image donnée par rapport aux deux structures IBP et IPP est obtenue. Les résultats expérimentaux ont également montré l'important gain en termes de débit par la structure proposée qui est d'environ 6,81% avec une qualité similaire mesurée par PSNR vis-à-vis de la structure de prédiction IBP.

Perspectives

Comme il est précisé au début de cette thèse, il existe plusieurs exigences pour la compression de la vidéo multi-vues autres que l'efficacité de la compression et l'accès aléatoire inter-vues comme par exemple la réduction de la complexité de calcul et le faible retard. Toutes les approches proposées se focalisent sur l'augmentation du nombre de vues-B. Toutefois, ces vues nécessitent une complexité de calcul accrue par rapport aux autres vues I et P. en effet, les vues-B utilisent jusqu'à quatre images de référence pour l'encodage de ses images non-clés.

Alors, nous pouvons envisager d'améliorer les méthodes proposées tout en diminuant la complexité de calcul et en assurant un faible retard. Ceci, peut être obtenu non seulement par la modification de la structure de prédiction, mais aussi par l'amélioration de la prédiction et la compensation de disparité ainsi que l'amélioration de la méthode de sélection du meilleur mode macro-block.

Bibliographie

[1]. P.Merkle, A.Smolic, K.Müller, and T.Wiegand, "Multi-view video plus depth representation and coding," in Proceedings of IEEE International Conference on Image Processing (ICIP'07), San Antonio, TX, USA, pp. 201-204, September 2007.

[2]. P. Kauff, N. Atzpadin, C. Fehn, K. Müller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," *Signal Processing: Image Communication*, vol. 22, no. 2, pp. 217-234, February 2007.

[3]. http://research.microsoft.com/en-us/projects/imv/.

[4]. L. Onural, A. Smolic, and T. Sikora, "An overview of a new European consortium: Integrated three-dimensional television—Capture, transmission and display (3DTV)," presented at the European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies (EWIMT04), London, U.K., November. 2004.

[5]. A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3D video and free viewpoint video—Technologies, applications and MPEG standards," presented at the IEEE International Conference on Multimedia and Expo (ICME 2006), Toronto, Ont, Canada, pp. 2161-2164, July 2006.

[6]. I. Sexton and P. Surman, "Stereoscopic and Autostereoscopic Display Systems," *IEEE Signal processing magazine*, vol. 16, no. 3, pp. 85–99, May 1999.

[7]. M. Halle, "Autostereoscopic displays and computer graphics," *Computer graphics, (ACM SIGGRAPH)*, vol. 31, no. 2, pp. 58-62, May 1997.

[8]. N.A. Dodgson, " Autostereoscopic 3D Displays ", *IEEE Computer Society*, vol. 38, no. 8, pp. 31-36, August 2007.

[9]. "Requirements on Multi-View Video Coding v.4," ISO/IEC JTC1/SC29/WG11, Poznan, Poland, Doc. N7282, July 2005.

[10]. G. Sullivan and T. Wiegand, "Video compression—From concepts to the H.264/AVC standard," *Proc. IEEE, Special Issue on Advances in Video Coding and Delivery*, Vol. 93, No. 1, pp. 18, January 2005.

[11]. Y.-L. Lee, J.-H. Hur, D.-Y. Kim, and Y.-K. Lee, "H.264/MPEG-4 AVC-based multiview video coding," presented at the MPEG2006/ M12871, 75th MPEG Meeting, Bangkok, Thailand, Doc. m12871, January 2006.

[12]. Y. Chen, P. Pandit, and S. Yea, "WD 4 reference software for MVC," ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-AD207, January 2009.

[13]. “H.261 : Video codec for audiovisual services at p x 64 kbit/s,”, Recommendation H.261 in ITU-T, March 1993.

[14]. “H.263 : Video coding for low bit rate communication,”, first Recommendation H.262 in ITU-T, March 1996.

[15]. “H.263+ : Video coding for low bit rate communication,”, second recommendation H.263 in ITU-T, February 1998.

[16]. “H.263++ : H.263 Annex U, V, W and X, ”, Complements Recommendation H.263 in ITU-T, January 2000.

[17]. Standard MPEG-1: ISO/IEC 11172-2, “Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s”,1996.

[18]. Standard MPEG-2: ISO/IEC 13818-2, “Information Technology – Generic coding of moving pictures and associated audio information,”, 2000.

[19]. Standard MPEG-4: ISO/IEC 14496-2, “Information Technology – Coding of Audio-Visual Objects-Part2,”, 2001.

[20]. D. Marpe, G. Blattermann, and T. Wiegand. "Adaptive Codes for H.26L," ITU-T/SG16 VCEG-L13, Eibsee, Germany, January 2001.

[21]. Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, “Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC),” 7th Meeting, DOC. JVTG050, Pattaya, Thailand, March 2003.

[22]. ITU-T rec. H.264 / ISO/IEC 11496-10, “Advanced Video Coding,” Final committee draft, document jvtf100, tech. rep, 2002.

[23]. T. Wiegand, G. J. Sullivan, G. Bjøntegaard and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–560, July. 2003.

[24]. A. Puria, X. Chenb and A. Luthrac, “Video coding using the .264/MPEG-4 AVC compression standard,” *Elsevier Science, Signal Processing: Image Communication,* vol. 19, no. 9, pp. 793–849, October 2004.

[25]. Iain E G Richardson, "H.264/MPEG-4 part 10: Transform quantization", H.264/MPEG-4 part 10 white paper,” tech. rep, 2002.

[26]. A. Luthra, P. Topiwala, “Overview of H.264/AVC video coding standard, ” in: Proceedings of the SPIE—Applications of Digital Image Processing XXVI, San Diego, Vol. 5203, August 2003.

[27]. J. Ribas-Corbera and S. Lei. ”Rate control in DCT video coding for low-delay communications”. *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 172–185, February 1999.

[28]. H. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, “Low-complexity transform and quantization in H.264/AVC,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 598 – 603, July 2003.

[29]. H. Nguyen and P. Duhamel. “Iterative joint source-channel decoding of variable length encoded video sequences exploiting source semantics,“ presented at the IEEE International Conference on Image Processing (ICIP '04), October 2004.

[30]. H. Nguyen, P. Duhamel, J. Brouet, and D. Rouffet, “Robust VLC sequence decoding exploiting additional video stream properties with reduced complexity,” presented at the IEEE International Conference on Multimedia and Expo (ICME 2004), Taipei, Taiwan, pp. 375–378, June 2004.

[31]. C. Marin, P. Duhamel, K. Bouchireb, and M. Kieffer. “Robust video decoding through simultaneous usage of residual source information and MAC layer CRC redundancy,” In IEEE Global telecommunication conference (Globecom’07), Washington, DC, pp. 2070–2074, November 2007.

[32]. P. T. Gary J. Sullivan and A. Luthra, “The H.264/AVC advanced video coding standard : Overview and introduction to the fidelity range extensions,” Presented at the SPIE Conference on Applications of Digital Image Processing XXVII, Special Session on Advances in the New Emerging Standard : H.*264*/AVC, August 2004.

[33]. S. Wenger, “H.264/AVC over IP,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 645 – 656, July 2003.

[34]. R. Schäfer, T.Wiegand, and H. Schwarz, ”H.264/AVC la norme qui monte”. *UER-Revue Technique*, Sélection, pp 1–10, 2003.

[35]. J. Lou, H. Cai, and J. Li, “A real time interactive multiview video system,” presented at the 13th ACM International Conference on Multimedia, Singapore, pp. 6–11, November 2005.

[36]. P. Merkle, A. Smolic, K. Müller, and T. Wiegand, “Efficient Prediction Structures for Multiview Video Coding,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1461–1473, November 2007.

[37]. A. Vetro, T. Wiegand, and G-J. Sullivan, “Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard,” *Proceedings of the IEEE*, vol. 99, no. 4, pp. 626-642, April 2011.

[38]. H. Schwarz, D. Marpe, and T. Wiegand, "Hierarchical B pictures," Joint Video Team, doc. JVT-P014, Poznan, Poland, July 2005.

[39]. H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MCTF," in Proceedings of IEEE International Conference on Multimedia and Expo *(*ICME 2006), Toronto, Ont, Canada, pp. 1929-1932, July 2006.

[40]. H. Zeng, K-K. Ma and C. Cai, "Fast Mode Decision for Multiview Video Coding Using Mode Correlation," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 21, no. 11, pp. 1659–1666, November 2011.

[41]. W. Zhu, X. Tian, F. Zhou and Y. Chen, "Fast Inter Mode Decision Based on Textural Segmentation and Correlations for Multiview Video Coding," *IEEE Trans. Consumer Electronics*, vol. 56, no. 3, pp. 1696–1704, August 2010.

[42]. E. Ekmekcioglu, S.T. Worrall, A. M. Kondoz, "Low-delay random view access in multi-view coding using a bit-rate adaptive downsampling approach," in Proceedings of IEEE International Conference on Multimedia and Expo (ICME 2008), Hannover, pp. 745–748, June 2008.

[43]. Y. Zhang, G. Jiang, M. Yu, and Y. S. Ho, "Adaptive multiview video coding scheme based on spatio-temporal correlation analyses," *ETRI Journal*, vol. 31, no. 2, pp. 151–161, April 2009.

[44]. ITU-T Rec. & ISO/IEC, "Advanced Video Coding for Generic Audiovisual Services," 14496-10 AVC, 2005.

[45]. ISO/IEC JTC1/SC29/WG11, "Description of Core Experiments in MVC," N8019, April 2006.

[46]. ITU-T Rec. H.264 &ISO/IEC JTC 1, "Advanced Video Coding for Generic Audiovisual Services," 14496-10, January 2011.

[47]. "Joint Multiview Video Model (JMVM) 1.0," JVT-T208, Klagenfurt, Austria, July 2006.

[48]. Yo-Sung Ho, and Kwan Jung Oh, "Overview of Multi-view Video Coding," in the Proceedings of Systems, Signals and Image Processing, 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services, pp. 5-12, June 2007.

[49]. Y. Chen, Y.-K. Wang, and M. M. Hannuksela, "Comments on MVC JD 2.0," JVT-W035, San Jose, Calif, USA, April 2007.

[50]. G. J. Sullivan, A. M. Tourapis, T. Yamakage, and C. S. Lim, "Draft AVC amendment text to specify constrained baseline profile, stereo high profile, and frame packing SEI message", Joint Video Team (JVT), London, U.K., Doc. JVT-AE204, July. 2009.

[51]. TK Tan, G. Sullivan and T. Wedi , "Recommended Simulation Common Conditions for Coding Efficiency Experiments Revision 1". DOC No: VCEG-AE010, 2007.

[52]. T. Wiegand, X. Zhang, and B. Girod, "Long-term memory motion compensated prediction," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 70-84, February 1999.

[53]. H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, Special Issue on Scalable Video Coding, no. 9, pp. 1103–1120, September 2007.

[54]. Y. Chen, Y.-K. Wang, K. Ugur, M. Hannuksela, J. Lainema, and M. Gabbouj, "The Emerging MVC Standard for 3D Video Services," *EURASIP Journal on Advances in Signal Processing*, Vol. 2009, No. 1, January 2009.

[55]. Y.-K. Wang, Y. Chen, and M. M. Hannuksela, "Time-first coding for multi-view video coding," JVT-U104, Hangzhou, China, October 2006.

[56]. "Methodology for the subjective assessment of the quality of television pictures," ITU-R BT.500-11, 2002.

[57]. A.Bekhouch, N.Doghmane. "Compression de la vidéo multi-vues basée sur le modèle de référence JMVM 8.0". In ICESTI'12, Annaba, Algeria, Novembre 2012.

[58]. A.Bekhouch, N.Doghmane. "Multiview video coding with an improved prediction structure for faster random access," *Journal of Electronic Imaging*, vol. 22, no. 4, 043010 (Oct–Dec 2013).

[59]. A.Bekhouch, N.Doghmane. "Une nouvelle structure de prédiction inter-vues pour la compression de la vidéo multi-vues". The Third International Conference on Image and Signal Processing and their Applications, accepted, Mostaganem, Algeria, from 2 – 4 December 2012.