

Ministère de l'enseignement Supérieur et de la recherche Scientifique

وزارة التعليم العالي والبحث العلمي

Badji Mokhtar Annaba University  
Université Badji Mokhtar – Annaba  
Faculté de Technologie



جامعة باجي مختار – عنابة

كلية التكنولوجيا

Département Electronique

قسم الإلكترونيك

## Thèse

Présentée pour obtenir le diplôme de

### Doctorat Troisième Cycle

Filière : Télécommunication

Spécialité : Multimédia et communication numérique

Par :

**Khalfa Sara**

Thème :

### Protection des données au niveau source d'un streaming vidéo dans les réseaux sans fil

Thèse soutenue le 02-06-2024 devant le jury composé de :

N°	Nom et prénom	Grade	Etablissement	Qualité
01	Benouaret Mohamed	Prof.	Université Badji Mokhtar -Annaba	Président
02	Harize Saliha	Prof.	Université Badji Mokhtar -Annaba	Rapporteur
03	Kouadria Nasreddine	MCA	Université Badji Mokhtar -Annaba	Co-rapporteur
04	Boumahrez Farouk	MCA	Université Abbes Laghrour -Khenchela	Examineur
05	Sahour Abdelhakim	MCA	Université Abbes Laghrour -Khenchela	Examineur

## حماية البيانات على مستوى المصدر في بث الفيديو في شبكات الاتصال اللاسلكية

### الملخص:

هذه أطروحة الدكتوراه تركز على تحسين جودة وحماية بيانات بث الفيديو في بيئات الشبكات اللاسلكية. تبدأ بتحليل عميق لتأثير فقد حزم البيانات على جودة الفيديو المستلم، مسلطة الضوء على التحديات الكامنة في النقل اللاسلكي. لحل هذه التحديات، تقدم عدة مساهمات مبتكرة المساهمة الأولى تعتمد على طريقة لحماية الفيديو المرسل عبر قناة ضوضاء. مبنية على ترميز الوصف المتعدد، تعمل هذه الطريقة مباشرة على مستوى المصدر لتعزيز مرونة الفيديو أمام اضطرابات القناة. وبالتالي، توفر حلاً واعدًا لضمان نقل فيديو عالي الجودة في بيئات لاسلكية غير متوقع المساهمة الثانية تعتمد على استخدام تقنيات التعلم العميق لتحسين جودة الفيديو المستلم. يتم دمج مُرمز تلقائي تحويلي ضمن عملية ما بعد المعالجة لتحسين جودة الصور المفككة. تُظهر النتائج المحصلة أن هذا النهج للتعلم العميق يسمح بتحسين كبير في الجودة المدركة للفيديو، حتى في حالة تشويش شروط النقل. يسلط هذا العمل الضوء على أهمية كبيرة لجودة وحماية بيانات بث الفيديو في الشبكات اللاسلكية. المساهمات المقترحة، من ترميز الوصف المتعدد إلى استخدام المُرمز تلقائي التحويلي، تفتح الباب أمام حلاً مبتكراً لضمان تجربة مشاهدة فيديو مثلى على الرغم من القيود المرتبطة بالشبكات اللاسلكية. تُسلط التوجهات المستقبلية المحددة الضوء على نطاق الفرص البحثية في هذا المجال المتطور باستمرار، مع تأثيرات على مجموعة متنوعة من التطبيقات تتراوح من التلفزيون المحمول إلى الواقع الافتراضي.

**كلمات مفتاحية:** ضغط الفيديو HEVC، ترميز الوصف المتعدد (MDC)، جودة التجربة (QoE)، معلمة التكميم (QP)، نسبة فقدان الحزم (PLR)، شبكات أدهوك اللاسلكية.

## « Protection des données au niveau source d'un streaming vidéo dans les réseaux sans fil »

### Résumé :

Cette thèse de doctorat se concentre sur l'amélioration de la qualité et la protection des données de streaming vidéo dans les environnements de réseaux sans fil. Elle débute par une analyse approfondie de l'impact des pertes de paquets sur la qualité vidéo reçue, mettant en avant les défis inhérents à la transmission sans fil. Pour résoudre ces défis, elle propose plusieurs contributions novatrices. La première contribution repose sur une méthode de protection de la vidéo transmise dans un canal bruité. Fondée sur le codage à description multiple, cette méthode agit directement au niveau de la source pour renforcer la résilience de la vidéo face aux perturbations du canal. Elle offre ainsi une solution prometteuse pour garantir une transmission vidéo de haute qualité dans des environnements sans fil imprévisibles. La deuxième contribution se fonde sur l'utilisation de techniques d'apprentissage profond pour améliorer la qualité vidéo reçue. Un auto-encodeur convolutionnel est intégré dans le processus de post-traitement pour améliorer la qualité des images décodées. Les résultats obtenus mettent en évidence que cette approche d'apprentissage profond permet une amélioration significative de la qualité perçue des vidéos, même en cas de conditions de transmission altérées. Ce travail met en lumière l'importance primordiale de la qualité et de la protection des données de streaming vidéo dans les réseaux sans fil. Les contributions proposées, du codage à description multiple à l'utilisation des auto-encodeurs convolutionnels, ouvrent la voie à des solutions innovantes pour assurer une expérience de visualisation vidéo optimale malgré les contraintes inhérentes aux réseaux sans fil. Les perspectives futures identifiées soulignent l'étendue des opportunités de recherche dans ce domaine en constante évolution, avec des implications pour une gamme d'applications allant de la télévision mobile à la réalité virtuelle.

**Mots clés :** HEVC, codage par description multiple (MDC) , Qualité d'expérience (QoE), Réseau de neurones convolutionnelle (CNN), Taux de perte des paquets (PLR), réseaux Ad-Hoc sans fil.

## **«Data protection at the source level of video streaming in wireless networks»**

### **Abstract:**

This doctoral thesis focuses on improving the quality and data protection of video streaming in wireless network environments. It begins with a thorough analysis of the impact of packet loss on received video quality, highlighting the challenges inherent in wireless transmission. To address these challenges, it proposes several innovative contributions. The first contribution is based on a method for protecting transmitted video in a noisy channel. Based on multiple description coding, this method acts directly at the source level to enhance video resilience to channel disturbances, offering a promising solution to ensure high-quality video transmission in unpredictable wireless environments. The second contribution relies on the use of deep learning techniques to enhance received video quality. A convolutional autoencoder is integrated into the post-processing to improve the quality of decoded images. The results demonstrate that this deep learning approach significantly enhances the perceived quality of videos, even under adverse transmission conditions. This work highlights the paramount importance of video streaming data quality and protection in wireless networks. The proposed contributions, from multiple description coding to the use of convolutional autoencoders, pave the way for innovative solutions to ensure an optimal video viewing experience despite the constraints of wireless networks. The identified future prospects underscore the range of research opportunities in this ever-evolving field, with implications for a variety of applications, from mobile television to virtual reality.

**Key words:** HEVC/H.265, Multiple Description Coding (MDC), Quality of Experience (QoE), convolutional neural network (CNN), Packet Loss ratio (PLR), Wireless Ad-Hoc Networks.

## Remerciement

D'abord et avant tout, je remercie mon DIEU « Allah » pour m'avoir accordé la santé, la patience, la connaissance et tout ce qui est nécessaire pour terminer cette thèse.

Je tiens à exprimer ma profonde gratitude envers toutes les personnes qui ont contribué à la réalisation de cette thèse. Le parcours de recherche a été long et exigeant, mais il n'aurait pas été possible sans le soutien inestimable de nombreux individus.

Tout d'abord, je souhaite remercier sincèrement mes parents, Ali Settah Rachida et Khalfa Mohamed. Leur amour inconditionnel, leur encouragement constant et leur soutien moral ont été les piliers sur lesquels j'ai pu construire ma réussite académique. Leurs sacrifices et leur confiance en moi ont été une source d'inspiration quotidienne.

Je souhaite remercier sincèrement ma directrice de thèse, le Professeur Harize Saliha, pour sa patience, sa sagesse et son dévouement tout au long de cette aventure académique. Ses conseils éclairés, ses discussions stimulantes et son soutien infaillible ont été essentiels pour orienter mes recherches et atteindre ce jalon.

Je souhaite également adresser mes remerciements sincères à mon co-directeur de thèse, Dr. Kouadria Nassreddine, pour son mentorat précieux et son expertise inestimable tout au long de cette recherche. Ses conseils éclairés, sa patience et son dévouement ont été d'une importance capitale pour la réalisation de cette thèse.

Je tiens à exprimer ma reconnaissance envers les membres du jury, le Professeur Benouaret Mohamed, le Professeur Boumahrez Farouk et le Professeur Sahour Abdelhakim, pour leurs précieux commentaires et leurs suggestions constructives qui ont contribué à améliorer ce travail de recherche.

Je souhaite également remercier chaleureusement mes collègues du laboratoire LASA de l'université Badji Mokhtar Annaba qui ont partagé cette aventure avec moi. Leurs idées, leurs discussions et leur camaraderie ont enrichi mon expérience de recherche, en particulier mon époux Docteur Kebir Abdeldjalil pour son précieux soutien tout au long de mon parcours de recherche en partageant ses connaissances techniques de manière généreuse, et en m'offrant des conseils avisés qui ont grandement contribué à l'aboutissement de cette thèse.

Ce travail représente l'aboutissement de nombreuses heures de travail acharné, de réflexion et de dévouement. Je suis honoré d'avoir eu l'opportunité de mener cette recherche, et je suis reconnaissant envers tous ceux qui ont rendu cette réalisation possible.

## Dédicaces

Cette these est dédié A mes chers parents

Mes filles Razane et Alicia Mayar ma source d'Energie

Mes sœurs Samira et Rania

## Liste des figures

Figure 1. Pourcentage du trafic mondial des données mobiles vidéo et non-vidéo .....	2
Figure 2. Historique des standards de codage vidéo [19].....	12
Figure 3. Les étapes de la compression video .....	14
Figure 4. Schéma synoptique d'un encodeur HEVC doté d'un décodeur intégré représenté en gris [22] .....	17
Figure 5. Illustration du partitionnement d'une image de taille 1280 x 720 (a) Partition en macroblocs de 16 x 16, telle qu'on la retrouve dans toutes les normes précédentes de codage vidéo de l'UIT-T et de l'ISO/CEI ; (b) Partition de l'image en CTU de 64 x 64, prise en charge dans le profil principal de HEVC.....	18
Figure 6. Illustration du regroupement des CTBs en CTU .....	18
Figure 7. La division d'un CTB de 64 x 64 en CB .....	19
Figure 8. Modes de prédiction intra angulaire dans le HEVC .....	20
Figure 9. Concept général de PCM basée sur un modèle de mouvement translationnel ..	21
Figure 10. Exemples d'artefacts de compression dans le codage vidéo. a Effet de halo. b Flou. c Blocage [30] .....	24
Figure 11. Exemples d'un réseau streaming serveur client [32] .....	27
Figure 12. Exemple d'architecture de streaming P2P en couches [34]. .....	28
Figure 13. Schéma de principe de l'encodeur/décodeur .....	29
Figure 14. QoS vs. QoE dans la diffusion de contenu multimédia [35]. .....	31
Figure 15. Illustration du phénomène de gigue. Les flux du haut et du bas ont respectivement une distance inter-paquet constante et variable.....	32
Figure 16. Illustration du phénomène de la perte de paquets.....	32
Figure 17. Des exemples d'échelles utilisées lors des tests subjectifs d'évaluation de qualité. De gauche à droite : une échelle comparative, une échelle catégorielle qualitative et son échelle discrète correspondante .....	36
Figure 18. Entête d'une unité NAL HEVC .....	39
Figure 19. Scénario de transmission vidéo HEVC.....	41
Figure 20. Le partitionnement de données de H.264/AVC [54] .....	44
Figure 21. Découpage en slice d'un frame de la séquence Parkjoy, (a) 2 slices, (b) 4 slices, (c) 6 slices.....	45
Figure 22. MDC avec échantillonnage spatial polyphasé [13]. .....	47
Figure 23. MDC temporel avec 2 descriptions.....	48

Figure 24. Système de communication basé sur l'apprentissage automatique. ....	49
Figure 25. Répartition des vidéos tests choisies dans le plan (SI, TI).....	57
Figure 26. Système d'essai proposé .....	58
Figure 27. Les courbes RD des 2 modes de codage.....	59
Figure 28. Impact du PLR sur la qualité vidéo en mode all intra.....	61
Figure 29. Impact du PLR sur la qualité vidéo en mode Low Delay P.....	63
Figure 30. Schéma d'adaptation pour l'évaluation objective d'une séquence vidéo : (a) SDC, (b) MDC .....	65
Figure 31. Modèle MDC spatial proposé .....	65
Figure 32. Plan SI TI des séquences video.....	66
Figure 33. Stratégies de suppression intégrale de trame proposées .....	69
Figure 34. Sélection des images de référence pour le POC16 en configuration LDP .....	69
Figure 35. Impact des différentes stratégies sur la qualité vidéo avec les schémas SDC et MDC en termes de PSNR (QP=22).....	71
Figure 36. Impact des différentes stratégies sur la qualité vidéo avec les schémas SDC et MDC en termes de PSNR (QP=37).....	72
Figure 37. Impact des différentes stratégies sur la qualité vidéo avec les schémas SDC et MDC en termes de MOS_Estimated (QP=22) .....	73
Figure 38. Impact des différentes stratégies sur la qualité vidéo avec les schémas SDC et MDC en termes de MOS_Estimated (QP=37) .....	73
Figure 39. Approche de post-traitement intégrée basée sur l'apprentissage profond dans le flux de travail de codage typique. La couleur jaune indique la focalisation de ce travail [101] .....	76
Figure 40. Architecture de l'auto-encodeur .....	77
Figure 41. Architecture de l'auto-encodeur utilisé.....	78
Figure 42. Résultat de l'apprentissage avec (modèle 1 : 50, modèle 2 :150) itérations en utilisant 700 images (few data) .....	80
Figure 43. Résultat de l'apprentissage avec (modèle 3 : 50, modèle 4 :150) itérations en utilisant 1200 images (Big data).....	81

## Liste des tableaux

Tableau 1. Récapitulatif des fonctionnalités utilisées dans HEVC et VVC [29].....	22
Tableau 2. Séquences de test issues des conditions de test communes JCT-VC.....	55
Tableau 3. Caractéristiques des séquences video utilisées.....	66
Tableau 4. Résultat du SDC .....	67
Tableau 5. Résultat du MDC (description paire) .....	67
Tableau 6. Résultat du MDC (description impaire) .....	67
Tableau 7. Échelle de qualité et de dégradation.....	70
Tableau 8. Pourcentage moyen de gain pour les 4 séquences.....	75
Tableau 9. Informations détaillées sur le modèle utilisé .....	78
Tableau 10. Résultat des 2 approches ‘few data’ et ‘big data’ avec 50 et 150 itérations.....	83
Tableau 11. Comparaison des résultats avec les données initiales c’est les vidéos avec perte de paquets PLR= 3% .....	83

## Liste des abréviations

AMVP	Prédiction avancée des vecteurs de mouvement
ARQ	Automatic Repeat request
AVI	Audio Video Interleave
CAGR	Compound Annual Growth Rate
CEI	Commission électrotechnique internationale
CIF	Format intermédiaire commun
CNN	Réseaux neuronaux convolutifs profonds
CSCC	Codage source-canal conjoint
CTU	Coding Tree Unit
DPB	Tampon d'image décodée
DCT	Transformée en cosinus discrète
DWT	Transformation en ondelettes discrètes
FEC	Forward Error Correction
GOP	Group d'images
HD	Haute définition
HDR	High Dynamic Range
HEVC	High efficiency video coding
HT	Transformation de Hadamard
HTTP	Hyper Text Transfer Protocol
HVS	Système visuel humain
LDP	Low delay P
IMR	Intra macroblock refresh
IoT	Internet of things
ISO	Organisation internationale de normalisation
ITU-T	Union internationale des télécommunications - Secteur de la normalisation des télécommunications
IWT	Transformation en ondelettes entières
MANEs	Media-aware network elements
MDC	Codage à descriptions multiple
MIMO	Multiple-Input, Multiple-Output
MOS	Mean Opinion Score
MOV	QuickTime Movie
MP4	MPEG-4
MPEG	Moving Picture Experts Group
MV	Vecteur de mouvement
NAL	Network Abstraction Layer
NTSC	National Television System Committee
P2P	Streaming vidéo pair à pair
PLR	Paquet loss rate
POC	Numéro de compte d'image
QoE	Qualité d'expérience
QP	Paramètres de quantification
RPS	Reference picture selection

RSVP	Protocole de réservation de ressources
RTP	Real-time Transport Protocol
RTSP	Real-Time Streaming Protocol
SAO	Sample Adaptive Offset filter
SD	Standard Definition
SDR	Plage dynamique standard
TCAC	Taux de Croissance Annuel Composé
TCP	Transport Control Protocol
UDP	User Datagram Protocol
UHD	Ultra Haute Définition
VLC	Code à longueur variable
VoD	Vidéo sur demande
VP 9	Codec Vidéo 9
VVC	Versatile Video Coding
WMV	Windows Media Video

## Table des matières

Introduction générale.....	1
<b>1. Les principes de base de la compression et la transmission vidéo .....</b>	<b>6</b>
1.1 Introduction .....	6
1.2 Définition de la vidéo .....	6
1.2.1 Histoire de la vidéo numérique .....	7
1.2.2 Scènes vidéo .....	7
1.2.3 Acquisition .....	8
1.2.4 Format vidéo et codecs.....	9
1.2.5 Différence entre vidéo brute et vidéo compressée .....	12
1.3 Contenu vidéo.....	13
1.4 Les étapes de la compression vidéo.....	13
1.4.1 Prédiction .....	14
1.4.2 Transformation .....	15
1.4.3 Quantification.....	15
1.4.4 Codage entropique.....	15
1.5 Décodage vidéo .....	15
1.5.1 Principe du codage HEVC .....	16
1.6 Versatile Video Coding.....	21
1.6.1 Phase de normalisation.....	22
1.7 Les artefacts liés à la compression vidéo .....	24
1.8 Transmission vidéo.....	25
1.8.1 Streaming vidéo traditionnel .....	26
1.8.2 Le streaming vidéo pair à pair (P2P).....	27
1.9 Conclusion.....	28
<b>2. Problématique des transmissions vidéo sur les réseaux avec perte .....</b>	<b>29</b>
2.1 Introduction .....	29
2.2 Qualité de service (QoS) .....	31
2.2.1 Effets perceptuels de la perte de paquets .....	33
2.2.2 Techniques pour améliorer la QoS.....	33

2.3 Qualité d'expérience (QoE).....	34
2.3.1 Les améliorations de la QoE dans le streaming vidéo.....	34
2.3.2 Moyens de mesurer la QoE .....	35
2.4 Relation entre la QoS et la QoE .....	38
2.5 L'influence des erreurs de transmission sur un flux H.265/HEVC.....	39
2.5.1 Ajustement du flux HEVC au réseau de paquets .....	39
2.5.2 Vulnérabilité du flux HEVC aux erreurs de transmission.....	41
2.6 Conclusion.....	42
<b>3. État de l'art sur les mécanismes d'amélioration de la qualité du streaming vidéo</b>	<b>43</b>
3.1 Introduction .....	43
3.2 Codage résilient aux erreurs .....	43
3.2.1 Codage source .....	43
3.2.2 Codage canal .....	48
3.2.3 Codage conjoint source et canal.....	49
3.2.4 L'apprentissage machine.....	49
3.3 Dissimulation d'erreurs du décodeur .....	50
3.3.1 Dissimulation d'erreur intra-frame .....	50
3.3.2 Dissimulation d'erreur inter-frame .....	50
3.3.3 Dissimulation d'erreur hybride .....	51
3.4 Les outils de résilience aux erreurs dans le HEVC .....	51
3.5 Conclusion.....	52
<b>4. Analyse des performances et résultat</b> .....	<b>54</b>
4.1 Introduction .....	54
4.2 Impacte de la perte des paquets sur la qualité vidéo reçu .....	55
4.2.1 Bases de données vidéo.....	55
4.2.2 Configuration du codeur et du décodeur du HEVC .....	57
4.3 Codage a description multiple spatial pour l'amélioration de la qualité video reçue ..	63
4.3.1 Description du système .....	64
4.3.2 Expérience.....	65
4.3.3 Résultat et discussion .....	70
4.4 Amélioration de la qualité video codé en HEVC basé sur l'apprentissage profond ....	75
4.4.1 Méthode proposée .....	76
4.4.2 Résultat et discussion de l'apprentissage .....	80

4.4.3 Résultat et discussion des tests.....	83
4.5 Conclusion.....	84
Conclusion et perspectives.....	86
Références.....	87

# Introduction générale

L'utilisation intensive des plateformes en ligne telles que Netflix, Disney+ ou bien myCANAL et les plateformes du live streaming tel que (YouTube live, Facebook live, Zoom) a considérablement augmenté ces dernières années en raison de la popularité croissante de la vidéo en streaming et de la disponibilité accrue d'internet haut débit. Les gens peuvent maintenant accéder à un large éventail de contenu vidéo à la demande à partir de n'importe où et à n'importe quel moment, ce qui les rend très pratiques pour les utilisateurs. De plus, avec des abonnements abordables et une large variété de contenu, telles que des films, des émissions de télévision et des documentaires, Netflix est devenu un choix populaire pour les divertissements à domicile.

Le streaming vidéo à faible latence est devenu une application clé pour la diffusion d'événements sportifs, la vidéoconférence, la présence virtuelle, Selon le rapport Cisco Visual Networking Index [1], et comme le montre la figure 1, qui représente actuellement 82 % du trafic global des consommateurs d'Internet [1], alors que ce pourcentage de vidéos n'était que de 59 % en 2017. En réponse à la pandémie du COVID-19, les écoles et les universités du monde entier ont adopté des cours en ligne, ce qui a entraîné une forte utilisation des plateformes de communication vidéo en temps réel telles que Zoom, recommandée par l'UNESCO<sup>1</sup>, qui est passée de la 180<sup>ème</sup> place en termes d'utilisation fin janvier à la 28<sup>ème</sup> place fin février 2020 [2]. Désormais les appareils conçus pour afficher des résolutions de haute définition (4K, 8K, 360<sup>0</sup>) et distribuer ces services vidéo deviennent facilement accessible au grand public ce qui conduit à une augmentation considérable du volume du contenu. Par exemple un clip vidéo Ultra Haute Définition (UHD) non compressée de quelques secondes peut occuper plusieurs giga-octets. De plus, les infrastructures de réseau ne peuvent pas transmettre une quantité énorme de données à l'état brut en raison de la bande passante limitée ce qui pose de nouveaux défis à la compression.

---

<sup>1</sup> UNESCO. (April 2020). Distance learning solutions. [Online].

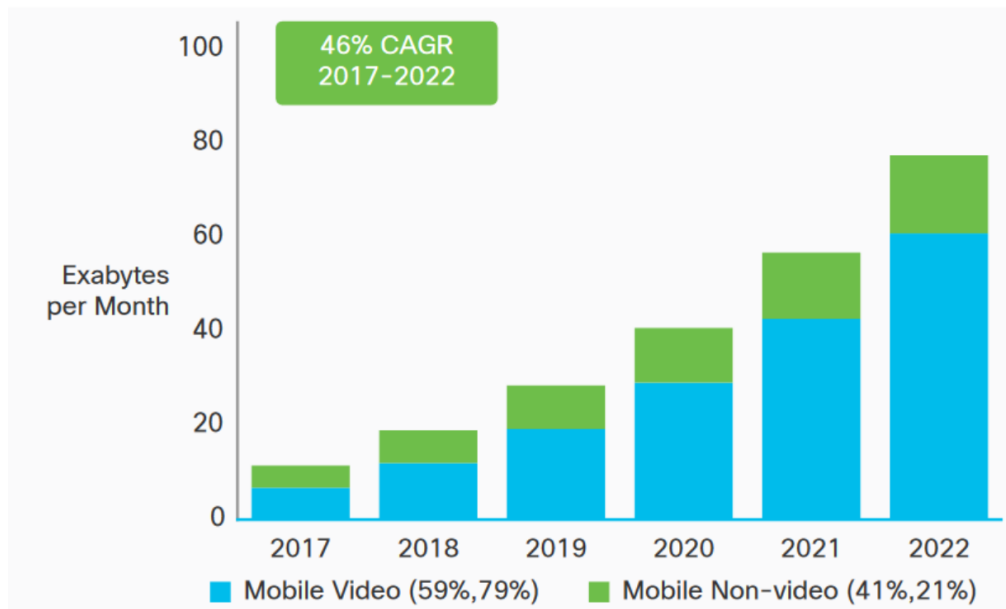


Figure 1. Pourcentage du trafic mondial des données mobiles vidéo et non-vidéo

Le CAGR est l'acronyme de "Compound Annual Growth Rate" en anglais, ce qui signifie en français le "Taux de Croissance Annuel Composé" (TCAC). Cette mesure est utilisée dans ce cas pour évaluer le taux de croissance en pourcentage du trafic mondial des données mobiles vidéo et non-vidéo.

Le HEVC ou H.265, successeur du H.264/AVC, offre un taux de compression nettement supérieur avec la même qualité par rapport à ses prédécesseurs pour être stocké ou envoyé via un service de streaming. Conçu pour les vidéos haute résolution (HD) et Ultra HD (4K et 8K), il s'est avéré efficace même pour les vidéos de faible résolution, y compris le format intermédiaire commun (CIF) comme présenté dans [3].

L'amélioration de près de 50 % de l'efficacité du codage de la norme HEVC est obtenue grâce à plusieurs caractéristique clé comme la structure Quad-tree Coding Tree Unit (CTU), la prédiction avancée de vecteur de mouvement (AMVP), des modes de codage intra/inter supplémentaires et des techniques de filtrage sophistiquées tel que le filtre SAO (Sample Adaptive Offset filter). Cependant, cette complexité supplémentaire des flux binaires codés en HEVC les rend plus vulnérables aux erreurs réseau par rapport aux flux binaires H.264. Toute perte d'information dans le flux binaire qui est un phénomène assez courant dans les réseaux sans fil, affectera non seulement la qualité perceptive du frame en cours, mais aussi celle des frames suivants qui dépendent indirectement de l'information perdue [4]. La diffusion de contenus vidéo avec une qualité satisfaisante pour les utilisateurs finaux est un véritable enjeu.

Une approche possible pour coder des flux vidéo robustes consiste à augmenter la résistance aux erreurs du HEVC en incluant des mécanismes de détection des erreurs dans le processus de codage. En conséquence, l'effet de la propagation des erreurs, due à une erreur de prédiction au niveau du décodage, sera limité. L'augmentation de la résistance aux erreurs des flux vidéo hautement compressés est un défi de recherche pertinent. Dont les solutions seront utiles pour accroître l'adaptabilité de la transmission dans diverses conditions de réseau et également pour faire face aux erreurs de communication. Elles permettront également d'offrir des niveaux accrus de qualité d'expérience (QoE) aux utilisateurs finaux. Il est donc d'actualité pour les opérateurs de calculer la QoE afin d'estimer la perception de l'utilisateur quant à la qualité de ces services. La QoE est faite par une évaluation subjective d'une personne ou d'un groupe de personnes dans les laboratoires où les vidéos passent à la télévision pour être jugées [5]. Cette évaluation est exprimée par les sentiments humains comme "bon", "excellent", "moyen", "médiocre" et "mauvais". Le Mean Opinion Score (MOS) est une métrique définie par l'Union internationale des télécommunications (l'ITU-T) dans la recommandation P.800.1 [6] très utilisée dans le domaine de la QoE. C'est un score qui peut varier de 1 à 5 pour refléter la qualité de restitution vidéo d'un codec.

#### Objectif de la thèse :

De ce contexte, cette thèse vise plusieurs objectifs :

- Maitriser et utiliser un codec tel que le HEVC pour la compression vidéo.
- Tester plusieurs types de descriptions (temporelles, spatiales)
- Etudier l'impact de différents paramètres de quantification (QP), diverses tailles du groupe d'images (Group Of Pictures ou GOP)
- Etudier l'impact des pertes de paquets isolés et en rafales sur la reconstruction de la vidéo.
- Rechercher les frames/parties de la vidéo les plus sensibles aux pertes de paquets, en particulier les effets des pertes de paquets lorsqu'elles se produisent :
  - ❖ Dans les frames I
  - ❖ Dans les frames P
  - ❖ Dans les frames B
- Proposer une architecture de transmission permettant de mieux protéger les frames les plus sensibles.

## Etat de l'art :

Plusieurs travaux ont été réalisés et publiés pour lutter contre la dégradation de la vidéo causée par la perte de paquets dans les canaux de transmission non fiables tels que les réseaux sans fil en général. Du côté de la réception, il y a la "dissimulation d'erreur", qui remplace les pixels perdus par les pixels adjacents en exploitant la similarité spatiale pour les petites zones ou la redondance temporelle pour la perte de l'image entière. Dans [4], Kulupana et al ont utilisé une méthode simple et pratique de copie de mouvement pour dissimuler l'erreur du côté récepteur. Kazemi et al [7] ont présenté une étude sur les techniques de dissimulation des erreurs vidéo temporelles pour HEVC et Versatile Video Coding (VVC ou H.266).

Les techniques de résilience aux erreurs sont appliquées du côté de la source pour protéger les données ou la vidéo à l'étape de la compression en créant un flux binaire robuste aux erreurs. Parmi ces techniques, il existe un "canal de retour" [8] où le récepteur peut demander la retransmission des parties perdues, mais cette technique n'est pas recommandée pour les applications en temps réel. Les méthodes « intra-refreshing » sont un groupe de techniques de résilience aux erreurs qui consistent à introduire des points d'intra-refreshing dans le flux binaire en forçant le codage de certaines trames ou blocs en intra. Cela permet d'éviter la propagation des erreurs dans la trame ainsi que sur plusieurs trames successives. Plusieurs travaux dans la littérature utilisent ces techniques pour améliorer la robustesse des vidéos [9, 10]. Le MDC est une solution prometteuse pour combattre la perte de paquets dans les réseaux best-effort tels qu'Internet [11]. Cette approche est compatible avec les applications en temps réel telles que la vidéoconférence, le streaming vidéo et les réseaux vidéo. Le MDC, contrairement au SDC, où un seul flux binaire est créé, permet de générer plusieurs flux binaires différents où chaque flux binaire est une description et où chaque description emprunte un chemin différent dans le canal de transmission. Du côté de la réception, les descriptions seront décodées indépendamment les unes des autres et le signal reçu peut être reconstruit à partir de plusieurs descriptions, ce qui réduit l'effet de la perte de paquets. Kazemi et al [12] ont présenté l'utilisation du H.264/AVC basé sur le MDC pour la transmission vidéo sans fil. Dans [13], ils ont publié des articles de synthèse sur le MDC. Boumahrez et al [14] ont utilisé la technique MDC temporelle pour améliorer la qualité vidéo codée HEVC d'un flux vidéo dans un réseau sans fil soumis à la perte de paquets. Ibrahim et al [15] ont utilisé la même méthode MDC pour protéger la vidéo codée H.264 transmise dans un réseau radio cognitif assisté par relais. Le travail de [16] a proposé un MDC à couche mixte basé sur H.264 pour l'optimisation adaptative du canal.

De nos jours, les réseaux de neurones profonds ont gagné en popularité en tant que solution polyvalente pour résoudre une multitude de problèmes informatiques, y compris le dé-bruitage d'images causé par des erreurs de transmission ou d'acquisition. Notamment, [17] ont exploré l'utilisation d'auto-encodeurs dans leur recherche pour éliminer le bruit gaussien à partir d'images dégradées. Leur travail a jeté les bases d'une des approches proposées dans cette thèse, qui s'inspire des concepts et des méthodes qu'ils ont développés. Notre recherche s'inscrit dans la continuité de ces efforts précédents et vise à apporter de nouvelles améliorations dans la suppression du bruit dans les images. Une autre catégorie de modèles de réseaux neuronaux profonds, qui a obtenu des résultats remarquables, est celle des auto-encodeurs de dé-bruitage [18]. Ces auto-encodeurs sont spécialement conçus pour le dé-bruitage d'images, prenant en entrée des images bruitées et cherchant à générer une version débruitée.

### Organisation de la thèse :

Dans cette thèse, nous explorerons en détail les problématiques liées à la transmission vidéo sur des réseaux sujets à des pertes de données. Pour ce faire, notre travail est organisé en quatre chapitres. Tout d'abord, le Chapitre 1 établira les bases de la compression vidéo et de la transmission. Ensuite, le Chapitre 2 définira la problématique de la transmission vidéo sur des réseaux avec pertes. Le Chapitre 3 constituera un état de l'art complet sur les mécanismes d'amélioration de la qualité du streaming vidéo. Dans le Chapitre 4, nous analyserons les performances et les résultats de nos recherches. Enfin, nous conclurons notre travail en résumant nos conclusions, en discutant des perspectives futures, et en mettant en lumière l'impact de nos découvertes dans le domaine de la transmission vidéo. Chaque chapitre s'inscrit dans une progression logique visant à résoudre la problématique centrale de cette thèse.

# Chapitre 1

## 1. Les principes de base de la compression et la transmission vidéo

### 1.1 Introduction

La vidéo est un média qui permet la capture et la reproduction d'images animées et de son simultanément. Depuis sa création, la vidéo est devenue un élément important dans notre culture, notre divertissement et notre vie quotidienne. Elle est utilisée dans de nombreux domaines, tels que la télévision, le cinéma, la publicité, l'éducation, la surveillance, la recherche scientifique et plus récemment, les réseaux sociaux et les plateformes de partage de vidéos en ligne. La vidéo peut être enregistrée et diffusée sous différentes formes et formats, allant de la vidéo analogique à la vidéo numérique haute définition en passant par la vidéo en 360 degrés. Ce chapitre présente les éléments fondamentaux de la codification vidéo, notamment la chaîne d'outils qui permet d'acquérir une séquence vidéo, de la coder, de la transmettre et de l'afficher. L'accent est mis sur les aspects liés au processus de codage et de décodage, tels que le format de représentation des séquences vidéo et la représentation des couleurs. Il présente également le concept fondamental et les principaux éléments constitutifs de la codification vidéo hybride, afin de comprendre comment ils interagissent entre eux. Les prochains chapitres permettront de bien connaître les différents blocs de traitement de la vidéo dans le cadre du HEVC.

### 1.2 Définition de la vidéo

D'un point de vue technique, une vidéo est une séquence d'images fixes appelées "frames" (ou images clés) qui sont affichées successivement à un rythme régulier, créant ainsi l'illusion de mouvement. Chaque image est constituée d'un ensemble de pixels (points) qui contiennent des informations sur la couleur et la luminosité de chaque point de l'image.

Les images de la vidéo sont généralement stockées dans un format numérique qui permet de les compresser afin d'occuper moins d'espace de stockage et de faciliter leur transfert et leur lecture. Les formats de compression vidéo les plus courants incluent le Moving Picture Experts Group (MPEG), Audio Video Interleave (AVI), Windows Media Video (WMV), QuickTime Movie (MOV) et MPEG-4 (MP4). Outre les images, la vidéo contient également des informations audios qui sont généralement enregistrées en même temps que les images. Les informations audios sont stockées dans un format numérique séparé et peuvent être synchronisées avec les images lors de la lecture.

La qualité de la vidéo dépend de plusieurs facteurs tels que la résolution, le taux de fréquence d'images (frame rate), le format de compression utilisé, le débit binaire (bit rate) et les paramètres d'encodage. Les vidéos hautes définitions (HD) ont une résolution supérieure à celle de la vidéo standard, offrant ainsi une meilleure qualité d'image et de détails.

### 1.2.1 Histoire de la vidéo numérique

L'histoire de la vidéo numérique remonte aux années 1950, lorsque les premiers enregistreurs vidéo ont été développés pour la télévision. À l'époque, ces enregistreurs utilisaient des bandes magnétiques pour stocker les signaux vidéo et audio.

Dans les années 1970, l'industrie de la vidéo a connu une évolution majeure avec l'introduction des premiers magnétoscopes domestiques. Ces magnétoscopes utilisaient des cassettes vidéo analogiques pour stocker les signaux vidéo et audio, permettant aux gens d'enregistrer et de regarder des émissions de télévision à leur convenance.

Au cours des années 1980 et 1990, la technologie de la vidéo numérique a commencé à émerger avec l'introduction des premiers caméscopes numériques et des premiers enregistreurs vidéo numériques. Ces appareils ont utilisé des signaux numériques pour enregistrer et stocker les images vidéo et audio, offrant une qualité supérieure et une plus grande flexibilité dans la post-production.

Dans les années 2000, la vidéo numérique a connu une croissance exponentielle avec la popularité de l'Internet et des sites de partage de vidéos en ligne tels que YouTube. Les caméras vidéo numériques sont devenues plus abordables et plus accessibles, permettant à tout le monde de produire et de partager des vidéos en ligne.

Aujourd'hui, la vidéo numérique est omniprésente dans notre vie quotidienne et est utilisée dans une grande variété de domaines, allant de la production de films et de télévision à la surveillance et à la formation en ligne. La technologie continue de s'améliorer, offrant des résolutions vidéo toujours plus élevées, une plus grande qualité d'image et une plus grande efficacité de stockage.

### 1.2.2 Scènes vidéo

Les séquences vidéo peuvent provenir de diverses sources, telles que la capture de scènes naturelles, la numérisation de films photographiques, l'enregistrement analogique de vidéos ou la génération synthétique de matériel vidéo tel que des animations générées par ordinateur. Chacune de ces sources peut présenter des caractéristiques et des défis différents en matière de compression vidéo. Par

exemple, les vidéos naturelles peuvent contenir des mouvements rapides et imprévisibles, tandis que les films photographiques peuvent présenter du grain et des artefacts. En outre, les caméras et les capteurs peuvent avoir des propriétés différentes, telles que la résolution et la fréquence d'images, qui influencent également les besoins de compression vidéo. Dans cette introduction, nous examinerons les défis et les opportunités associés à la compression vidéo de différentes sources de séquences vidéo, ainsi que les techniques et les outils de compression utilisés dans la recherche actuelle sur la compression vidéo.

### 1.2.3 Acquisition

Les vidéos peuvent être issues de diverses sources, notamment la capture de scènes naturelles, la numérisation de films photographiés, l'enregistrement analogique de vidéos ou la génération synthétique de matériel vidéo, comme les animations générées par ordinateur. Il est également possible de combiner ces différentes sources. Selon l'origine de la vidéo, le signal présente des caractéristiques propres en ce qui concerne le contenu de la scène représentée et la présence d'autres composants de signal qui interfèrent avec celui-ci, tels que le bruit de la caméra, le grain du film, le flou de mouvement ou les artefacts résultant de la numérisation de sources analogiques ou physiques. Dans un modèle de signal simple, une séquence vidéo acquise peut être décrite comme un signal vidéo non déformé, auquel est ajoutée une composante de bruit. En outre, les artefacts non additifs dans la séquence vidéo peuvent être liés à des distorsions géométriques ou des problèmes d'alignement des couleurs dus aux propriétés de l'objectif de la caméra. Lorsque des scènes naturelles sont capturées, les propriétés de la caméra, telles que la construction et la résolution de la puce du capteur ou la vitesse d'obturation, jouent un rôle important dans les propriétés de la séquence vidéo résultante et la qualité de l'image.

#### 1.2.4 Format vidéo et codecs

Ces résolutions d'images, allant du CIF à l'Ultra HD (4K), ont une incidence significative sur la qualité et la clarté visuelle des images et vidéos, jouant un rôle essentiel dans divers domaines allant de la vidéosurveillance aux médias de divertissement.

Dans le domaine de la compression et de la transmission de vidéos, le format CIF est largement utilisé comme une résolution standard pour les vidéos. Le CIF possède une résolution de 352x288 pixels, avec un rapport d'aspect de 4:3. Cette résolution est souvent utilisée comme un point de départ pour les images vidéo de qualité intermédiaire et est fréquemment rencontrée dans les systèmes de vidéosurveillance et les communications vidéo en temps réel.

Une résolution dérivée du CIF est le 2CIF, également connu sous le nom de Half D1. Le 2CIF offre une résolution de 704x288 pixels avec un rapport d'aspect de 2:1. Cette résolution améliorée est couramment utilisée dans les systèmes de vidéosurveillance pour obtenir une meilleure qualité d'image sans augmenter considérablement les exigences en bande passante.

Une autre résolution dérivée du CIF est le 4CIF, également appelé Full D1. Le 4CIF a une résolution de 704x576 pixels, toujours avec un rapport d'aspect de 4:3. Cette résolution est utilisée pour des applications vidéo plus exigeantes, telles que les enregistrements vidéo professionnels et certaines applications de vidéosurveillance haut de gamme.

En ce qui concerne les résolutions standard définies pour la télévision et la vidéo, nous trouvons les résolutions SD (Standard Definition). Ces résolutions incluent le 480p National Television System Committee (NTSC) avec une résolution de 720x480 pixels et le 576p (PAL) avec une résolution de 720x576 pixels. Les résolutions SD sont couramment utilisées pour les émissions de télévision et les vidéos grand public.

Les résolutions HD (High Definition) ont gagné en popularité en raison de leur qualité d'image améliorée. Le 720p offre une résolution de 1280x720 pixels avec un rapport d'aspect de 16:9, tandis que le 1080p (Full HD) offre une résolution de 1920x1080 pixels, également avec un rapport d'aspect de 16:9. Ces résolutions HD sont couramment utilisées pour les téléviseurs, les moniteurs d'ordinateur, les vidéos en streaming et les appareils photo numériques haut de gamme.

Une résolution supérieure à la HD est l'Ultra HD, également connue sous le nom de 4K. Le 2160p offre une résolution de 3840x2160 pixels avec un rapport d'aspect de 16:9, fournissant une qualité d'image exceptionnelle pour les téléviseurs haut de gamme, les moniteurs d'ordinateur et les vidéos

professionnelles. Le 8K (4320p) est une autre résolution de l'Ultra HD, offrant une résolution de 7680x4320 pixels, mais elle est encore moins répandue dans le grand public en raison de la disponibilité limitée des écrans et du contenu.

Les codecs vidéo jouent un rôle crucial dans la compression et la transmission efficaces des vidéos dans diverses applications, allant de la diffusion en continu et du divertissement aux systèmes de vidéosurveillance et aux communications en ligne. Le choix du codec dépend des besoins spécifiques de chaque utilisation, en équilibrant la qualité vidéo, la taille des fichiers et la compatibilité avec les appareils et les plateformes cibles. Parmi eux on trouve :

❖ MPEG-1 et MPEG-2 :

Les codecs MPEG-1 et MPEG-2 ont été parmi les premiers codecs vidéo largement utilisés. Le MPEG-1 est souvent utilisé pour la compression de vidéos de qualité VCD (Video CD) avec une résolution de 352x240 (NTSC) ou 352x288 (PAL). Le MPEG-2, quant à lui, est utilisé pour la compression de vidéos de qualité DVD avec des résolutions telles que 720x480 (NTSC) ou 720x576 (PAL). Ces codecs ont été fondamentaux pour la diffusion de vidéos sur disques optiques et ont servi de base à de nombreux autres codecs ultérieurs.

❖ H.264 (MPEG-4 AVC) :

L'H.264, également connu sous le nom de MPEG-4 Partie 10 Advanced Video Coding (AVC), est l'un des codecs vidéo les plus populaires et largement utilisés. Il offre une excellente qualité vidéo avec une compression élevée, ce qui le rend adapté à une variété d'applications, y compris les vidéos en streaming, les vidéos haute définition (HD) et les vidéos enregistrées sur les appareils mobiles. L'H.264 est devenu le codec de facto pour la vidéo en ligne et la diffusion en continu.

❖ H.265 (HEVC) :

L'H.265, également connu sous le nom de High Efficiency Video Coding (HEVC), est une évolution de l'H.264. Il offre une compression plus efficace, permettant de réduire davantage la taille des fichiers vidéo tout en maintenant une qualité visuelle élevée. Le HEVC est de plus en plus utilisé pour les vidéos en 4K et 8K, ainsi que pour les vidéos haute résolution nécessitant une bande passante réduite.

❖ H.266 (VVC)

Le VVC (Versatile Video Coding) est un codec vidéo de nouvelle génération, également connu sous le nom de H.266. Il succède à l'HEVC (H.265) et offre une compression vidéo encore plus efficace, réduisant la taille des fichiers tout en maintenant une haute qualité visuelle. Le VVC est particulièrement adapté pour les vidéos en streaming en 4K, 8K et au-delà. Bien que récent, le VVC devrait jouer un rôle clé dans l'avenir du streaming vidéo haute résolution, mais son adoption dépendra de sa compatibilité avec les appareils grand public et les plateformes de streaming.

❖ VP9 :

VP9 est un codec vidéo open-source développé par Google et utilisé notamment pour la diffusion en streaming sur YouTube. Il offre une bonne qualité vidéo avec une compression efficace et est une alternative à H.264 et H.265 pour la diffusion de vidéos en ligne. VP9 est soutenu par de nombreux navigateurs web modernes, ce qui facilite la lecture de vidéos VP9 en streaming.

❖ AV1 :

AV1 est un autre codec vidéo open-source développé par l'Alliance for Open Media (AOMedia). Il vise à offrir une qualité vidéo élevée tout en réduisant considérablement la taille des fichiers vidéo. AV1 est principalement utilisé pour les vidéos en streaming, en particulier pour les contenus en 4K et 8K. En raison de sa récente adoption, il gagne en popularité et sa prise en charge est de plus en plus répandue. Figure 2 résume l'historique des standards de codage vidéo.

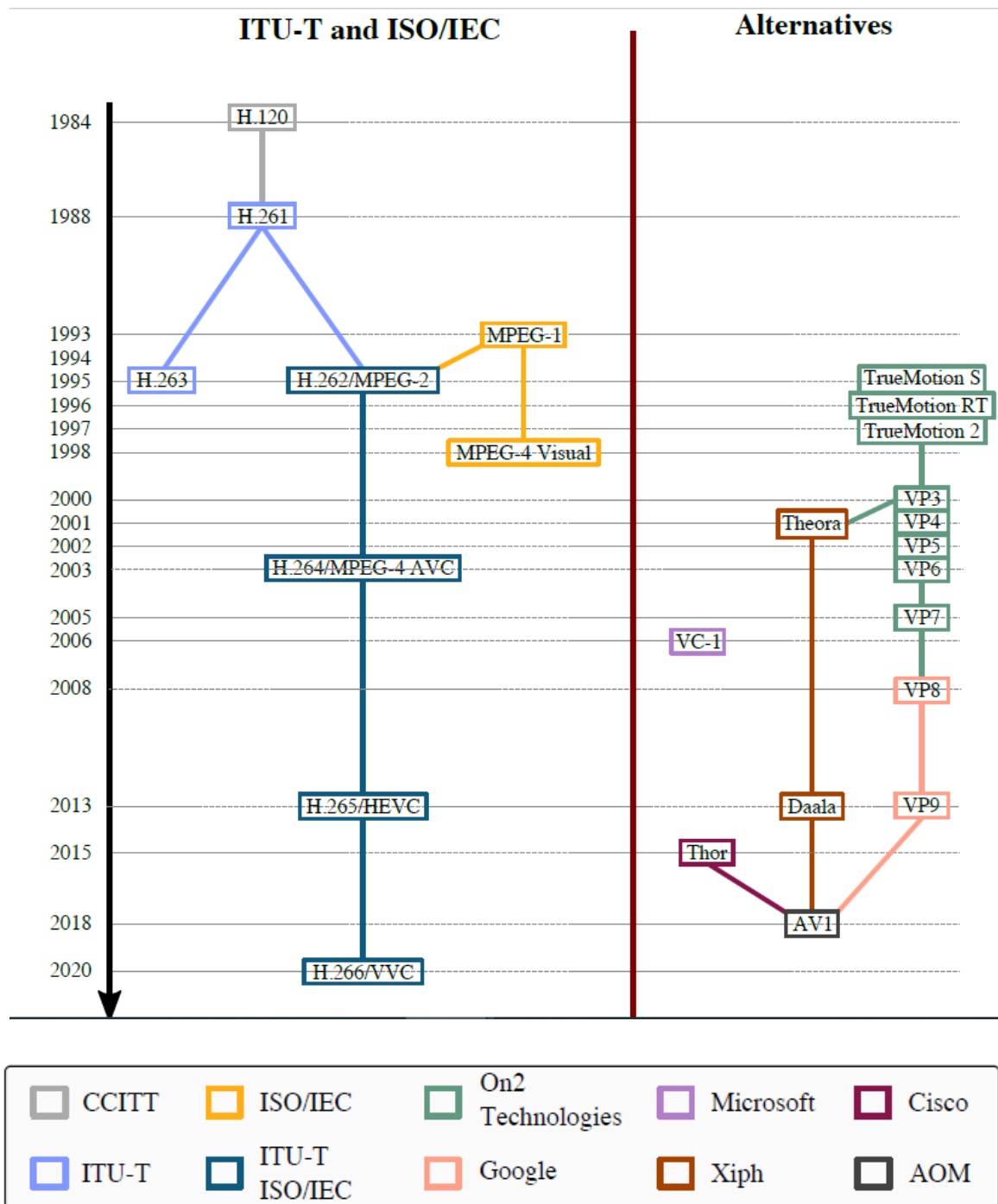


Figure 2. Historique des standards de codage vidéo [19]

### 1.2.5 Différence entre vidéo brute et vidéo compressée

La principale différence entre une vidéo brute et une vidéo compressée réside dans la quantité de données qu'elles contiennent et la qualité de l'image. Une vidéo brute, également appelée vidéo non

compressée, est une vidéo qui contient toutes les données d'origine capturées par l'appareil photo ou la caméra, sans aucune modification. Ces vidéos sont volumineuses et occupent énormément d'espace de stockage. Elles ont, cependant, une qualité d'image supérieure à celle d'une vidéo compressée, car elles ne sont pas affectées par la perte de données due à la compression.

En revanche, une vidéo compressée est une vidéo qui a été traitée pour réduire la quantité de données qu'elle contient. La compression vidéo est utilisée pour réduire la taille du fichier, afin de faciliter le stockage, le partage et la diffusion sur les réseaux. Cependant, la compression peut entraîner une perte de qualité d'image, en particulier lorsqu'elle est compressée de manière significative. La qualité d'une vidéo compressée dépend de plusieurs facteurs, tels que le taux de compression, la méthode de compression utilisée et le codec utilisé.

### 1.3 Contenu vidéo

Les vidéos présentent une grande variété de contenu, allant des scènes naturelles aux dessins animés en passant par les sports, les émissions et les vidéos générées par ordinateur, chacun pouvant être subdivisé en sous-types distincts. Cette diversité pose des défis à la recherche basée sur le contenu, car chaque type de contenu présente des différences importantes qui doivent être prises en compte dans plusieurs domaines. Premièrement, la sélection de la source vidéo pour les expériences subjectives doit tenir compte des caractéristiques spécifiques du contenu. Deuxièmement, l'analyse des caractéristiques du contenu est importante pour améliorer les mesures objectives de qualité vidéo. Des travaux antérieurs ont montré que les caractéristiques du contenu ont un impact significatif sur les mesures objectives de qualité vidéo [20]. Troisièmement, les caractéristiques du contenu doivent être prises en compte pour améliorer l'efficacité du codage vidéo. Des expériences ont montré que le choix de la configuration optimale de codage vidéo dépend des caractéristiques du contenu. Enfin, l'utilisation d'algorithmes de commutation adaptatifs sensibles au contenu peut améliorer la résilience aux erreurs en choisissant la technique de dissimulation d'erreur la plus appropriée en fonction des caractéristiques du contenu [14].

Certaines caractéristiques du contenu vidéo notamment celles liées au mouvement et à la texture seront utilisées dans cette thèse dans la sélection d'un sous-ensemble de contenus pour mener des expériences et pour l'analyse des résultats subjectivement et objectivement.

### 1.4 Les étapes de la compression vidéo

L'objectif de la compression est de convertir un contenu d'entrée en un flux de bits contenant le moins d'informations possible tout en permettant la reconstruction du contenu. Pour rendre ce flux de

bits plus léger que le contenu d'entrée, les techniques consistent principalement à exploiter les limites du système visuel humain (HVS) et à supprimer les informations redondantes dans l'entrée. Les étapes de la compression vidéo sont illustré dans la figure 3.

Il existe deux principaux types de compression utilisés par les codecs vidéo :

Compression avec perte (Lossy) : Dans ce type de compression, certaines informations jugées moins importantes pour la perception visuelle humaine sont supprimées pour réduire la taille du fichier. Bien que cela entraîne une perte de qualité, la compression avec perte permet d'obtenir des fichiers beaucoup plus petits, adaptés à la transmission et au stockage.

Compression sans perte (Lossless) : Contrairement à la compression avec perte, la compression sans perte permet de conserver toutes les informations originales de la vidéo, garantissant ainsi une qualité vidéo parfaite. Cependant, les fichiers résultants sont moins compressés et donc plus volumineux que les fichiers compressés avec perte.

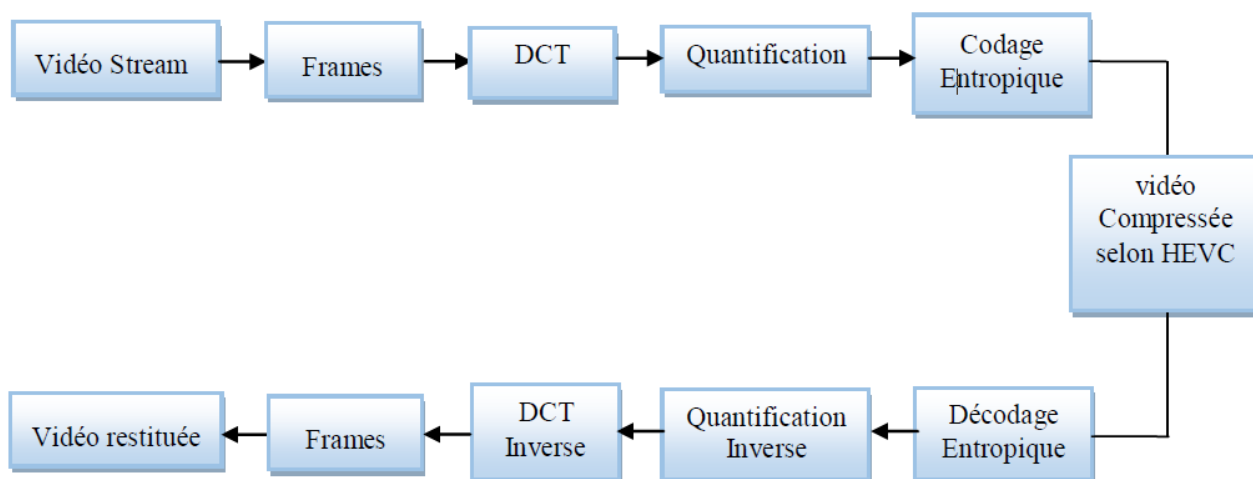


Figure 3. Les étapes de la compression video

#### 1.4.1 Prédiction

La vidéo est constituée d'une série d'images qui présentent des redondances spatiales et temporelles élevées exemple dans un match de football entre deux séquences, le terrain et le stade ne changent pas, le public bouge légèrement, mais la vraie distinction se fait dans les mouvements des joueurs et du ballon. Pour éliminer les dépendances spatiales et temporelles, appelées respectivement prédiction intra et inter, des prédictions sont utilisées dans la vidéo. La prédiction intra est réalisée entre les blocs de pixels dans une même image, tandis que la prédiction inter est effectuée entre le bloc actuel

et le meilleur match du bloc prédit ou reconstruit qui peut être présent dans les images précédentes ou futures décodées. La différence entre les blocs originaux et prédits, appelée erreur de prédiction, est ensuite transmise au processus de transformation.

#### 1.4.2 Transformation

La transformation permet de décorrélérer les pixels d'une image et de compacter l'énergie dans un nombre restreint de coefficients. Le choix de la méthode de transformation est une étape essentielle lors du processus de conception d'un nouvel encodeur de compression. Les propriétés de la transformation choisie caractérisent les performances et les caractéristiques de l'encodeur. Plusieurs méthodes de transformation réversibles sont utilisées dans le domaine de la compression d'images et de vidéos. Les plus répandues sont : la transformation de Karhunen Lovev (KL), la transformée en cosinus discrète (DCT), la transformation en ondelettes entières (IWT), la transformation de Hadamard (HT), également connue sous le nom de transformation de Walsh-Hadamard et la transformation en ondelettes discrètes (DWT).

#### 1.4.3 Quantification

L'étape de quantification attribue des valeurs, prises à partir d'un ensemble fini dénombrable, à tous les coefficients de transformation. Contrairement aux étapes précédentes, l'étape de quantification est irréversible et introduit une distorsion, donc cette étape n'est appliquée qu'à un codage avec perte.

#### 1.4.4 Codage entropique

L'encodage d'entropie opère sur les statistiques des coefficients quantifiés pour éliminer la redondance statistique entre eux. De plus, un codeur d'entropie utilise le code à longueur variable (VLC) pour les mots de code, de sorte que les coefficients les plus fréquents sont représentés par des codes binaires courts, et les coefficients les moins fréquents sont représentés par des codes binaires longs. Considérons une variable aléatoire discrète (imprévisible)  $X$  avec des réalisations  $x_i$  où  $i = 1 ; 2 ; \dots ; n$  qui correspondent aux coefficients de l'image transformée, et  $P(x_i)$  est la probabilité d'apparition du coefficient  $x_i$ . L'entropie de l'image transformée est définie par la valeur attendue de la quantité d'information.

### 1.5 Décodage vidéo

Le décodage d'un flux de bits vidéo HEVC suit le même processus de base que pour tout autre système de codage vidéo. Le flux de bits encodé est reçu par le décodeur et stocké dans un tampon. Le décodeur reconstruit ensuite les données encodées en une séquence vidéo dans le format indiqué

par l'encodeur. Dans le cas de HEVC, le processus de décodage est particulièrement complexe car il utilise des techniques avancées telles que la prédiction inter-image, la prédiction intra-image, le codage de la résiduelle transformée et le filtrage en boucle. Le décodeur HEVC utilise également des algorithmes sophistiqués pour réduire le débit binaire du flux vidéo, tels que la quantification adaptative et la réduction des bits résiduels. Enfin, la résilience aux erreurs est un aspect clé du décodage HEVC car il est capable de traiter les pertes de transmission et les erreurs pour fournir une vidéo de haute qualité malgré les perturbations du canal de communication. Le décodage de flux vidéo HEVC est donc un processus complexe mais efficace, qui permet une compression élevée du flux vidéo tout en maintenant une qualité d'image élevée.

### 1.1 High Efficiency Video Coding

Le HEVC, est une norme de compression vidéo élaborée conjointement par le groupe d'experts en codage vidéo de l'ITU-T et le groupe d'experts en images animées du MPEG de l'organisation internationale de normalisation (ISO) et la Commission électrotechnique internationale (CEI). Elle est parfois appelée H.265 et a été publiée en 2013. Conçue pour améliorer la qualité d'image tout en réduisant la quantité de données requises pour stocker ou transmettre les vidéos, elle utilise des techniques de compression plus avancées que celles utilisées dans les normes précédentes, telles que MPEG-2, MPEG-4 et H.264/AVC.

Grâce à son efficacité de compression accrue, HEVC peut compresser les vidéos jusqu'à deux fois plus que H.264/AVC, ce qui se traduit par des fichiers vidéo deux fois plus petits pour une qualité d'image équivalente. Cette norme est particulièrement utile pour les vidéos haute résolution telles que la 4K et la 8K, ainsi que pour la diffusion en continu et la vidéo à la demande sur Internet. HEVC utilise des techniques de compression telles que la prédiction inter-image, la transformation de blocs, la quantification et l'encodage entropique. Elle prend également en charge des fonctionnalités avancées telles que la gamme dynamique élevée (HDR) technologie utilisée dans la photographie, la vidéo et l'affichage numérique pour reproduire une plage de luminosité plus étendue que celle obtenue avec des techniques traditionnelles et les couleurs 10 bits. HEVC est de plus en plus couramment utilisé dans les applications de diffusion en continu et les télévisions 4K et 8K.

#### 1.5.1 Principe du codage HEVC

Le HEVC, comme toutes les normes précédentes de codage vidéo de l'UIT-T et de l'ISO/CEI JTC depuis H.261 [21], utilise une approche de codage vidéo hybride classique basée sur des blocs (comme illustré à la figure 4). L'algorithme de codage source de base est hybride, combinant une

prédiction inter-image pour exploiter les dépendances statistiques temporelles, une prédiction intra-image pour exploiter les dépendances statistiques spatiales, et un codage par transformation des signaux résiduels de prédiction pour exploiter davantage les dépendances statistiques spatiales.

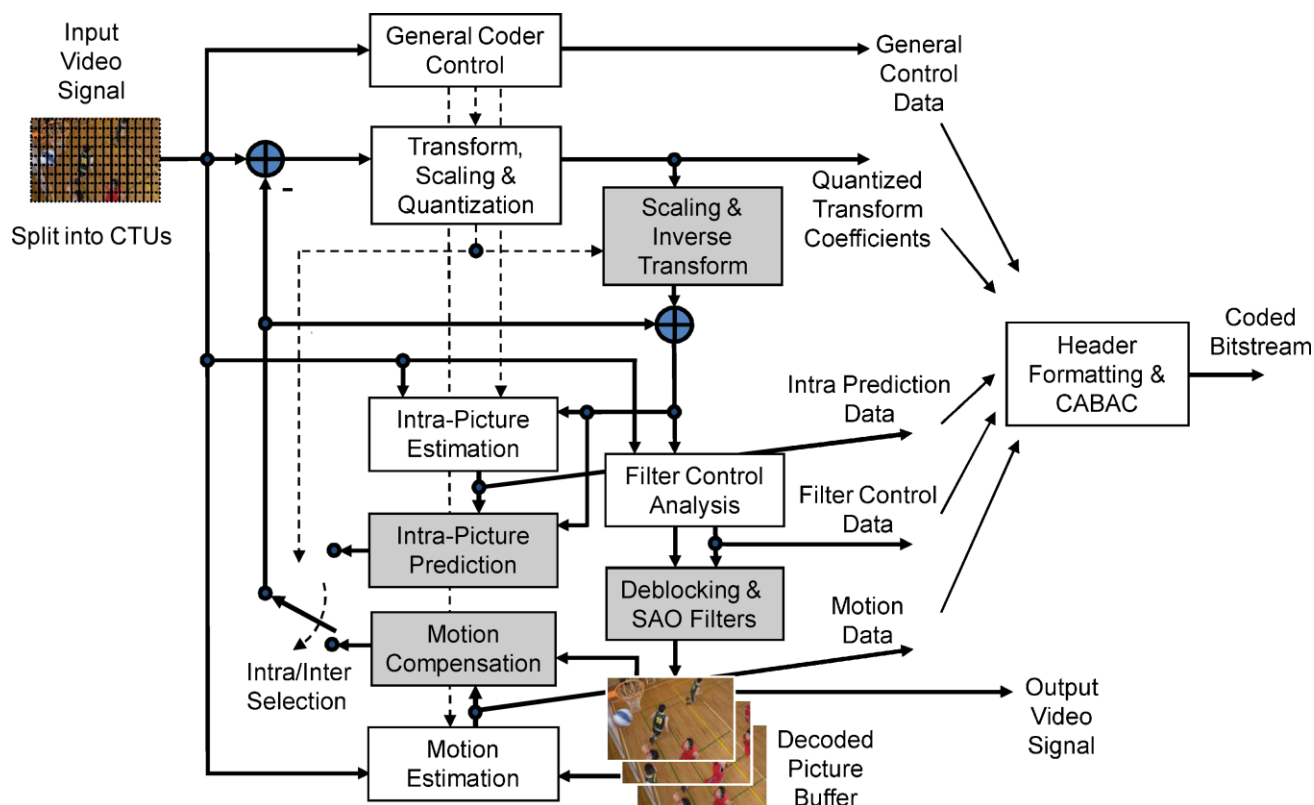


Figure 4. Schéma synoptique d'un encodeur HEVC doté d'un décodeur intégré représenté en gris [22].

### 1.5.1.1 Représentation échantillonnée d'images

Le HEVC traite des vidéos représentées en espace colorimétrique tristimulus  $YCbCr$ , avec un échantillonnage 4:2:0, 4:2:2 et 4:4:4, avec 8 et 10 bits par échantillon pour la représentation du signal vidéo d'entrée et de sortie décodé. L'espace colorimétrique qui permet de séparer la couleur en trois composantes distinctes : Y, Cb et Cr. La composante Y, également appelée luminance, quantifie la luminosité tandis que les deux composantes de chrominance Cb et Cr décrivent la mesure dans laquelle la couleur s'éloigne du gris vers les teintes de bleu et de rouge, respectivement. Ce choix de sous-échantillonnage de chrominance permet de réduire la quantité de données à stocker ou à transmettre, sans sacrifier considérablement la qualité visuelle.

### 1.5.1.2 Partitionnement de blocs

HEVC découpe chaque image en blocs de carrés disjoints de taille identique. Chacun de ces blocs sert de racine à une structure arborescente de quadtree de premier niveau, appelé bloc de l'arbre de codage (CTB). Les CTBs peuvent ensuite être subdivisés le long de la structure de l'arbre de codage en blocs de codage (CBs). Ce sont ces derniers que l'encodeur doit prendre en compte lorsqu'il doit décider entre une prédiction intra-image ou une prédiction inter-image.

Chaque bloc CTB des échantillons de luminance, ainsi que ses deux CTB correspondants d'échantillons de chrominance et la syntaxe associée à ces blocs d'échantillons, sont regroupés en une seule unité de codage appelée unité d'arbre de codage (CTU) qui représente l'unité de traitement de base dans HEVC comme le représente la figure 5 et 6 et est à cet égard similaire au concept de macrobloc dans les normes de codage vidéo précédentes [23-27]. La largeur et la hauteur de l'unité CTU sont signalées dans un ensemble de paramètres de séquence (SPS), ce qui signifie que toutes les CTU dans une séquence vidéo ont la même taille : 64x64, 32x32 ou 16x16.

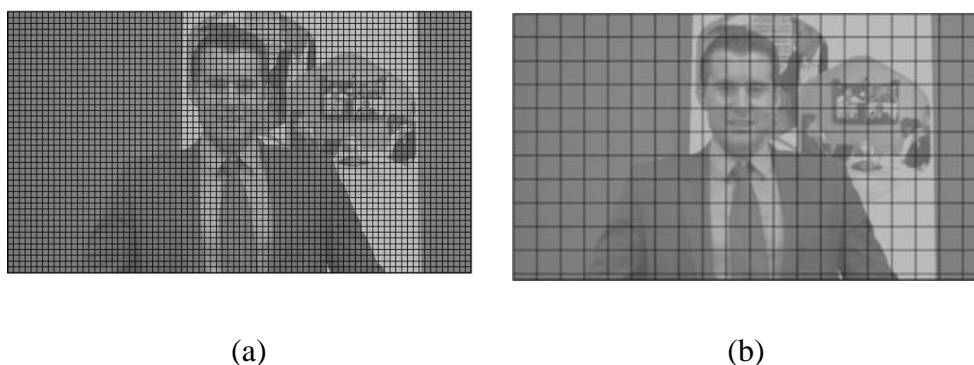


Figure 5. Illustration du partitionnement d'une image de taille 1280 x 720 (a) Partition en macroblocs de 16 x 16, telle qu'on la retrouve dans toutes les normes précédentes de codage vidéo de l'UIT-T et de l'ISO/CEI ; (b) Partition de l'image en CTU de 64 x 64, prise en charge dans le profil principal de HEVC.

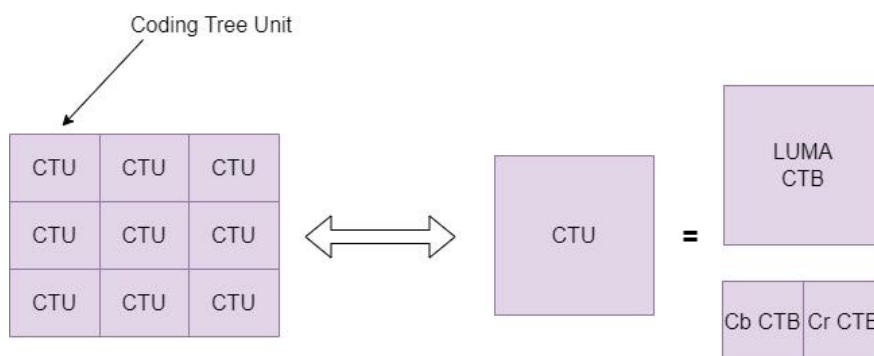


Figure 6. Illustration du regroupement des CTBs en CTU



### 2.2.3.2. Prédiction intra image

HEVC prend en charge des méthodes de prédiction intra-image, telles que la prédiction directionnelle avec 33 orientations différentes pour les blocs de transformation de taille variable (TB) allant de 4×4 à 32×32. On utilise les échantillons de frontière décodés des TB voisins pour former le signal de prédiction. Les modes de prédiction planaire et DC sont également disponibles. Pour les composantes chroma, les modes de prédiction horizontale, verticale, planaire et DC peuvent être signalés ou liés aux modes de prédiction de luminance. Chaque bloc de codage (CB) peut être codé par différents types selon le type de tranche, et HEVC supporte les techniques de codage prédictif Intra-Angular, Intra-Planar et Intra-DC, avec plusieurs autres méthodes courantes. La figure 8 montre les Modes de prédiction intra angulaire dans le HEVC.

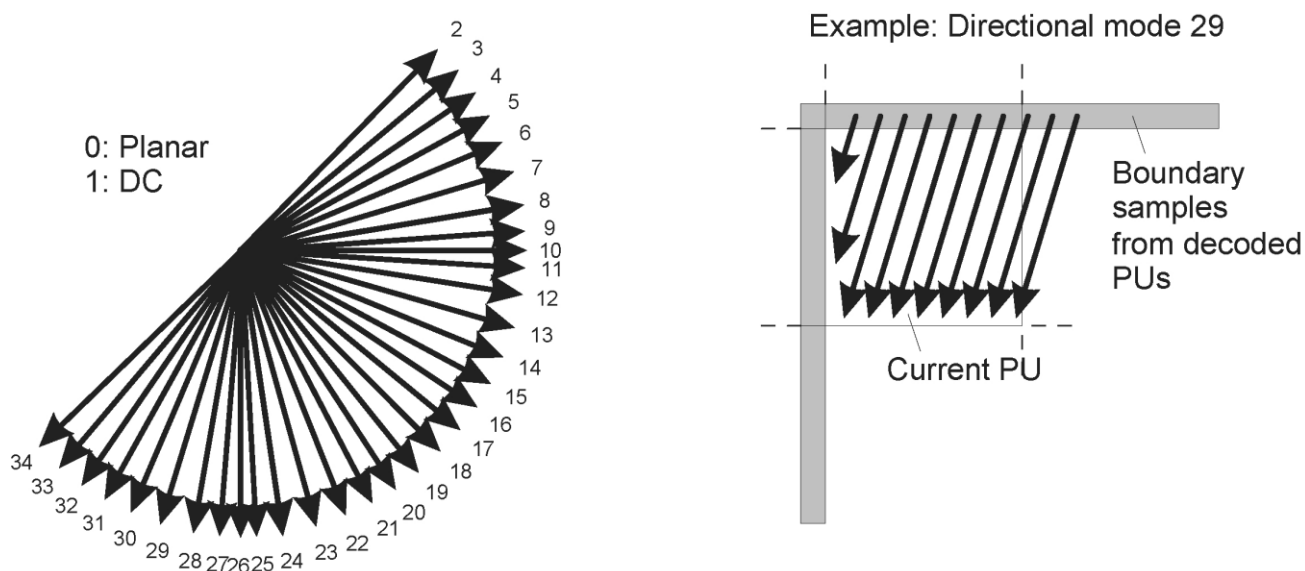


Figure 8. Modes de prédiction intra angulaire dans le HEVC

### 2.2.3.3. Prédiction inter image

Alors que la prédiction intra-image exploite la corrélation entre les échantillons voisins dans l'espace, la prédiction inter-image utilise la corrélation temporelle entre les images pour dériver une prédiction compensée par mouvement (PCM) pour un bloc d'échantillons d'image. Pour cette PCM basée sur des blocs, une image vidéo est divisée en blocs rectangulaires. En supposant un mouvement homogène à l'intérieur d'un bloc et que les objets en mouvement sont plus grands qu'un bloc, pour chaque bloc, on peut trouver un bloc correspondant dans une image préalablement décodée qui sert de prédicteur. Le concept général de PCM basée sur un modèle de mouvement translationnel est illustré dans la figure 9.

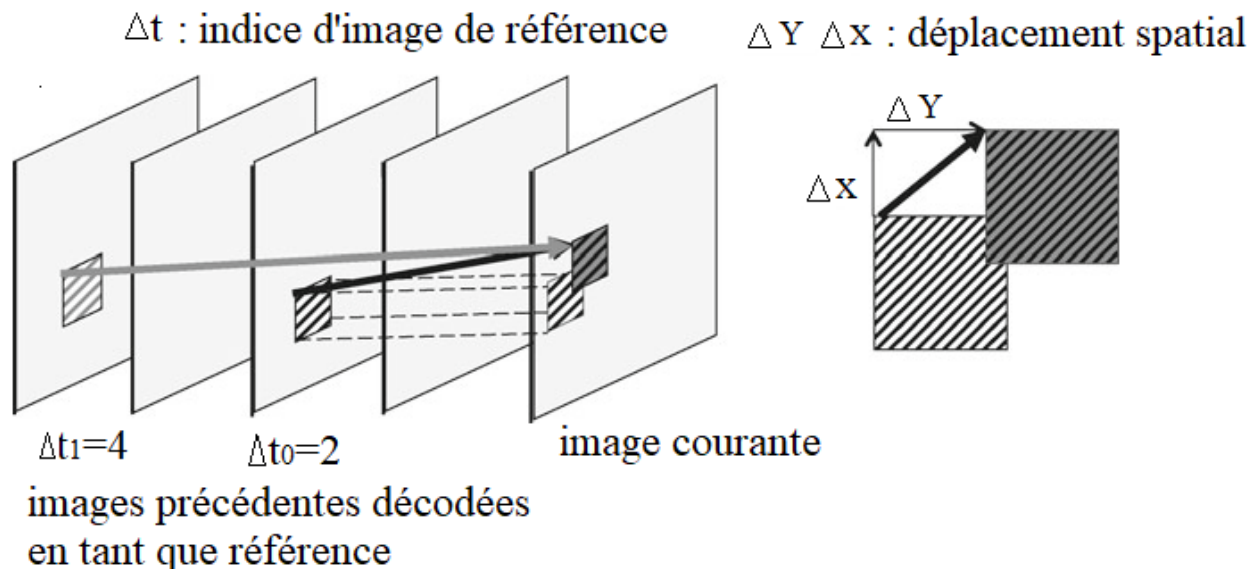


Figure 9. Concept général de PCM basée sur un modèle de mouvement translationnel.

Dans le codec HEVC, la prédiction inter-image améliore les normes vidéo précédentes comme H.264/AVC. Elle utilise une compétition de vecteurs de mouvement avancée et une fusion de blocs de prédiction pour simplifier la signalisation des données de mouvement. L'interpolation des échantillons d'images de référence utilise des filtres de haute précision pour améliorer le filtrage, tandis que la signalisation de la prédiction pondérée peut se faire par des poids explicites ou une moyenne de prédictions. Ce chapitre explique en détail ces aspects du standard HEVC, mettant en avant leur efficacité de codage et leur complexité, tout en distinguant entre la prédiction intra-image basée sur la corrélation spatiale et la prédiction inter-image basée sur la corrélation temporelle.

## 1.6 Versatile Video Coding

VVC, est un nouveau standard de compression vidéo qui a été développé pour offrir une qualité d'image supérieure et une meilleure efficacité de compression par rapport aux normes précédentes telles que le H.264 et le H.265/HEVC. Le VVC a été développé conjointement par l'ITU-T VCEG et le MPEG-ISO/IEC JTC1, et a été publié en juillet 2020. Il est conçu pour répondre aux exigences croissantes de la transmission vidéo à haute résolution, ainsi qu'à la demande croissante de contenus vidéo en ligne et de diffusion en continu. Il utilise des techniques de compression vidéo plus avancées, à savoir la Sélection de Transformée Multiple (MTS), la Transformée Non-Séparable Basse Fréquence (LFNST), la Ligne de Référence Multiple (MRL) et l'Arbre Quadtree avec Arbre Multi-Type Imbriqué (QTMT). Ces améliorations augmentent l'efficacité de codage du VVC par rapport à son prédécesseur, HEVC, au détriment d'une augmentation de la complexité de calcul. Le VVC peut compresser les vidéos jusqu'à deux fois plus efficacement que le H.265/HEVC, tout en

offrant une qualité d'image supérieure. VVC a été conçu pour offrir et améliorer les fonctionnalités et l'efficacité de codage pour une gamme plus étendue d'applications existantes notamment celles allant au-delà des définitions standard et haute définition, le streaming à très faible latence, le contenu d'écran, la vidéo 360°, ainsi que le support du codage multicouche dès sa première version.

### 1.6.1 Phase de normalisation

32 organisations ont soumis des propositions à la demande de propositions pour coder trois types de contenu vidéo : plage dynamique standard (SDR), plage dynamique élevée (HDR) et vidéo à 360 degrés [28]. Dans la plupart des scénarios de test, toutes les contributions ont dépassé HEVC en termes de qualité subjective, et certaines soumissions ont été supérieures à la technologie précédemment explorée dans le cadre de JEM dans un nombre significatif de situations, tel qu'indiqué dans un examen subjectif indépendant effectué en avril 2018. Le développement de VVC a commencé en avril 2018 avec la première ébauche du document de spécification et du logiciel de modèle de test, après l'examen des idées les plus performantes parmi toutes les soumissions. Suite aux demandes de propositions, où un nombre significatif d'outils de codage ont été soumis, il a été décidé d'adopter une approche "à partir de zéro". Cette première ébauche ne comprenait que la partition avancée de blocs Quad-Tree avec Multi-Type Tree (QT+MTT), qui a été identifiée comme un élément commun dans presque toutes les propositions, en particulier parce que sa mise en œuvre aurait un impact majeur sur la conception de tous les autres outils de codage basés sur les blocs. En plus de cela, des outils de codage supplémentaires provenant des réponses à la demande de propositions, ainsi que de nouveaux outils, ont été minutieusement étudiés dans des "expériences de base" en termes d'efficacité de codage et de complexité de mise en œuvre. Ces outils ont été ajoutés à l'architecture de VVC dans les cas où un compromis raisonnable entre l'efficacité de codage et la complexité a été observé.

Les fonctionnalités utilisées dans le HEVC et le VVC sont résumés dans le tableau 1.

**Tableau 1.** Récapitulatif des fonctionnalités utilisées dans HEVC et VVC [29].

Building block of the hybrid scheme	HEVC version 1	VVC version 1
Block Partitioning	64x64 max CTU size	128x128 max CTU size
	Quadtree (QT)	Quadtree & Multi-Type Tree (QT+MTT)
	Coding Units (CU)	Coding Units (CU)
	Prediction Units (PU)	Chroma Separate Tree (CST)
	Residual Quadtree	Local Dual Tree
	Transform (RQT)	Virtual Pipeline Data Units (VPDUs)
Merge Mode	Merge Mode	Extended Merge Mode and MVP
	Advanced MVP	- History-based MV Prediction (HMVP) - Pairwise Average MVP candidate - Sub-block-based Temporal MVP (SbTMVP)

Motion Compensated or Inter-Picture Prediction		- Merge with MVD (MMVD) - Symmetric MVD (SMVD)
		Adaptive MV Resolution (AMVR)
	8-taps Interpolation Filters	8-taps Interpolation Filters (IF) 1 6-tap alternative Half-pel IF
		Geometric Partitioning Mode (GPM)
		Bi-prediction with CU-level (BCW)
		Combined Intra/Inter-picture Prediction (CIIP)
		Decoder-side MV Refinement (DMVR)
		Bi-Directional Optical Flow (BDOF)
		Affine Motion
	Pred Refinement with OPTical Flow (PROF)	
Intra-Picture Prediction	33 Angles	93 Angles
	Linear interpolation	4-tap IFs (2 sets of filters)
	DC, Planar	DC, Planar
		Position-Dependent Pred. Combination (PDPC)
		Multiple Reference Lines (MRL)
		Matrix-based Intra-picture Prediction (MIP)
		Cross-Component Linear Model (CCLM)
		Intra Sub-Partitions (ISP)
Transform and Quantization	Square transforms(max 32x32)	Non-square transforms (max 64x64)
		Multiple Transform Selection (MTS)
		Non-Separable Secondary Transform (LFNST)
		Sub-Block Transform (SBT)
		Adaptive chroma QP offset
	Sign Data Hiding	Sign Data Hiding (SDH)
		Dependent Quantization (DQ)
		Joint coding of chroma residuals (JCCR)
Entropy Coding	CABAC	CABAC with high-accuracy multi-hypothesis probability estimates
	Coefficient Groups	Additional coefficient group sizes
	Reverse diagonal, hor. and ver. coefficient scan	Reverse diagonal coefficient scan only
		Improved probability model selections for absolute transform coefficient levels
In-loop Filtering	Deblocking	- Luma Mapping with Chroma Scaling (LMCS) - Deblocking Boundary Handling Modifications - Deblocking Long Filter Luma Adaptive Deblocking
	Sample Adaptive Offset (SAO)	- SAO - Adaptive Loop Filter (ALF) - Cross Component ALF (CC-ALF)
Special Modes	PCM 4x4 TS Trans. Quant. Bypass Quantization Bypass	4x4-32x32 Transform skip (TS)
Screen Content Coding		Block-Level Differential PCM (BDPCM)
		Transform Skip Residual Coding (TSRC)
		Intra-picture Block Copy (IBC)
		Palette Mode
		Adaptive Color Transform (ACT)
		MV Wrap-Around

## 1.7 Les artefacts liés à la compression vidéo

Pour obtenir une représentation codée compacte d'un signal permettant une reconstruction parfaite, il est possible d'appliquer la prédiction, la décorrélation et le codage entropique. Cependant, la quantification du signal provoque une perte d'informations qui entraîne des artefacts de compression courants.

Trois types d'artefacts peuvent apparaître dans une seule image du codage prédictif : l'effet de halo, le flou et le blocage. Ces artefacts sont considérés comme "locaux" à l'image car ils sont causés par la prédiction et la reconstruction de l'erreur de prédiction quantifiée de l'image. Ils peuvent être observés dans les schémas de codage d'images fixes ainsi que de codage vidéo. Cependant, pour les séquences vidéo, des artefacts supplémentaires peuvent se produire, entraînant une altération significative de la qualité visuelle de la vidéo reconstruite. Les mouvements saccadés peuvent également apparaître lorsque les vecteurs de mouvement codés ne correspondent pas aux mouvements observés dans la scène. Un exemple avec une capture d'écran d'une vidéo reconstruite et une illustration unidimensionnelle pour chaque artefact est présenté dans la figure 10.

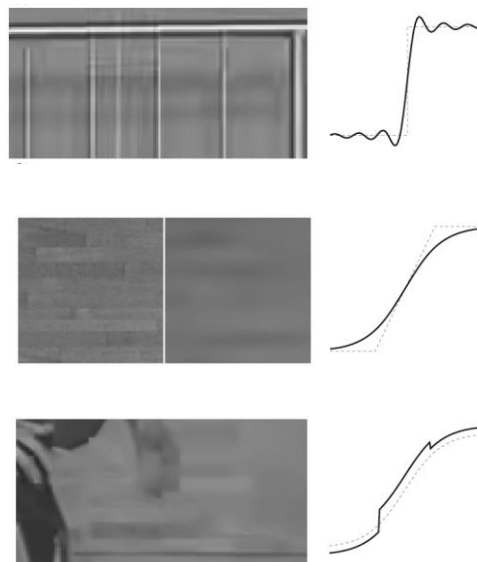


Figure 10. Exemples d'artefacts de compression dans le codage vidéo. a Effet de halo. b Flou. c Blocage [30].

Le décodage des vidéos sans artefacts est un défi majeur dans la compression vidéo, le HEVC utilise plusieurs filtres pour atteindre cet objectif. Le premier filtre est le filtre de dé-ringing (Halo), qui

atténue les anneaux de Gibbs qui se produisent autour des bords des objets lorsqu'ils sont compressés.

L'effet de halo est un artefact de compression qui survient lors de la transformation et de la quantification d'un signal. Il est causé par la perte d'informations à hautes fréquences lors de la quantification des coefficients de transformation, ce qui entraîne des artefacts dans le signal reconstruit. Ces artefacts se caractérisent par des dépassements et des sous-dépassements qui donnent l'impression d'un contour flou ou d'un halo autour des objets présents dans l'image ou la vidéo. Ce phénomène peut être observé dans différents types de codage d'images et de vidéos et peut avoir un impact négatif sur la qualité de l'image ou de la vidéo reconstruite.

Le deuxième filtre est le filtre de déblocage [31], qui réduit les blocs d'artefacts qui se produisent dans les zones de transition entre les blocs de la grille de codage. Le troisième filtre est le filtre de dé-contournement, qui lisse les contours de l'objet pour éliminer les défauts de contour. Le dernier filtre est le filtre de dé-bruitage sample adaptive offset (SAO), qui élimine les bruits de fond indésirables dans les images, améliorant ainsi la qualité de l'image globale. Ces filtres sont essentiels pour améliorer la qualité visuelle des vidéos compressées, en particulier à des taux de bits plus bas.

## 1.8 Transmission vidéo

La transmission vidéo est un processus complexe qui implique l'envoi d'une séquence d'images en mouvement à partir d'un émetteur vers un récepteur. Cette séquence d'images est généralement encodée en utilisant des normes de compression vidéo pour réduire la quantité de données à transmettre. Elle peut être utilisée dans une variété d'applications telles que la diffusion en direct d'événements sportifs, la vidéoconférence, la télévision, les systèmes de sécurité et de surveillance, la vidéo à la demande et le stockage vidéo.

La qualité de la transmission vidéo dépend de plusieurs facteurs tels que la bande passante du canal de transmission, la qualité de l'encodage vidéo, la résolution de l'image, le taux de rafraîchissement de l'image et le temps de latence entre l'émetteur et le récepteur.

La conception d'un système de transmission vidéo doit prendre en compte ces facteurs pour garantir une transmission fiable et de haute qualité de la séquence vidéo. Les aspects clés de la conception d'un système de transmission vidéo incluent l'accessibilité du signal, la robustesse du signal de transmission contre les perturbations, les distorsions et les pertes. Dans les scénarios de transmission orientés paquets, une transmission sans erreur des paquets est souvent requise. Les distorsions dans

le signal de transmission peuvent entraîner la perte des paquets concernés dans de tels systèmes. Par conséquent, des mécanismes de protection contre les pertes, et éventuellement la récupération des paquets perdus, sont souhaitables.

Dans cette optique, des normes et des protocoles de transmission vidéo ont été développés pour garantir l'interopérabilité et l'efficacité des systèmes de transmission vidéo, tels que MPEG, H.264, H.265, Real-time Transport Protocol (RTP), Real-Time Streaming Protocol (RTSP).

### 1.8.1 Streaming vidéo traditionnel

Le streaming vidéo (ou vidéo en continu) est une méthode de diffusion de vidéos sur Internet qui permet aux utilisateurs de regarder du contenu audiovisuel en temps réel, sans avoir besoin de télécharger le fichier vidéo complet sur leur appareil. Contrairement au téléchargement traditionnel, où l'intégralité du fichier doit être récupérée avant d'être lue, le streaming vidéo permet une lecture quasi instantanée grâce à la transmission continue des données vidéo. Il fonctionne selon le principe du téléchargement progressif : les vidéos sont divisées en petits paquets de données (segments) qui sont envoyés et reçus séquentiellement par les appareils des utilisateurs. Ces paquets sont stockés dans un tampon (buffer) pendant la lecture, ce qui permet de faire face aux variations de la vitesse de téléchargement et d'éviter les interruptions ou les temps de chargement.

Il existe différents mécanismes pour livrer les flux vidéo aux clients. Le streaming traditionnel utilise le protocole RTP en combinaison avec le protocole RTSP (Real Time Streaming Protocol) et fournit le contenu aussi longtemps que le client regarde la vidéo. Les inconvénients de cette méthode de livraison pour les services vidéo sur demande (VoD) sont la fiabilité, l'utilisation de l'UDP (User Datagram Protocol) comme protocole de transport et l'utilisation de serveurs de streaming propriétaires. Bien que conçu à l'origine pour le transfert de fichiers texte sur Internet, le protocole HTTP (Hyper Text Transfer Protocol) est devenu un protocole pratique pour la livraison de vidéos. Étant donné qu'il utilise le numéro de port bien connu 80, les contenus vidéo peuvent facilement traverser les pare-feux, même les plus restrictifs. Un autre avantage de ce protocole est qu'il n'exige pas de serveur de streaming dédié et utilise les mêmes serveurs pour héberger les contenus Web. Le protocole est également très fiable car il utilise le protocole TCP (Transport Control Protocol) comme protocole de transport, garantissant que les données entre le serveur et le client arriveront dans l'ordre et sans erreurs. Ce type de streaming est appelé streaming progressif car il tend à télécharger autant de la vidéo que possible. De par la nature du protocole TCP, qui a tendance à utiliser au maximum la bande passante disponible, les clients téléchargent une partie significativement plus importante de la vidéo au tout début du transfert. Par conséquent, il est très

courant que la vidéo soit entièrement téléchargée au moment où le client n'a visionné qu'une petite partie. Si le client décide d'arrêter de regarder, les données téléchargées restantes sont gaspillées. Un autre inconvénient vient du fait que la vidéo téléchargée peut être distribuée ultérieurement par le client, ce qui pose un problème de droits d'auteur. La figure 11 montre un Exemples d'un réseau streaming serveur client.

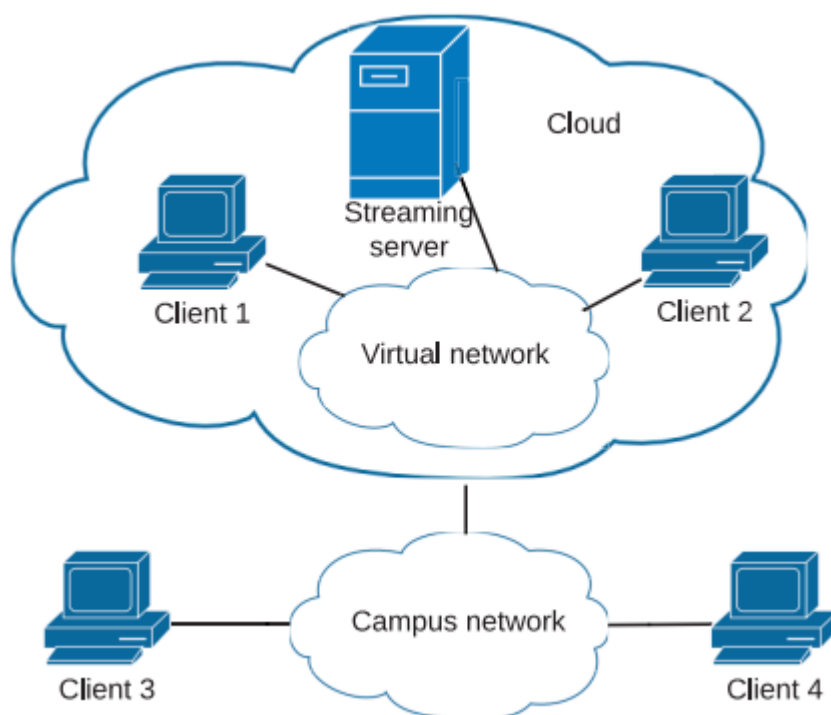


Figure 11. Exemples d'un réseau streaming serveur client [32].

### 1.8.2 Le streaming vidéo pair à pair (P2P)

Est une méthode de diffusion de contenu en continu qui repose sur un réseau décentralisé d'ordinateurs connectés les uns aux autres. Contrairement au streaming traditionnel, où le contenu est diffusé à partir d'un serveur centralisé vers les utilisateurs, le streaming P2P permet à chaque utilisateur de devenir à la fois un client et un fournisseur de contenu. Dans un réseau de streaming P2P, chaque utilisateur appelé « pair » qui visionne une vidéo en streaming télécharge simultanément de petits morceaux de contenu (paquets de données) à partir d'autres utilisateurs qui ont déjà visionné ou qui sont en train de visionner la même vidéo. Une fois ces paquets téléchargés, l'utilisateur peut les visionner tout en les partageant à son tour avec d'autres utilisateurs qui souhaitent accéder au même contenu. le streaming d'une vidéo en couches sur une architecture P2P est une approche prometteuse pour la diffusion de vidéos à grande échelle à un grand nombre de récepteurs hétérogènes [33] . Un exemple d'une telle architecture est présenté dans la Figure 12. Dans cet

exemple, la source vidéo diffuse une vidéo en couches composée de trois couches. Les pairs s'abonnent à un certain nombre de couches en fonction de leurs capacités, et chaque pair récepteur devient à son tour pair émetteur en fournissant une ou plusieurs couches.

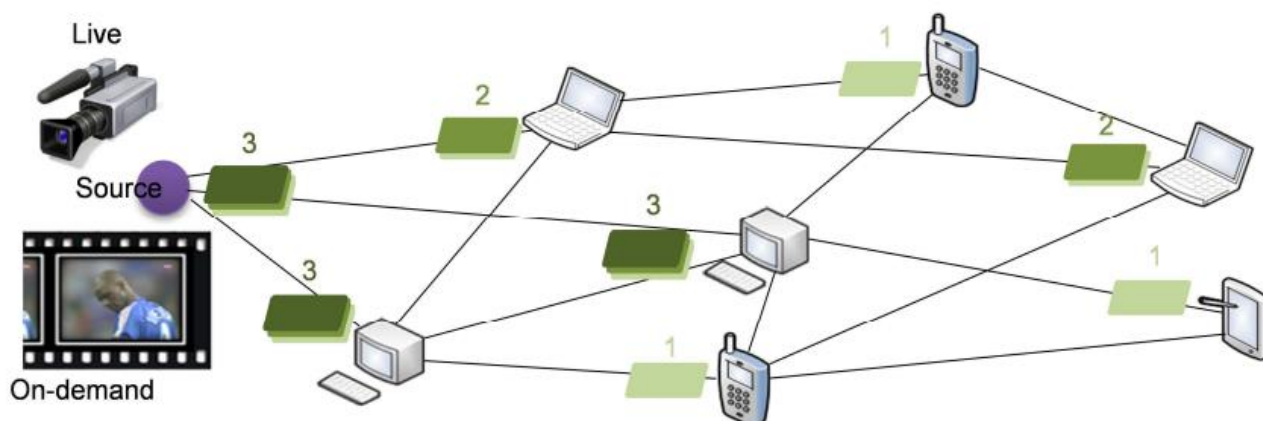


Figure 12. Exemple d'architecture de streaming P2P en couches [34].

## 1.9 Conclusion

En conclusion, ce chapitre a présenté les principes de base de la compression vidéo, qui sont essentiels pour comprendre le fonctionnement des techniques de compression vidéo modernes telles que le standard HEVC et le VVC. Nous avons commencé par expliquer la différence entre la vidéo brute et la vidéo compressée, puis nous avons examiné les concepts clés tels que la redondance spatiale, temporelle et perceptuelle, ainsi que les différentes étapes du processus de compression vidéo, y compris l'encodage et le décodage. Nous avons également abordé les différents types de compression vidéo, tels que la compression sans perte et la compression avec perte, ainsi que les artefacts liés à la compression vidéo tel que l'effet de halo.

# Chapitre 2

## 2. Problématique des transmissions vidéo sur les réseaux avec perte

### 2.1. Introduction

Depuis plus de deux décennies, la transmission vidéo sur des canaux sujets à des pertes de données est devenue un problème majeur. Les exigences de bande passante nécessaires pour transmettre des vidéos brutes rendent cette méthode impraticable, obligeant ainsi à utiliser la compression vidéo. Toutefois, la vidéo compressée est très vulnérable aux pertes de paquets qui sont courantes sur les réseaux "best-effort" tels que l'Internet. Ces pertes peuvent être causées par divers facteurs tels que la congestion, les délais, la bande passante limitée, etc. La Figure 13 illustre un modèle de distorsion dans un canal de transmission avec pertes.

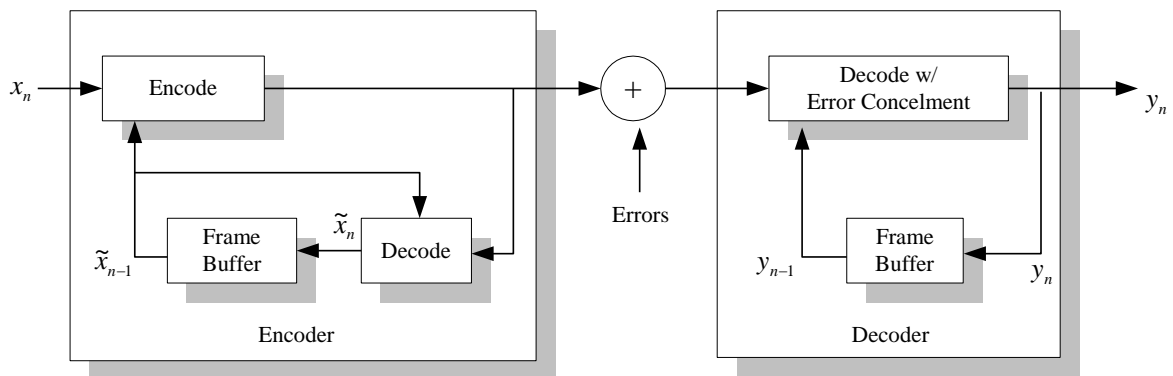


Figure 13. Schéma de principe de l'encodeur/décodeur

Les frames de référence au niveau du codeur sont restitués comme suit :

$$\tilde{x}_n = \begin{cases} [x_n - \tilde{x}_{n-1}(mv) + q_r] + \tilde{x}_{n-1}(mv) & INTER \\ x_n + q_a & INTRA \end{cases} \quad (1)$$

Où

$q_r$  : Résidu inter frame

$q_a$  : Résidu intra frame

Le processus de décodage est fondamentalement l'inverse du processus de codage :

$$y_n = \begin{cases} [x_n - \tilde{x}_{n-1}(mv) + q_r] + y_{n-1}(mv) & INTER \\ x_n + q_a & INTRA \end{cases} \quad (2)$$

Toutefois, si des erreurs sont détectées, le décodeur doit passer en mode de dissimulation d'erreurs. En supposant une simple dissimulation d'erreur temporelle, le pixel décodé,  $y_n$ , peut être modélisé comme suit :

$$y_n = y_{n-1} \quad (3)$$

On peut considérer qu'il existe une différence entre  $\tilde{x}_{n-1}$  et  $y_{n-1}$  qui sont les images de référence pour la compensation de mouvement dans l'encodeur et le décodeur, respectivement. Nous définissons cette différence comme le bruit de dérive (dû à la perte de paquets) :

$$d_{n-1} = y_{n-1} - \tilde{x}_{n-1} \quad (4)$$

Dans un canal sujet aux erreurs, en supposant que la dérive et le bruit de quantification ne sont pas corrélés, la distorsion globale entre les unités décodés et originaux peut être exprimée comme suit :

$$\begin{aligned} D_{total}^n &= E(x_n - y_n)^2 \\ &= E(x_n - \tilde{x}_n + \tilde{x}_n - y_n)^2 \\ &= E(x_n - \tilde{x}_n)^2 + (\tilde{x}_n - y_n)^2 = D_q^n + D_{drift}^n \end{aligned} \quad (5)$$

Où  $D_q^n$  et  $D_{drift}^n$  représentent la distorsion due à la quantification et le bruit de dérive, respectivement. C'est-à-dire,

$$\begin{aligned} D_q^n &= E(x_n - \tilde{x}_n)^2 \\ D_{drift}^n &= E(\tilde{x}_n - y_n)^2 \end{aligned} \quad (6)$$

Dans ce chapitre, nous allons aborder la problématique des services multimédia plus précisément la vidéo sur des réseaux peu fiables tel que le réseau IP (Internet Protocol). La chaîne de diffusion vidéo se compose principalement de deux services : la compression et la transmission, qui sont les sources de distorsions dues à l'utilisation de la quantification et des réseaux de type "best-effort" respectivement.

La qualité du service vidéo peut être jugée à l'aide de deux facteurs : le facteur technique donne une indication objective sur l'état du réseau, tant dis que le facteur humain reflète le degré de satisfaction de l'utilisateur final du service multimédia. Le facteur technique est appelé qualité de service (QoS) tandis que le facteur humain est appelé qualité d'expérience (QoE). La figure 14 illustre la principale différence entre les deux termes. Les sous-sections suivantes présentent un bref aperçu de chaque

terme. Nous revoyons aussi dans ce chapitre l'influence des erreurs de transmission sur un flux H.265/HEVC autour du quel est concentré notre travail.

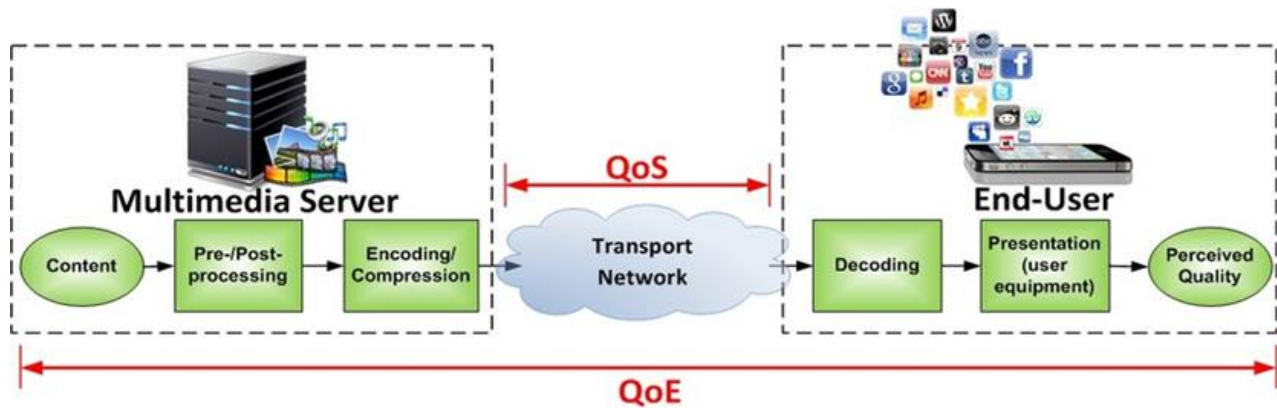


Figure 14. QoS vs. QoE dans la diffusion de contenu multimédia [35].

## 2.2. Qualité de service (QoS)

La notion de la QoS a été proposée par l'Union internationale des télécommunications (UIT) en 1994. La mesure de la QoS de bout en bout joue un rôle indispensable dans la prise de décision des services cloud et des services Internet des objets (IoT). En effet, elle fournit une indication objective de la performance d'un réseau en fonction des caractéristiques suivantes :

- **Le délai ou bien la latence** : Le temps écoulé lors du passage d'un paquet réseau de la source jusqu'à la destination, également appelé latence, désigne essentiellement les délais perceptibles par les utilisateurs. Ces délais sont généralement causés par un ralentissement du paquet au niveau des routeurs en raison d'une faible bande passante. L'objectif de la gestion de la QoS est d'optimiser les ressources d'un réseau et de garantir de bonnes performances aux applications telles que la vidéo et l'audio pour améliorer le service client selon les besoins. Pour un même service Web, différents utilisateurs auront des expériences différentes en fonction de leurs préférences respectives, ce qui entraîne des évaluations de la qualité de service variées [36].
- **La gigue** : c'est une variation irrégulière et imprévisible du temps de transit des paquets de données sur un réseau. Elle mesure les écarts de délais entre l'envoi et la réception des paquets d'information. Comme l'illustre la figure 15.

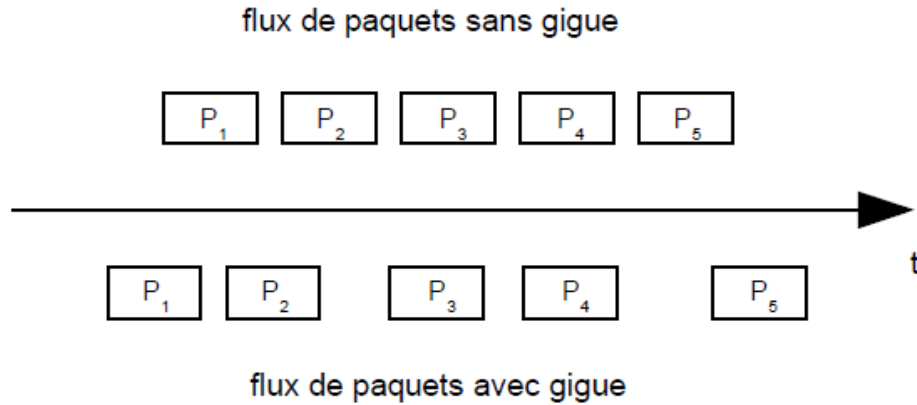


Figure 15. Illustration du phénomène de gigue. Les flux du haut et du bas ont respectivement une distance inter-paquet constante et variable.

- **La perte de paquets** : La perte de paquets illustré dans la figure 16 se produit lorsque certains paquets de données ne parviennent pas à destination ou sont abandonnés en cours de route. Dans le cas de la transmission vidéo, cela peut entraîner une diminution de la qualité de l'image, des artefacts et des sauts dans la vidéo, ainsi qu'une perte de synchronisation audio-vidéo. On peut mesurer la perte de paquets sous forme de taux de perte de frames (PLR), qui représente le pourcentage de trames qui auraient dû être transmises par un réseau, mais qui ont été perdues en cours de route. Ce taux est largement utilisé pour évaluer la qualité de la transmission de données, notamment pour les applications en temps réel telles que la vidéo et la voix sur IP. Si le taux de perte de trames est élevé, cela peut causer des interruptions dans la transmission de données, ce qui peut entraîner une dégradation de la qualité de la vidéo ou du son. En règle générale, une perte de paquets aléatoire inférieure à 1 % est considéré comme étant "bonne" pour la diffusion en continu d'audio ou de vidéo, et de 1 à 2,5 % comme étant "acceptable" [37].

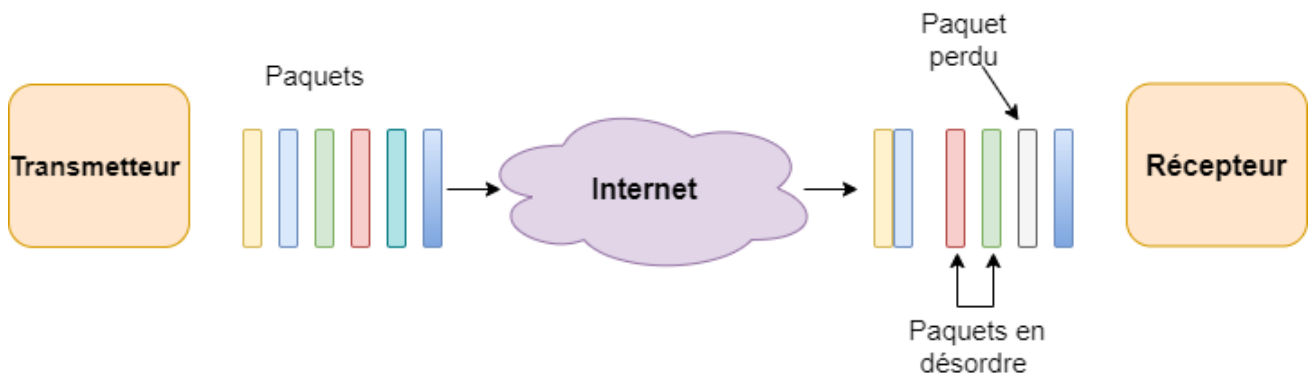


Figure 16 .Illustration du phénomène de la perte de paquets

### 2.2.1. Effets perceptuels de la perte de paquets

De nombreuses études sont effectuées pour analyser l'effet de la perte de paquets sur la qualité vidéo qui survient lors de la transmission de cette dernière dans un canal peu fiable. Selon l'étude [38] plusieurs paramètres entrent en jeu pour varier l'effet de la perte de paquets sur la vidéo à savoir : l'emplacement de la perte dans le bit Stream, les paramètres de codage, les caractéristiques de la vidéo elle-même et la stratégie de dissimulation d'erreurs. Dans [39], les auteurs ont démontré que l'effet de la perte est plus visible lorsque celle-ci affecte un frame de référence.

### 2.2.2. Techniques pour améliorer la QoS

Il existe plusieurs techniques qui peuvent être utilisées pour garantir une bonne performance des applications les plus sensibles telles que les applications en temps réel qui sont considérées comme un vrai challenge en raison de leur sensibilité accru à la perte de paquets ou à la gigue. Si un paquet est perdu ou retardé lors d'une communication vidéo en temps réel l'utilisateur final peut très vite apercevoir une gigue.

Parmi ces techniques il y a :

**Hiérarchisation du trafic voix sur IP (VoIP, en anglais) sensible aux délais via les routeurs :** lors d'une situation de forte congestion, la hiérarchisation permet de classer le trafic et de lui attribuer différentes priorités en fonction de son type. Les paquets classés prioritaires sont envoyés en premier

**Réservation de ressources :** La réservation de ressources permet de diviser les ressources du réseau en fonction du trafic de différents types et origines, de définir des limites et de garantir la bande passante. Le protocole de réservation de ressources (RSVP, en anglais) est un protocole de la couche transport qui assure cette réservation.

**Mise en file d'attente :** La mise en file d'attente est le processus de création de politiques qui fournissent un traitement préférentiel à certains flux de données par rapport à d'autres. Les files d'attente sont des tampons de mémoire de haute performance dans les routeurs et les commutateurs, dans lesquels les paquets qui passent sont conservés dans des zones de mémoire dédiées. Lorsqu'un paquet se voit attribuer une priorité plus élevée, il est déplacé vers une file d'attente dédiée qui achemine les données à un rythme plus rapide, ce qui réduit les risques de perte.

**Marquage du trafic :** Lorsque les applications qui nécessitent une priorité sur la bande passante du réseau ont été identifiées, le trafic doit être marqué. Cela est possible grâce à des processus tels que

la classe de service, qui marque un flux de données dans l'en-tête de la trame de la couche 2, et le point de code des services différenciés, qui marque un flux de données dans l'en-tête du paquet de la couche 3.

### 2.3. Qualité d'expérience (QoE)

Se réfère à l'expérience effective de l'utilisateur sur la qualité de la vidéo diffusée. C'est une mesure subjective de la qualité de la diffusion de vidéo qui prend en compte la satisfaction de l'utilisateur et la perception de la qualité de la vidéo.

La qualité d'expérience est définie par l'IUT-T étant la mesure globale de l'acceptabilité d'une application ou d'un service, tel qu'il est perçu subjectivement par l'utilisateur final. C'est un concept essentiel pour les systèmes de communication vidéo.

En conclusion, le concept de qualité d'expérience décrit la qualité subjective perçue par les utilisateurs finaux par rapport à un service, habituellement un service vidéo. Il peut offrir un indicateur complémentaire orienté utilisateur aux mesures objectives de la QoS.

#### 2.3.1. Les améliorations de la QoE dans le streaming vidéo

Elles peuvent inclure :

**Qualité d'image** : Une qualité d'image élevée est importante pour une expérience de streaming de qualité. Les algorithmes de compression vidéo tels que le VVC et HEVC peuvent aider à améliorer la qualité d'image en réduisant les erreurs de compression.

**Stabilité de la lecture** : Une lecture stable est essentielle pour une expérience de streaming de qualité. Les algorithmes de dissimulation d'erreurs peuvent aider à masquer les erreurs qui peuvent se produire lors de la transmission de la vidéo sur un réseau.

**Latence** : La latence est le temps qui s'écoule entre le moment où la vidéo est diffusée et le moment où elle est affichée sur l'écran. Une diffusion en direct avec une faible latence est importante pour une expérience de streaming de qualité.

**Stabilité du réseau** : désigne la capacité d'un système de réseau à fournir un service de manière fiable et sans interruption. Cela comprend la capacité du réseau à être accessible et à fonctionner correctement en tout temps pour les utilisateurs. C'est un élément clé pour garantir une expérience de streaming de vidéo fluide et sans défauts pour les utilisateurs.

**Facilité d'utilisation** : La facilité d'utilisation des plateformes de streaming est importante pour une expérience de streaming de qualité. Elles doivent être simples à utiliser et intuitives pour les utilisateurs.

### 2.3.2. Moyens de mesurer la QoE

L'évaluation de la qualité d'expérience (QoE) dans le streaming vidéo peut se faire en plusieurs manières : Évaluation subjective qui est une évaluation de la qualité d'expérience basée sur les commentaires et les perceptions des utilisateurs. Cela peut inclure des enquêtes et des tests de qualité de l'expérience. D'autres part, l'évaluation objective c'est une évaluation de la qualité d'expérience basée sur des mesures objectives telles que la qualité d'image, la stabilité de la lecture, la latence, la disponibilité et la facilité d'utilisation de la plateforme de streaming.

#### 2.3.2.1. Méthodes subjectives

Les tests subjectifs d'évaluation de la qualité de streaming vidéo suivent des protocoles standardisés élaborés par le corps ITU-T [40-42]. Les protocoles incluent l'environnement d'enregistrement du signal source ainsi que les caractéristiques de la scène ou encore les conditions d'évaluation tel que la luminosité de la pièce et la distance de visionnage. Ces tests subjectifs permettent de mesurer l'impact subjectif de différents facteurs tels que la résolution, la fluidité, les défauts visuels, etc.

Les sujets évaluent les données visuelles de différentes manières en fonction de la méthode de test utilisée. Par exemple, la notation de catégorie absolue (ACR, Absolute Category Rating) est une méthode subjective de mesure de la qualité vidéo qui utilise une échelle de notation pour évaluer la qualité de la vidéo. Cette méthode se concentre sur la façon dont les utilisateurs perçoivent la qualité de la vidéo et consiste à demander aux utilisateurs de noter la qualité de la vidéo à l'aide d'une échelle prédéfinie.

L'échelle peut varier en fonction de l'application, mais généralement, elle comporte un nombre de catégories distinctes (par exemple, excellent, très bon, bon, moyen, etc.). Les utilisateurs sont invités à noter la qualité de la vidéo en choisissant la catégorie qui leur semble la plus appropriée. Des exemples de plusieurs types d'échelles sont donnés figure 17.

- **Échelle comparative** : Cette échelle utilise des termes relatifs pour évaluer la qualité de la vidéo, tels que "meilleure" ou "moins bon".
- **Échelle catégorielle qualitative** : Cette échelle utilise des termes qualitatifs pour décrire la qualité de la vidéo, tels que "excellent", "bonne", "mauvaise", etc.

- **Échelle discrète correspondante** : Cette échelle associe une valeur numérique à chaque catégorie de l'échelle catégorielle qualitative pour en faciliter la quantification et l'analyse.

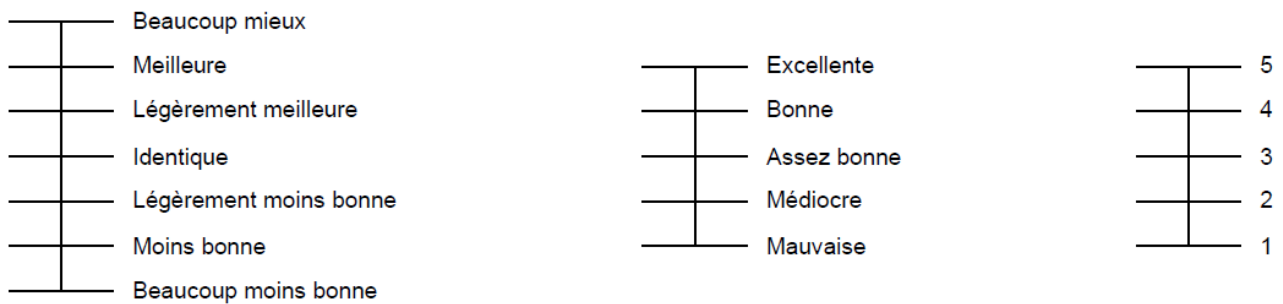


Figure 17. Des exemples d'échelles utilisées lors des tests subjectifs d'évaluation de qualité. De gauche à droite : une échelle comparative, une échelle catégorielle qualitative et son échelle discrète correspondante.

La notation de catégorie absolue est souvent utilisée pour évaluer la qualité de la vidéo en situation réelle, où les utilisateurs peuvent visualiser la vidéo dans des conditions similaires à celles d'utilisation normale. Cette méthode peut être utile pour comprendre les préférences et les attentes des utilisateurs en matière de qualité vidéo et pour évaluer les impacts subjectifs de différents facteurs sur la qualité perçue de la vidéo.

Les méthodes subjectives de mesure de la qualité vidéo présentent plusieurs limitations :

- **Biais subjectif** : les opinions des utilisateurs peuvent être influencées par leur expérience personnelle, leurs préférences et leurs attentes, ce qui peut conduire à des biais subjectifs dans les résultats.
- **Variabilité** : les perceptions et les opinions des utilisateurs peuvent varier considérablement, ce qui peut rendre difficile la mesure de la qualité vidéo de manière fiable.
- **Coût et complexité** : les méthodes subjectives de mesure de la qualité vidéo peuvent être coûteuses et complexes à mettre en place et à interpréter. En effet, ils nécessitent au moins 15 personnes ayant une bonne acuité visuelle pour chaque série de contenus multimédias évaluée.
- **Bruit** : les méthodes subjectives sont soumises à un certain nombre de sources de bruit, telles que les erreurs de mesure, les erreurs de codage, les défauts de la vidéo, etc., qui peuvent rendre difficile l'interprétation des résultats.

- **Temps** : les méthodes subjectives nécessitent souvent plus de temps pour obtenir des résultats fiables que les méthodes objectives, ce qui peut rendre difficile leur utilisation pour des tâches de mesure de la qualité rapides.

Par conséquent, les méthodes subjectives de mesure de la qualité vidéo peuvent être utiles pour comprendre les préférences et les attentes des utilisateurs en matière de qualité vidéo, mais elles ne doivent pas être utilisées seules pour évaluer la qualité vidéo. Il est souvent préférable de les combiner avec des méthodes objectives pour obtenir une évaluation complète de la qualité vidéo.

#### *2.3.2.2. Méthodes objectives*

Les métriques objectives de mesure de la qualité vidéo sont des algorithmes qui permettent d'évaluer la qualité d'une vidéo de manière quantifiable. Elles se basent sur des critères techniques tels que la résolution d'image, la qualité du mouvement, le taux de compression, le rapport signal sur bruit, etc. Elles permettent de faire des comparaisons précises entre différentes vidéos et de déterminer objectivement la qualité de la vidéo. Ainsi, Elles peuvent être utilisées pour mesurer la qualité de la vidéo au cours du processus de production, pour évaluer les différences entre différents codecs vidéo, ou pour comparer les performances de différents systèmes de diffusion vidéo. Les métriques objectives de mesure de la qualité vidéo sont essentielles pour les professionnels du secteur de la vidéo, car elles leur permettent d'améliorer la qualité de la vidéo et de s'assurer que les vidéos sont diffusées avec la meilleure qualité possible.

Parmi les métriques objectives pour mesurer la qualité vidéo les plus couramment utilisés on trouve :

**PSNR (Peak Signal-to-Noise Ratio)** : mesure la qualité d'une vidéo en comparant les pixels originaux à ceux dans la vidéo codée.

**SSIM (Structural Similarity Index Measure)** : une métrique qui mesure la similitude structurelle entre deux images pour évaluer la qualité de la vidéo.

**VMAF (Video Multimethod Assessment Fusion)** : une métrique qui combine plusieurs métriques pour donner une évaluation complète de la qualité vidéo.

## 2.4. Relation entre la QoS et la QoE

Il existe plusieurs modèles de corrélation entre la QoS (Quality of Service) et la QoE pour les services multimédias. Voici un exemple de modèle fréquemment utilisé :

**Modèle de la chaîne de causalité :** Ce modèle considère que la QoS est un prédicteur direct de la QoE. Selon ce modèle, une bonne QoS (par exemple, une bande passante élevée et une faible latence) entraîne une bonne QoE (par exemple, une vidéo en streaming fluide et de bonne qualité).

**Modèle multi-facteur :** Ce modèle considère que la QoE est influencée par plusieurs facteurs, tels que la QoS, les attentes des utilisateurs, le contenu de la vidéo, etc. Selon ce modèle, une bonne QoS peut améliorer la QoE, mais elle n'est pas la seule variable déterminante.

**Modèle de l'adaptation :** Ce modèle considère que les utilisateurs s'adaptent aux conditions de qualité de la vidéo au fil du temps. Selon ce modèle, une mauvaise QoS peut entraîner une mauvaise QoE au départ, mais les utilisateurs peuvent s'y habituer avec le temps.

Dans ce contexte, il y a plusieurs travaux dans la littérature qui ont été élaborés comme la métrique audio-visuelle propriétaire V-Factor<sup>2</sup> établit une corrélation entre les paramètres de la QoS et leur impact sur la QoE. Cette métrique est inspirée du modèle du système visuel humain (SVH). Le modèle prend en compte les aspects spatiaux et temporels de la vision humaine et leur interaction.

Dans [43], un modèle pour la corrélation entre la QoS et la QoE d'un service multimédia est proposé. Ce modèle inclut les paramètres suivants : délai, la gigue, taux de perte, taux d'erreur et la bande passante. Cependant, le modèle n'a pas été validé par des tests expérimentaux, donc nous ne pouvons pas être certains de son efficacité.

Il est important de noter que ces modèles ne sont que des approches générales de la corrélation entre la QoS et la QoE. La corrélation exacte peut varier en fonction de différents contextes et des préférences uniques des utilisateurs. Il est donc important d'effectuer des études subjectives et objectives pour évaluer la corrélation entre la QoS et la QoE pour un service multimédia particulier.

---

<sup>2</sup> <http://www.symmetricom.com/products/qoe-assurance/v-factor/>



vidéo sur les réseaux IP. Ce protocole permet une diffusion fluide des médias sur les réseaux, en assurant une transmission efficace des paquets.

L'en-tête d'unité NAL est optimisé pour être facilement traité par les MANEs (media-aware network elements), ces éléments réseau conscients des médias. La RFC 7798 [44] définit les différentes méthodes d'encapsulation des unités de transport du flux HEVC dans les paquets RTP, permettant ainsi une meilleure gestion et transmission des flux vidéo HEVC sur les réseaux. En utilisant l'en-tête d'unité NAL en conjonction avec RTP, les flux vidéo HEVC peuvent être acheminés de manière plus efficace et fluide, ce qui est essentiel pour les applications de streaming en temps réel. Cette combinaison de protocoles facilite la diffusion de contenus multimédias tout en optimisant les performances et la qualité de l'expérience utilisateur.

RTP utilise le port UDP (User Datagram Protocol) pour le transport des paquets de données. UDP est un protocole de transport non fiable. Cela signifie que les paquets RTP envoyés via UDP ne sont pas garantis d'arriver à destination, et il n'y a aucune garantie quant à l'ordre dans lequel ils arriveront. Cependant, malgré son manque de fiabilité, UDP est préféré pour le transport de données en temps réel, comme l'audio et la vidéo, car il offre une latence plus faible et une vitesse de transmission plus rapide par rapport à TCP. Dans les applications de streaming multimédia en temps réel, la priorité est souvent donnée à la rapidité de la transmission plutôt qu'à la fiabilité, car quelques paquets perdus peuvent être mieux tolérés que des délais de transmission plus longs.

La figure 19 illustre un scénario de transmission vidéo HEVC typique impliquant des MANEs (media-aware network elements). Dans ce cas, le MANE, agissant en tant que "passerelle media", est situé entre l'expéditeur et le destinataire du flux vidéo. Pour optimiser la gestion et la manipulation du flux vidéo, le MANE doit identifier le type d'unité NAL (Network Abstraction Layer) contenu dans les paquets RTP. En effet, connaître le type d'unité NAL est crucial pour le MANE, car cela lui permet de prendre des décisions éclairées concernant le traitement des flux vidéo. Par exemple, il peut détecter des unités NAL critiques et décider de les dupliquer afin d'améliorer la fiabilité de la transmission. Ces unités NAL critiques pourraient contenir des informations essentielles pour la qualité de l'expérience utilisateur, et leur duplication vise à réduire les risques de pertes de paquets qui pourraient altérer la qualité globale de la vidéo.

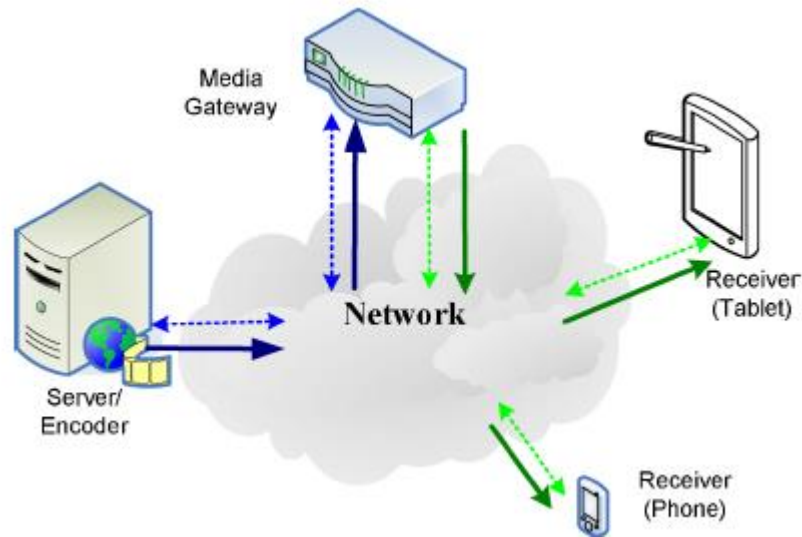


Figure 19. Scénario de transmission vidéo HEVC

### 2.5.2. Vulnérabilité du flux HEVC aux erreurs de transmission

Bien que la norme HEVC représente une progression significative en matière de compression par rapport à ses prédécesseurs, cela se fait au prix d'une augmentation de l'activité d'inter-prédiction et de compensation de mouvement lors du processus de codage. Cette stratégie consiste principalement à augmenter l'efficacité de codage en utilisant plus d'informations redondantes temporelles et spatiales[45]. Toutefois, il est important de noter que cette haute efficacité de codage entraîne une complexité de calcul élevée. En d'autres termes, un flux de bits hautement compressé contient davantage d'informations vidéo redondantes encodées, ce qui rend le contenu vidéo compressé plus sensible aux erreurs de bits de canal que les flux de bits encodés avec les normes précédentes.

En HEVC, deux techniques sont utilisées pour améliorer le codage prédictif des vecteurs de mouvement. La première est la prédiction avancée des vecteurs de mouvement (AMVP), qui signale au décodeur le meilleur prédicteur pour chaque bloc de mouvement. La seconde est la fusion de blocs de prédiction inter, qui permet de dériver toutes les données de mouvement d'un bloc à partir des blocs voisins. Cette technique remplace les modes directs et skip utilisés dans H.264/AVC

L'utilisation de la prédiction avancée des vecteurs de mouvement (MV) dans HEVC contribue considérablement à augmenter les dépendances spatiales et temporelles. Cette technique utilise des MV prédits temporellement, On obtient donc une grande efficacité de compression, mais la résistance aux erreurs de transmission est pénalisée. En outre, l'utilisation de la prédiction avancée

des vecteurs de mouvement dans HEVC [4] [46] contribue également à augmenter les dépendances spatiales et temporelles dans une large mesure. Bien que les candidats MV spatiaux soient plus souvent utilisés, les MV prédits temporellement ont un impact plus important sur la propagation des erreurs.

Dans [45] et [47] a été rapporté qu'en dépit de son efficacité de codage supérieure, le HEVC présente en général une mauvaise performance de résilience aux erreurs. Ainsi, lorsque les flux de bits HEVC subissent des pertes de transmission, cela entraîne une dégradation significative de la qualité à la fois objective et subjective. Il a été aussi rapporté dans [48] que L'augmentation de la dépendance temporelle dans HEVC réduit sa capacité à être résistant aux erreurs par rapport à H.264/AVC.

## 2.6. Conclusion

En conclusion, la transmission de vidéo sur un réseau avec perte de paquets peut entraîner une dégradation significative de la qualité de la vidéo. La perte de paquets peut être mesurée sous forme de taux de perte de paquets, qui indique le pourcentage de paquets perdues lors de la transmission. Un PLR élevés peut entraîner des interruptions dans la transmission de données, ce qui peut causer une détérioration de la qualité de la vidéo. Bien que des techniques telles que la prédiction avancée des vecteurs de mouvement dans HEVC puisse contribuer à améliorer la compression des vidéos, elles augmentent également les dépendances spatiales et temporelles, ce qui réduit sa résilience aux erreurs.

### 3. État de l'art sur les mécanismes d'amélioration de la qualité du streaming vidéo

#### 3.1. Introduction

Lorsqu'une vidéo est transmise à travers un canal bruyant, la qualité peut être considérablement altérée, entraînant une dégradation de la qualité d'expérience de l'utilisateur. Pour pallier à ce problème, plusieurs mécanismes d'amélioration de la qualité vidéo ont été développés. Parmi ces mécanismes, on peut citer le codage à plusieurs descriptions qui permet de transmettre plusieurs versions de la même vidéo à différents débits ou résolutions pour une adaptation de la qualité en fonction des capacités du canal de transmission. Un autre mécanisme important est la dissimulation d'erreur, également appelée 'error concealment' en anglais, qui permet de masquer les erreurs de transmission et de réduire leur impact sur la qualité de la vidéo. Ainsi, différents algorithmes ont été développés pour reconstruire les images manquantes ou endommagées afin de minimiser les effets négatifs sur l'expérience utilisateur. Dans ce chapitre, nous explorerons l'état de l'art des mécanismes d'amélioration de la qualité du streaming vidéo.

#### 3.2. Codage résilient aux erreurs

##### 3.2.1. Codage source

**Partitionnement de données** : plusieurs codecs utilisent cette technique tels que H.263 [49], MPEG-4 [50] et H.264/AVC [51, 52]. Les flux binaires ne portent pas systématiquement la même quantité de donnée et n'ont pas la même importance. Le H.264 permet le codage des slices en 3 parties A, B et C comme l'indique la figure 20. Chaque partie est associée à une protection aux erreurs adaptée pendant la transmission. Chaque partie peut contenir les informations d'entête, les blocs codés en inter et en intra et les vecteurs de mouvement. Il faut noter que cette technique n'est pas prise en charge dans le HEVC [53].

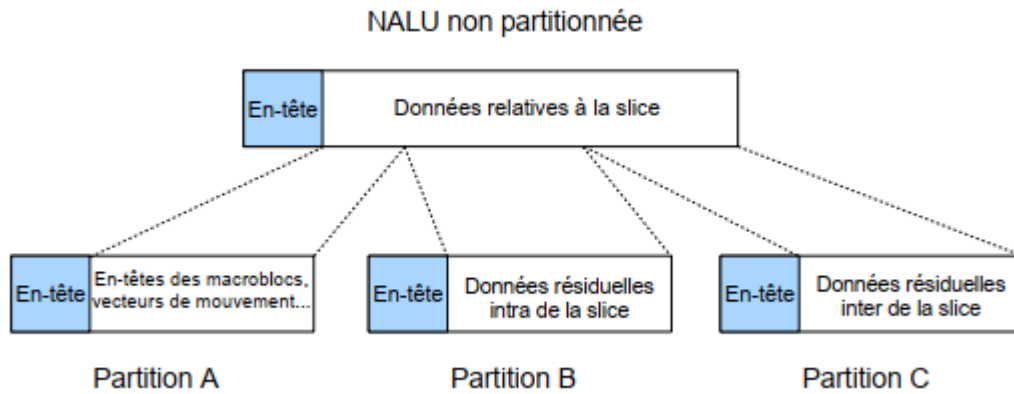


Figure 20. Le partitionnement de données de H.264/AVC [54].

**Le découpage d'image (Slicing) :** Le concept du slicing ou l'image est découpée en plusieurs tranches (slices) a été introduit dans plusieurs codecs tels que le H264 ou bien le HEVC. La figure 21 montre le découpage en slice de la séquence Parkjoy. Ce concept permet de réduire la longueur des paquets transmis en diminuant le nombre de CTU encodées dans chaque slice, ce qui réduit la propagation des erreurs au niveau du décodeur. Toutefois, lorsque le nombre de slices encodées augmente, cela a un impact négatif sur l'efficacité du codage. Aux limites des tranches, les prédictions intra et de mouvement sont interdites, ce qui limite la capacité de prédiction spatiale de l'image. De plus, cela entraîne une augmentation de la surcharge due à la structure des tranches.





(c)

Figure 21. Découpage en slice d'un frame de la séquence Parkjoy, (a) 2 slices, (b) 4 slices, (c) 6 slices.

Dans [55], l'auteur a utilisé le concept du découpage en tranche comme un outil de résilience aux erreurs de transmission. Cette technique a donné un résultat satisfaisant pour la séquence Foreman, et un bon résultat subjectif par rapport à la technique intra macroblock refresh (IMR) pour la séquence paris.

T. Alfaqheri et al [45], a proposé un algorithme introduit dans l'encodeur dont l'objectif est de réduire la propagation des erreurs au niveau des tranches pour prendre en charge des applications de diffusion vidéo à faible délai. Cet algorithme extrait les slices qui contiennent la région d'intérêt pour la coder en mode intra.

**Tranches redondantes :** L'idée est d'avoir une tranche principale codée et une ou plusieurs représentations de tranches redondantes de différentes qualités. En cas de perte de la tranche principale, le décodeur peut reconstruire la tranche perdue à l'aide des tranches redondantes de qualité inférieure [56].

**Rafraîchissement intra :** cette technique vise à réduire l'impact des erreurs en insérant périodiquement des images intra (I-frames) dans le flux vidéo compressé. Les images intra sont des images codées de manière indépendante, ce qui signifie qu'elles ne dépendent pas des images précédentes pour leur reconstruction. En insérant régulièrement des images intra, l'effet des erreurs est limité à la zone entourant chaque image intra, plutôt que de se propager sur plusieurs images consécutives. Cela permet au décodeur de récupérer plus facilement des erreurs en reconstruisant les images manquantes à partir des images intra. Elle est souvent utilisée dans les applications de diffusion vidéo en direct ou à faible délai, où la correction rapide des erreurs est essentielle pour

garantir une expérience de visionnage fluide. Cette technique contribue à améliorer la résilience du flux vidéo face aux erreurs de transmission, offrant ainsi une meilleure qualité visuelle.

Dans [57], l'auteur a utilisé un rafraichissement intra adaptatif en se basant sur le mécanisme de quantification pour améliorer la qualité de la vidéo reçue de type « screen content » largement utilisée dans les téléconférences et le partage d'écran. La méthode a permis une amélioration significative de la qualité subjective de la vidéo reçue. [58] a proposé un rafraichissement intra adaptatif basé sur les régions d'intérêts. Les résultats expérimentaux démontrent que la méthode proposée est nettement meilleure que les schémas existants en termes d'économie de bits et de qualité visuelle perçue. Dans [59], les performances de plusieurs structures de codage en présence de différentes techniques de codage intra ont été testé. Les résultats pour les séquences vidéo moins actives ont montré que la structure de codage IBP avec une période intra de 18 obtenait les meilleures performances à des taux de perte plus faibles.

**Codage à descriptions multiples :** Le codage à descriptions multiples est une technique avancée utilisée dans la compression vidéo pour améliorer l'efficacité de la transmission. Contrairement aux approches traditionnelles qui se contentent d'encoder une seule description du contenu, le codage à description multiple permet de générer deux [60] ou plusieurs descriptions indépendantes du même contenu. Chaque description, appelée couche ou canal, peut être encodée à des débits et avec des paramètres différents. L'idée principale derrière le codage à description multiple est d'adapter la transmission et la restitution du contenu en fonction des contraintes du réseau, des capacités des appareils et des préférences de l'utilisateur. Lors de la transmission, les différentes descriptions peuvent être envoyées simultanément ou séparément, offrant ainsi une certaine résilience aux erreurs de transmission et une adaptabilité aux variations du réseau. Du côté de la restitution, le récepteur peut sélectionner et combiner les différentes descriptions pour obtenir la meilleure qualité possible en fonction des conditions de lecture.

Cette approche présente plusieurs avantages, notamment une meilleure tolérance aux erreurs de transmission, une meilleure adaptabilité aux variations de débit du réseau et une meilleure qualité perçue par l'utilisateur final.

#### ❖ *MDC dans le domaine spatial :*

Dans ce contexte, le processus de MDC est appliqué dans le domaine des pixels. Une approche simple consiste à partitionner l'image ou la trame en plusieurs sous-images, puis à les encoder

individuellement. La Figure 22 illustre le procédé d'échantillonnage spatial polyphasé (PSS) d'une trame [61] pour générer quatre sous-images permettant quatre descriptions. Lorsque toutes les descriptions sont reçues du côté du décodeur, les sous-images sont fusionnées pour reconstituer l'image en résolution complète. Dans le cas contraire, toute description manquante doit être récupérée à l'aide de techniques d'interpolation ou similaires. Cependant, cette approche de base ne permet pas d'ajouter de la redondance ni d'améliorer la qualité. Afin de résoudre ce problème, différentes solutions ont été proposées. Il est important de noter que l'échantillonnage spatial ou temporel réduit la corrélation entre les données, ce qui entraîne une réduction de la compressibilité de la source. Par conséquent, bien que cet échantillonnage introduise indirectement une certaine quantité de redondance par rapport au codage sans perte, cette forme de redondance n'est pas avantageuse en termes de qualité latérale.

Plusieurs travaux ont proposé une méthode de MDC spatiale tel que le zero padding et le zero padding adaptative basé sur le contenu vidéo [62-64] ou bien le filtrage [65, 66].

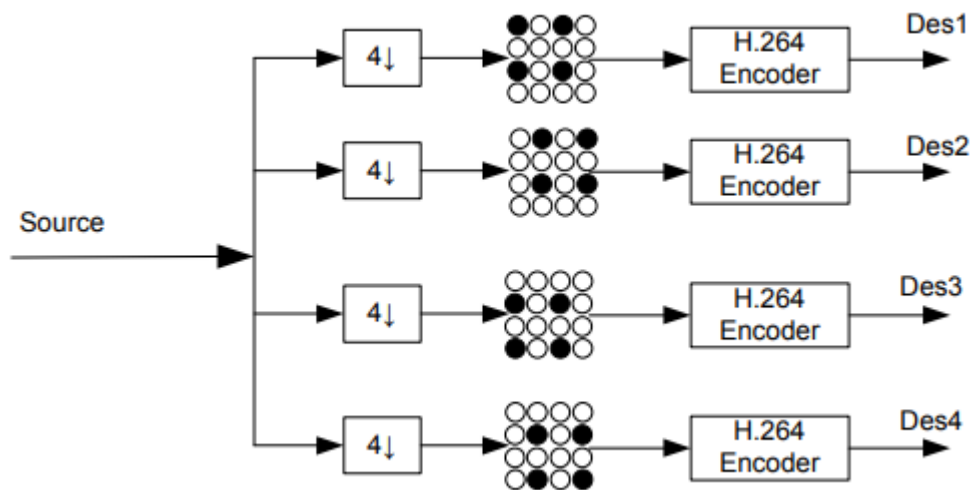


Figure 22. MDC avec échantillonnage spatial polyphasé [13].

❖ *MDC dans le domaine temporel :*

La granularité de cette catégorie de méthodes est une trame c'est-à-dire les descriptions sont générées par un processus effectué au niveau de la trame. La méthode la plus simple est de diviser la vidéo en deux descriptions une contient les trames paires et l'autre les trames impaires comme est

proposé dans [14] et illustré dans la figure 23. Cette méthode a permis une amélioration de la qualité des vidéos codé en HEVC.

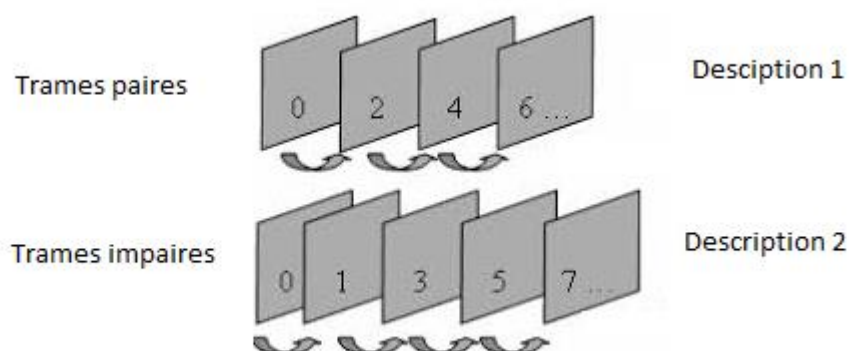


Figure 23. MDC temporel avec 2 descriptions.

#### ❖ *MDC dans le domaine fréquentiel :*

Il existe aussi des approches de MDC dans le domaine fréquentiel. On cite principalement le MDC avec quantification scalaire en utilisant de différentes méthodes de quantification comme le décalage d'intervalle de quantification [67], l'indexation des tables [68] et le MDC par transformation.

#### 3.2.2. Codage canal

Dans un réseau sans fil, le canal de communication est souvent exposé à des perturbations et présente un niveau de bruit élevé, ce qui nécessite une codification de canal efficace. La codification de canal comprend deux processus principaux. Le premier consiste à convertir les données d'entrée en une séquence d'entrée de canal grâce à une correspondance ou à un encodage. Le deuxième processus est l'inverse de la correspondance, également appelé décodage, qui permet de récupérer les données d'origine transmises. Les deux techniques de base utilisées pour la transmission vidéo sont le FEC (Forward Error Correction) et l'ARQ (Automatic Repeat reQuest) et Différents types de codes sont utilisés pour cette tâche :

- Les codes linéaires en blocs
- Les codes de convolution
- Les codes Turbo

➤ Les codes de canal Multiple-Input, Multiple-Output (MIMO)

Dans [69] une synthèse détaillée sur les codes canal cité ci-dessus est présentée. Une autre synthèse est présentée dans [70] sur les approches du codage canal assistées par l'intelligence artificielle.

3.2.3. Codage conjoint source et canal

Cette approche est utilisée en communication pour optimiser la transmission de données à la fois au niveau de la source et du canal de communication. Cette technique combine le codage de source, qui compresse les données sources pour réduire la redondance, et le codage de canal, qui introduit de la redondance pour améliorer la fiabilité de la transmission dans un canal bruité.

3.2.4. L'apprentissage machine

L'utilisation de l'apprentissage automatique dans les systèmes de communication sans fil comme il est illustré dans la figure 24 s'est largement répandue en raison de la forte similitude entre l'architecture des auto-encodeurs et les systèmes de communication [71] [72]. Ils ont été utilisés pour résoudre les problèmes liés aux sources dégradées et repousser les limites de la communication [73] [74].

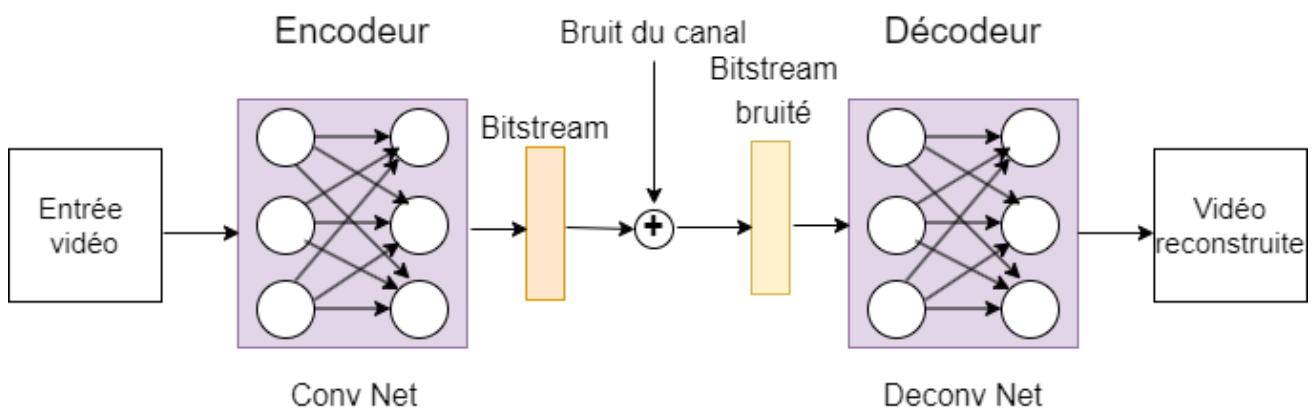


Figure 24. Système de communication basé sur l'apprentissage automatique.

Dans [75], des modèles de dé-bruitage de canal par diffusion pour les communications sans fil. Ces modèles peuvent être utilisés comme un nouveau module de couche physique après l'égalisation du canal pour apprendre la distribution du signal d'entrée du canal, puis utiliser cette connaissance acquise pour éliminer le bruit du canal. Ces modèles ont été appliqués à un système de communications sémantiques basé sur le codage source-canal conjoint (CSCC).

### 3.3. Dissimulation d'erreurs du décodeur

La dissimulation des erreurs ou error concealment en anglais (EC) est un domaine essentiel dans le contexte de l'internet of things (IoT). En raison de la nature sans fil et souvent instable des réseaux IoT, les erreurs de transmission peuvent se produire fréquemment. La dissimulation des erreurs vise à minimiser les effets de ces erreurs sur la qualité et la fiabilité des données transmises. L'objectif principal de la dissimulation des erreurs est de restaurer les données corrompues ou manquantes en utilisant diverses techniques de récupération d'erreurs. Ces techniques peuvent inclure la prédiction, l'interpolation, la répétition ou la substitution de données manquantes.

#### 3.3.1. Dissimulation d'erreur intra-frame

La dissimulation d'erreur intra-trame également appelée dissimulation d'erreur spatiale fait référence au processus de dissimulation des erreurs ou des artefacts au sein d'une seule trame d'une vidéo sans se fier aux informations des trames voisines c'est-à-dire que seuls les pixels dans le même frame sont utilisés pour estimer le pixel perdu. Dans [76] Un modèle de texture basé sur une interpolation géométrique est utilisé pour récupérer les pixels endommagés en utilisant les textures d'image des pixels voisins dans [77] Les pixels dans les blocs manquants sont reconstruits successivement en fonction d'un prédicteur linéaire adaptatif. L'article [78] propose une méthode rapide de dissimulation d'erreurs spatiales utilisant le mode de prédiction des blocs voisins.

Dans les nouveaux codecs tel que HEVC et VVC la zone perdue est relativement grande en raison de la taille des CTUs qui peut atteindre  $64 \times 64$  pixels pour le HEVC [22] et  $128 \times 128$  pixels pour le VVC [79]. Ces raisons font que les techniques spatiales de dissimulation d'erreurs sont moins efficaces que les techniques temporelles [7].

#### 3.3.2. Dissimulation d'erreur inter-frame

La dissimulation d'erreur inter-trame, désigne le processus de dissimulation des erreurs ou des artefacts au sein d'une trame vidéo en utilisant les informations des trames voisines. Ce type de dissimulation d'erreur est généralement utilisé pour récupérer les vecteurs de mouvement (MV) perdus en considérant les MV des blocs correspondants dans une trame de référence. La technique *Motion-copy* est une des solutions les plus faciles. Implémentée dans le logiciel de référence H.264/AVC. Dans [80] la technique consiste à utiliser les vecteurs de mouvement moyens des blocs environnants disponibles du MB perdu Les blocs environnants sont les blocs situés au-dessus, en dessous, à gauche et à droite du MB perdu. Ce qui a fourni meilleure dissimulation des pertes. Dans ce travail [81], l'auteur a proposé une dissimulation d'erreurs compensée par le mouvement pour

HEVC basée sur la fusion de blocs et l'énergie résiduelle. Les résidus du bloc Co-localisé sont fusionnés ensemble. Ensuite, le MV du bloc fusionné est défini comme la moyenne des MV des blocs environnants.

Une deuxième approche également très pratique pour la dissimulation d'erreurs temporelle c'est *l'extrapolation du vecteur de mouvement* ou les MVs des blocs de la trame précédente sont projetés sur les blocs de la trame actuelle, et le MV dont le bloc pointé présente la plus grande zone de chevauchement avec le bloc actuel est sélectionné comme le MV récupéré [82].

L'algorithme proposé dans [83] vise à dissimuler les erreurs des vidéos codées en H.264 basé sur la détection de changement de scène et le modèle d'analyse en composantes principales (PCA). Les images perdues sont dissimulées par l'algorithme de Projection sur des Ensembles Convexes avec le modèle PCA.

Benjak et al [84] ont présenté une méthode de dissimulation d'erreurs pour VVC qui dissimule les trames B à l'aide du réseau de interpolation de trames neuronales. Il est estimé que cette approche est actuellement l'algorithme de dissimulation d'erreurs le plus performant pour les paramètres d'une seule slice par trame B.

### 3.3.3. Dissimulation d'erreur hybride

Les techniques de dissimulation d'erreurs utilisées dans les domaines Intra et Inter présentent des avantages et des inconvénients spécifiques. Il peut être préférable, dans certains cas, de combiner les techniques de dissimulation d'erreur établies dans ces deux domaines, voire d'explorer d'autres techniques provenant d'autres domaines de traitement du signal et de l'image pour améliorer la performance de la dissimulation d'erreurs.

Dans [83], une EC jointe spatio-temporelle combinée à un MDC est utilisée pour estimer les informations perdues. Une approche similaire dans [85] combine des techniques EC temporelles et spatiales sans utiliser le MDC pour estimer les informations perdues. Une dissimulation d'erreurs spatiale hybride utilisant le visage humain et les contours d'image sur des séquences vidéo H.264/AVC est proposée par C Nam et al [86].

## 3.4. Les outils de résilience aux erreurs dans le HEVC

Cette sous-section résume les outils de résilience aux erreurs pris en charge par le HEVC [87] :

- **L'ensemble d'images de référence: Reference picture selection en anglais (RPS)** L'ordre de décodage des images et le Tampon d'images décodées (DPB) dans HEVC sont modifiés

par rapport à H.264/AVC afin d'exploiter les avantages de RPS et d'améliorer la résilience aux erreurs. Dans HEVC, le RPS contient le Compteur d'ordre d'images (POC) de toutes les images de référence, ce qui améliore la possibilité de créer une bonne dissimulation des images de référence perdues.

- **Le rafraîchissement intra** : le HEVC prend en charge l'insertion périodique d'une trame intra pour atténuer la propagation des erreurs.
- **Les slices et les tuiles (tile en anglais)** : le HEVC permet le partitionnement de la trame en plusieurs slices et tuiles ce qui peut aider à la résilience aux erreurs [88].
- **Le message d'informations supplémentaires d'amélioration (SEI)** : qui peut être utilisé pour détecter les erreurs vu qu'il prend en charge des métadonnées qui fournissent des informations sur le timing des images vidéo et l'interprétation appropriée de l'espace colorimétrique utilisé dans le signal vidéo.
- **Désactivation de la fonction Prédiction adaptative des vecteurs de mouvement (AMVP)** : dans le cas où la perte d'une trame qui utilise AMVP se produit cela provoque une défaillance du décodage entropique avec une importante détérioration de la qualité en raison de la propagation de l'erreur [89, 90].

### 3.5. Conclusion

Dans ce chapitre, nous avons examiné en détail l'état de l'art des mécanismes d'amélioration de la qualité du streaming vidéo, en nous concentrant spécifiquement sur les trois principaux domaines : le codage source, le codage canal et le codage conjoint source-canal.

Tout d'abord, nous avons étudié les techniques de codage source qui visent à créer un bit stream robuste, capable de résister aux perturbations introduites lors de la transmission et de garantir une qualité de vidéo élevée pour les utilisateurs. Parmi ces techniques, les méthodes MDC ont été identifiées comme une approche prometteuse pour améliorer la résilience du streaming vidéo. Ensuite, nous nous sommes penchés sur les techniques de codage canal, qui sont utilisées pour protéger les données vidéo contre les erreurs introduites lors de la transmission sur un canal de communication. Nous avons également étudié le codage conjoint source-canal, qui cherche à optimiser conjointement le codage source et le codage canal pour une transmission vidéo de haute qualité. Nous avons exploré des approches récentes basées sur l'apprentissage machine notamment l'apprentissage profond qui sont de plus en plus appliqués aux systèmes de communication ces dernières années. En outre, nous avons également étudié les méthodes de dissimulation d'erreur du côté du décodeur. Ces méthodes visent à reconstruire les zones endommagées d'une séquence vidéo

en utilisant des techniques de prédiction ou d'interpolation, afin de minimiser l'impact des erreurs de transmission sur la qualité visuelle.

Enfin, nous avons abordé les outils de résilience aux erreurs pris en charge par le HEVC, qui propose des fonctionnalités telles que le rafraichissement intra, Les slices et les tuiles, qui contribuent à améliorer la résistance aux erreurs et à maintenir une qualité de vidéo élevée dans des conditions de transmission difficiles.

# Chapitre 3

## 4. Analyse des performances et résultat

### 4.1. Introduction

L'évolution des réseaux IP a entraîné une demande croissante de transmissions de contenu vidéo de haute qualité, ce qui a suscité le besoin de solutions pour améliorer la transmission de la vidéo et garantir une qualité d'expérience et de service optimales. Des problèmes tels que le délai, la gigue, la variation du retard et la perte de paquets peuvent affecter la qualité de la vidéo sur un réseau IP, et il est essentiel de trouver des solutions pour les atténuer. Dans ce contexte, l'émetteur joue un rôle crucial.

Dans ce chapitre, nous allons aborder deux aspects essentiels de l'amélioration de la transmission vidéo sur un réseau sans fil utilisant la norme de compression HEVC.

Tout d'abord, nous allons examiner et analyser l'impact de la perte de paquets sur les vidéos encodées en HEVC. La transmission vidéo sur un réseau sans fil peut être sujette à des pertes de paquets, ce qui peut entraîner une dégradation de la qualité vidéo perçue par l'utilisateur final. Nous étudierons en détail les effets de cette perte de paquets sur la vidéo encodée en HEVC et évaluerons les stratégies possibles pour atténuer ces effets indésirables. L'objectif est de mieux comprendre les problèmes liés à la transmission de vidéos HEVC dans un réseau sans fil et d'identifier des solutions pour minimiser l'impact de la perte de paquets sur la qualité vidéo.

Ensuite, nous présenterons une nouvelle technique de codage source qui vise à améliorer la qualité vidéo perçue par l'utilisateur final. Cette technique, basée sur notre analyse précédente, vise à compenser les pertes de paquets et à optimiser la qualité de la vidéo décodée en HEVC. Nous explorerons les différentes stratégies de codage source utilisées pour améliorer la qualité vidéo et discuterons leurs applications dans le contexte spécifique de la norme HEVC et de la transmission vidéo sur un réseau sans fil. L'objectif est de proposer une approche novatrice qui permettra d'améliorer de manière significative la qualité vidéo disponible pour les utilisateurs finaux. nous explorerons également une approche novatrice pour améliorer la qualité vidéo dans un environnement sujet à des pertes de paquets. Cette méthode repose sur les principes de l'apprentissage profond, en particulier les auto-encodeurs convolutionnels. Nous montrerons comment cette approche peut être utilisée pour compenser les pertes de paquets et optimiser la qualité de la vidéo décodée en HEVC. Cette stratégie, basée sur une analyse préalable, vise à offrir une solution pratique pour garantir une expérience utilisateur optimale, même dans des conditions de

réseau difficiles. Nous explorerons en détail les applications de cette méthode spécifiquement dans le contexte de la norme HEVC et de la transmission vidéo sur un réseau sans fil, mettant en lumière les avantages et les limites de cette approche innovante.

## 4.2. Impacte de la perte des paquets sur la qualité vidéo reçu

### 4.2.1. Bases de données vidéo

La mesure de l'efficacité des techniques de traitement de vidéo s'est vue facilitée par la disponibilité de plusieurs bases de données. Ces bases de données jouent un rôle crucial dans la comparaison des performances, le développement et le test de nouveaux modèles. Actuellement, plusieurs bases de données populaires existent dans le domaine de l'évaluation de la qualité. Cependant, la plupart des bases de données de pointe se concentrent principalement sur le taux de perte de paquets, les artefacts de codage et la compression. Dans cette étude, nous avons utilisé les séquences et les conditions de test communes JCT-VC [91], qui ont été utilisées lors du développement de la norme HEVC. Ces séquences sont répertoriées dans le tableau 2. Elles sont toutes en format YUV, présentent un sous-échantillonnage de 4:2:0 et une profondeur de bits de 8 bits.

**Tableau 2.** Séquences de test issues des conditions de test communes JCT-VC.

Classe	Nom de la séquence	Nombre de frame	Résolution	Frame rate
<b>B</b>	Kimono	240	1920x1080	24
	ParkScene	240	1920x1080	24
	BasketballDrive	500	1920x1080	50
	Cactus	500	1920x1080	50
	BQTerrace	600	1920x1080	60
<b>E</b>	FourPeople	600	1280 x 720	60
	KristenAndSara	600	1280 x 720	60
	Johnny	600	1280 x 720	60

Ces séquences vidéo ont été choisies avec différentes informations perceptuelles spatiales (SI) et temporelles (TI) conformément à l'ITU-T [41]. Pour SI, l'écart type ( $std_{space}$ ) des pixels de chaque trame (plan de luminance) au temps  $n$  ( $F_n$ ) filtré par le filtre Sobel [ $Sobel(F_n)$ ] est calculé, et la valeur maximale de la série temporelle est choisie pour représenter l'indice d'information spatiale de la vidéo [26]. SI est défini comme suit :

$$SI = \max_{time}\{std_{time}[sobel(F_n)]\} \quad (1)$$

La différence des valeurs de pixels (plan de luminance) au même emplacement entre des trames successives de la séquence vidéo est utilisée pour calculer l'information perceptuelle temporelle. L'indice d'information perceptuelle temporelle de la vidéo est déterminé par la valeur maximale dans le temps ( $max_{time}$ ) de l'écart type dans l'espace ( $std_{space}$ ) de  $M_n(i, j)$  sur l'ensemble de  $i$  et  $j$ , ce qui est défini comme suit :

$$M_n(i, j) = F_n(i, j) - F_{n-1}(i, j) \quad (2)$$

$$TI = \max_{time}\{std_{space}[M_n(i, j)]\} \quad (3)$$

$M_n(i, j)$  représente la caractéristique de différence de mouvement et  $F_n(i, j)$  représente le pixel à la  $i^{\text{ème}}$  ligne et  $j^{\text{ème}}$  colonne de la  $n^{\text{ème}}$  trame dans le temps. Les séquences vidéo choisies couvrent les quatre coins de l'espace (SI, TI) comme illustré dans la figure 25.

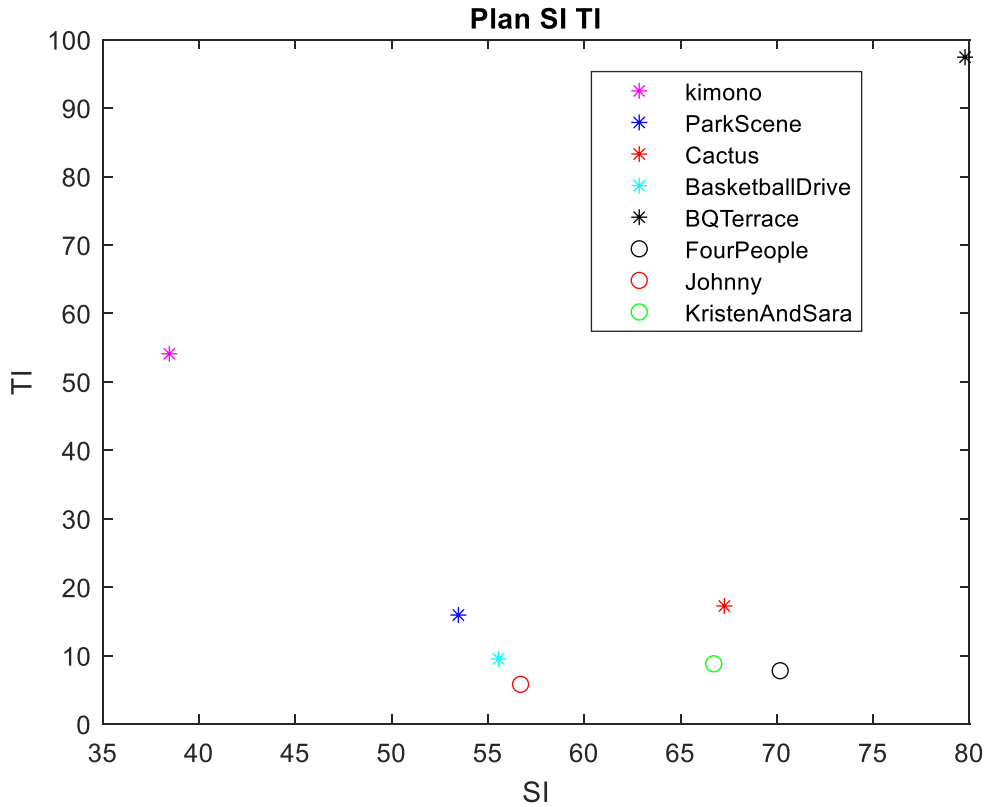


Figure 25. Répartition des vidéos tests choisies dans le plan (SI, TI).

#### 4.2.2. Configuration du codeur et du décodeur du HEVC

Nous avons réalisé les tests sur 8 vidéos de 2 classes différentes en fonction de leur résolution (classe B et E). les vidéos originales ont été codées à l'aide du modèle de test HEVC version 16.11 [92] avec deux modes de prédiction soit ALL intra et Low Delay P.

**All intra (AI)** : Dans cette configuration, chaque image est encodée en tant qu'image I. Étant donné qu'aucune prédiction d'image inter n'est utilisée, elle convient donc aux applications à faible délai et à débit binaire plus élevé. Le décalage de QP dans cette configuration est de 0 car le QP est maintenu constant sur l'ensemble de la séquence. Dans notre cas cette configuration est utilisée comme référence.

**Low delay P** : Dans cette configuration, similaire à la configuration précédente, la réorganisation des images n'est pas autorisée. La première image est encodée en tant qu'image I, et les images suivantes sont encodées en tant qu'images prédictive (P) (PIII...) avec une taille de groupe d'image

(GOP) de 4. Cette configuration est souvent utilisée dans des applications en temps réel où la latence est un facteur critique, comme la vidéoconférence et la diffusion en direct.

La compression des séquences vidéo a été réalisée avec 4 niveaux, c'est-à-dire 4 valeurs de paramètre de quantification (QP) (22, 27, 32, 37). Ffmpeg [93] est utilisé pour diffuser le flux HEVC en utilisant les 2 protocoles UDP et RTP dans un canal soumis à des pertes de paquets ainsi que pour le décodage du flux binaire reçu.

La perte des paquets a été introduite d'une façon aléatoire en utilisant l'émulateur réseau (NetEM) qui est une amélioration des fonctionnalités de contrôle du trafic Linux qui permet d'ajouter du délai, des pertes de paquets, des duplications et d'autres caractéristiques aux paquets sortants d'une interface réseau sélectionnée. NetEm est construit en utilisant les fonctionnalités existantes de QoS et de Services Différenciés (diffserv) du noyau Linux [94].

Dans la configuration expérimentale, une architecture de réseau local sans fil Wireless Local Area Network en anglais (WLAN) a été utilisée pour diffuser une vidéo avec le fichier d'extension HEVC (.hevc) depuis l'émetteur vers le récepteur. De plus, le domaine sans fil est configuré en mode "ad hoc" avec la fonctionnalité du point d'accès désactivée.

Plus précisément, nous avons examiné les taux de perte de paquets de 1, 2 et 3 %. La figure 26 montre le système d'essai proposé.

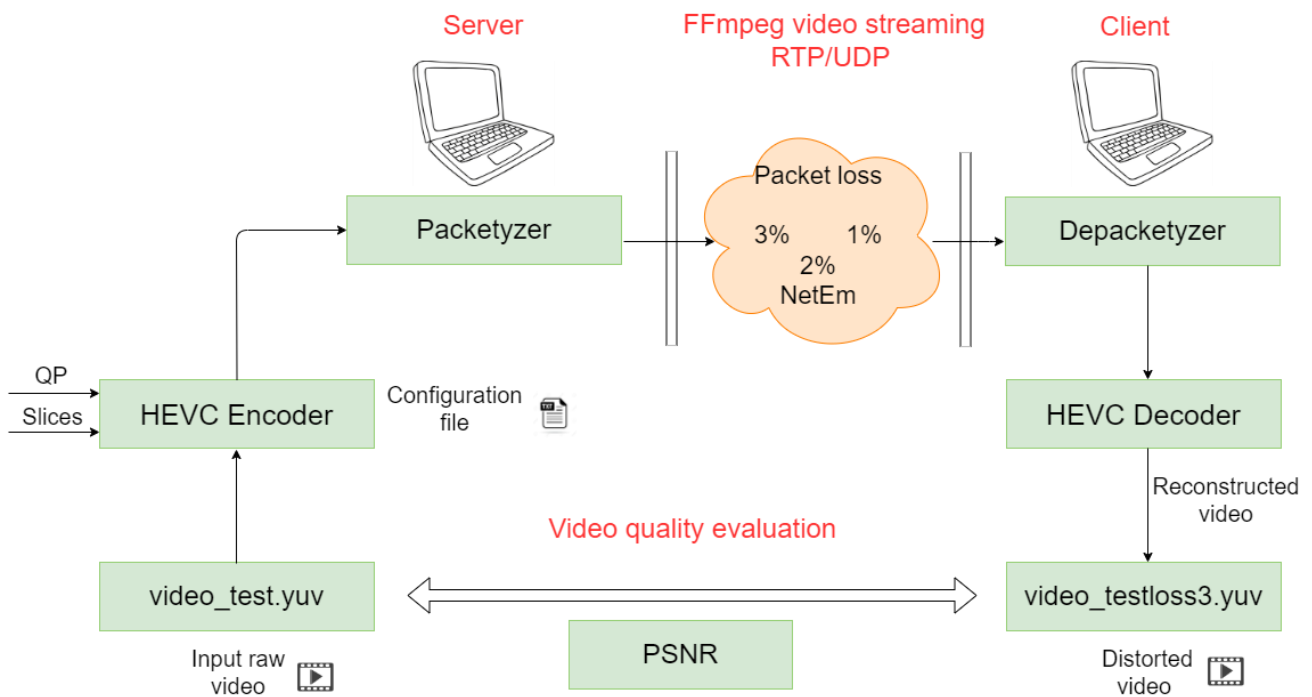


Figure 26. Système d'essai proposé

Le PSNR a été utilisé comme métrique d'évaluation qui est défini comme suit :

$$PSNR = 10 \cdot \log_{10} \frac{d^2}{EQM}$$

$d$  : la valeur maximum possible pour un pixel

EQM : erreur quadratique moyenne définie pour 2 images  $I_o$  et  $I_r$  de taille  $m \times n$  par la formule suivante :

$$EQM = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (I_o(i,j) - I_r(i,j))^2$$

Résultat codage :

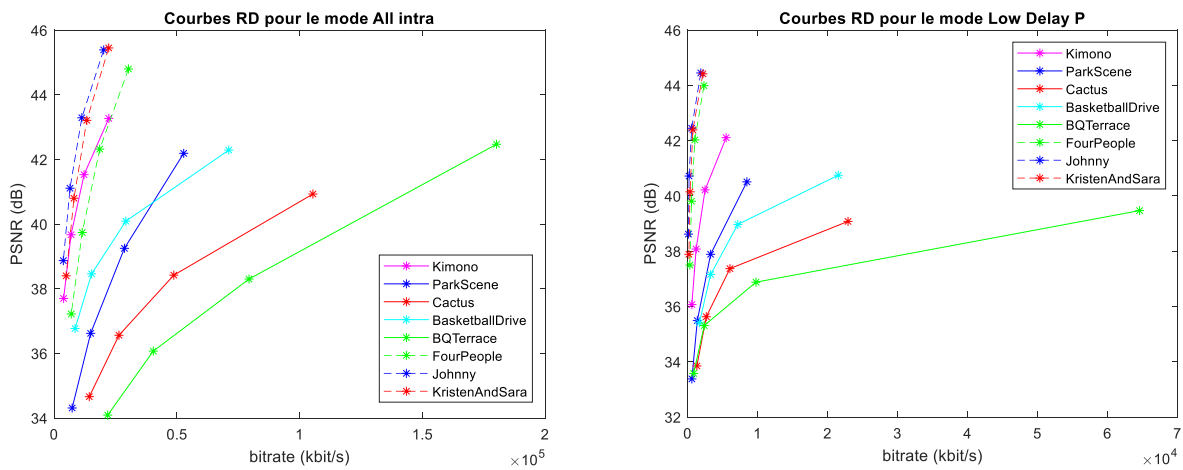


Figure 27. Les courbes RD des 2 modes de codage.

La figures 27 représente les courbes de taux-distorsion (RD) des séquences vidéo utilisées.

Dans cette étude comparative des courbes RD entre les configurations "low delay P" et "All Intra" en HEVC, plusieurs observations pertinentes peuvent être faites. Lors de l'analyse des courbes RD pour les deux configurations, il est clairement observable que la configuration "low delay P" présente un avantage distinct en termes de réduction du débit binaire par rapport à la configuration "All intra" par exemple la séquence kimono de la classe B présente une économie de débit binaire de 75% avec QP=22 avec une différence de 1dB et de 85% pour QP=37 avec une différence de 1.5 dB. De même pour la séquence Johnny de la classe E qui présente une économie de débit binaire 90% pour QP=22 avec une différence de 1dB et de 97% pour QP=37 avec une différence de 0.25 dB. Cela indique que,

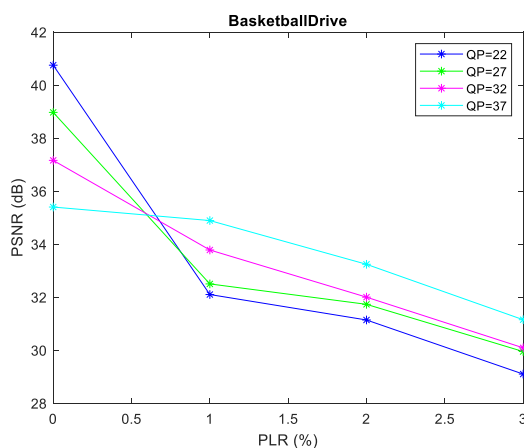
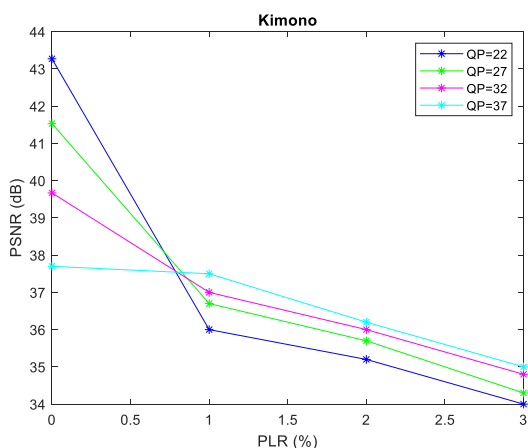
pour une qualité visuelle donnée, la configuration "low delay P" permet de gagner en bande passante quel que soit la résolution des séquences vidéo, ce qui peut être crucial dans des scénarios de diffusion en direct ou de vidéoconférence en temps réel.

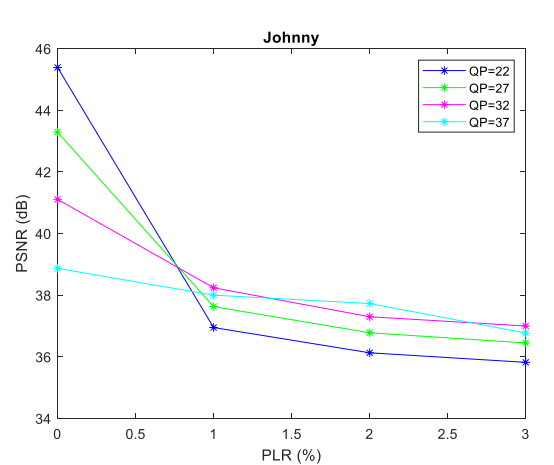
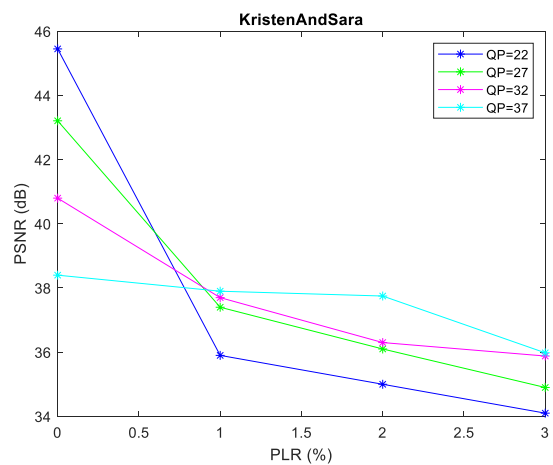
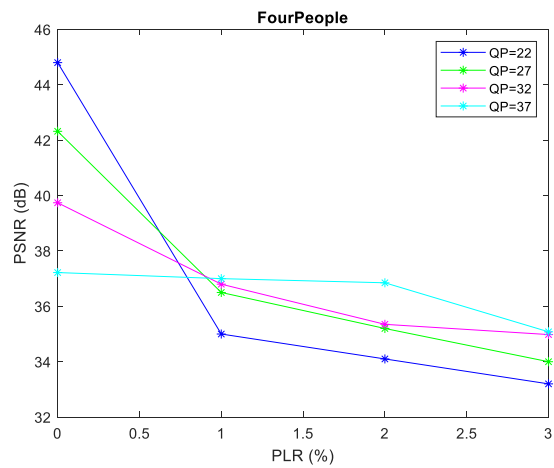
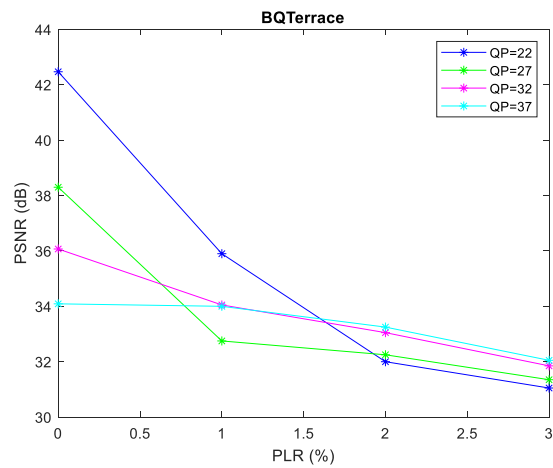
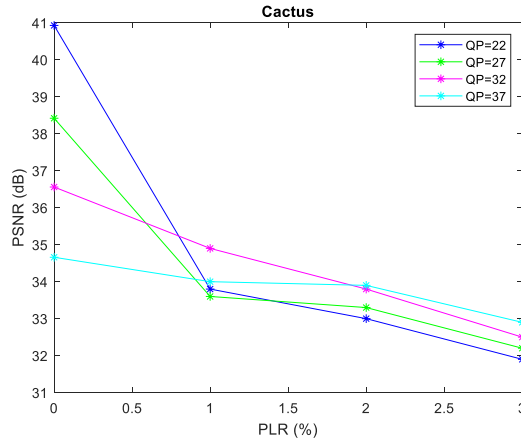
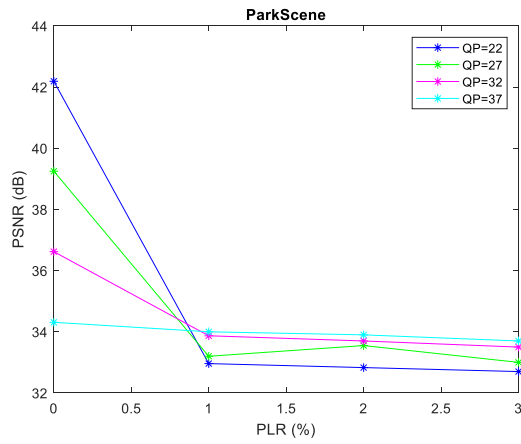
L'analyse comparative des courbes RD pour différentes séquences vidéo, en mettant l'accent sur les différences d'activités, a révélé des tendances significatives. Plus précisément, les courbes RD ont mis en évidence que les séquences présentant des activités de mouvement élevé sont plus sensibles à la compression et présentent une dégradation rapide de la qualité que celles à mouvement modéré. Par exemple la séquence BQTerrace qui est une séquence à mouvement rapide et se trouve dans le coin supérieur droit du plan SI TI présente le PSNR le plus bas avec le débit binaire le plus élevé tant dis que la séquence qui est caractérisée par un mouvement modéré présente le PSNR le plus élevé avec le débit binaire le plus bas.

Cette observation nous aide à prendre de meilleures décisions lors du codage et de la configuration en fonction du niveau de mouvement dans les vidéos. Cela améliore l'efficacité de la compression pour différents types de séquences vidéo.

Résultat transmission :

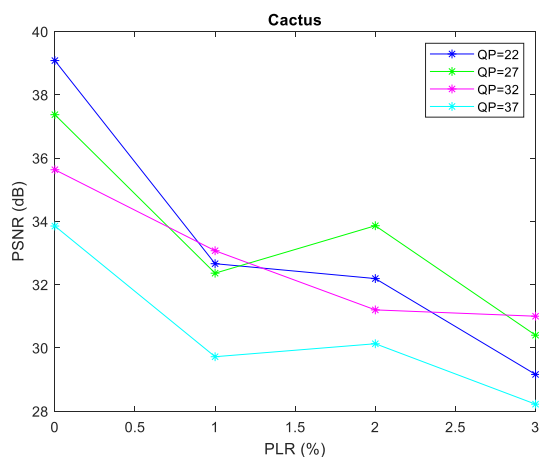
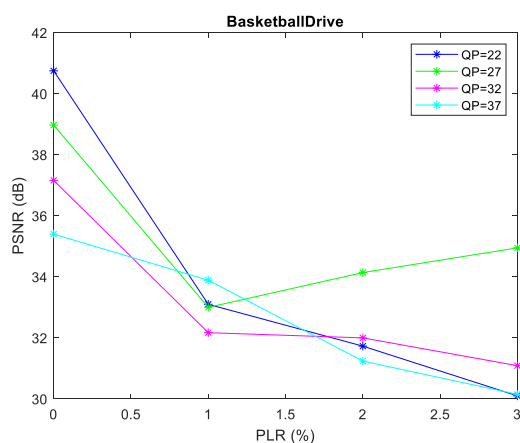
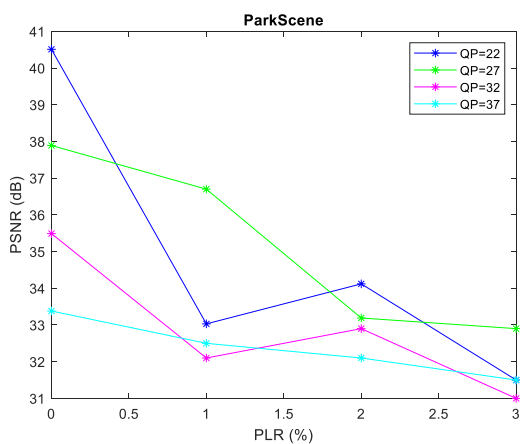
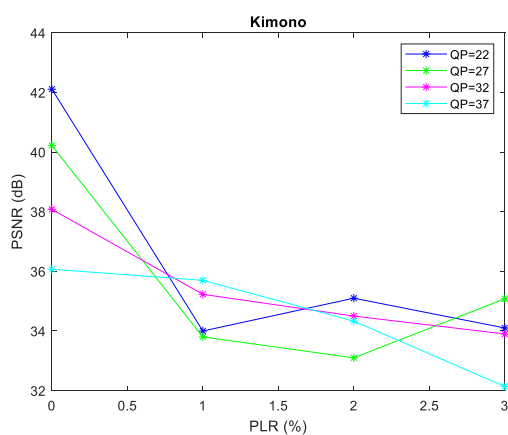
All Intra :





Les courbes de PSNR pour la configuration "All Intra" dans la figure 28 montrent une tendance générale de décroissance progressive du PSNR à mesure que le taux de perte de paquets augmente. Cela s'explique par le fait que chaque image est encodée indépendamment et n'est pas influencée par d'autres images de référence, ce qui rend la séquence plus résistante aux pertes de paquets. Cependant, au fur et à mesure que les pertes de paquets s'accumulent, des artefacts visuels et des distorsions deviennent visibles, entraînant une baisse du PSNR.

Low Delay P :



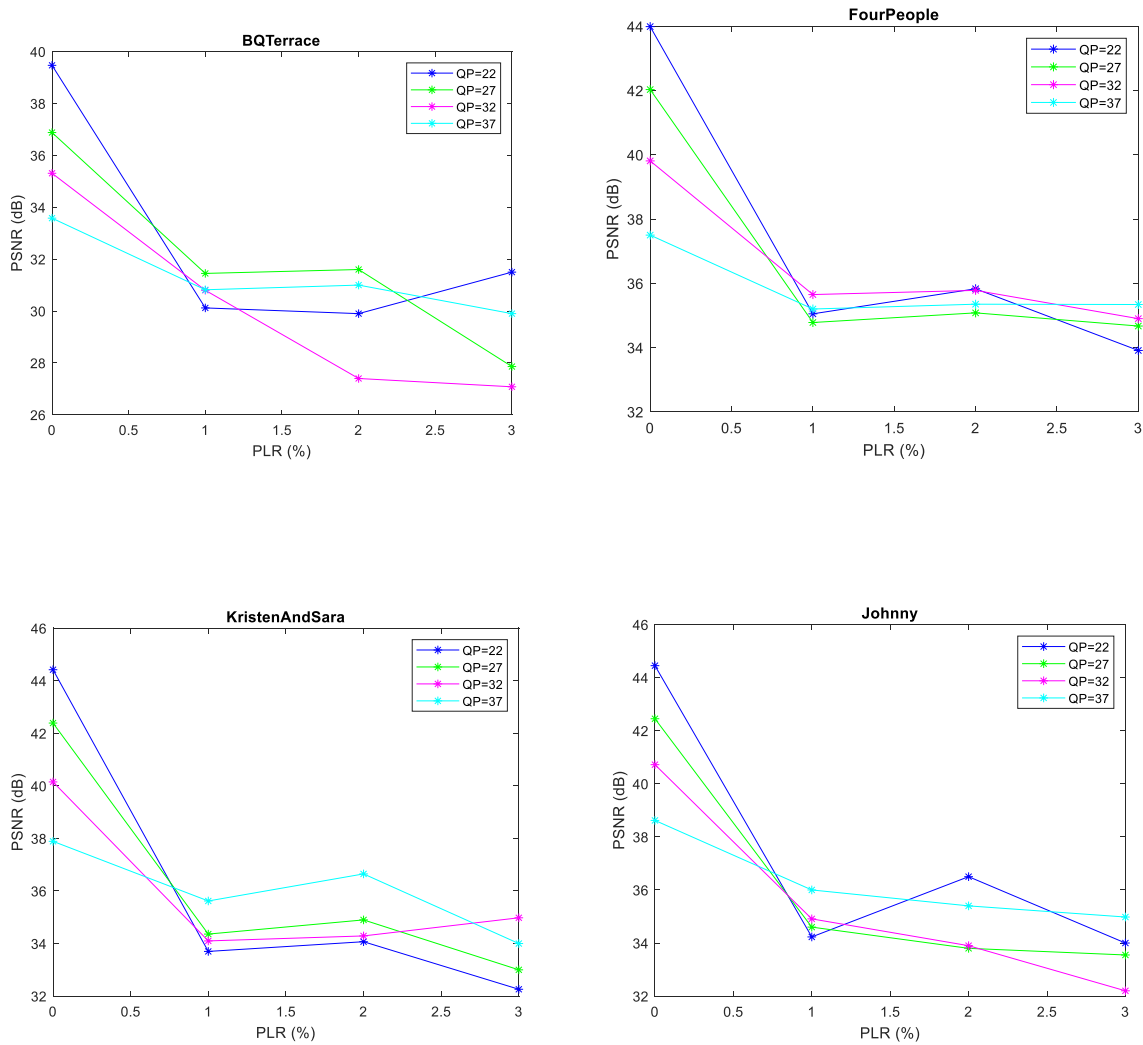


Figure 28. Impact du PLR sur la qualité vidéo en mode Low Delay P.

Les courbes de PSNR de la configuration "Low Delay P" dans la figure 29 en plus de montrer des baisses de PSNR plus marquées à mesure que le taux de perte augmente elles montrent aussi des résultats de PSNR imprévisible c'est-à-dire que le PSNR du PLR=3% peut parfois être plus élevé que le PSNR mesuré pour le taux de le taux de PLR= 1%. Cette dégradation plus rapide de la qualité peut être attribuée aux caractéristiques spécifiques de la configuration "low delay P". Étant conçue pour réduire la latence, cette configuration peut utiliser des prédictions inter-images plus fréquentes et des images de référence provenant de trames précédentes pour générer les images prédites. Lorsque des paquets sont perdus pendant la transmission, les images de référence manquantes peuvent entraîner des erreurs de prédiction plus importantes, provoquant des distorsions visibles.

#### 4.3. Codage a description multiple spatial pour l'amélioration de la qualité video reçue

Le but de ce travail est, d'une part, d'étudier l'impact des différents types de pertes et leur effet sur la qualité de la vidéo reconstituée. Dans ce contexte, cinq scénarios différents de perte de trame complète ont été créés, pouvant être des "pertes en rafale" ou des "pertes uniformes". D'autres facteurs qui influencent la qualité de la vidéo reconstituée ont également été pris en compte, notamment le contenu de la vidéo, la valeur du Paramètre de Quantification (QP) du côté de l'encodeur qui détermine le niveau de détails spatiaux conservés [95], et la méthode de dissimulation d'erreur du côté du décodeur qui a été conçue non seulement pour éviter le blocage du décodeur, mais aussi pour minimiser les effets de la perte. D'autre part, un schéma d'adaptation des paramètres de l'encodeur HEVC basé sur la MDC spatiale est proposé comme solution à la dégradation de la qualité vidéo causée par la perte de paquets.

#### 4.3.1. Description du système

La Figure 1(a) décrit un SDC utilisant la norme HEVC (High Efficiency Video Coding) et deux paramètres de quantification (QP=22 et QP=37) pour obtenir respectivement un débit élevé et un faible débit. Le Framework de suppression des unités de couche d'abstraction réseau HEVC est le processus qui permet de contrôler la perte de paquets ou de trames à partir du flux de bits. Pour cette raison, une version modifiée de HM 16.11 est utilisée du côté du décodeur de la chaîne de transmission.

La Figure 1(b) montre le modèle de QoE proposé en utilisant le schéma d'adaptation de l'encodeur HEVC. Cette proposition permet une protection au niveau source d'un signal vidéo transmis dans un canal présentant des pertes en utilisant la technique MDC spatiale, qui consiste à diviser chaque trame de la vidéo en 2 sous-images. Une sous-image contient les lignes impaires tandis que la deuxième contient les lignes paires. Par conséquent, cela génère deux descriptions du signal source comme illustré dans la Figure 2. Chaque description de la vidéo est décodée séparément, puis les deux sont entrelacées pour former une seule vidéo qui sera comparée à la vidéo originale (brute) en utilisant deux types de métriques : le PSNR et le score moyen d'opinion estimé (Estimated Mean Opinion Score, MOS\_Estimated). Ces deux processus de division de la vidéo avant l'encodage et d'entrelacement après le décodage ont été réalisés avec le logiciel FFmpeg, une plate-forme croisée, en utilisant le filtrage via la bibliothèque "libavfilter".

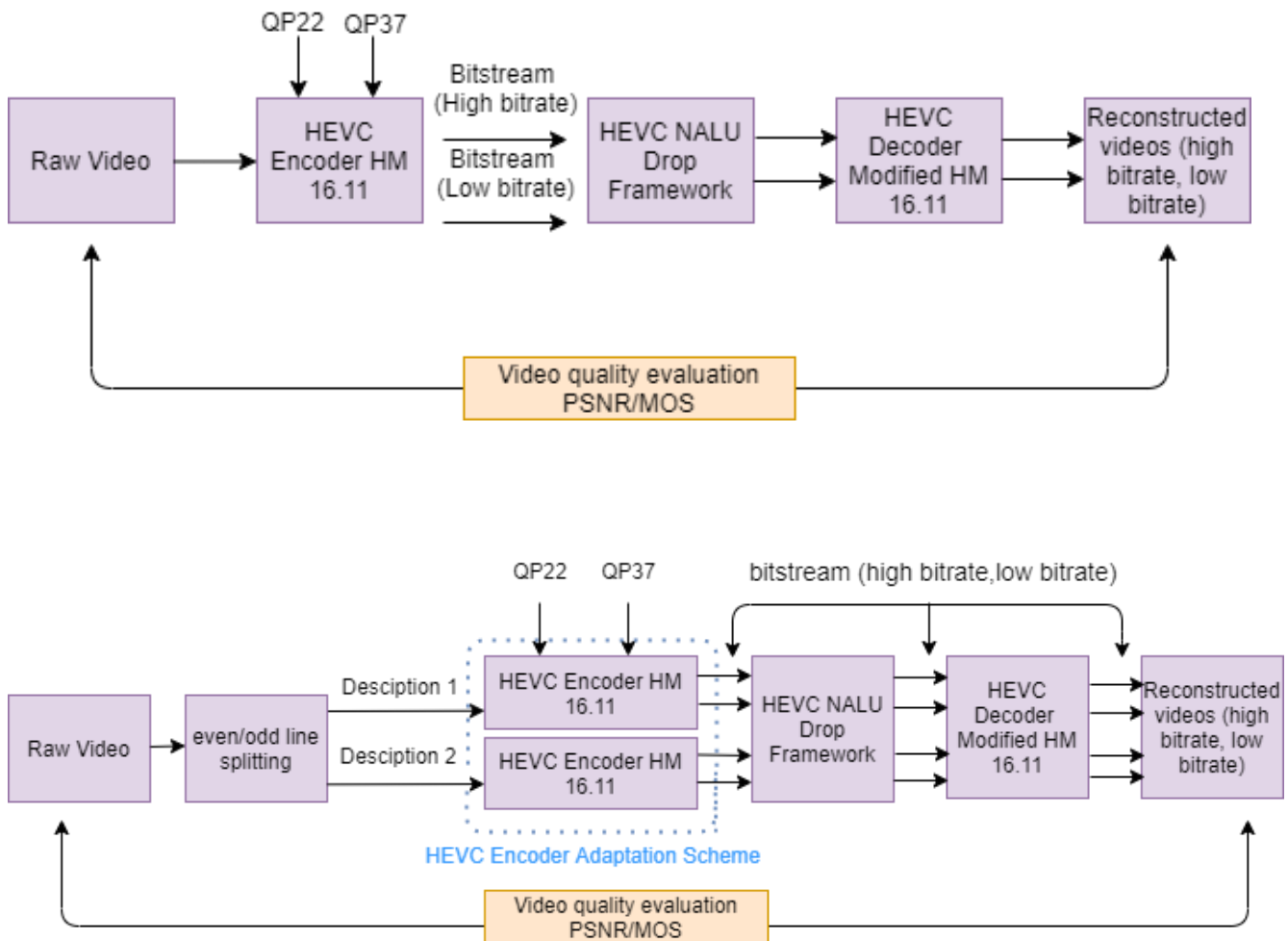


Figure 29. Schéma d'adaptation pour l'évaluation objective d'une séquence vidéo : (a) SDC, (b) MDC.

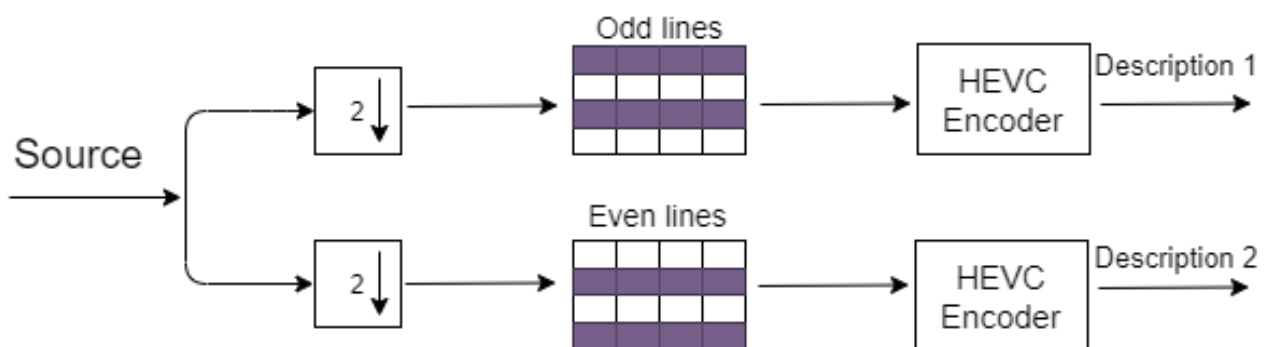


Figure 30. Modèle MDC spatial proposé.

### 4.3.2. Expérience

#### a. Séquences vidéo de test

Quatre séquences vidéo de la base de données de qualité vidéo ReTRiEVED [96] [97], avec la résolution spatiale couramment utilisée actuellement, le format commun intermédiaire à quatre

canaux (4CIF) (704×576 pixels/image), ont été sélectionnées. Elles sont énumérées dans le Tableau 3. Et représenter dans le plan SI TI dans la figure 32.

**Tableau 3.** Caractéristiques des séquences video utilisées.

Séquence videos	Sous- échantillonnage	Durée (s)	Nombre d'images	Rate (frame/s)
Duckstakeoff	4:2:0	9	250	25
Ice	4:2:0	7	240	30
Soccer	4:2:0	7	300	30
Crowdrun	4:2:0	9	250	25

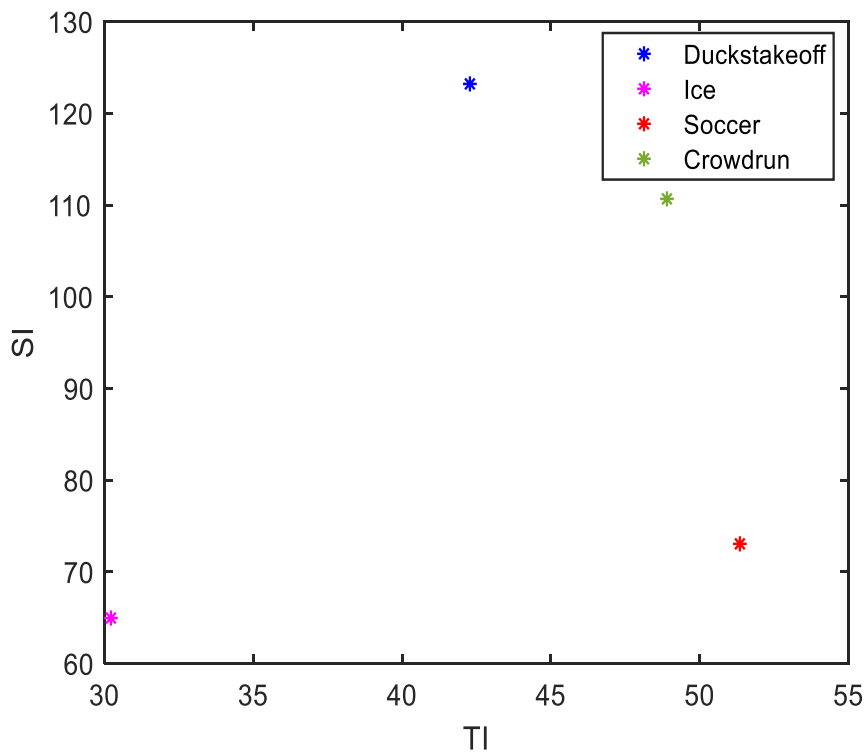


Figure 31. Plan SI TI des séquences video

b. Les paramètres de codage

Les expériences ont été menées en utilisant HM 16.11 et les deux types de configurations de faible latence, Prédictive (LDP) et Bidirectionnelle (LDB), comme recommandé par les conditions de test communes du JCT-VC [28]. Dans ces configurations, la première trame est encodée en tant que trame intra (I) et les trames suivantes sont encodées en tant que trames prédites simples ou trames bidirectionnelles. Le codage est effectué sans délai structurel (l'ordre de codage est égal à l'ordre de sortie), de sorte que le réarrangement des trames n'est pas autorisé. Par conséquent, les expériences abordent les applications de livraison vidéo à faible délai ou de vidéo conversationnelle. Les

conditions de test d'encodage sélectionnées sont également basées sur les recommandations du JCT-VC. En particulier, deux paramètres de quantification (QP=22 et QP=37) ont été utilisés, ce qui a conduit à deux flux de bits de deux taux de compression différents. Pour chaque vidéo, la taille du groupe d'images (GOP) a été fixée à 4. Pour vérifier la correction du processus de compression, les descriptions SDC et MDC ont été utilisées sans aucune perte. Les résultats sont présentés dans les tableaux 4, 5 et 6. Les expériences ont été menées sur un CPU Intel(R) Core(TM) i7-9700K, vitesse du processeur @ 3.60GHz, RAM 32 Go, système 64 bits, architecture du processeur exécutant Ubuntu 18.04LTS Linux.

**Tableau 4.** Résultat du SDC

Video	Bit rate (kbps)				PSNR (dB)				Encoding time (s)			
	22		37		22		37		22		37	
	LDP	LDB	LDP	LDB	LDP	LDB	LDP	LDB	LDP	LDB	LDP	LDB
Duckstakeoff	2280.28	14778.26	2039.08	1620.84	37.05	37	26.57	26.86	4013.13	4312.22	1917.06	2956.15
ice	1438.32	1399.65	188.5	187.39	43.91	43.94	36.24	36.27	1160.07	1513.24	791.91	993.9
Soccer	3611.17	3247.87	289.57	281.07	41.35	41.51	32.7	32.79	2229.55	2670.97	1457.49	1568.28
crowdrun	11612.14	11216.32	1132.29	1125.26	37.47	37.52	27.91	27.92	3000.06	3396.36	1688.81	1773.34

**Tableau 5.** Résultat du MDC (description paire)

Video	Bit rate (kbps)				PSNR (dB)				Encoding time (s)			
	22		37		22		37		22		37	
	LDP	LDB	LDP	LDB	LDP	LDB	LDP	LDB	LDP	LDB	LDP	LDB
Duckstakeoff	12612.04	12545.9	1308.9	1224.56	36.96	36.97	25.73	25.8	2029.31	2254.81	992.93	1203.14
ice	1051.69	1030.66	135.82	135.05	43.46	43.37	34.86	34.87	631.7	817.64	405.75	570.24
Soccer	2706.38	2477.9	204.6	197.61	40.83	40.93	31.45	31.54	1253.21	1593.63	588.83	803.11
crowdrun	8380.13	8053.73	770.36	762.38	37.35	37.35	25.44	26.5	1671.72	1866.1	707.08	924.96

**Tableau 6.** Résultat du MDC (description impaire)

QP	Bit rate (kbps)				PSNR (dB)				Encoding time (s)			
	22		37		22		37		22		37	
	LDP	LDB	LDP	LDB	LDP	LDB	LDP	LDB	LDP	LDB	LDP	LDB
Duckstakeoff	12613	12545.9	1307.86	1224.02	36.96	36.97	25.71	25.8	2018.9	2260.14	987.56	1224.83
ice	1052.04	1030.35	136.07	135.67	43.47	43.49	34.85	34.87	636.7	810.28	415.6	581.64
Soccer	2708.57	2478.34	204.76	197.68	40.82	40.92	31.44	31.54	1274.06	1469.4	600	805.32
crowdrun	8372.36	8048	768.54	760.2	37.35	37.35	26.5	26.5	1688.72	1862.23	712.87	915.67

D'après le tableau ci-dessus, on peut noter que les résultats de compression sont conformes aux attentes. Un encodage avec une petite valeur de QP donne un débit binaire élevé et un PSNR élevé, c'est-à-dire une meilleure qualité vidéo, bien que le temps de codage augmente. Une valeur élevée de QP donne un faible débit binaire et un faible PSNR, entraînant une diminution de la qualité vidéo et du temps de codage. L'encodage de différentes séquences avec différents contenus vidéo avec la même valeur de QP donne des débits binaires et des valeurs de PSNR différentes. La séquence Duckstakeoff, qui est caractérisée par un fort SI (Indice Spatial), a un PSNR de 37,05 dB pour

QP=22 et la configuration LDP, tandis que la séquence Ice a un PSNR de 43,91 dB pour la même valeur de QP et la même configuration. Cette différence de près de 6 dB montre que la séquence Duckstakeoff ne tolère pas beaucoup la compression. Elle nécessite également un temps de codage élevé par rapport à la vidéo Ice, qui est une séquence avec un faible SI et TI (Indice Temporel). La configuration LDB permet d'économiser du débit binaire tout en maintenant le même PSNR, c'est-à-dire de conserver la même qualité vidéo par rapport à la configuration LDP, mais la configuration LDP nécessite un temps d'encodage plus élevé.

Les tableaux 4, 5 et 6 indiquent que le temps de codage dépend de plusieurs paramètres, notamment la valeur de QP. Plus la valeur de QP est petite, plus le temps de codage est long. La configuration choisie influence également le temps de codage. La configuration LDB prend plus de temps de codage que la configuration LDP. La nature de la vidéo affecte de manière significative le temps de codage. La séquence Duckstakeoff, qui est la plus complexe spatialement (a le plus haut SI), prend plus de temps de codage que la séquence Ice, qui a le plus bas SI et TI.

#### c. Scénario de suppression intégrale de trame proposé

Dans la transmission vidéo à faible débit binaire, lorsque le phénomène de perte de paquets en rafale se produit, il y a une forte probabilité que la trame entière soit perdue [98]. Cela affecte la qualité de toute la vidéo reçue lorsque le mode de codage utilisé est à faible délai. En effet, cette configuration est très sensible à la perte de paquets en raison de la forte dépendance entre les trames. La perte d'une trame affecte les trames suivantes qui utilisent la trame défectueuse comme référence, et la position de la perte est un facteur qui influence la qualité de la vidéo reconstituée. Pour ces raisons, la configuration à faible délai est mise en évidence dans ce travail.

Pour analyser la propagation des erreurs dans les configurations LDP et LDB, des suppressions intégrales de trame ont été appliquées aux flux de bits sous cinq formes différentes : suppression en rafale de trame complète ou suppression uniforme de trame complète, comme indiqué dans la Figure 33. À notre connaissance, il s'agit de la première étude qui aborde ces cas dans la littérature.



Figure 32. Stratégies de suppression intégrale de frame proposées.

La stratégie une a été mise en œuvre pour démontrer l'importance de la dernière trame du GOP (Group of Pictures) et l'impact de sa perte pendant la transmission sur la qualité de la vidéo reconstituée. Dans un GOP, cette trame est considérée comme une trame centrale car elle a un débit binaire plus élevé, déterminé par le facteur QP. Dans la configuration typique à faible délai P ou B utilisée dans cette étude, la dernière trame du GOP a un facteur QP de 0,578, tandis que les autres trames du même GOP ont un facteur QP de 0,4624. Par conséquent, la trame centrale, qui a une meilleure qualité, est conservée plus longtemps dans le tampon d'image décodée (DPB) et utilisée dans la prédiction des autres trames, comme illustré dans la figure 34 qui montre la sélection de l'image de référence pour le numéro de compte d'image (POC) 16 en mode LDP.

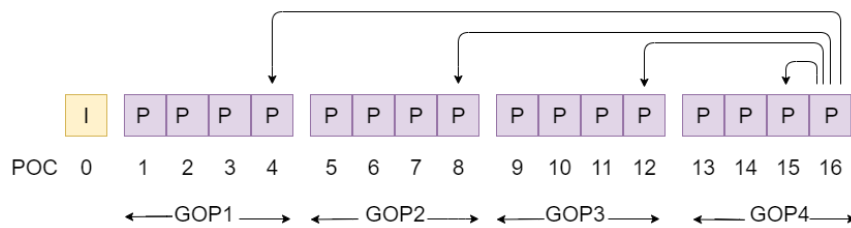


Figure 33. Sélection des images de référence pour le POC16 en configuration LDP.

Le résultat sera comparé avec la stratégie 2, qui consiste à supprimer la première trame d'un GOP, qui est une trame ordinaire. La stratégie 3 implique la mise en place d'une perte de type en rafale, c'est-à-dire la suppression de 4 trames consécutives (un GOP complet). Le résultat sera comparé avec la stratégie 4, qui simule une perte de type uniforme, c'est-à-dire la suppression d'une trame de

chaque GOP. La dernière stratégie consiste à produire une perte de type en rafale vers la fin de la vidéo. Le résultat sera analysé avec la stratégie 3, qui est une perte de type en rafale survenant au début de la vidéo.

d. Les paramètres du décodage

Du côté du décodeur, dans le cas du SDC, chaque flux de bits est décodé avec la norme HEVC, tandis que dans le cas du MDC, chaque description est décodée indépendamment puis sera ensuite entrelacée.

Un décodeur HM modifié a été utilisé, avec « motion-copy » comme technique de dissimulation d'erreur pour reconstituer les trames manquantes. Cette méthode a été choisie pour sa capacité à approximer les informations de pixel et de vecteur de mouvement perdues avec une complexité de calcul réduite.

e. Les méthodes d'évaluation

Dans cet article, deux types de métriques ont été utilisés : le PSNR et le MOS\_Estimated, qui est une estimation des scores MOS à partir du PSNR mesuré, comme proposé par [99], en utilisant une fonction d'approximation :

$$MOS_{Estimated} = 0,0935 PSNR + 0,152 \quad (4)$$

Les résultats sont interprétés dans le Tableau 7.

**Tableau 7.** Échelle de qualité et de dégradation.

Résultat	Quality scale	Degradation scale
5	Excellent	Imperceptible
4	Good	Perceptible, but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

4.3.3. Résultat et discussion

Les Figures 7, 8, 9 et 10 montrent les résultats des cinq cas de perte complète de trame en termes de PSNR et de MOS\_Estimated. Ils sont comparés pour discuter les points suivants :

- Les effets du contenu vidéo.

- L'effet de la valeur QP.
- L'effet de l'importance des trames dans le même GOP (1ère et 4ème trame du GOP).
- Les effets de la suppression uniforme et en rafale.
- L'impact de notre schéma d'adaptation de l'encodeur proposé basé sur la MDC spatiale sur la qualité de la vidéo reconstituée.

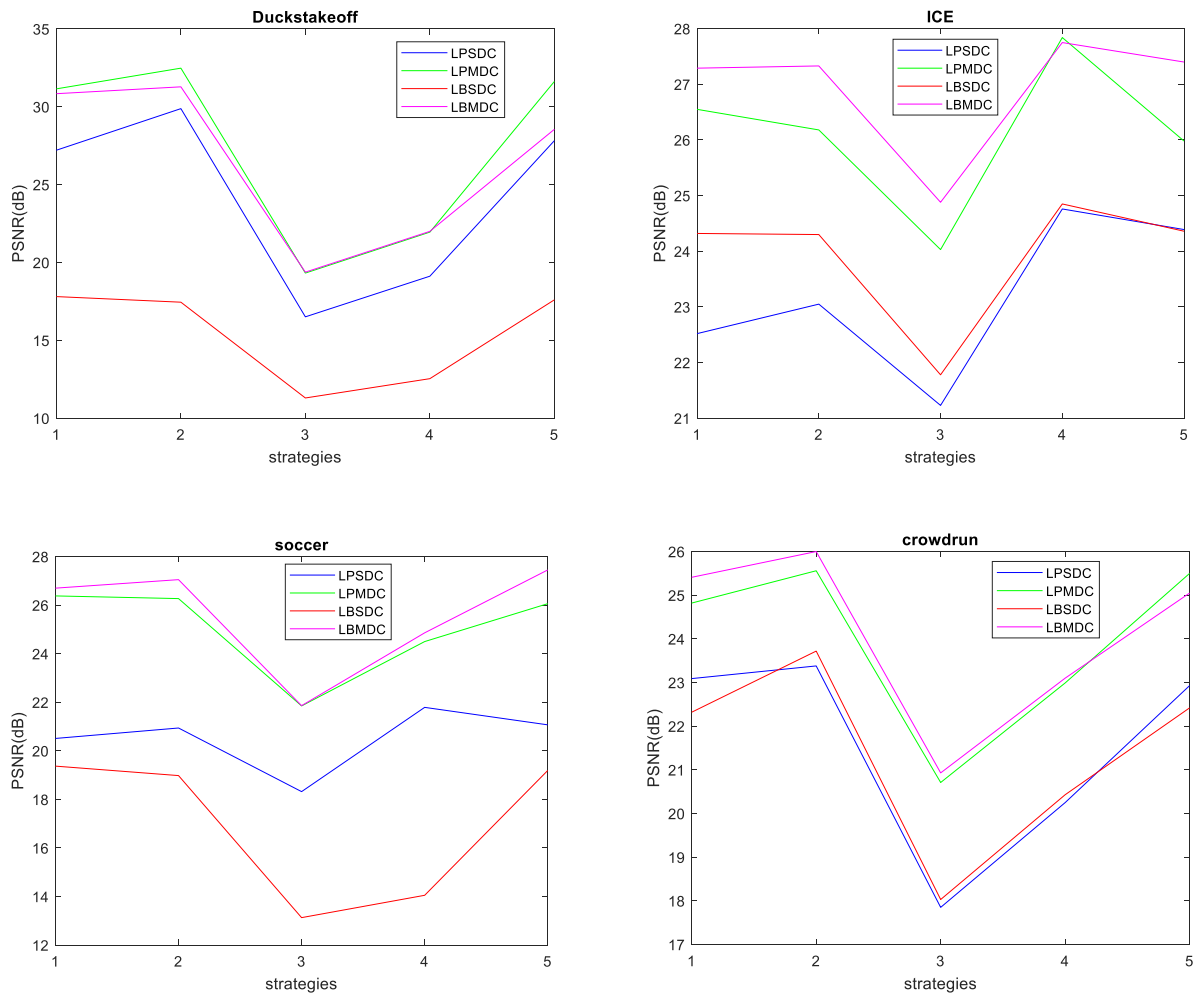


Figure 34. Impact des différentes stratégies sur la qualité vidéo avec les schémas SDC et MDC en termes de PSNR (QP=22).

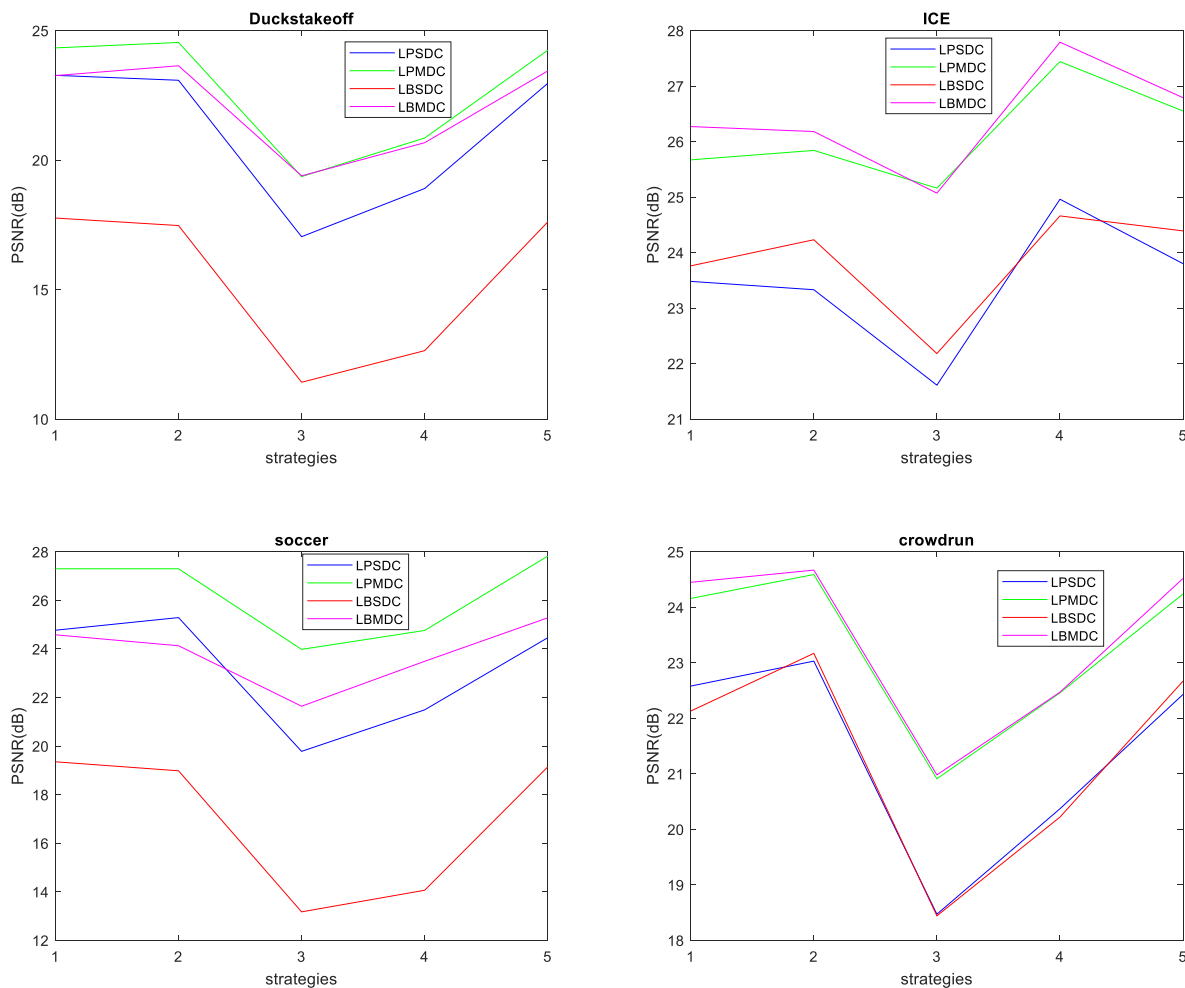
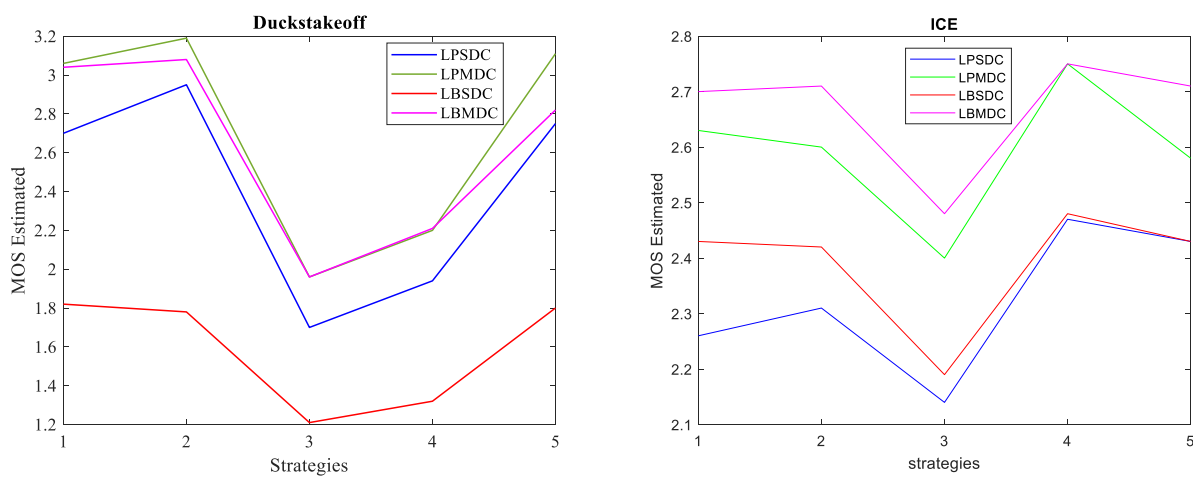


Figure 35. Impact des différentes stratégies sur la qualité vidéo avec les schémas SDC et MDC en termes de PSNR (QP=37).



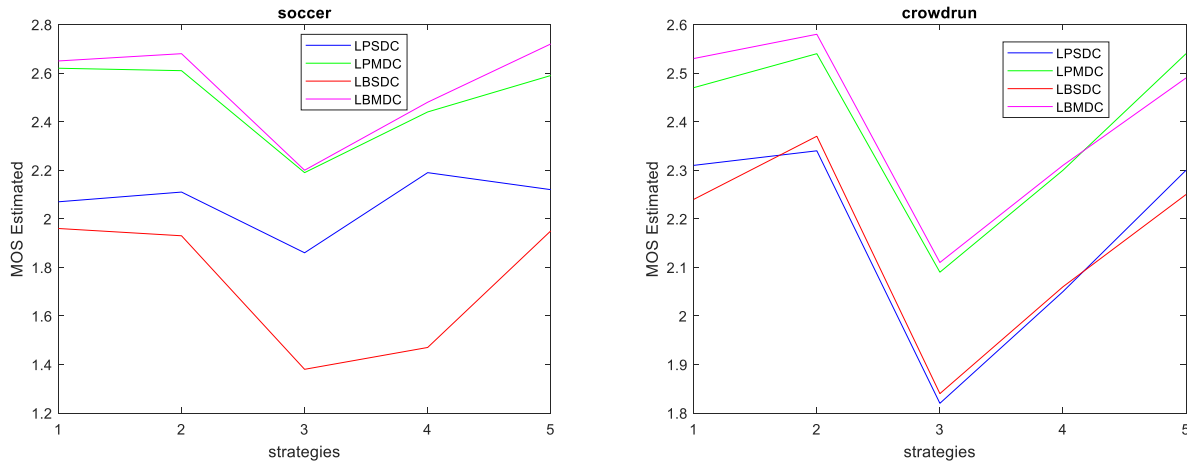


Figure 36. Impact des différentes stratégies sur la qualité vidéo avec les schémas SDC et MDC en termes de MOS\_Estimated (QP=22).

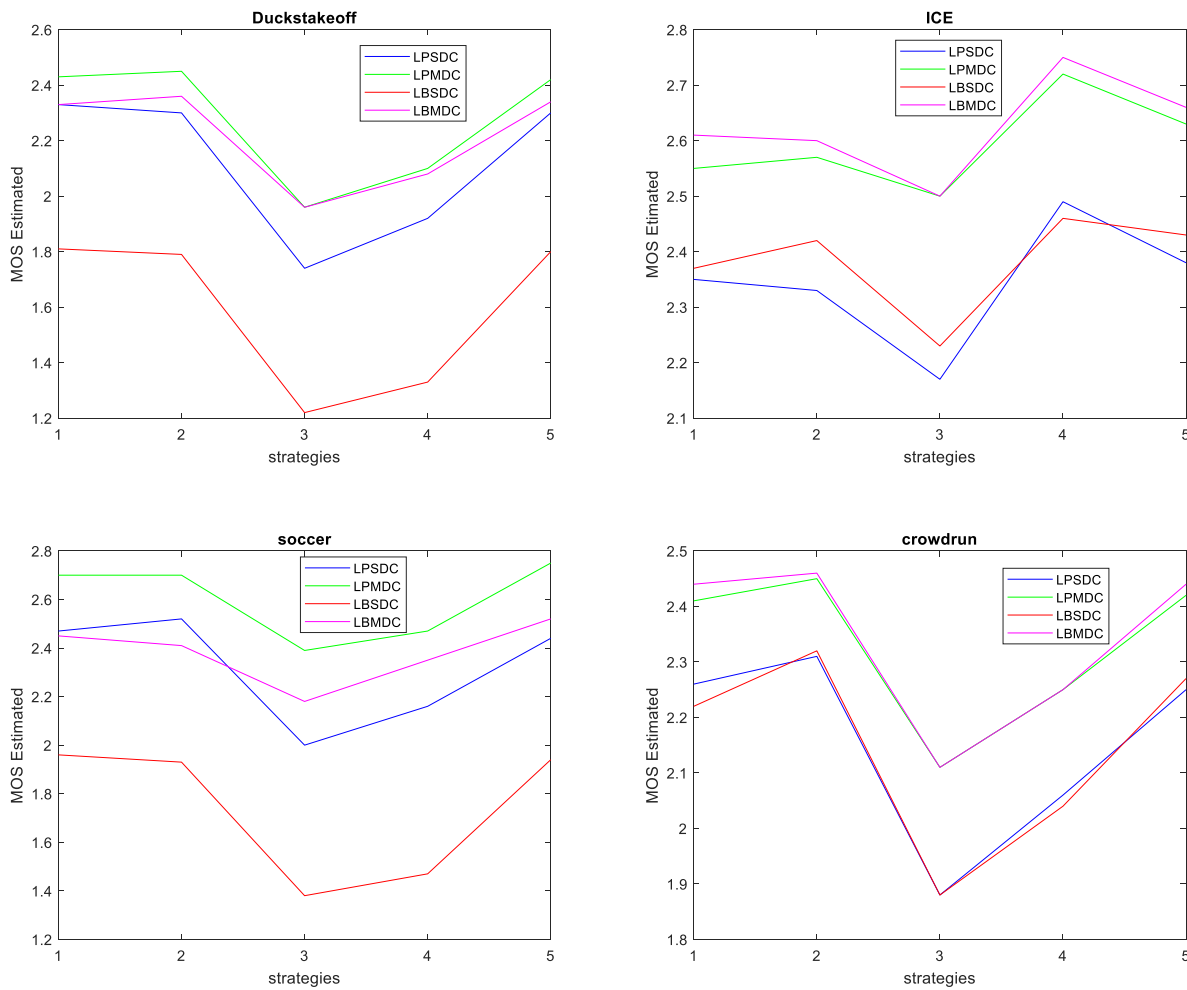


Figure 37. Impact des différentes stratégies sur la qualité vidéo avec les schémas SDC et MDC en termes de MOS\_Estimated (QP=37).

Selon les résultats, on peut en déduire que :

La valeur de QP et le contenu vidéo affectent directement le PSNR et le MOS\_Estimated et ont un impact considérable sur le maintien d'une bonne qualité vidéo. D'après les figures précédentes, on peut observer que les vidéos Ice et Soccer, qui présentent moins d'activités de mouvement, sont les plus résistantes à la perte de paquets. Les séquences Duckstakeoff et Crowdrun, qui ont un SI (Indice Spatial) et un TI (Indice Temporel) élevés, sont plus sensibles à la perte de paquets. Il est évident qu'une petite valeur de QP donne un PSNR et un MOS\_Estimated élevés, tandis qu'une valeur élevée de QP donne un PSNR et un MOS\_Estimated faibles, et cela pour toutes les séquences vidéo.

Pour démontrer l'importance de la localisation de la perte au niveau des trames au sein du même GOP, la stratégie 1 sera comparée à la stratégie 2. Pour toutes les séquences et les deux valeurs de QP utilisées, il a été constaté que la perte de la 4ème trame du GOP a un effet plus important sur la qualité de la vidéo reconstituée que la perte de la première trame. La raison en est que dans la configuration à faible délai P ou B, la 4ème trame du GOP a un débit binaire plus élevé que les autres et cette même trame est utilisée dans la prédiction des autres trames, de sorte que sa perte génère une propagation d'erreur aux autres trames.

En comparant les stratégies 3 et 5, il a été observé que la stratégie 3, qui consiste à éliminer le GOP entier, affecte davantage la qualité de la vidéo reconstituée que la stratégie qui consiste à éliminer uniformément 4 trames, malgré le même nombre de trames perdues, c'est-à-dire 4 trames, ce qui est le résultat du décodeur HM modifié avec la copie de mouvement comme dissimulation d'erreur. La technique ne remplace que la dernière trame lorsqu'il s'agit d'une perte en rafale, entraînant ainsi la perte des trois autres trames, tandis que dans le cas d'une perte uniforme, le décodeur remplace toutes les trames perdues. Les Figures 35, 36, 37 et 38 démontrent que la stratégie 3 est le pire cas pour les différentes vidéos utilisées dans les expériences.

Contrairement à la stratégie 3, la stratégie 4 donne les meilleurs résultats car la perte est localisée au début de la vidéo, tandis que dans la stratégie 4, la perte est localisée à la fin de la vidéo, ce qui provoque la propagation de l'erreur tout au long de la séquence en raison de la forte dépendance entre les trames dans la configuration à faible délai P ou B. Les résultats montrent que pour le même taux de perte, nous pouvons obtenir différentes qualités de vidéo reconstituée.

Pour prouver l'efficacité du MDC, l'impact de la perte complète de trame sur la qualité de la vidéo reconstituée basée sur le MDC a été comparé à celle de la qualité de la vidéo reconstituée codée en SDC. Les résultats en termes de PSNR et de MOS\_Estimated pour les 2 valeurs de QP (22 et 37) montrent une nette amélioration dans le cas du MDC avec un gain important en PSNR, ce qui prouve l'efficacité de cette approche pour lutter contre la perte de paquets. Selon le Tableau 8 qui résume les

résultats en termes du pourcentage de gain moyen, la méthode MDC utilisée dans ce travail permet un gain de 14% pour QP=22 et 13% pour QP=37 en mode LDP. Les gains de qualité sont encore plus importants en mode LDB avec 35,57% pour QP=22 et 30,32% pour QP=37.

**Tableau 8.** Pourcentage moyen de gain pour les 4 séquences.

Configuration	LDP		LDB	
	QP		QP	
Video	22	37	22	37
Duckstakeoff	14%	8.1%	72.3%	47%
Ice	13%	12%	13%	11%
Soccer	22%	23%	44%	53.3%
Crowdrun	7%	16%	13%	10%
Mean value	14%	13%	35.57%	30.32%

Dans cette étude [100], l'impact des paramètres d'encodage tels que la valeur QP choisie, le contenu vidéo et la méthode de correction d'erreur du côté du décodeur, ainsi que la perte qui affecte le flux de bits sous différentes formes en rafales ou de manière uniforme, a été analysé. Les résultats expérimentaux ont montré que chacun des facteurs mentionnés ci-dessus joue un rôle important dans la qualité de la vidéo reconstruite. De plus, une méthode MDC spatiale permettant de générer 2 flux de bits à partir de la vidéo originale a été proposée pour minimiser la dégradation de la qualité vidéo causée par le phénomène de perte de paquets. Les résultats en termes de PSNR (Peak Signal-to-Noise Ratio) et MOS\_Estimated (MOS estimé) ont été satisfaisants, et la méthode proposée a permis un gain moyen de 14 % pour QP = 22, 13 % pour QP = 37 en mode LDP, 35,57 % pour QP = 22 et 30,32 % pour QP = 37 en mode LDB. Le gain est plus important pour les vidéos encodées avec la configuration LDB.

#### 4.4. Amélioration de la qualité video codé en HEVC basé sur l'apprentissage profond

L'utilisation croissante des méthodes d'apprentissage profond, en particulier des réseaux neuronaux convolutifs profonds (CNN), a considérablement propulsé le domaine de la compression vidéo et a ouvert la voie au développement de réseaux dédiés à l'élimination des artefacts, basés sur des architectures CNN. Ces réseaux peuvent être intégrés en tant que filtres en boucle ou utilisés en post-traitement. D'ailleurs, des recherches récentes ont suggéré l'application de filtres en boucle basés sur les CNNs [99] [100] [101] [102]. Alors que les filtres en boucle requièrent une prise en compte à la fois lors de l'encodage et du décodage, les approches de post-traitement se montrent plus adaptables puisqu'elles sont appliquées après l'étape de décodage. La Figure 39 illustre une approche de post-

traitement basée sur un CNN. Dans ce travail, nous proposons une méthode de post-traitement basée sur les réseaux neuronaux pour améliorer la qualité des images décodées.

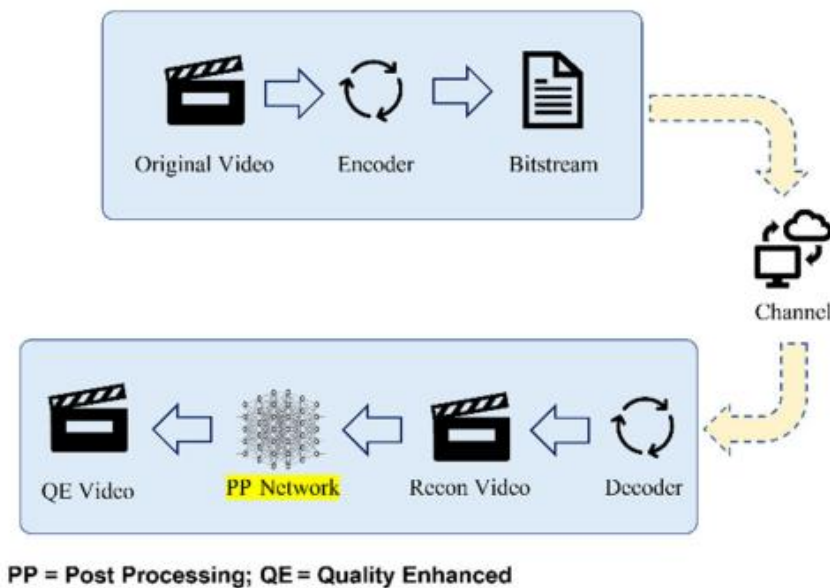


Figure 38. Approche de post-traitement intégrée basée sur l'apprentissage profond dans le flux de travail de codage typique. La couleur jaune indique la focalisation de ce travail [101].

#### 5.4.1 Méthode proposée

Dans cette section, l'algorithme et l'architecture proposés sont présentés et expliqués. La Figure 39 illustre une approche de post-traitement basée sur un CNN intégrée dans un pipeline de traitement conventionnel. Dans ce pipeline de post-traitement, le flux de bits transmis est décodé pour produire des images reconstruites, et un modèle d'apprentissage profond est appliqué pour améliorer la qualité vidéo des images reconstruites. Nous avons choisi l'auto-encodeur qui est utilisé pour la segmentation [102] ainsi que pour le débruitage et l'amélioration de la qualité d'images [103].

##### 5.4.1.1 L'auto encodeur convolutionnel

Les auto encodeurs sont un type de réseau neuronal qui est formé de manière non supervisée pour apprendre la représentation compressée des données brutes. La figure 1 présente la représentation de base de l'auto encodeur. Un auto encodeur est composé de trois parties : l'encodeur, le bottleneck (point de compression) et le décodeur. L'encodeur compressé l'image en une dimension inférieure et le décodeur reconstitue l'image à partir de cette représentation. La dimension compressée, également appelée point de compression, est la plus basse possible des données d'entrée. L'architecture de l'auto-encodeur utilisée pour l'expérience est présentée dans la figure 40 et la figure 41.

Le modèle est composé de 10 couches de convolutions, 5 couches pour l'encodeur et 5 couches de convolution transposée coté décodeur. La taille des filtres et le nombre de paramètres utilisé dans chaque couche est présenté dans le tableau 9.

Dans ce modèle la couche de convolution se termine par la couche de correction ou la fonction d'activation 'rectified linear unit' en anglais (ReLU) a été utilisé pour remplacer toute les valeurs négative par des zéros, la fonction est définie comme suit :

$$f(x) = \max(0, x) \quad (5)$$

"x" représente l'entrée du neurone, et la fonction ReLU renvoie soit la valeur de  $x$  si  $x$  est positive, soit zéro si  $x$  est négatif.

La fonction de perte 'Mean Squared Error' en anglais (MSE) a été utilisé pour mesurer la qualité des prédictions du modèle par rapport au valeurs attendues.

L'optimiseur Adam, ou Adaptive Moment Estimation en anglais a été utilisé dans l'entraînement du modèle pour permettre de mettre à jour les poids des réseaux de neurones de manière à minimiser la fonction de perte.

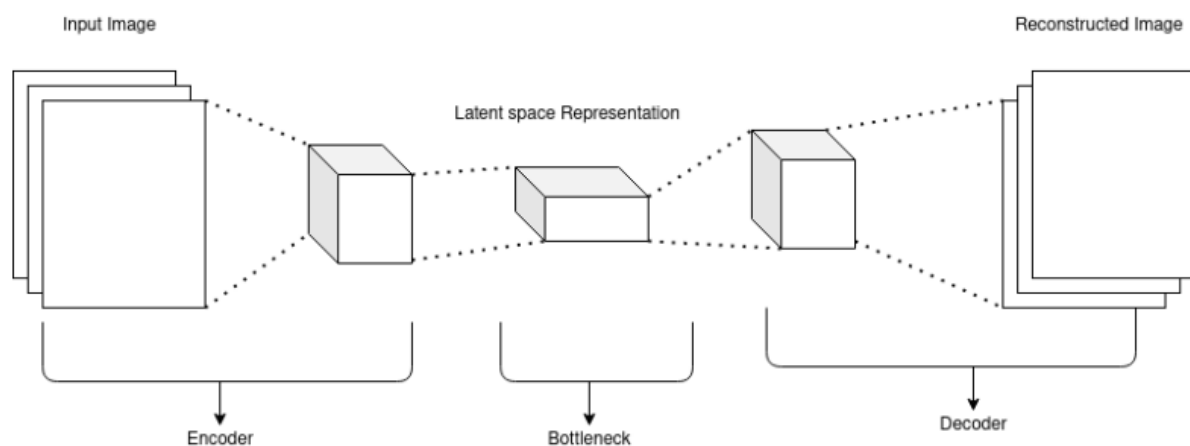


Figure 39. Architecture de l'auto-encodeur

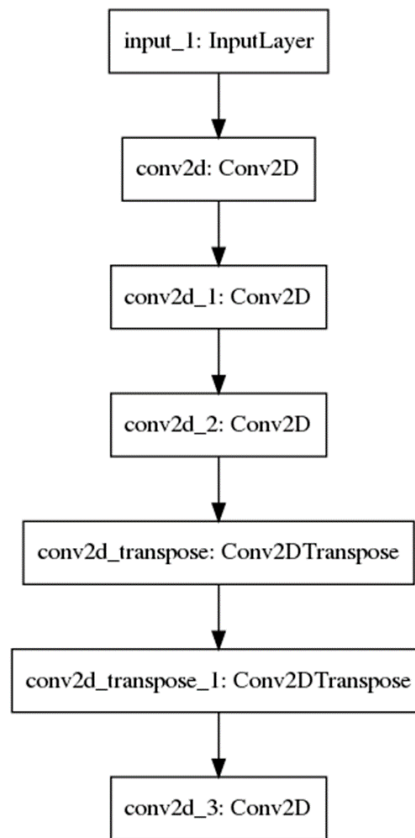


Figure 40. Organigramme de l'auto-encodeur utilisé

**Tableau 9.** Informations détaillées sur le modèle utilisé

Type de couche	Taille de l'image de sortie	nombre de filtres	Nombre de paramètres
input_1	704x576	3	0
conv2d	704x576	16	448
conv2d_1	704x576	32	4640
conv2d_2	704x576	64	18496
conv2d_3	704x576	128	73856
conv2d_4	704x576	128	147584
conv2d_transposé	704x576	128	147584
conv2d_transposé_1	704x576	64	73792
conv2d_transposé_2	704x576	32	18464
conv2d_transposé_3	704x576	16	4624
conv2d_transposé_4	704x576	3	435
Total paramètres			489923

#### 5.4.1.2 Apprentissage

Dans ce travail, l'apprentissage supervisé a été utilisé comme une méthode fondamentale pour la formation du modèle. Cinq vidéos issues de la base de données "Retrieved" ont été sélectionnées pour la phase d'apprentissage, à savoir les vidéos intitulées "Ice", "Duckstakeoff", "Soccer", "Crowdrun" et "Harbour". Chacune de ces vidéos a été encodée en utilisant le codec HEVC (HM 16.11) avec quatre valeurs de QP (22, 27, 32, 37) en mode low delay P. de plus une perte de paquets d'un taux de 3% a été introduite aux vidéos.

Au total une base de 10720 images a été construite entre images originales et images avec artefacts.

Une partie de ces images a été choisie pour faire l'apprentissage en se basant sur la méthode de validation croisé K-Fold, couramment utilisée pour évaluer la performance d'un modèle d'apprentissage automatique [104]. Elle permet de diviser un ensemble de données en k sous-ensembles (appelés plis ou "folds"), puis d'entraîner et de tester le modèle, en utilisant un pli différent comme ensemble de test à chaque itération (epoch). Cette approche permet d'obtenir une estimation plus fiable de la performance du modèle, en réduisant les biais potentiels liés à la division des données. Le training a été fait avec deux approches une appelée 'few data training' et l'autre 'big data training' dans few data training la taille du sous ensemble est de 35 images c'est-à-dire 35 images de chaque vidéo compressé avec une valeur différente de QP dans la même vidéo, tandis que pour big data 60 images ont été pris en compte. Ce qui fait un total de 700 images et 1200 images dans few data training et big data training respectivement, 20 % de ce totale ont été réservé pour la validation de l'apprentissage, chaque approche est encore entraînée avec 50 itérations et 150 itérations. Idem pour les images des vidéos originales puisque c'est un système supervisé. Ce qui fait qu'au totale 4 modèles ont été créé.

En ce qui concerne le test, la vidéo complète a été utilisé pour tester le modèle. Durant la phase de l'apprentissage l'image complète a été exploité contrairement à d'autres travaux où l'image a été divisé en block de 35x35 [105] ou bien 65x65 sub-images [106] par faute de mémoire.

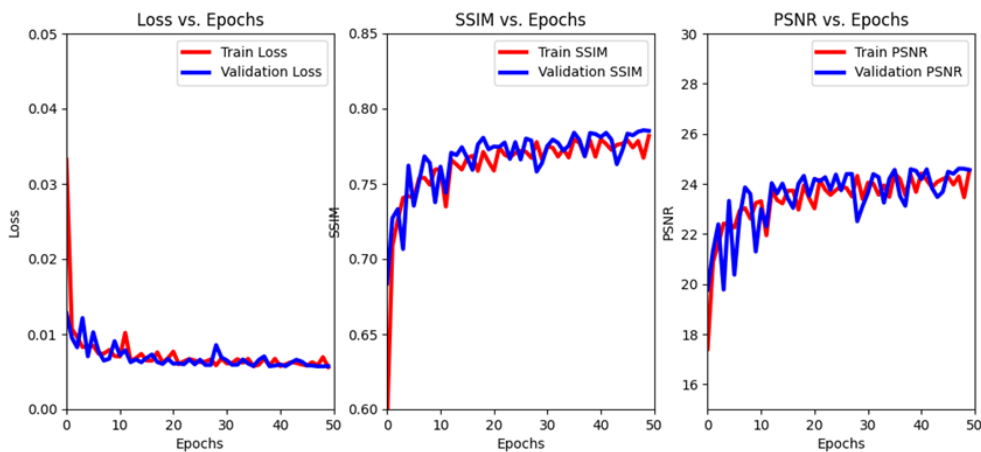
#### 5.4.1.3 Implémentation (logiciels et librairies)

Pour la mise en œuvre de nos modèles, nous avons opté pour le langage de programmation Python, reconnu pour sa polyvalence et sa richesse en bibliothèques dédiées à l'apprentissage automatique. Plus spécifiquement, nous avons utilisé les bibliothèques Keras et TensorFlow [107]. Keras offre une interface conviviale pour la création de réseaux de neurones et est largement utilisé pour le développement rapide de modèles d'apprentissage profond. TensorFlow, quant à lui, fournit une

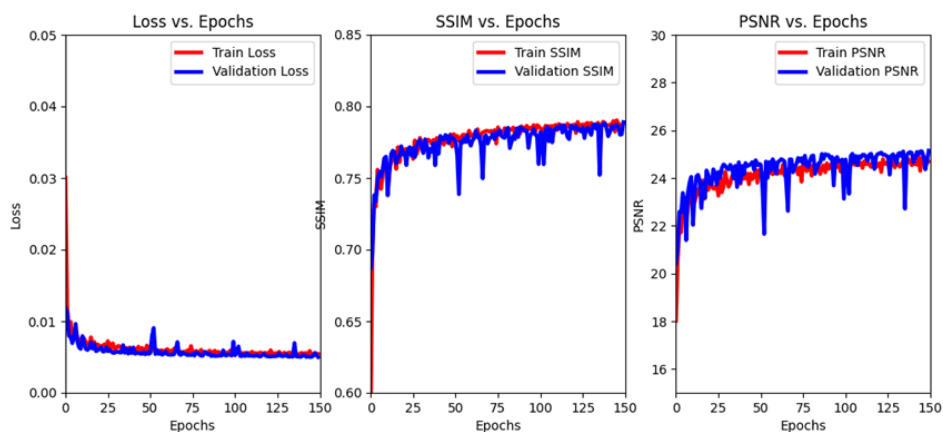
infrastructure puissante pour la construction et l'entraînement de réseaux de neurones, offrant une grande flexibilité et des performances élevées. Et une carte graphique GeForce GTX 1080 Ti conçue pour les applications de calcul intensives, dispose de 3584 cœurs CUDA et 11Go de mémoire GDDR5X.

#### 5.4.2 Résultat et discussion de l'apprentissage

Les métriques PSNR, SSIM (Structural Similarity Index) et la perte (loss) ont été employées pour évaluer les résultats de l'apprentissage. En fonction du nombre d'itération (epochs).

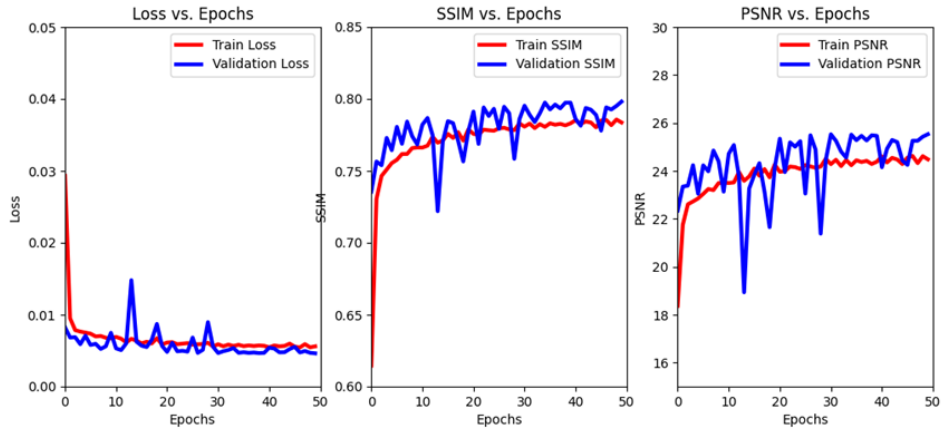


(1)

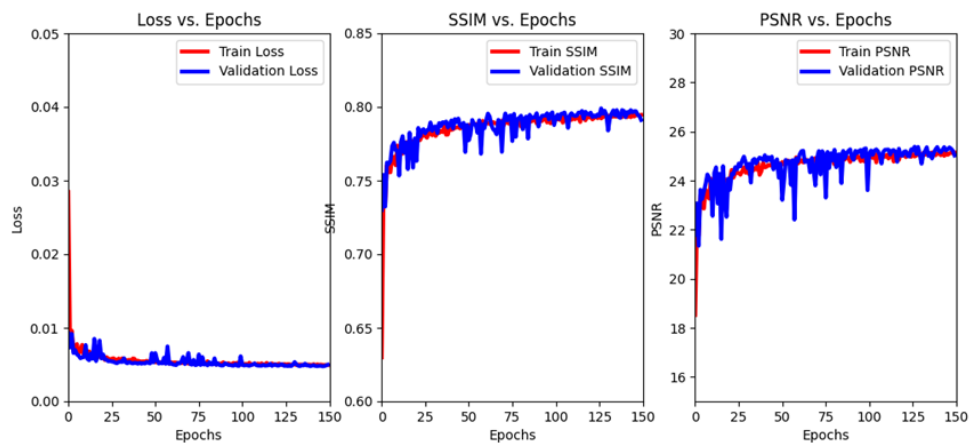


(2)

Figure 41. Résultat de l'apprentissage avec (modèle 1 : 50, modèle 2 : 150) itérations en utilisant 700 images (few data).



(3)



(4)

Figure 42. Résultat de l'apprentissage avec (modèle 3 : 50, modèle 4 :150) itérations en utilisant 1200 images (Big data).

D'après la figure 42 et la figure 43 :

**Modèle 1 (Few Data, 50 itérations):**

- Loss vs. Epochs: La courbe de perte pour l'entraînement diminue de manière significative, tandis que la perte de validation a des fluctuations, ce qui peut suggérer un possible surapprentissage.
- SSIM vs. Epochs: Le SSIM d'entraînement augmente alors que le SSIM de validation fluctue, indiquant une possible divergence entre les données d'entraînement et de validation.
- PSNR vs. Epochs: Le PSNR montre une amélioration, mais il y a des écarts notables entre les courbes d'entraînement et de validation.

### **Modèle 2 (Few Data, 150 itérations):**

- Loss vs. Epochs: La courbe de perte d'entraînement semble converger alors que la validation montre des fluctuations, indiquant potentiellement du surapprentissage.
- SSIM vs. Epochs: Les courbes SSIM d'entraînement et de validation sont proches l'une de l'autre avec une tendance à la hausse.
- PSNR vs. Epochs: Le PSNR d'entraînement et de validation montre une tendance à la hausse avec quelques fluctuations.

### **Modèle 3 (Big Data, 50 itérations):**

- Loss vs. Epochs: La courbe de perte d'entraînement diminue rapidement au début puis se stabilise, tandis que la perte de validation connaît des fluctuations mais reste relativement stable. Cela suggère une convergence rapide du modèle.
- SSIM vs. Epochs: Le SSIM d'entraînement et de validation semble suivre une tendance à la hausse, ce qui est bon. Toutefois, la validation SSIM a des fluctuations.
- PSNR vs. Epochs: Le PSNR d'entraînement et de validation augmente également, indiquant une meilleure qualité de reconstruction.

### **Modèle 4 (Big Data, 150 itérations):**

- Loss vs. Epochs: La courbe de perte montre une convergence stable à long terme pour l'entraînement et la validation. Le modèle ne semble pas surapprendre malgré un plus grand nombre d'itérations.
- SSIM vs. Epochs: Le SSIM d'entraînement et de validation continue d'augmenter, suggérant une amélioration continue de la qualité de la reconstruction.
- PSNR vs. Epochs: Le PSNR pour l'entraînement et la validation montre également une tendance à la hausse avec quelques fluctuations.

D'après les résultats on peut conclure que :

- Les modèles entraînés sur big data montrent de meilleures performances globales par rapport à ceux entraînés sur peu de données. C'est attendu car davantage de données d'entraînement offrent généralement une meilleure généralisation et performance.
- L'augmentation du nombre d'itérations de 50 à 150 a un impact positif sur la performance. Cela est particulièrement visible dans les graphiques SSIM et PSNR.

- Étant donné que le modèle contient 10 couches de convolution, il est possible que l'entraînement avec plus d'itérations permette une meilleure convergence, surtout si vous utilisez big data.
- Pour les modèles avec peu de données, l'augmentation des itérations ne semble pas avoir un impact aussi prononcé, ce qui suggère qu'ils peuvent avoir atteint leur performance optimale en termes de données disponibles.

### 5.4.3 Résultat et discussion des tests

#### Few data training data = 700 frames

#### Big data training data = 1200 frames

**Tableau 10.** Résultat des 2 approches 'few data' et 'big data' avec 50 et 150 itérations.

Fe w dat a	PSNR (dB)																				
	Crowd				Ducks				Harbour				Ice				Soccer				
	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	
	22	27	32	37	22	27	32	37	22	27	32	37	22	27	32	37	22	27	32	37	
50	30,42	28,50	26,31	24,10	29,40	27,16	25,19	23,26	35,35	32,98	30,05	27,50	38,88	37,28	35,29	33,00	35,15	34,22	31,61	29,35	
150	31,01	28,86	26,54	24,24	30,80	27,71	25,57	23,50	35,70	33,31	30,31	27,31	39,03	37,50	35,57	33,28	36,63	34,55	32,13	29,69	
Big dat a	PSNR (dB)																				
	Crowd				Ducks				Harbour				Ice				Soccer				
	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	QP	
	22	27	32	37	22	27	32	37	22	27	32	37	22	27	32	37	22	27	32	37	
50	30,81	28,70	26,41	24,15	29,98	27,35	25,29	23,27	35,36	33,03	30,10	27,53	38,94	37,33	35,35	33,06	35,92	34,32	31,77	29,45	
150	31,33	28,99	26,61	24,28	31,03	27,85	25,60	23,59	35,87	33,45	30,32	27,77	39,19	37,61	35,64	33,32	36,81	34,65	32,18	29,71	
Gain moyen en %	<b>50</b>	<b>1,29</b>	<b>0,68</b>	<b>0,38</b>	<b>0,19</b>	<b>1,97</b>	<b>0,67</b>	<b>0,37</b>	<b>0,01</b>	<b>0,03</b>	<b>0,15</b>	<b>0,15</b>	<b>0,10</b>	<b>0,15</b>	<b>0,13</b>	<b>0,16</b>	<b>0,17</b>	<b>2,19</b>	<b>0,27</b>	<b>0,59</b>	<b>0,33</b>
	<b>150</b>	<b>1,02</b>	<b>0,47</b>	<b>0,24</b>	<b>0,10</b>	<b>0,76</b>	<b>0,51</b>	<b>0,11</b>	<b>0,37</b>	<b>0,45</b>	<b>0,44</b>	<b>0,05</b>	<b>0,25</b>	<b>0,41</b>	<b>0,29</b>	<b>0,20</b>	<b>0,10</b>	<b>0,48</b>	<b>0,28</b>	<b>0,15</b>	<b>0,09</b>

**Tableau 11.** Comparaison des résultats avec les données initiales c'est les vidéos avec perte de paquets PLR= 3%.

	PSNR en dB avec PLR=3%				Gain moyen en %			
QP	22	27	32	37	22	27	32	37
crowd	28,02	27,61	24,08	21,48	10,25	4,17	9,91	12,65
Ducks	27,27	25,76	24,12	20,14	11,12	6,82	5,35	16,20
Harbour	34,6	31,35	29	29,1	2,80	5,87	4,12	5,83
Ice	38,1	35	32,9	30,09	2,39	6,94	7,78	10,22
Soccer	34,9	33,07	29,41	28	3,52	4,13	8,53	5,53

D'après le tableau 10 on peut dire que Les modèles Big Data semblent avoir de meilleurs résultats que ceux entraînés avec Few Data, en particulier lorsque le nombre d'itérations augmente. Le modèle entraîné avec Big Data et 150 itérations semble offrir le meilleur PSNR, en particulier pour la vidéo

"Ice". En ce qui concerne l'analyse du Gain moyen en après reconstitution par l'auto encodeur ou les résultats sont présenté dans le tableau 11 on peut dire que :

- "Duckstakeoff" présente le gain moyen le plus élevé à QP 37 avec 16,20%, indiquant une amélioration substantielle après la reconstitution.
- "Ice" a le gain le plus bas à QP 22 avec 2,39%.
- Pour la vidéo "Crowd", le gain varie de 4,17% à 12,65%.
- "Duckstakeoff" varie de 6,82% à 16,20%.
- Les autres vidéos montrent également des variations notables, avec des gains tendant à augmenter à mesure que QP augmente.
- L'auto-encodeur semble apporter des améliorations notables, en particulier à des valeurs de QP élevées (comme QP 37). Cela est mis en évidence par le gain moyen en pourcentage, qui montre comment la qualité est améliorée après la reconstitution.

Les résultats de l'apprentissage mettent en lumière l'importance de la taille de l'ensemble de données et du nombre d'itérations dans l'entraînement des auto-encodeurs. Ils soulignent également la nécessité de trouver un équilibre entre la complexité du modèle, la quantité de données et le nombre d'itérations pour obtenir des performances optimales.

Les résultats des tests suggèrent que l'auto-encodeur apporte des améliorations notables à la qualité de la reconstitution vidéo, en particulier à des valeurs de QP plus élevées, comme QP 37. Le gain moyen en pourcentage souligne clairement comment la qualité est améliorée après la reconstitution, offrant des perspectives positives pour l'application de ces modèles dans des contextes de compression vidéo.

## 5.5 Conclusion

Dans ce dernier chapitre de la thèse, les deux aspects essentiels liés à l'amélioration de la transmission vidéo sur un réseau sans fil en utilisant la norme de compression HEVC ont été abordés. Tout d'abord, l'impact de la perte de paquets sur les vidéos encodées en HEVC a été analysé, mettant en lumière les conséquences potentielles de ces pertes sur la qualité vidéo perçue par l'utilisateur final. Cette analyse approfondie a permis une meilleure compréhension des problèmes liés à la transmission de vidéos HEVC dans un environnement sans fil, tout en identifiant des solutions potentielles pour atténuer ces effets indésirables. Une technique de codage source appelée le codage à description multiple spatiale a été utilisée pour protéger la vidéo contre la perte

de paquets. De plus, une approche d'apprentissage profond en utilisant l'auto-encodeur a été proposée pour améliorer la qualité des vidéos affectées par une perte de paquets, dans notre cas où le PLR est de 3%.

Cette recherche répond à la demande croissante de vidéos de haute qualité sur les réseaux IP, en soulignant l'importance de l'innovation dans la transmission vidéo et en proposant des solutions concrètes pour garantir une expérience vidéo optimale. Ces travaux ouvrent la porte à de futures recherches et développements dans ce domaine en constante évolution.

## Conclusion et perspectives

Dans le cadre de cette thèse de doctorat, nous avons examiné de manière approfondie les défis et les solutions liés à la protection des données de streaming vidéo dans les réseaux sans fil. Notre recherche a débuté par une analyse minutieuse de l'impact des pertes de paquets sur la qualité vidéo reçue, mettant en évidence l'importance cruciale de la fiabilité de la transmission dans les environnements sans fil. Ensuite, nous avons introduit une contribution novatrice en proposant une méthode de protection de la vidéo transmise à travers un canal bruité grâce au codage à description multiple. Cette approche, en agissant au niveau de la source, a démontré son efficacité pour renforcer la résilience de la vidéo contre les perturbations du canal.

De plus, notre recherche s'est étendue à l'amélioration de la qualité vidéo reçue, où nous avons exploré l'application des techniques d'apprentissage profond. Notre contribution à cet égard a été l'intégration d'un auto-encodeur convolutionnel, une méthode de l'apprentissage profond, pour améliorer la qualité vidéo après la transmission. Les résultats ont montré que cette approche a le potentiel de significativement accroître la qualité perçue des vidéos, même dans des scénarios où la transmission peut être altérée.

Alors que cette thèse a jeté les bases pour la protection des données de streaming vidéo dans les réseaux sans fil, de nombreuses perspectives de recherche restent ouvertes pour l'avenir :

**Apprentissage profond avancé :** Poursuivre les recherches sur des architectures d'apprentissage profond plus avancées, telles que les réseaux adverses génératifs (GANs), pour une amélioration plus fine de la qualité vidéo.

**Applications spécifiques :** Appliquer les méthodes développées dans cette thèse à des domaines spécifiques tels que la télévision mobile, la surveillance vidéo sans fil et la réalité augmentée/virtuelle.

## Références

- [1] V. Cisco, "Cisco visual networking index: Forecast and trends, 2017–2022 white paper," *Cisco Internet Report*, vol. 17, p. 13, 2019.
- [2] Z. R. Alashhab, M. Anbar, M. M. Singh, Y.-B. Leau, Z. A. Al-Sai, and S. A. Alhayja'a, "Impact of coronavirus pandemic crisis on technologies and cloud computing applications," *Journal of Electronic Science and Technology*, vol. 19, p. 100059, 2021.
- [3] S. Hojati, M. Kazemi, and P. Moallem, "Error concealment with parallelogram partitioning of the lost area," *Multimedia Tools and Applications*, vol. 79, pp. 7449-7469, 2020.
- [4] G. Kulupana, D. S. Talagala, H. K. Arachchi, and A. Fernando, "End user video quality prediction and coding parameters selection at the encoder for robust HEVC video transmission," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, pp. 3367-3381, 2018.
- [5] D. Soldani, M. Li, and R. Cuny, *QoS and QoE management in UMTS cellular systems*: John Wiley & Sons, 2007.
- [6] I. Rec, "P. 800.1, mean opinion score (mos) terminology," *International Telecommunication Union, Geneva*, 2006.
- [7] M. Kazemi, M. Ghanbari, and S. Shirmohammadi, "A review of temporal video error concealment techniques and their suitability for HEVC and VVC," *Multimedia Tools and Applications*, vol. 80, pp. 12685-12730, 2021.
- [8] A. Kramer, "Improving communication reliability by use of an intermittent feedback channel," *IEEE Transactions on Information Theory*, vol. 15, pp. 52-60, 1969.
- [9] P. P. Garrido Abenza, M. P. Malumbres, P. Piñol, and O. López-Granado, "Source coding options to improve HEVC video streaming in vehicular networks," *Sensors*, vol. 18, p. 3107, 2018.
- [10] H. Chen, C. Zhao, M.-T. Sun, and A. Drake, "Adaptive intra-refresh for low-delay error-resilient video coding," *Journal of Visual Communication and Image Representation*, vol. 31, pp. 294-304, 2015.
- [11] C. Lin, Y. Zhao, T. Tillo, and J. Xiao, "Multiple description coding for stereoscopic videos with stagger frame order," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, pp. 1016-1025, 2014.
- [12] M. Kazemi, R. Iqbal, and S. Shirmohammadi, "Redundancy allocation based on the weighted mismatch-rate slope for multiple description video coding," *IEEE Transactions on Multimedia*, vol. 19, pp. 54-66, 2016.
- [13] M. Kazemi, S. Shirmohammadi, and K. H. Sadeghi, "A review of multiple description coding techniques for error-resilient video delivery," *Multimedia Systems*, vol. 20, pp. 283-309, 2014.
- [14] F. Boumehrez, R. Brai, N. Doghmane, and K. Mansouri, "Quality of experience enhancement of high efficiency video coding video streaming in wireless packet networks using multiple description coding," *Journal of electronic imaging*, vol. 27, pp. 013028-013028, 2018.
- [15] N. Ibrahim, A. Sali, H. A. Karim, A. Ramli, N. Ibrahim, and D. Grace, "Multiple description coding for enhancing outage and video performance over relay-assisted cognitive radio networks," *IEEE Access*, vol. 10, pp. 11750-11762, 2022.
- [16] M. Kazemi, "End-to-end Distortion Modeling and Channel Adaptive Optimization of Mixed Layer Multiple Description Coding Scheme," *Computational Intelligence in Electrical Engineering*, vol. 12, pp. 31-42, 2021.
- [17] K. Bajaj, D. K. Singh, and M. A. Ansari, "Autoencoders based deep learner for image denoising," *Procedia Computer Science*, vol. 171, pp. 1535-1541, 2020.
- [18] T. Larrue, Y. Li, X. Meng, and C.-M. Han, "Denoising Videos with Convolutional Autoencoders," 2018.
- [19] T. Amestoy, "Optimisation du Codec VVC basé sur la Réduction de Complexité et le Traitement Parallèle," INSA de Rennes, 2021.
- [20] Z. Shang, J. P. Ebenezer, Y. Wu, H. Wei, S. Sethuraman, and A. C. Bovik, "Study of the subjective and objective quality of high motion live streaming videos," *IEEE Transactions on Image Processing*, vol. 31, pp. 1027-1041, 2021.

- [21] I. T. Union, "ITU-T Recommendation H. 261 Version 1: Video Codec for Audiovisual Services at P× 64 Kbits," *Technical Report ITU-T*, p. E 1934, 1990.
- [22] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1649-1668, 2012.
- [23] I. O. f. S. I. E. Commission, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s," *ISO/IEC 11172*, 1993.
- [24] S. CCITT and H. Recommendation, "261–Video Codec for Audiovisual Services at px64 kbit/s," *The International Telegraph and Telephone Consultative Committee*, vol. 115, 1990.
- [25] K. Rijkse, "H. 263: Video coding for low-bit-rate communication," *IEEE Communications magazine*, vol. 34, pp. 42-45, 1996.
- [26] I. Rec, "H. 262 | ISO/IEC 13818-2," *Information technology—Generic coding of moving pictures and associated audio information—Video*, 2000.
- [27] I. Richardson, "White paper: an overview of H. 264 advanced video coding," *Vcodex/OneCodec*, vol. 2011, 2007.
- [28] B. Bross, K. Andersson, M. Bläser, V. Drugeon, S.-H. Kim, J. Lainema, J. Li, S. Liu, J.-R. Ohm, and G. J. Sullivan, "General video coding technology in responses to the joint call for proposals on video compression with capability beyond HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, pp. 1226-1240, 2019.
- [29] A. Nasrallah, "Nouvelles techniques de compression pour le codage vidéo prochaine-génération," Institut Polytechnique de Paris, 2021.
- [30] M. Wien, "High efficiency video coding," *Coding Tools and specification*, vol. 24, 2015.
- [31] J. R. Atencia, O. L. Granado, M. P. Malumbres, M. O. Martínez-Rach, and G. Van Wallendael, "Analysis of the perceptual quality performance of different HEVC coding tools," *IEEE Access*, vol. 9, pp. 37510-37522, 2021.
- [32] S. Gramatikov, "Analysis of server and network performance for HTTP-based streaming," 2017.
- [33] N. Ramzan, E. Quacchio, T. Zgaljic, S. Asioli, L. Celetto, E. Izquierdo, and F. Rovati, "Peer-to-peer streaming of scalable video in future Internet applications," *IEEE Communications magazine*, vol. 49, pp. 128-135, 2011.
- [34] A. Bradai, T. Ahmed, R. Boutaba, and R. Ahmed, "Efficient content delivery scheme for layered video streaming in large-scale networks," *Journal of network and computer applications*, vol. 45, pp. 1-14, 2014.
- [35] R. Trestian, I.-S. Comsa, and M. F. Tuysuz, "Seamless multimedia delivery within a heterogeneous wireless networks environment: Are we there yet?," *IEEE Communications Surveys & Tutorials*, vol. 20, pp. 945-977, 2018.
- [36] M. You, X. Xin, W. Shangguang, L. Jinglin, S. Qibo, and Y. Fangchun, "QoS evaluation for web service recommendation," *China Communications*, vol. 12, pp. 151-160, 2015.
- [37] C. Kenneth, J. Mansfield, and J. Antonakos, "Computer Networking from LANs to WANs: Hardware, Software, and Security," ed: Boston, MA, USA: Course Technology, 2010.
- [38] S. Kanumuri, P. C. Cosman, A. R. Reibman, and V. A. Vaishampayan, "Modeling packet-loss visibility in MPEG-2 video," *IEEE Transactions on Multimedia*, vol. 8, pp. 341-355, 2006.
- [39] A. R. Reibman, S. Kanumuri, V. Vaishampayan, and P. C. Cosman, "Visibility of individual packet losses in MPEG-2 video," in *2004 International Conference on Image Processing, 2004. ICIP'04.*, 2004, pp. 171-174.
- [40] I. Rec, "P. 800: Methods for subjective determination of transmission quality," *International Telecommunication Union, Geneva*, vol. 22, 1996.
- [41] T. Installations and L. Line, "Subjective video quality assessment methods for multimedia applications," *Networks*, vol. 910, p. 5, 1999.
- [42] R. I.-R. BT, "Methodology for the subjective assessment of the quality of television pictures," *International Telecommunication Union*, 2002.

- [43] H. J. Kim, D. H. Lee, J. M. Lee, K. H. Lee, W. Lyu, and S. G. Choi, "The QoE evaluation method through the QoS-QoE correlation model," in *2008 Fourth International Conference on Networked Computing and Advanced Information Management*, 2008, pp. 719-725.
- [44] Y.-K. Wang, Y. Sanchez, T. Schierl, S. Wenger, and M. M. Hannuksela, "RTP payload format for high efficiency video coding (HEVC)," 2070-1721, 2016.
- [45] T. T. Alfaqheri and A. H. Sadka, "Low delay error resilience algorithm for H. 265| HEVC video transmission," *Journal of Real-Time Image Processing*, vol. 17, pp. 2047-2063, 2020.
- [46] W. Guo, Y. Zhang, and B. Li, "Constrained motion estimation-based error resilient coding for HEVC," in *Ninth International Conference on Graphic and Image Processing (ICGIP 2017)*, 2018, pp. 1161-1167.
- [47] D. Flynn, D. Marpe, M. Naccari, T. Nguyen, C. Rosewarne, K. Sharman, J. Sole, and J. Xu, "Overview of the range extensions for the HEVC standard: Tools, profiles, and performance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, pp. 4-19, 2015.
- [48] B. Oztas, M. T. Pourazad, P. Nasiopoulos, and V. C. Leung, "A study on the HEVC performance over lossy networks," in *2012 19th IEEE International Conference on Electronics, Circuits, and Systems (ICECS 2012)*, 2012, pp. 785-788.
- [49] S. Wenger, G. D. Knorr, J. Ott, and F. Kossentini, "Error resilience support in H. 263+," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 867-877, 1998.
- [50] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proceedings of the IEEE*, vol. 86, pp. 974-997, 1998.
- [51] S. Kumar, L. Xu, M. K. Mandal, and S. Panchanathan, "Error resiliency schemes in H. 264/AVC standard," *Journal of Visual Communication and Image Representation*, vol. 17, pp. 425-450, 2006.
- [52] X. Zhang, X. Peng, S. Fowler, and D. Wu, "Robust h. 264/avc video transmission using data partitioning and unequal loss protection," in *2010 10th IEEE International Conference on Computer and Information Technology*, 2010, pp. 2471-2477.
- [53] T. Schierl, M. M. Hannuksela, Y.-K. Wang, and S. Wenger, "System layer integration of high efficiency video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1871-1884, 2012.
- [54] F. Boulos, "Transmission d'images et de vidéos sur réseaux à pertes de paquets: Mécanismes de protection et optimisation de la qualité perçue," Université de Nantes, 2010.
- [55] S. Wenger, "H. 264/avc over ip," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 645-656, 2003.
- [56] P. Ferré, D. Agrafiotis, and D. Bull, "A video error resilience redundant slices algorithm and its performance relative to other fixed redundancy schemes," *Signal Processing: Image Communication*, vol. 25, pp. 163-178, 2010.
- [57] J. Yang, Z. Yin, and T. Tang, "Adaptive Intra Refresh for Screen Content Video Transmission in Dynamic Network," *Mobile Networks and Applications*, pp. 1-11, 2023.
- [58] H. Li, W. Lei, and W. Zhang, "Perceptual video coding based on adaptive region-level intra-period," in *2022 7th International Conference on Computer and Communication Systems (ICCCS)*, 2022, pp. 387-392.
- [59] D. I. MUSTAFA and I. A. ALI, "ERROR RESILIENCE OF H. 264/AVC CODING STRUCTURES FOR DELIVERY OVER WIRELESS NETWORKS," *Journal of Duhok University*, vol. 25, pp. 114-128, 2022.
- [60] H. P. Cong, X. H. Van, and D. T. Dinh, "HEVC Compatible Multiple Description Coding for Robust Video Transmission over Lossy Networks," *VNU Journal of Science: Computer Science and Communication Engineering*, 2022.
- [61] Z. Wei, K.-K. Ma, and C. Cai, "Prediction-compensated polyphase multiple description image coding with adaptive redundancy control," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 465-478, 2011.
- [62] S. Shirani, M. Gallant, and F. Kossentini, "Multiple description image coding using pre-and post-processing," in *Proceedings International Conference on Information Technology: Coding and Computing*, 2001, pp. 35-39.

- [63] D. Wang, N. Canagarajah, D. Redmill, and D. Bull, "Multiple description video coding based on zero padding," in *2004 IEEE International Symposium on Circuits and Systems (IEEE Cat. No. 04CH37512)*, 2004, pp. II-205.
- [64] T. Tillo and G. Olmo, "Data-dependent pre-and postprocessing multiple description coding of images," *IEEE Transactions on Image Processing*, vol. 16, pp. 1269-1280, 2007.
- [65] C. Ates, Y. Urgan, B. Demir, O. Urhan, and S. Erturk, "Polyphase downsampling based multiple description image coding using optimal filtering with flexible redundancy insertion," in *2008 International Conference on Signals and Electronic Systems*, 2008, pp. 193-196.
- [66] Y. Yapıcı, B. Demir, S. Ertürk, and O. Urhan, "Downsampling-based multiple description image coding using optimal filtering," *Journal of electronic imaging*, vol. 17, pp. 033018-033018-9, 2008.
- [67] V. Parameswaran, A. Kannur, and B. Li, "Adapting quantization offset in multiple description coding for error resilient video transmission," *Journal of Visual Communication and Image Representation*, vol. 20, pp. 491-503, 2009.
- [68] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Transactions on Information Theory*, vol. 39, pp. 821-834, 1993.
- [69] T. Venugopal and S. Radhika, "A survey on channel coding in wireless networks," in *2020 International Conference on Communication and Signal Processing (ICCSP)*, 2020, pp. 0784-0789.
- [70] T.-M. Dang, E. Cho, and S.-H. Kim, "Current Trends on Deep Learning-aided Channel Coding," in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 2022, pp. 1222-1225.
- [71] T. O'shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, pp. 563-575, 2017.
- [72] T. J. O'Shea, K. Karra, and T. C. Clancy, "Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention," in *2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2016, pp. 223-228.
- [73] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. Be'ery, "Deep learning methods for improved decoding of linear codes," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, pp. 119-131, 2018.
- [74] A. Caciularu and D. Burshtein, "Blind channel equalization using variational autoencoders," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2018, pp. 1-6.
- [75] T. Wu, Z. Chen, D. He, L. Qian, Y. Xu, M. Tao, and W. Zhang, "CDDM: Channel Denoising Diffusion Models for Wireless Communications," *arXiv preprint arXiv:2305.09161*, 2023.
- [76] W. Kumwilaisak and C.-C. J. Kuo, "Spatial error concealment with sequence-aligned texture modeling and adaptive directional recovery," *Journal of Visual Communication and Image Representation*, vol. 22, pp. 164-177, 2011.
- [77] J. Liu, G. Zhai, X. Yang, B. Yang, and L. Chen, "Spatial error concealment with an adaptive linear predictor," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, pp. 353-366, 2014.
- [78] P.-S. Hwang, J.-H. Ri, and Y.-H. Yun, "Method for speeding up spatial error concealment using prediction mode of the neighboring blocks on H. 264 video communication," *Multimedia Tools and Applications*, vol. 82, pp. 13733-13743, 2023.
- [79] J. Chen, Y. Ye, and S. Kim, "Algorithm description for versatile video coding and test model 10 (VTM 10) JVET-S2002," in *19th Meeting of the Joint Video Exploration Team (JVET)*, 2020, pp. 1-67.
- [80] P. Haskell and D. Messerschmitt, "Resynchronization of motion compensated video affected by ATM cell loss," in *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1992, pp. 545-548.
- [81] Y.-L. Chang, Y. A. Reznik, Z. Chen, and P. C. Cosman, "Motion compensated error concealment for HEVC based on block-merging and residual energy," in *2013 20th International Packet Video Workshop*, 2013, pp. 1-6.

- [82] Q. Peng, T. Yang, and C. Zhu, "Block-based temporal error concealment for video packet using motion vector extrapolation," in *IEEE 2002 International Conference on Communications, Circuits and Systems and West Sino Expositions*, 2002, pp. 10-14.
- [83] G. Choe, C. Nam, and C. Chu, "An effective temporal error concealment in H. 264 video sequences based on scene change detection-PCA model," *Multimedia Tools and Applications*, vol. 77, pp. 31953-31967, 2018.
- [84] M. Benjak, N. Aust, Y. Samayoa, and J. Ostermann, "Neural Network-based Error Concealment for B-Frames in VVC," in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2022, pp. 511-515.
- [85] J. Seiler, M. Schober, and A. Kaup, "Spatio-temporal error concealment in video by denoised temporal extrapolation refinement," in *2013 IEEE International Conference on Image Processing*, 2013, pp. 1613-1616.
- [86] C. Nam, S. Ryu, C. Chu, T. Kim, S. Choe, and J. Choe, "A hybrid spatial error concealment using human face and image edge over H. 264/AVC video sequences," *Multimedia Tools and Applications*, vol. 82, pp. 3769-3799, 2023.
- [87] A. Aldahdooh, "Content-Aware Video Transmission in HEVC Context: Optimization of compression, of error resilience and concealment, and of visual quality," Nantes, 2017.
- [88] P. Piñol, M. Martínez-Rach, P. Garrido, O. Lopez-Granado, and M. P. Malumbres, "Error resilient coding techniques for video delivery over vehicular networks," *Sensors*, vol. 18, p. 3495, 2018.
- [89] B. Li, J. Xu, and H. Li, "Parsing robustness in high efficiency video coding-analysis and improvement," in *2011 Visual Communications and Image Processing (VCIP)*, 2011, pp. 1-4.
- [90] J.-L. Lin, Y.-W. Huang, C.-M. Fu, C.-Y. Chen, Y.-P. Tsai, and S. Lei, "Syntax for AMVP Parsing Error Control," *JCTVC-D126, Daegu, KR*, 2011.
- [91] F. Bossen, "Common test conditions and software reference configurations," *JCTVC-L1100*, vol. 12, 2013.
- [92] HM. *reference software for HEVC (version 16.11)*.  
 . Available: [https://hevc.hhi.fraunhofer.de/svn/svn\\_HEVCSoftware/branches/](https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/branches/)
- [93] F. Developers. (2020). *Ffmpeg tool (Version 4.3.1)*. Available: <https://ffmpeg.org/>
- [94] T. L. Foundation. (2017). *Network emulation with NetEM*. Available: <http://www.linuxfoundation.org/collaborate/workgroups/networking/netem>
- [95] K. E. Psannis, "HEVC in wireless environments," *Journal of Real-Time Image Processing*, vol. 12, pp. 509-516, 2016.
- [96] P. Paudyal, F. Battisti, and M. Carli, "Evaluation of the effects of transmission impairments on perceived video quality by exploiting ReTRiEVED dataset," *Journal of Electronic Imaging*, vol. 26, pp. 023003-023003, 2017.
- [97] P. Paudyal, F. Battisti, and M. Carli, "Impact of video content and transmission impairments on quality of experience," *Multimedia Tools and Applications*, vol. 75, pp. 16461-16485, 2016.
- [98] C. Liu, R. Ma, and Z. Zhang, "Error concealment for whole frame loss in HEVC," in *Advances on Digital Television and Wireless Multimedia Communications: 9th International Forum on Digital TV and Wireless Multimedia Communication, IFTC 2012, Shanghai, China, November 9-10, 2012. Proceedings*, 2012, pp. 271-277.
- [99] P. Dymarski, S. Kula, and T. N. Huy, "QoS conditions for VoIP and VoD," *Journal of Telecommunications and Information Technology*, pp. 29-37, 2011.
- [100] S. Khalfa, S. Harize, and N. Kouadria, "Source Level Protection for HEVC Video Coded in Low Delay Mode for Real-Time Applications," *Automatic Control and Computer Sciences*, vol. 57, pp. 305-316, 2023.
- [101] T. Das, K. Choi, and J. Choi, "High Quality Video Frames from VVC: A Deep Neural Network Approach," *IEEE Access*, 2023.
- [102] A. Kebir and M. Taibi, "End-to-end deep auto-encoder for segmenting a moving object with limited training data," *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 12, 2022.

- [103] K. Cho, "Simple sparsification improves sparse denoising autoencoders in denoising highly corrupted images," in *International conference on machine learning*, 2013, pp. 432-440.
- [104] H. L. Vu, K. T. W. Ng, A. Richter, and C. An, "Analysis of input set characteristics and variances on k-fold cross validation for a Recurrent Neural Network model on waste disposal rate estimation," *Journal of environmental management*, vol. 311, p. 114869, 2022.
- [105] Y. Dai, D. Liu, and F. Wu, "A convolutional neural network approach for post-processing in HEVC intra coding," in *MultiMedia Modeling: 23rd International Conference, MMM 2017, Reykjavik, Iceland, January 4-6, 2017, Proceedings, Part I 23*, 2017, pp. 28-39.
- [106] S. Kuanar, C. Conly, and K. Rao, "Deep learning based HEVC in-loop filtering for decoder quality enhancement," in *2018 Picture Coding Symposium (PCS)*, 2018, pp. 164-168.
- [107] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*: " O'Reilly Media, Inc.", 2022.