Transmission d'images et de vidéos sur réseaux à pertes de paquets : mécanismes de protection et optimisation de la qualité perçue

Fadi Boulos

UNIVERSITE DE NANTES

ÉCOLE DOCTORALE

« SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE MATHEMATIQUES »

Année : 2010

Thèse de Doctorat de l'Université de Nantes

Spécialité : AUTOMATIQUE ET INFORMATIQUE APPLIQUEE

Présentée et soutenue publiquement par

Fadi BOULOS

le 12 février 2010

à Polytech'Nantes

Transmission d'images et de vidéos sur réseaux à pertes de paquets : mécanismes de protection et optimisation de la qualité perçue

Jury

Rapporteurs	:	Rik VAN DE WALLE	Professeur, université de Ghent, Belgique
		François-Xavier COUDOUX	Professeur, université de Valenciennes
Examinateurs	:	Gerardo RUBINO	Directeur de recherche INRIA, IRISA Rennes
		Patrick LE CALLET	Professeur, université de Nantes
		Benoît PARREIN	Maître de conférences, université de Nantes
Membre invité	:	David HANDS	Responsable d'équipe, British Telecommunications plc, Angleterre

Directeur de Thèse : Patrick LE CALLET
Co-encadrant : Benoît PARREIN
Laboratoire : Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN)
Composante de rattachement du directeur de thèse : Polytech'Nantes

N° ED 503 -082

Remerciements

Je voudrai tout d'abord remercier MM. François-Xavier Coudoux et Rik Van de Walle d'avoir accepté la tâche de rapporter ma thèse et d'avoir contribué à l'amélioration de la qualité de cette thèse.

Je remercie M. Gérardo Rubino d'avoir accepté d'examiner ma thèse et M. David Hands d'avoir fait le déplacement pour participer au jury de thèse. Je remercie également "Dave" de m'avoir donné la chance de vivre une belle expérience en Angleterre.

Je tiens à exprimer ma sincère gratitude envers mes encadrants Benoît Parrein et Patrick Le Callet pour leurs conseils avisés et pour avoir rendu ces années agréables et fructueuses.

Un grand merci aux membres de l'équipe IVC pour l'ambiance qui règne dans notre laboratoire : Jean-Pierre pour ses conseils paternels ; Florent pour les différents services informatiques rendus (et les nouveautés Mac) ; Olivier, Sylvain et Romain pour la séance de sport hebdomadaire ; Romuald pour les tests subjectifs et les expressions d'argot que j'ai apprises ; David (Chen) pour le travail qu'il a fait et pour l'équipe efficace qu'on a formée ; Guillaume pour les "lifts" du soir et les discussions para-académiques ; Eddy pour les échanges philosophiques qu'on a eus ; Nico pour avoir toujours réussi à régler mes problèmes d'informatique ; Marcus pour les discussions scientifiques poussées ; Jiazi pour sa puissante machine de calcul ; Vincent pour les leçons sur la société française ; Yohann pour son humeur agréable ; Hélène et Karine pour leur aide toujours accompagnée d'un sourire.

Je remercie mes amis à Nantes (que je n'oserai pas commencer à nommer par peur d'en oublier quelqu'un) pour avoir rendu mon séjour très agréable ; Khalil pour son éternelle bonne humeur, son humour et les discussions existentielles qu'on a eues ; toute ma famille pour son encouragement continu et sa confiance en moi ; mes parents pour m'avoir toujours motivé à atteindre l'excellence ; Manal pour être dans ma vie.

À mes parents

Table des matières

Introduction générale

Internet multimédia

Le multimédia (image, vidéo, musique, *etc.*) occupe une place importante au sein de la société d'aujourd'hui grâce à la diversité des équipements électroniques capables de créer et de lire des contenus multimédias. Le protocole Internet (IP) n'est cependant pas parfaitement adapté aux services multimédias. En effet, les données multimédias sont volumineuses et imposent des contraintes qui ne correspondent pas nécessairement au mode d'opération de l'IP tel qu'il a été conçu.

Les pertes de paquets sont inévitables sur un réseau à dimension planétaire. Des protocoles comme TCP (*Transmission Control Protocol*) assurent généralement la retransmission des paquets perdus. Cette solution n'est toutefois pas idéale car elle n'est pas compatible avec la contrainte temps-réel qui caractérise des services multimédias comme la visioconférence, les jeux en réseaux ou la diffusion télé.

Les solutions mises en place pour garantir un niveau de qualité acceptable interviennent à différentes étapes de la chaîne de transmission. Elles peuvent ainsi agir en prévention (codes correcteurs par anticipation) ou en réponse aux problèmes (compensation de pertes) ; elles peuvent être implantées aux deux bouts de la chaîne de transmission (codage et décodage robustes) ou dans le réseau (classification de trafic). De plus, nous pouvons distinguer entre les méthodes qui opèrent au niveau du contenu de la source (prise en compte de la hiérarchie) et celles, plus génériques, qui s'appliquent à toutes sortes de données (le codage réseau par exemple).

Le succès des solutions visant l'amélioration de la qualité de service peut être mesuré par plusieurs moyens. Un moyen est la mesure des paramètres de qualité de service tels que le taux de pertes de paquets ou la gigue. Ces mesures sont en effet indicatives mais ne donnent pas à elles seules une information suffisante sur la qualité perçue par l'utilisateur final.

La qualité d'usage reflète le niveau de satisfaction des utilisateurs d'un service. Le moyen le plus fiable de la mesurer est la conduite d'expérimentations appelées tests subjectifs d'évaluation de qualité. Pendant ces tests réalisés dans un environnement normalisé (selon la recommandation BT.500-11 [itu02] par exemple), un panel d'observateurs évaluent la qualité d'images ou de vidéos en leur attribuant une note choisie sur une échelle donnée.

Protection fondée sur la perception

Différentes approches existent pour améliorer la qualité d'usage d'un service vidéo. Parmi ces approches, nous retrouvons les méthodes d'amélioration sur la base d'indicateurs de la perception humaine de qualité. Ces méthodes trouvent leur justification dans le fait que l'utilisateur final du service est généralement un être humain et qu'il est donc raisonnable d'améliorer la qualité d'usage en fonction de sa perception. Cependant, une bonne compréhension des facteurs qui font varier la qualité d'usage est nécessaire au préalable. Il faut ainsi identifier l'effet de gêne créé par les erreurs de transmission. Ceci se fait en étudiant de près la réaction des utilisateurs face

aux dégradations engendrées par les motifs de pertes qui typiquement caractérisent un réseau de paquets.

Au niveau de la perception de qualité, le Système Visuel Humain (SVH) joue un rôle essentiel dans l'acquisition de l'information. Le SVH a toutefois une capacité limitée de traitement de l'information visuelle. Ceci se traduit par des mécanismes de sélection de l'information visuelle la plus importante dans notre champ visuel. Dans un contexte de protection des données multimédias contre les erreurs de transmission, une approche intéressante serait d'allouer plus de protection aux données susceptibles d'attirer l'attention de l'utilisateur.

Pour déterminer les parties de l'information perceptuellement importantes, des méthodes de mesure peuvent être utilisées. Pour la vidéo par exemple, le suivi des mouvements des yeux renseigne sur les régions de l'image qui attirent le plus l'attention de l'observateur. Une fois la hiérarchie de la source identifiée, la robustesse des régions d'intérêt contre les erreurs de transmission doit être augmentée pour garantir une qualité d'usage satisfaisante.

Les données multimédias d'un flux étant généralement compressées, il existe de fortes dépendances entre elles. Ces dépendances augmentent la vulnérabilité du flux contre les erreurs de transmission. Ainsi, la corruption d'une partie des données ou leur perte engendre une propagation en cascade de l'erreur qui peut dégrader sévèrement la qualité du flux décodé. Il paraît donc raisonnable d'essayer de réduire cet effet de propagation notamment dans les régions perceptuellement significatives.

Organisation du mémoire

Ce mémoire est organisé en quatre parties.

Dans la première partie, nous expliquons la problématique des transmissions multimédias sur IP en tous ses aspects. Nous introduisons ainsi dans le chapitre 1 les notions de qualité de service et de qualité d'usage en insistant sur leur interaction. Nous décrivons également l'influence des erreurs de transmission qui peuvent typiquement avoir lieu dans un réseau de paquets sur la qualité d'usage. Dans le chapitre 2, nous établissons un état de l'art des principales méthodes utilisées dans la littérature pour atteindre des niveaux acceptables de qualité de service et d'usage. Nous classons ces méthodes en fonction de leur niveau d'intervention dans la chaîne de transmission. Nous distinguons ainsi les méthodes opérant dans les composants du réseau (classification de trafic, codage réseau), celles qui font partie du codage canal (codes correcteurs par anticipation), les techniques de robustesse intrinsèques au codage source et enfin les méthodes de compensation des erreurs de transmission qui sont mises en œuvre à la réception du flux multimédia. Cette classification orientera le travail présenté dans les parties suivantes du mémoire.

Dans la deuxième partie, nous nous intéressons à l'application de la transformation Mojette dans des contextes d'amélioration de la qualité de service et de la qualité d'usage. Cette transformation de Radon discrète a des propriétés intéressantes qui facilitent son utilisation dans des

applications réseaux. Nous proposons ainsi dans le chapitre 3 d'utiliser la transformation Mojette comme opérateur de codage réseau (*network coding*). Cette approche [Ahl00] du domaine de la théorie de l'information vise l'optimisation de l'utilisation de la bande passante en vue de l'amélioration de la qualité de service. D'autre part, nous exploitons la redondance offerte par la transformation Mojette pour appliquer une protection inégale d'information sur une image codée en JPEG 2000 dans le chapitre 4. Ce chapitre met en avant l'importance de la construction d'une hiérarchie perceptuelle de la source en vue de l'application d'une protection adéquate.

La troisième partie traite deux aspects de la perception humaine de la qualité de vidéos. Dans le chapitre 5, nous menons une étude approfondie des effets perceptuels des pertes de paquets sur la qualité visuelle de vidéos codées en H.264/AVC. Cette étude vise premièrement l'établissement d'une relation entre un indicateur de qualité de service particulier (la perte de paquet) et la qualité d'usage. Le second objectif de cette étude est l'identification d'une hiérarchie au niveau du contenu de la vidéo. Cette hiérarchie est essentiellement représentée par la différence d'importance perceptuelle entre les parties d'une même image. Plus précisément, nous montrons que la qualité visuelle est fortement influencée par la position des dégradations par rapport à la région d'intérêt de l'image. Ce constat nous mène à réaliser des expérimentations de suivi d'attention visuelle que nous décrivons dans le chapitre 6. Le résultat de ces expérimentations est une carte de saillance pour chaque image de la vidéo reflétant le degré d'attraction de l'attention visuelle de chaque pixel.

Dans la quatrième partie de ce mémoire, nous développons deux méthodes de protection des régions d'intérêt utilisant les données de saillance obtenues précédemment. Le but de cette protection est d'atténuer la propagation spatio-temporelle des dégradations dans une vidéo. Ces méthodes d'amélioration de la robustesse des flux vidéos contre les pertes de paquets sont guidées par la hiérarchie de la source. La première méthode, décrite dans le chapitre 7, agit au moment du codage de la vidéo en modifiant les modes de codage des macroblocs dans la région d'intérêt de l'image. La seconde méthode, développée à partir du même algorithme que la première, est détaillée dans le chapitre 8. Cette méthode couple l'arrêt de la propagation spatio-temporelle avec une technique de codage robuste offerte par la norme H.264/AVC : l'organisation flexible de macroblocs (FMO). Nous évaluons les performances de cette méthode à l'aide de tests subjectifs de qualité.

Nous terminons le mémoire par une conclusion générale reprenant les thématiques abordées dans ce travail de recherche et nous proposons des perspectives pour l'extension des contributions.

Première partie

Problématique des transmissions multimédias sur IP

Chapitre 1

Services multimédias sur réseaux à qualité de service non garantie

Introduction

L'*Internet Protocol* (IP) a été initialement conçu en tant que simple réseau de données. À l'origine, le fonctionnement "pour le mieux" ou *best effort* de ce réseau des réseaux pouvait convenir à un seul type de trafic. Avec la croissance du trafic multimédia sur IP, le réseau a dû évoluer pour s'adapter aux besoins des services multimédias. Généralement exigents en termes de délai et de bande passante, ces services souffrent de baisses de qualité sévères si les conditions de transmission se dégradent.

Dans ce premier chapitre, nous abordons la problématique des services multimédias sur les réseaux IP au travers de deux volets : la Qualité de Service (QdS) et la Qualité d'Usage (QdU). La QdS donne une indication objective sur l'état du réseau. La QdU reflète le degré de satisfaction de l'utilisateur final du service multimédia. Nous présentons ainsi les différents paramètres qui déterminent la QdS et la QdU et nous essayons d'identifier la relation qui les lie.

Enfin, nous nous intéressons à un service multimédia particulier : la vidéo. Nous choisissons l'exemple de la vidéo car elle permet d'illustrer clairement les répercussions d'une qualité de service non garantie sur la qualité visuelle. Sa dépendance spatio-temporelle en fait un service multimédia particulièrement sensible aux conditions de transmission. Nous revoyons les principes de la norme de codage vidéo H.264/AVC autour de laquelle se concentre notre travail et nous expliquons l'influence des erreurs de transmission sur un flux H.264/AVC.

1.1 Qualité de service

Un réseau IP *best effort* traite tous les paquets transmis de la même façon. Il ne tient pas compte des contraintes que peuvent avoir quelques applications pour des services particuliers. Les réseaux *best effort* sont cohérents avec la philosophie initiale de l'Internet qui a été conçu comme un réseau très simple avec des applications plus ou moins évoluées. Mais, les services multimédias ont besoin du maintien des paramètres qui définissent la qualité de service.

1.1.1 Les paramètres de qualité de service

La QdS est une indication objective de la performance d'un réseau. Les paramètres caractérisant la QdS sont : la perte de paquets, le délai et la variation de délai (gigue) [Sus00, Tan03].

Les pertes de paquets sont chroniques et font partie intégrante de la transmission IP. Dans les moments de congestion extrême, le taux de pertes dépasse le seuil de tolérance et remet en cause la qualité et l'intégralité du flux transmis. À titre d'exemple, des taux de pertes compris entre 3% et 20% ont été enregistrés sur le backbone de l'Internet en 1999 [Wen99]. Actuellement, les taux de pertes journaliers moyens aux États-Unis sont généralement inférieurs à 1%[1].

Le délai est le temps mis par un paquet pour arriver de la source à sa destination finale. Il est essentiellement dû au délai de propagation et au temps d'attente dans les files au niveau des routeurs.

La variation de délai appelée gigue représente la différence entre les temps d'arrivée de deux paquets consécutifs d'un même flux. Ainsi, pour des paquets transmis à une période de t ms, une gigue existe si la différence entre leurs temps d'arrivée est égale à $(t \pm \Delta t)$ ms. La figure 1.1 illustre ce phénomène de gigue. Le paramètre de désordonnancement de paquets s'ajoute à celui de la gigue pour exprimer les désynchronisations possibles à la réception.

[1] Internet Health Report, `http://www.internetpulse.net/Main.aspx?Metric=PL`.

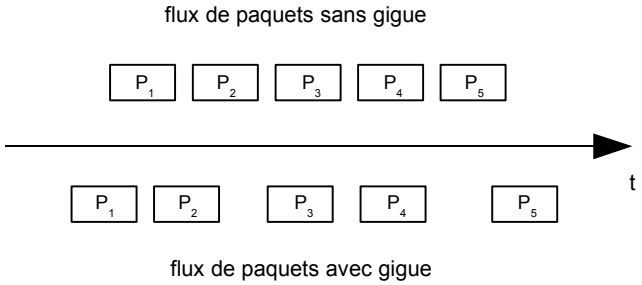


FIG. 1.1 – Illustration du phénomène de gigue. Les flux du haut et du bas ont respectivement une distance inter-paquet constante et variable.

Applications	**Contraintes**	**Explications**
Asservissement téléchirurgie	Délai critique : <50ms Gigue nulle	L'action doit être synchronisée avec le retour image.
Jeux en réseaux	Délai : <100 ms Gigue nulle.	Temps de réaction bref.
Audioconférence	Délai : 150 ms	Avec plusieurs participants, l'interactivité doit être parfaite
Téléphonie	Délai : < 200 ms	Duplex interactif est une condition du confort de la communication
Travail coopératif, communautés virtuelles	Délai moyen : 500 ms	La voix est une composante du dialogue mais le semi-duplex est acceptable
Diffusion de médias	Délai de l'ordre de la sec.	Mémoire tampon compense le délai, la gigue voire la perte
Web, transfert de fichiers	Délai élastique	Quelques secondes sont acceptables

FIG. 1.2 – Quelques applications et leurs contraintes de QdS [Sus00].

Les applications sur IP ont des contraintes différentes en ce qui concerne les valeurs des paramètres de QdS. Quelques exemples d'applications sont donnés dans le tableau de la figure 1.2 inspiré de [Sus00]. Par exemple, le mail ou le transfert de fichiers nécessitent une très haute fiabilité mais tolèrent des délais assez longs (de l'ordre de la seconde). C'est pour ce type de service nécessitant des délais élastiques que l'IP a été initialement conçu.

D'autre part, les services multimédias sont généralement gourmands en bande passante mais acceptent jusqu'à un certain degré les pertes de paquets. Pour un service de visioconférence sur IP, un délai dépassant les 150 ms rend le service inutilisable.

Une solution aux problèmes du délai et de sa variation est l'utilisation d'une mémoire tampon (*buffer*) dans le décodeur pour la lecture du contenu multimédia. Cette mémoire tampon sert à assembler les paquets de données, éventuellement les réordonnancer, les décoder et compenser ceux qui sont perdus. Elle agit comme un véritable amortisseur des variations temporelles des paramètres de QdS. Le but de cette configuration est de permettre une lecture fluide du contenu décodé indépendamment de la variation du délai d'arrivée des paquets. La longueur de la mémoire tampon est généralement réglée de façon à largement couvrir la somme du délai de réception des paquets et de sa variation maximale. Dans un contexte de *streaming* de vidéos via TCP, une mémoire tampon de dix secondes au plus permet une lecture fluide de la vidéo [She09]. Il est communément acquis que les paramètres de QdS pour une session multimédia temps-réel peuvent généralement être réduits au délai et à la perte de paquets si une mémoire tampon de taille convenable est utilisée [Jia99].

La conséquence ultime d'une combinaison de mauvaises valeurs de ces paramètres QdS est la perte de paquets. En effet, une congestion de paquets à un nœud particulier cause l'élimination

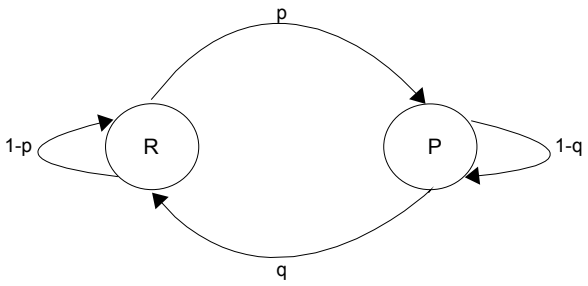


FIG. 1.3 – Modèle de Gilbert-Elliott. P et R indiquent respectivement la perte et la réception d'un paquet.

de paquets dans la file d'attente. De même, un paquet faisant partie d'un flux vidéo peut être considéré comme étant perdu par le décodeur s'il arrive après un délai supérieur à une limite fixée. Les motifs de pertes constatés sur l'Internet montrent que la probabilité de perte d'un paquet est plus grande si le paquet précédent a été perdu [Jia99]. Les pertes surviennent généralement en rafale. L'explication est la suivante : la perte d'un paquet laisse supposer la présence d'une congestion qui probablement va affecter un groupe de paquets. Ceci a conduit à la modélisation des pertes par des canaux à mémoire notamment par le modèle de Gilbert-Elliott.

Ce modèle, illustré sur la figure 1.3, est largement utilisé dans la transmission de flux multimédias. Il implique que le passage à un nouvel état dépend de l'état courant du canal. p est la probabilité de perdre un paquet sachant que le paquet précédent a été reçu. Cette probabilité est différente de la probabilité $1 - q$ de perdre un paquet sachant que le précédent a été également perdu. Des traces réelles enregistrées sur des réseaux en fonctionnement peuvent venir alimenter les probabilités de manière à augmenter le réalisme du modèle[2].

Les pertes de paquets affectent directement la qualité d'usage comme nous le verrons à la sous-section 1.2 de ce chapitre et plus en détail dans le chapitre 5 de ce mémoire.

1.1.2 Évolution de l'Internet multimédia

La transformation de l'Internet en un réseau capable d'offrir des services multimédias acceptables implique l'utilisation de protocoles pouvant relever les défis imposés par ce type de service. Une des contraintes majeures est le respect des contraintes de délai assez strictes. TCP garantit la réception des paquets par la retransmission des paquets non acquittés. Cependant, les services interactifs comme la visioconférence ne peuvent tolérer les délais de retransmission. Ceci nous mène à distinguer entre deux types de pertes : les pertes "directes" et "indirectes".

[2]Simulateur de pertes Telchemy. Disponible sur `http://www.telchemy.com/telsim.html`.

Les pertes directes désignent les paquets perdus dans le réseau et qui n'atteignent pas leur destination. Cette définition se rapproche de la vision de la QdS par les moniteurs de réseaux.

Les pertes indirectes, au niveau de l'application, représentent les paquets du flux qui ne sont plus considérés au moment du décodage. Ceci peut être dû à leur arrivée avec un délai conséquent ou bien à leur désordonnancement. Un décodeur qui attend un paquet donné pour poursuivre le décodage ne l'utilisera plus s'il arrive longtemps après. De même, le décodeur peut considérer un paquet comme perdu s'il arrive plus tôt que les paquets en cours de décodage.

Le protocole UDP (*User Datagram Protocol*) formalisé dans la RFC (*Request For Comments*) 768 [Pos80b] est un autre protocole de transport (couche 4 du modèle OSI - *Open System Interconnection*). Contrairement à TCP qui nécessite l'établissement d'une connexion entre deux ordinateurs en communication, UDP est en mode non connecté. Il se présente comme un protocole de transport léger (en-tête cinq fois plus petite que celle de TCP) mais ne garantit pas la livraison des paquets de données. De plus, UDP n'offre aucune possibilité de régulation de flux. Son utilisation ne peut donc être généralisée sur tous les types de trafic. L'utilisation d'UDP a cependant été proposée pour les applications qui ne tolèrent pas les longs délais comme par exemple les applications multimédias. Pour assurer un service de qualité acceptable, il faut cependant le coupler avec un protocole de niveau supérieur capable de détecter les pertes de paquets et d'y réagir.

Le protocole RTP, standardisé dans la RFC 3550 [Sch03], a été introduit pour satisfaire aux contraintes de temps de certains services, en particulier les services de visioconférence multi-utilisateurs. Il propose des fonctionnalités d'horodatage et des mécanismes de synchronisation de flux qui rendent possible le transport de données ayant des propriétés temps-réel. D'autre part, le protocole RTCP (*Real-time Transport Control Protocol*), qui assure le contrôle et la gestion de RTP, permet la mesure de la qualité de service à travers les rapports d'émission (SR pour *Sender Report*) et de réception (RR pour *Receiver Report*). RTP rend possible le réordonnancement du flux multimédia dans le cas de variation de délai des paquets et la détection d'éventuelles pertes de ces derniers. Il est indépendant des couches sous-jacentes. Toutefois, RTP est le plus souvent utilisé au-dessus du protocole UDP formant ainsi la pile RTP/UDP/IP dont l'en-tête globale a une taille égale à 40 octets (réductible à 2 ou 3 octets par compression d'en-tête [Bor01]).

La RFC 3984 définit l'encapsulation des NALU produites par un codeur H.264/AVC dans les paquets RTP pour des services multimédias variés (visioconférence, *streaming*, VOD pour *Video On Demand*). Les modes de mise en paquets comprennent la mise en paquet RTP d'une NALU individuelle, de plusieurs NALU et la fragmentation d'une NALU dans plusieurs paquets. La pile RTP/UDP/IP est utilisée dans plusieurs travaux de la littérature pour les services *streaming* de vidéos H.264/AVC comme par exemple dans [Hil06] et [Goh08].

1.1.3 Limitations

Alors que l'Internet se transforme de plus en plus en un réseau multimédia, la transmission de tels contenus sur un réseau IP n'est pas sans limitations. Il reste des défis techniques à surmonter pour que la qualité d'usage des services multimédias soit acceptable. Les défis les plus importants sont rappelés ci-dessous.

QdS non garantie Les paramètres de QdS sur l'Internet ne sont pas toujours garantis. Les valeurs de perte de paquets, de délai et de variation de délai peuvent varier brusquement influençant ainsi la qualité d'usage. Ces changements imprévisibles sont généralement dus à une congestion de paquets au niveau d'un ou de plusieurs routeurs ou à cause de la rupture d'un lien.

La contrainte temps-réel La QdS sur l'Internet étant non garantie, la contrainte temps-réel exigée par certaines applications se trouve difficilement satisfaite. Ainsi, sur des connexions résidentielles typiques, les applications nécessitant des délais assez courts comme celles de commande à distance ou de jeux en réseaux n'offrent pas toujours une bonne QdU. Même avec l'amélioration de l'infrastructure du réseau depuis quelques années, le délai et la gigue restent les défis majeurs à surmonter pour obtenir une QdU constante dans un contexte temps-réel.

Disponibilité de bande passante À mesure que la capacité du réseau augmente, les besoins applicatifs explosent. Avec l'avènement de la vidéo 3D et de la télévision HD, les besoins en bande passante ont largement augmenté. Ce constat implique une évolution du codage source et du codage canal pour permettre une qualité d'usage satisfaisante.

La perte directe de paquets Le résultat inévitable d'une mauvaise QdS est la perte de paquets. Les pertes directes sont généralement dues aux comportements des routeurs et à leur gestion des files d'attente. Elles sont minimales au cœur du réseau en temps normaux mais peuvent être significatives sur certains réseaux d'accès. Par exemple, la présence d'obstacles ou l'interférence entre signaux peuvent engendrer des pertes de paquets sur un lien sans fil. Les pertes de paquets sur l'Internet sont caractérisées par leur irrégularité et leur occurrence en rafales [Bor98, Pax99]. De plus, elles varient fortement avec le temps : elles sont plus fortes pendant la journée que la nuit [Tho97]. Les moyens d'atténuer les conséquences des pertes directes font l'objet d'étude de cette thèse.

1.2 Qualité d'usage

Les applications multimédias sont plus vulnérables aux problèmes qui peuvent survenir dans le réseau que d'autres applications comme le transfert de fichiers par exemple. L'acceptabilité

d'un service multimédia sur un réseau est conditionnée par le degré de satisfaction de l'utilisateur final. Cette satisfaction est mesurée par la Qualité d'Usage (QdU), traduction en français du terme *Quality of Experience*. Plusieurs paramètres influencent la QdU, notamment ceux de la QdS.

1.2.1 Définition

La QdU est une mesure de la qualité de l'expérience vécue par l'utilisateur. Pereira décompose la QdU en trois couches : la sensation, la perception et l'émotion [Per05].

La sensation représente le premier contact entre l'être humain et son environnement. Ce contact stimule les perceptions qui sont liées au processus cognitif. Ce dernier représente la façon dont l'homme acquiert les informations visuelles en se basant sur ses connaissances antérieures.

L'émotion est le sentiment éprouvé par l'utilisateur face au contenu multimédia. En d'autres termes, l'émotion reflète le degré d'intéressement que le contenu apporte à l'utilisateur.

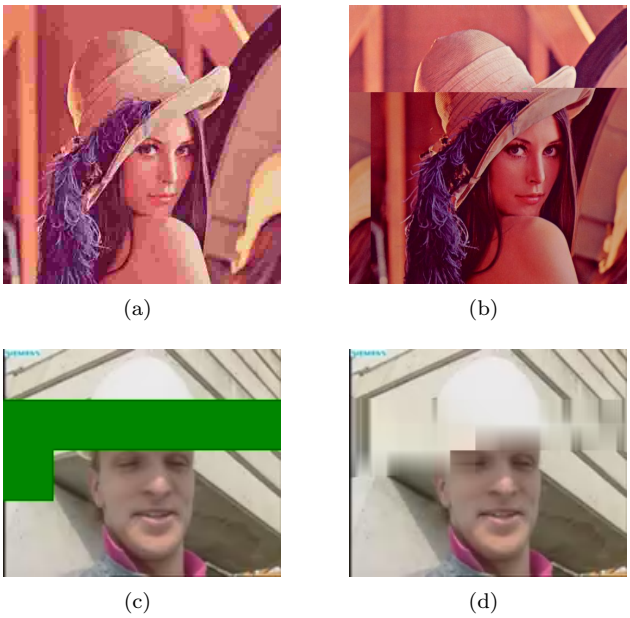
Gulliver et Ghinea [Gul07] proposent une définition de la QdU comme étant la qualité de perception. Cette dernière est la combinaison de trois composantes : l'assimilation, le jugement de qualité et la satisfaction. La qualité d'assimilation est une mesure de la clarté du contenu d'un point de vue informatif. Le jugement de qualité reflète la qualité de présentation et est indépendant du contenu multimédia lui-même. La satisfaction indique le degré d'appréciation globale de l'utilisateur.

Plusieurs facteurs contribuent à la QdU et peuvent être classés dans deux catégories.

La première catégorie est celle des facteurs psycho-sociologiques tels le vécu culturel de l'utilisateur ou son interactivité avec le contenu. Ainsi, un utilisateur habitué à regarder des films sévèrement codés peut être tolérant vis-à-vis des défauts de qualité d'une vidéo. L'état d'esprit de l'utilisateur et son appréciation du contenu d'une vidéo par exemple peuvent aussi influencer la QdU [Yam05]. Ces facteurs ne font pas l'objet principal de cette thèse. Nous limitons dans la suite notre étude aux facteurs dépendant des différentes étapes de la chaîne de transmission.

Nous retrouvons dans la seconde catégorie les paramètres liés au codage du contenu multimédia, à sa transmission et à son décodage. Durant la phase de codage, un bas débit peut rendre le contenu d'une image ou d'une vidéo difficilement reconnaissable. Au contraire, l'incorporation de techniques de codage qui augmentent la robustesse du flux contre les erreurs de transmission améliore la QdU. Pendant la transmission, des événements comme les erreurs binaires, la perte de paquets ou leur désordonnancement réduisent sévèrement la qualité du rendu final d'un flux multimédia. Enfin, au décodage, le support sur lequel le flux décodé est affiché et la stratégie de compensation des erreurs ou des pertes de paquets contribuent aussi à la qualité finale du contenu décodé.

Quelques exemples de dégradations de la qualité visuelle liées aux différentes étapes de la transmission sont donnés figure 1.4. La figure 1.4.a représente l'image *Lena* codée à bas débit



(a) (b)

(c) (d)

FIG. 1.4 – Des exemples de dégradations liées au codage (1.4.a), aux erreurs binaires (1.4.b) et aux pertes de paquets sans compensation (1.4.c) et avec compensation (1.4.d).

(0, 15 bpp) par un système de codage standard JPEG. À ce débit, la qualité de départ de l'image est mauvaise. Sur la figure 1.4.b, nous remarquons qu'une erreur binaire singulière (inversion de la valeur d'un bit de 0 à 1) résulte en une dégradation significative de la qualité de l'image. Deux exemples d'images extraites de la séquence vidéo *Foreman*, décodée à partir d'un flux ayant subi la perte de trois de ses paquets sont représentés sur les figures 1.4.c et 1.4.d. Les blocs de l'image colorés de manière uniforme constituent la partie de l'information perdue et qui n'a pas été remplacée, contrairement à l'image de la figure 1.4.d où le décodeur a appliqué une stratégie de compensation des blocs perdus. Cette stratégie consiste à remplacer les parties de l'image perdues par une représentation approximative calculée à partir des parties de l'image entourant celle qui est perdue. Les algorithmes de compensation de pertes de paquets sont détaillés dans le chapitre 2.

Ces trois exemples de dégradations montrent l'extrême sensibilité du contenu aux erreurs de transmission. L'amélioration de la robustesse du contenu sera abordée dans la quatrième partie de ce mémoire.

Une autre forme de dégradation particulièrement gênante dans le cas d'un flux multimédia est la désynchronisation entre les signaux audio et vidéo. Cette déformation s'illustre par exemple

par une non-correspondance entre les mouvements des lèvres d'une personne et le son entendu. Plusieurs causes peuvent être envisagées : délais de transmission différents pour les composantes audio et vidéo, processeur de faible puissance au décodage ou décodage à une fréquence d'images inférieure à celle de codage [Che98]. Selon l'ITU-R, une désynchronisation entre audio et vidéo de l'ordre de quelques dizaines de millisecondes dégrade fortement la qualité [itu98]. Cette conclusion n'est pas unanime, Steinmetz ayant montré qu'une désynchronisation égale à 80 ms reste acceptable [Ste96].

1.2.2 Évaluation de la qualité d'usage

La QdU peut être mesurée à l'aide de deux moyens : les tests subjectifs ou les métriques objectives. Les tests subjectifs d'évaluation de qualité consistent en un groupe de personnes qui utilisent le service et lui attribuent une note de qualité. Cette note de qualité doit refléter leur degré de satisfaction. D'autre part, les métriques objectives de qualité sont des algorithmes établis dans le but d'automatiser le processus d'évaluation. Leurs performances sont mesurées par rapport aux résultats des tests subjectifs.

1.2.2.1 Méthodes subjectives

L'utilisateur du service multimédia est le plus apte à évaluer la qualité d'usage. Les tests subjectifs se font dans des environnements normalisés. Les conditions de test tels l'illumination de la salle, son isolation accoustique, la distance entre l'observateur et l'écran et les caractéristiques de ce dernier doivent être conformes aux normes. Par exemple, les normes P.800 [itu96] et P.910 [itu99] de l'ITU-T spécifient respectivement les méthodologies d'évaluation de la qualité des signaux audio et vidéo d'un service de visioconférence sur des réseaux à commutation de paquets. L'inconvénient des tests subjectifs est qu'ils sont coûteux en termes de temps et de moyens humains. En effet, ils nécessitent au moins 15 personnes ayant une bonne acuité visuelle pour chaque série de contenus multimédias évaluée [itu99]. Les durées des tests dépendent du nombre d'échantillons à évaluer, de leurs durées et de la nature de la tâche assignée aux participants. Pour cette dernière, si le temps réservé à l'évaluation n'est pas fixé, le test peut prendre plus de temps que dans le cas où la durée de vote est limitée.

Les notes de qualité attribuées lors de tests subjectifs sont généralement choisies par les participants à partir d'une échelle de valeurs qui leur est proposée au début du test. Cette échelle peut être numérique ou sémantique, discrète ou continue, comparative ou absolue.

Une échelle sémantique contient des adjectifs indiquant l'appréciation comme "excellente", "bonne" ou "mauvaise". Une échelle continue couvre un intervalle (par exemple de 1 à 100) et le sujet peut choisir n'importe quelle valeur entière appartenant à cette intervalle. Une échelle comparative sert à comparer deux versions d'un même contenu ayant subi différents traitements. Des adjectifs comme "légèrement meilleure", "identique" et "moins bonne" sont utilisés sur de telles échelles.

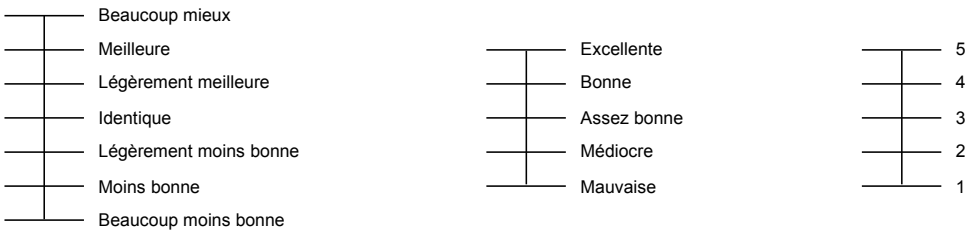


FIG. 1.5 – Des exemples d'échelles utilisées lors des tests subjectifs d'évaluation de qualité. De gauche à droite : une échelle comparative, une échelle catégorielle qualitative et son échelle discrète correspondante.

Des exemples de plusieurs types d'échelles sont donnés figure 1.5. L'échelle de gauche est une échelle comparative à sept niveaux tandis que l'échelle centrale est sémantique à cinq niveaux. L'échelle de droite représente les niveaux de cette dernière traduits simplement sous la forme de chiffres. Nous revenons en détails aux tests subjectifs dans le chapitre 5.

1.2.2.2 Métriques objectives

Le but de l'utilisation de métriques objectives est d'évaluer la qualité de signaux ayant subi un traitement particulier sans avoir besoin de recourir aux tests subjectifs. Les traitements peuvent être de nature dégradante comme par exemple la compression et la transmission ou améliorante tels les post-traitements d'affichage qui ont lieu après le décodage. Les métriques objectives, appelées aussi critères objectifs de qualité, peuvent être classifiées selon qu'elles exploitent ou non des propriétés du Système Visuel Humain (SVH). Stéphane Péchard [Péc08] a adopté cette classification (illustrée figure 1.6) pour les critères de qualité de vidéos. La classification de Péchard reste valide pour les autres contenus multimédias (image, audio, *etc.*).

Les approches signal prennent en compte uniquement les données brutes de la vidéo. Un exemple de métrique appartenant à cette catégorie est le PSNR (*Peak Signal-to-Noise Ratio*).

Les approches perceptuelles modélisent à différents degrés les mécanismes du SVH dans l'évaluation de la qualité d'une vidéo. Elles comprennent quatre groupes de modèles.

Les modèles basés sur les mécanismes bas niveau du SVH simulent la réponse du SVH aux stimuli ou sa sensibilité au contraste par exemple. La métrique DVQ [Wat01] est un exemple appartenant à cette catégorie.

Les modèles structurels mesurent les dégradations de l'information structurelle extraite par le SVH à partir d'une image. La métrique SSIM [Wan04] compare ainsi la luminance, le contraste et la structure de deux signaux pour fournir une note de qualité.

Les modèles avec connaissance du système dégradant sont basés sur une mesure des dégradations les plus communes d'un système donné. Par exemple, Farias [Far04] considère le codage

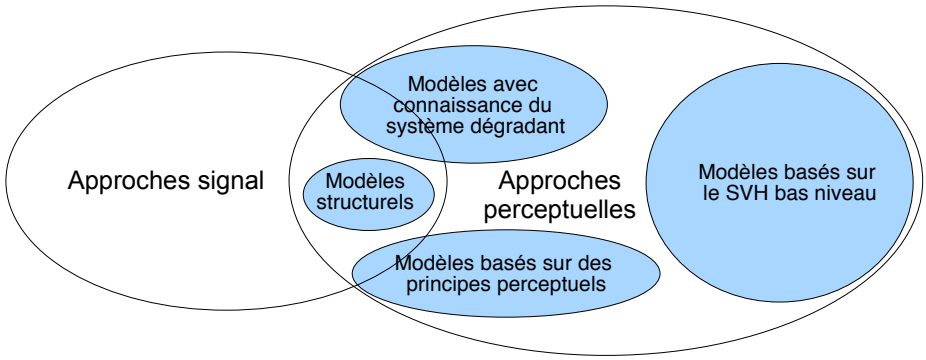


FIG. 1.6 – Classification des critères objectifs d'évaluation de la qualité visuelle de vidéos selon Péchard [Péc08].

MPEG-2 et mesure l'effet de blocs, le flou, le bruit et l'effet *ringing*.

Enfin, les modèles basés sur des principes perceptuels combinent les mécanismes bas niveau du SVH avec l'extraction de traits caractéristiques de l'image (à l'aide de filtres spatio-temporels par exemple).

La validation des performances d'un critère objectif de qualité se fait par comparaison de ses notes à celles recueillies lors de tests subjectifs. Cette comparaison est principalement établie à l'aide du coefficient de corrélation entre les notes objectives et les notes subjectives MOS (*Mean Opinion Score*).

1.2.3 Relation entre qualité de service et qualité d'usage

Les relations qui lient la QdS à la QdU sont complexes. Comprendre comment la variation des paramètres de QdS se répercute sur la QdU nécessite un contrôle rigide des conditions de transmission. Actuellement, les valeurs seuils de paramètres comme le délai ou la bande passante sont connues. Au-delà de ces valeurs, nous pouvons affirmer que la QdU sera au-dessous du seuil de satisfaction de l'utilisateur final. Par exemple, la transmission d'un flux vidéo codé et ayant une résolution HD (Haute Définition) sur un lien à 2 Mbit/s résultera en une QdU médiocre voire mauvaise. De même qu'un délai supérieur à 200 ms rendra un service de téléphonie inutilisable.

Par contre, si les paramètres de la QdS ont des valeurs variant autour de ces seuils, déduire le niveau de la QdU devient une tâche difficile. Ainsi, un taux de pertes de paquets autour de 10% tend généralement à indiquer une mauvaise qualité. Cependant, si la majorité des paquets perdus contient une partie redondante de l'information ou une partie de très faible importance, la qualité peut rester acceptable. De même, si l'information perdue est facilement remplaçable par le décodeur, l'impact de la perte peut être non significatif.

La métrique de qualité audio-visuelle propriétaire V-Factor[3], basée sur le modèle de [vdBL96], établit une relation entre les paramètres de QdS et leur influence sur la QdU. Les paramètres liés au réseau considérés par V-Factor sont la gigue, la perte de paquets et le désordonnancement des images. Les pertes unitaires (un seul paquet) et les pertes en rafales (plusieurs paquets) sont distinguées car leurs effets ne sont pas les mêmes. La contribution des paramètres de QdS à la qualité globale est modélisée sous la forme d'un facteur exponentiel.

Un modèle de corrélation entre la QdS et la QdU d'un service multimédia est proposé dans [Kim08]. Ce modèle intègre les paramètres suivants : délai, gigue, taux de pertes, taux d'erreurs, bande passante et taux de connexions établies. Cependant, le modèle n'est pas validé par des tests expérimentaux donc nous ne pouvons être sûrs de son efficacité.

Une relation entre les pertes de paquets et la qualité de vidéos est établie par Yamada *et al.* [Yam07]. Dans ce travail, les auteurs identifient les informations perdues à partir du flux vidéo et mesurent la diminution de qualité qui s'en suit au niveau de la vidéo décodée. Les dégradations sont mesurées au niveau des blocs de l'image qui remplacent les blocs perdus. Les résultats présentés montrent que cette métrique d'évaluation de la qualité visuelle corrèle bien (coefficient de corrélation de $0,95$) avec l'erreur quadratique moyenne entre la vidéo originale et la vidéo dégradée.

Une étude de l'impact de la gigue et des pertes de paquets sur la qualité de vidéos est réalisée dans [Cla99]. Les résultats montrent que l'influence de ces deux paramètres est presque identique sur la QdU. Gulliver et Ghinea étudient aussi la relation qui existe entre le délai et la gigue d'une part et la QdU d'autre part [Gul07]. Ils concluent que le délai et la gigue n'ont aucun effet sur la capacité de l'utilisateur à assimiler l'information relayée par le contenu multimédia. Par contre, leur effet est remarquable sur la qualité visuelle d'une vidéo et dépend fortement du contenu. Ce même résultat est déduit pour la satisfaction qui est étroitement liée à la qualité visuelle.

Pour des services multimédias à bas débit (type visioconférence), les pertes de paquets engendrent généralement la perte d'images entières. Pastrana-Vidal *et al.* étudient dans [PV04b] l'effet d'une mauvaise QdS sur la QdU. Ils montrent que la QdU est très sensible à la fluctuation de la QdS. En particulier, ils montrent que lors d'une visioconférence, une mauvaise QdS pendant une durée déterminée influence moins la QdU qu'une fluctuation de la QdS pendant toute la durée de la visioconférence.

Les modèles cités dans cette section contribuent à la compréhension de la relation entre les paramètres de QdS et leurs influences individuelle et collective sur la QdU. Cependant, cette problématique reste un sujet ouvert. Nous essayons d'ajouter quelques éléments de réponse dans le chapitre 5.

[3]http ://www.symmetricom.com/products/qoe-assurance/v-factor/

1.3 La vidéo comme service multimédia

Nous considérons la vidéo comme exemple de service multimédia dans cette section. L'intérêt de cet exemple est qu'il illustre bien les dégradations dont un service multimédia peut souffrir sur un réseau à qualité de service non garantie. En effet, les dimensions spatio-temporelles du signal vidéo le rendent particulièrement sensible aux dégradations subies dans le réseau.

Exemple d'un codage source : la norme H.264/AVC

La norme de codage vidéo H.264/MPEG-4 AVC [itu07] s'inscrit dans la continuité des normes précédentes de la famille MPEG (*Moving Picture Experts Group*). Cette norme a été mise en place en 2003 par le JVT (*Joint Video Team*), l'équipe de travail conjointe de l'ITU-T (secteur de normalisation) et de l'organisation internationale de normalisation ISO (*International Organization for Standardization*). L'apport principal de ce système de codage est son efficacité par rapport à ces prédécesseurs, permettant un gain en débit de 50% par rapport à MPEG-2 à qualité constante. Cependant, ce gain est achevé aux dépens d'une complexité de codage accrue due en partie à la forte dépendance entre les images de la vidéo. Les dépendances ainsi créées engendrent des dégradations de qualité en cascade quand une partie d'une image est perdue.

1.3.1 Les principes de codage H.264/AVC

Le fonctionnement d'un codeur H.264/AVC est illustré figure 1.7. Le codage d'une image commence par son découpage en blocs de taille 16×16 appelés macroblocs. Selon le type de l'image, ces derniers sont codés en mode intra ou inter. Le codage intra n'utilise que des échantillons de blocs déjà codés dans la même image. Le codage inter utilise des images décodées (appelées images de référence) pour obtenir la meilleure prédiction d'un bloc de l'image codée. Ensuite, la différence entre le bloc prédit (intra ou inter) et le bloc original est calculée et une transformée entière est appliquée à cette information résiduelle. Les coefficients de la transformée sont alors quantifiés et codés par un algorithme de codage entropique.

La norme H.264/AVC propose deux codages entropiques : CAVLC (*Context-Adaptive Variable-Length Coding*) [Wie03] et CABAC (*Context-Based Adaptive Binary Arithmetic Coding*) [Mar03]. Ces codages ont la particularité de s'adapter au contexte : la table de codes est changée en fonction des symboles déjà transmis. CABAC est plus performant que CAVLC mais au prix d'une complexité calculatoire plus élevée. Dans [Maz09], des tests effectués sur plusieurs résolutions de vidéos montrent que CABAC réduit le débit de $11,1\%$ par rapport à CAVLC pour une vidéo HD. Il est aussi montré que le temps de décodage est presque le même pour les deux algorithmes.

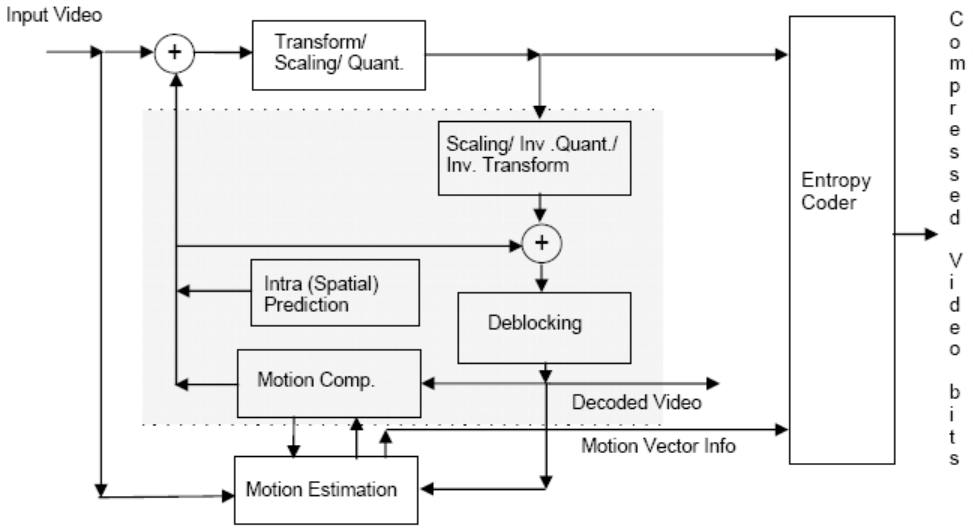


FIG. 1.7 – Schéma fonctionnel du codage H.264/AVC [Sul04].

1.3.2 Adaptation du flux H.264/AVC aux réseaux de paquets

H.264/AVC a été conçu pour être flexible et facilement adapté aux besoins de transmission. Ainsi, le système de codage comprend deux couches : VCL (*Video Coding Layer*) et NAL (*Network Abstraction Layer*). La couche VCL représente le cœur du codeur où sont effectués tous les processus de la prédiction au codage entropique décrits ci-dessus. La couche NAL traite les données issues de la couche VCL en vue de leur adaptation à un système de transport ou à un support de stockage particuliers. Ainsi, la RFC 3984 [Wen05] définit les différentes méthodes d'encapsulation des unités de transport du flux H.264/AVC dans les paquets RTP (*Real-time Transport Protocol*). Les propriétés du flux H.264/AVC qui permettent son utilisation efficace dans des environnements sans fil sont décrites dans [Sto03].

L'unité de base d'un flux binaire H.264/AVC est l'unité NAL que nous allons appeler NALU (*Network Abstraction Layer Unit*) dans la suite. Une NALU contient une en-tête de huit bits suivie par un nombre entier d'octets de données représentant une slice. Elle peut être considérée comme une NALU VCL ou non selon le type des données qu'elle contient. Les NALU VCL encapsulent des données liées au codage vidéo brut tandis que les autres NALU portent des éléments de syntaxe utiles au décodage du flux binaire et à l'affichage de la vidéo décodée. Des exemples de ce type de NALU syntaxique sont les NALU SPS (*Sequence Parameter Set*) et PPS (*Picture Parameter Set*) qui représentent respectivement les en-têtes de la séquence et de

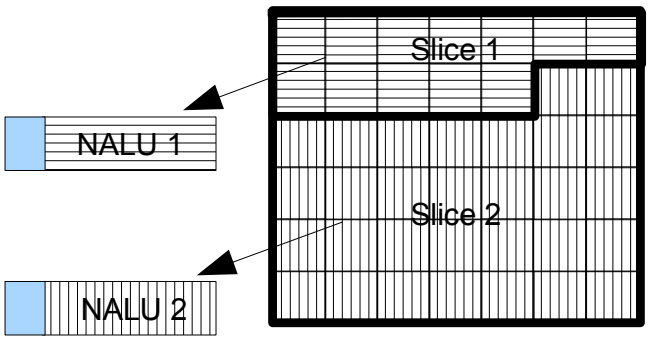


FIG. 1.8 – Encapsulation de deux slices d'une image dans deux NALU.

l'image. Ces NALU sont d'une importance majeure au processus de décodage et la perte d'une unité peut engendrer l'arrêt total du décodage.

Dans la perspective d'adaptation au contexte de transmission, la détermination du mode de codage se fait au niveau d'une "slice". Une slice contient un nombre entier de macroblocs codés formant une entité indépendamment décodable. Une image peut contenir une ou plusieurs slices. Les avantages de l'introduction de cette structure sont d'une part l'allégement de la dépendance intra-image et d'autre part la création d'entités de données à un niveau plus bas que l'image entière pour obtenir une plus grande flexibilité.

Des slices de petite taille dans une image I offrent une meilleure protection contre les pertes de paquets car la quantité de données perdues n'affecterait dans ce cas qu'une petite partie de l'image I. Mais ceci engendrerait des surcoûts de codage que nous expliquons dans la section 2.2.2.1 du chapitre 2. Au niveau NAL, chaque slice est encapsulée dans une NALU. La composition des slices et leur encapsulation dans les NALU sont illustrées figure 1.8.

1.3.3 Vulnérabilité du flux H.264/AVC aux erreurs de transmission

L'efficacité du codage H.264/AVC est principalement basée sur une utilisation augmentée de la prédiction intra-image et inter-images. Ceci crée une forte dépendance entre les différentes parties du flux vidéo. Cette dépendance s'avère coûteuse dans le cas où une information de référence est perdue car elle peut entraîner une dégradation de qualité en cascades. Par exemple, la perte d'un paquet contenant les données relatives à un bloc codé en mode intra se répercute sur tous les blocs qui en ont besoin pour leur décodage.

La propagation spatio-temporelle de l'effet d'une perte est généralement inévitable à cause de la nature même du codage vidéo. La propagation spatiale résulte des dépendances dues au codage entropique et de la prédiction intra-image utilisant des données elles-mêmes prédites à

partir d'une référence perdue. En effet, la partition en slices de données restreint le codage intra dans une image codée en intra aux macroblocs d'une même slice. Donc, la perte d'une slice d'une image n'affecte pas les autres slices de cette même image. D'autre part, il existe plusieurs degrés de propagation temporelle déterminés par l'ordre de prédiction d'un macrobloc. Une prédiction à partir d'un macrobloc perdu est dite de premier ordre de propagation. Une prédiction de second ordre concerne un macrobloc prédit à partir d'une référence elle-même prédite d'un macrobloc perdu. Les ordres de prédiction sont limités par la taille du GOP (*Group Of Pictures*) et par le nombre maximal de références possibles qui est un paramètre spécifié au codage.

Le degré d'influence qu'engendre une perte sur la qualité visuelle finale dépend, en plus de l'étendue de la propagation des dégradations, de plusieurs autres facteurs dont le codage source et la méthode de compensation des pertes au décodage. Ces facteurs sont détaillés dans le chapitre 5 de ce mémoire.

Sur la figure 1.9, nous illustrons une perte de slice dans l'image Im_N et le phénomène de propagation spatio-temporelle dans les images suivantes Im_{N+i} et Im_{N+j} où $j > i$. Nous supposons que l'image Im_N est une image codée en mode intra et que Im_{N+i} et Im_{N+j} sont codées en mode inter. Les macroblocs colorés en gris dans ces deux images représentent la propagation temporelle de la perte tandis que les macroblocs hachurés sont le résultat de la propagation spatiale. Ces derniers sont prédits à partir de macroblocs de la même image touchés par la propagation temporelle.

Conclusion

Dans ce premier chapitre, nous avons abordé la problématique des transmissions multimédias sur les réseaux IP. Nous avons introduit les notions de qualité de service et de qualité d'usage en exposant la relation complexe qui les lie. Les paramètres de qualité de service dont nous avons souligné l'importance sont le délai, la gigue et la perte de paquets. Ces paramètres ne permettent pas à eux seuls de donner une indication sur la qualité d'usage attendue. La définition de la qualité d'usage que nous avons proposée englobe des facteurs allant du niveau psychologique (émotion, sensation, *etc.*) au niveau technique (débit source, erreurs de transmission, *etc.*). Nous avons aussi décrit les différentes phases d'une communication multimédia sur l'Internet en précisant l'impact d'un dysfonctionnement au niveau de chacune des étapes de transmission sur la qualité. Dans ce cadre, nous avons revu les protocoles de transmission de référence facilitant la transmission de contenus multimédias sur IP.

De plus, nous avons présenté les classes de méthodes d'évaluation de la qualité d'usage ainsi que leurs avantages et leurs limitations. Nous retenons aussi que la qualité visuelle est une composante essentielle de la qualité d'usage. Elle est fortement influencée par la QdS et les paramètres imposés tout au long de la chaîne de transmission.

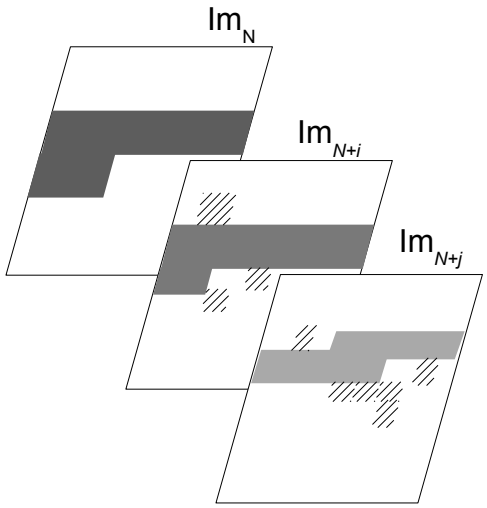


FIG. 1.9 – Un exemple de slice perdue (image Im_N) et de la propagation spatio-temporelle qui s'en suit (images Im_{N+i} et Im_{N+j}). La slice perdue et les slices affectées par cette perte sont respectivement colorées en noir et gris. La couleur grise est nuancée pour marquer l'atténuation de l'effet de perte avec la distance temporelle par rapport à la perte. Les macroblocs hachurés représentent la propagation spatiale de la perte.

Enfin, nous avons décrit le principe de fonctionnement de la norme de codage vidéo H.264/AVC en mentionnant sa spécificité dans un environnement de transmission non-fiable. Nous avons ainsi mis en évidence la propagation spatio-temporelle des dégradations sur un type de service multimédia.

Pour garantir un niveau satisfaisant de QdS et améliorer le rendu final d'un contenu multimédia, plusieurs solutions sont envisageables. Ces solutions agissent soit en prévention des erreurs de transmission, soit en réponse aux erreurs dans un contexte de compensation. Les différents mécanismes d'amélioration de la QdS et de la QdU font l'objet d'étude du chapitre 2.

Chapitre 2

État de l'art des mécanismes d'amélioration de la qualité de service et de la qualité d'usage

Introduction

La dégradation de la qualité d'usage est due à la perte d'information au codage et au cours de la transmission. Ce constat nécessite une protection des données dans le but de réduire voire d'éliminer les dégradations dues aux problèmes de transmission. Ceci peut se faire en améliorant la qualité de service du réseau, par exemple en réduisant le délai de transmission, sa variation (la gigue) et la perte de paquets et en augmentant la bande passante. L'augmentation de la QdS participe à l'augmentation de la QdU.

D'autres approches se sont intéressées à la problématique de l'amélioration directe de la QdU. Les solutions ainsi proposées génèrent toujours un surcoût qui peut prendre la forme d'un débit supplémentaire ou d'une réduction de l'efficacité du codage source et canal.

Dans ce chapitre, nous adoptons une classification générique des mécanismes d'amélioration de la QdS et de la QdU. Dans le volet de la QdS, nous décrivons tout d'abord une approche réseau représentée par les protocoles de classification de trafic. Ensuite, nous présentons brièvement une approche appartenant au domaine de la théorie de l'information et visant l'optimisation de l'utilisation de la bande passante disponible : le codage réseau (*network coding*).

Dans le volet de la QdU, nous commençons par les codes correcteurs d'erreurs et nous introduisons la notion de protection inégale de l'information. Passant du codage canal au codage source, nous présentons les principales techniques de codage robuste intrinsèques à la norme H.264/AVC ainsi que les algorithmes de compensation de pertes qui opèrent au décodage. Bien

que ces deux mécanismes opèrent au niveau du contenu, le premier se situe en amont de la transmission (au codage source) alors que le second agit au décodage. Nous évoquons ensuite les associations de techniques agissant aux différents points de la chaîne de transmission pour garantir une qualité d'usage satisfaisante.

Enfin, nous discutons de l'utilisation de chacune des méthodes présentées et proposons un schéma général de classification. À noter que la plupart des mécanismes proposés dans ce chapitre peuvent être utilisés pour les données multimédias génériques (image et vidéo), même si les travaux de la littérature présentés s'intéressent essentiellement à la vidéo.

2.1 Amélioration de la qualité de service

Les mécanismes de protection décrits dans cette section interviennent uniquement au niveau du réseau. Nous présentons tout d'abord un modèle de catégorisation du trafic sur l'Internet : DiffServ [Bla98]. DiffServ repose sur l'agrégation des différents flux ayant la même priorité en un seul flux. Ensuite, nous nous intéressons à un nouveau paradigme de routage de données : le codage réseau.

2.1.1 Classification de trafic

Le premier modèle de classification de trafic a été IntServ [Bra94]. Ce modèle est fondé sur une réservation préalable de la bande passante nécessaire à la transmission d'un flux ou d'une application. Cette demande se fait au travers du protocole RSVP (*Resource reSerVation Protocol*) [Bra97] qui réserve, à chaque nœud, les ressources nécessaires aux flux transmis. La complexité du modèle IntServ l'a néanmoins rendu peu utilisable. Un autre modèle moins complexe a alors été mis en place : DiffServ.

Le modèle DiffServ (*Differentiated Services*) catégorise le trafic de l'Internet en plusieurs classes de priorités différentes. Trois classes sont typiquement identifiées.

La première classe est la classe *Best Effort* qui n'assigne aucune priorité aux paquets d'un flux (comportement par défaut). La seconde classe est la classe *Expedited Forwarding* dédiée au trafic nécessitant des délais de transmission. La classe *Assured Forwarding* (AF) garantit une certaine quantité de bande passante et offre la possibilité de différencier l'allocation de cette quantité à quatre sous-classes de trafic.

La priorité d'un paquet est assignée en remplaçant le champ TOS (*Type Of Service*) de l'en-tête d'un paquet IP par le champ DS. Parmi les huit bits de ce champ, les six les plus significatifs sont utilisés par le DSCP (*Differentiated Services Code Point*) pour indiquer le niveau d'importance à assigner à ce paquet. Les deux bits restants ne sont pas couramment utilisés. Le champ TOS était prévu initialement pour indiquer la priorité et les différents critères de QdS dont doit bénéficier un paquet mais il n'a pas été utilisé intensivement en pratique.

Chaque paquet, dont l'importance est déterminée par la valeur de DSCP, subit un traitement différencié par les routeurs cœur (*core routers*) du réseau DiffServ. L'intérêt principal de DiffServ est qu'il apporte une différenciation des traitements appliqués à chaque service tout en conservant la simplicité du protocole IP et ses avantages, notamment le routage libre de toute contrainte.

L'idée de l'application de la différenciation de services à la vidéo a été avancée dans plusieurs travaux de la littérature. Une application visant un service de *streaming* de vidéos codées en H.264/AVC est proposée dans [Oro04]. Dans ce travail, l'acheminement assuré AF de DiffServ est exclusivement utilisé pour assigner et gérer les priorités des paquets. Ces priorités prennent la forme de "couleurs" (sous-classes) attribuées à chaque paquet et varient inversement à la préséance d'élimination (*drop precedence*). Ainsi, les paquets "verts" ont la priorité la plus élevée et sont donc les moins disposés à être éliminés. Les paquets de couleur "jaune" et "rouge" ont respectivement une priorité moyenne et basse. Les paquets contenant des slices de type I (codées en mode intra) sont marqués comme étant verts. Comme les slices de type B ne sont pas utilisées dans le cadre de ce travail, la couleur des slices P varie entre "jaune" et "rouge" selon une loi de Bernoulli d'espérance égale à $0,5$. Un réseau filaire sur lequel sont transportés des paquets UDP est simulé sur ns-2[1]. Des congestions sont générées sur un lien particulier pour forcer l'élimination intentionnelle de paquets par un routeur (*Random Early Detection*).

Deux critères sont utilisés dans [Oro04] pour comparer les performances de la méthode proposée à celles d'une diffusion sur un réseau IP *best effort* : la qualité de la séquence vidéo décodée et le taux de paquets perdus. La qualité est mesurée à l'aide d'une métrique perceptuelle objective de qualité : VQM (*Video Quality Metric*). Cette métrique sera détaillée dans le chapitre 7 de ce mémoire. Les résultats des simulations montrent que l'approche proposée est surtout efficace dans des situations de fortes congestions, pour lesquelles la qualité mesurée par VQM est améliorée de 20% par rapport à une approche *best effort*. L'amélioration de qualité est due à la perte des paquets les moins importants dans le cas d'une classification de trafic DiffServ.

Des résultats similaires en termes de qualités visuelle et objective (mesurée à l'aide du PSNR) sont rapportés dans [Shi01]. Pour le *streaming* de la séquence vidéo *Foreman* codée en H.263+, l'utilisation d'une seule file d'attente est comparée à l'utilisation de plusieurs files d'attente ayant chacune une priorité DiffServ. Une amélioration de qualité atteignant 4 dB (de 21 à 25 dB) est notée quand DiffServ est utilisé en présence d'un taux de pertes de paquets égal à 10%.

Les solutions citées précédemment apportent une solution au problème de l'utilisation de services multimédias sur IP avec un niveau satisfaisant de QdS. Cependant, le déploiement de DiffServ reste limité à des réseaux de petite taille comme les LAN (*Local Area Network*). De ce fait, l'ITU travaille sur la mise en place de réseaux de convergence multimédia à commutation de paquets : les NGN (*Next Generation Network*). Cette proposition, formalisée dans [itu04c], consiste à remplacer tous les réseaux de communication mobiles et fixes, cellulaires et télé-

[1] `http://www.isi.edu/nsnam/ns/`

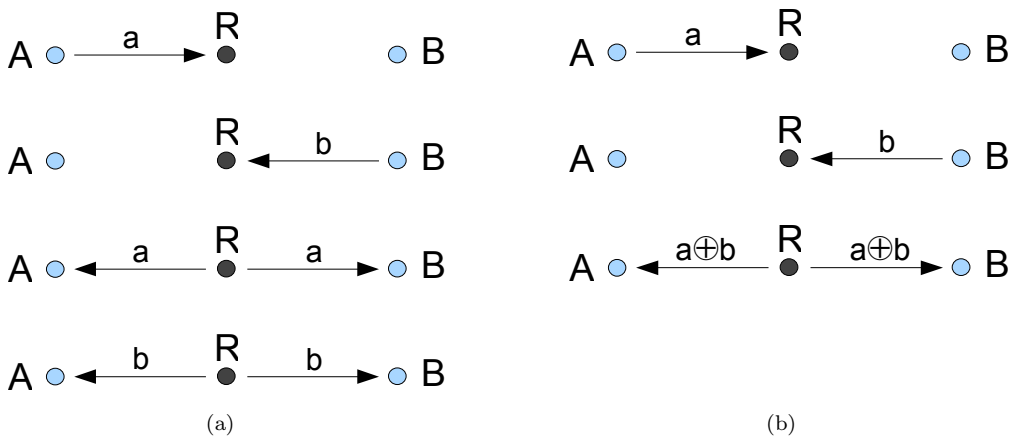


FIG. 2.1 – Diffusion d'informations radios (a) sans codage réseau et (b) avec codage réseau. A (respectivement B) veut envoyer un message a (resp. b) à B (resp. A) à travers le nœud relais R. Trois diffusions sont nécessaires au lieu de 4 quand le codage réseau est utilisé.

phoniques par un réseau basé sur le protocole Internet. Ce réseau est supposé permettre la transmission de flux multimédia avec indicateurs de QdS. IMS[2] (*IP Multimedia Subsystem*) a été proposé par l'organisme 3GPP (*3rd Generation Partnership Project*) pour être l'architecture de base des NGN multimédias.

Les approches que nous avons vues dans cette sous-section sont des approches réseaux au problème de l'amélioration de la QdS. Dans la section suivante, nous décrivons une approche agissant au niveau du réseau mais qui relève du domaine de la théorie de l'information.

2.1.2 Codage réseau

Le codage réseau (de l'anglais *network coding*), proposé par Ahlswede *et al.* en 2000 [Ahl00], repose sur l'idée que les éléments du réseau peuvent effectuer des opérations arithmétiques sur les paquets transmis pour optimiser l'utilisation de la bande passante disponible. Un exemple simple de codage réseau est donné figure 2.1.

A et B sont deux nœuds communiquant à travers le nœud relais R. Pour permettre la réception respective des messages a et b par B et A, quatre diffusions sont nécessaires dans le cas d'une transmission par diffusion classique. Si un codage réseau de type XOR est effectué dans le nœud relais R, la diffusion de $a \oplus b$ permet à A et B de décoder le message et retrouver l'élément qui leur manque.

[2] `http://www.3gpp.org/article/ims`

L'implantation du codage réseau au niveau d'un nœud permet à ce dernier de combiner des paquets ensemble de façon à ne former qu'un seul paquet. Ce paquet est ensuite transmis à d'autres nœuds qui à leur tour le combineront avec les autres paquets reçus des différents nœuds. Les paquets résultant des différentes combinaisons sont indiscernables et contribuent d'une manière équivalente au processus de décodage. Ainsi, un ensemble suffisant de paquets (sources ou combinaisons) autorise un décodage au niveau des nœuds récepteurs qui résolvent le système linéaire d'équations sous-jacent au codage et récupèrent les symboles sources.

Une des applications du codage réseau est l'amélioration de la QdS dans les réseaux sans fil. Dans [Sef07], des algorithmes sont proposés pour maximiser la qualité des vidéos décodées tout en préservant une utilisation optimale de la bande passante. Ces algorithmes sont basés sur des codes qui combinent des paquets appartenant à des flux vidéos différents. Le choix des paquets se fait en fonction de la contribution du code à l'amélioration de la qualité des vidéos. De plus, les délais (échéances) des paquets vis-à-vis du processus de décodage sont pris en compte. Les tests expérimentaux sur trois séquences vidéos ayant subi des pertes de paquets entre 1% et 20% montrent que la qualité augmente de 23 à 29 dB (en termes de PSNR) par rapport à des transmissions sans codage réseau.

La qualité des vidéos est mesurée dans ce travail à l'aide du PSNR. Sauf indication de notre part, toutes les mesures de qualité des travaux de la littérature présentés dans ce chapitre sont données en termes de PSNR.

Nguyen *et al.* [Ngu07] traitent aussi le problème des applications multimédias dans les réseaux sans fil. Ils proposent, dans le but de garantir une qualité maximale, un codage réseau qui augmente l'efficacité de l'utilisation de la bande passante. Ce codage est couplé au niveau du point d'accès à un algorithme d'ordonnancement des paquets basé sur un processus de décision markovien.

Un codage réseau prioritaire est proposé dans [Tho09] pour augmenter la robustesse des paquets vidéos contre les pertes dans les réseaux *overlay* filaires. En particulier, les paquets sont traités d'une façon différenciée suivant la classe d'importance à laquelle ils appartiennent. L'utilisation du codage réseau dans ce contexte permet d'augmenter la qualité de service sans avoir besoin d'un contrôle centralisé. Dans un souci de réduire le délai, le codage réseau est restreint aux paquets appartenant au même GOP. Les résultats présentés montrent une robustesse accrue pour des taux de pertes de 5% par rapport à un codage réseau classique.

Le codage réseau est décrit plus en détail dans le chapitre 3. Dans le reste de ce chapitre, nous décrivons les principaux mécanismes d'amélioration de la QdU.

2.2 Amélioration de la qualité d'usage

Les moyens utilisés pour améliorer la qualité d'usage sont présentés dans cette section. Dans le reste de la thèse, nous réduisons la qualité d'usage à sa composante visuelle uniquement car les contenus multimédias étudiés se réduisent aux images fixes et aux vidéos. Nous commençons notre état de l'art par les principaux mécanismes de codage canal.

2.2.1 Codes correcteurs d'erreurs

Les codes correcteurs d'erreurs par anticipation (FEC pour *Forward Error Correction*) représentent une évolution du contrôle d'erreur classique. C'est un codage canal dont le but est la protection des données sources en prévention des erreurs, des corruptions et des pertes de paquets. Ils sont surtout utilisés si la retransmission des données perdues n'est pas possible à cause de contraintes de délai ou de l'inexistence de canal de retour. Cette protection peut être soit répartie équitablement sur toutes les parties de l'information soit appliquée inégalement.

Dans le premier cas, les paquets reçoivent le même taux de protection sans prendre en compte une quelconque hiérarchie. Dans le second cas, une politique bien définie détermine l'attribution de la protection aux symboles ou paquets, le but étant de mieux protéger les flux importants. L'importance est déterminée à la source, principalement en fonction de l'impact sur la qualité d'usage que causerait la perte de telle ou telle donnée source.

2.2.1.1 Principe des codes correcteurs

Les codes FEC ajoutent de la redondance aux données sources pour les protéger des erreurs de transmission. Ils représentent une bonne solution pour les applications multimédias car ils permettent de s'affranchir de la retransmission des paquets perdus, coûteuses en termes de délai. Le principe général des codes FEC repose sur l'idée que la réception d'un sous-ensemble des données transmises doit être suffisante pour pouvoir reconstruire le message d'origine.

Soit k le nombre de symboles sources et n le nombre total de symboles transmis par bloc. La construction d'un code systématique permet la séparation des k symboles sources des $(n-k)$ symboles de redondance.

Le taux de codage (*code rate*) R représente la fraction d'information contenue dans chaque mot de code. Il est donné par l'équation suivante :

$$R = \frac{k}{n} \tag{2.1}$$

La valeur de R varie inversement avec le taux de redondance ρ du code donnée par :

$$\rho = 1 - \frac{k}{n} \tag{2.2}$$

Un code est d'autant plus efficace que la valeur de R est grande mais possède alors un faible pouvoir de correction.

Dans la famille des codes en blocs linéaires, les codes MDS (*Maximum Distance Separable*) sont des codes dont la distance de Hamming minimale d_{min} est égale à :

$$d_{min} = n - k + 1 \tag{2.3}$$

Ces codes sont optimaux : pour une quantité de redondance fixée, le nombre de symboles erronés que le code peut corriger est maximal. La capacité de correction d'un code en blocs linéaire dépend de d_{min} selon l'équation suivante :

$$t = \frac{d_{min} - 1}{2} \tag{2.4}$$

Pour un code MDS, l'équation (2.4) devient :

$$t = \frac{n - k}{2} \tag{2.5}$$

Le décodage de l'information initiale autorise au plus la perte de $(\frac{n-k}{2})$ symboles.

De la protection de symboles à la protection de paquets

La correction de symboles erronés nécessite au préalable leur détection ce qui réduit la capacité de correction du code. Dans un contexte de transmission de paquets de symboles, la détection des symboles perdus peut être facilement réalisée au niveau des paquets.

En effet, les paquets ont des identifiants contenus dans leur en-tête qui aident à la détection du paquet perdu. Par exemple, pour un flux multimédia transporté dans des paquets RTP, le paquet perdu est facilement identifiable au décodage grâce au champ *Sequence Number* de l'en-tête des paquets. L'application des codes correcteurs au niveau des paquets permet donc de profiter pleinement de la capacité de correction du code qui devient pour un code MDS :

$$t = n - k \tag{2.6}$$

En supposant l'identification des pertes, k paquets parmi n sont suffisants pour décoder. La perte d'au plus $(n - k)$ paquets est donc autorisée.

En mode paquet, l'utilisation des codes correcteurs peut se faire également sous forme systématique auquel cas les paquets sources ne sont pas modifiés lors de la génération des paquets redondants. Ceci peut être utile dans un scénario où les pertes n'impactent pas les paquets sources ou si aucune perte n'a lieu. Le décodeur peut ainsi directement utiliser les paquets sources sans nécessairement attendre la fin du décodage canal et/ou concentrer le décodage uniquement sur les symboles sources en cas d'impossibilité de reconstruction.

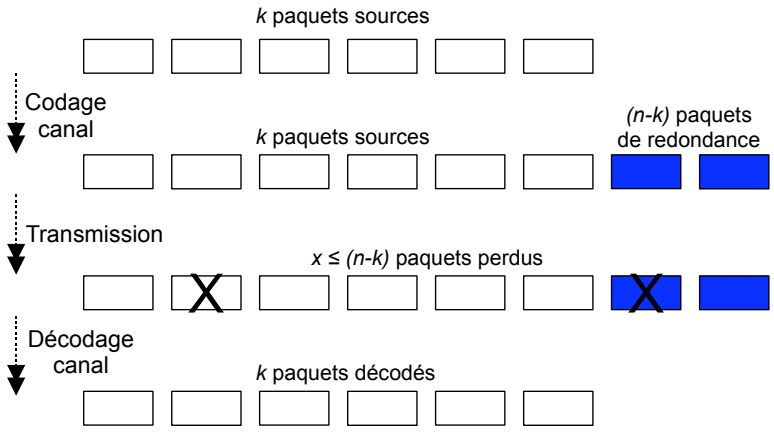


FIG. 2.2 – Principe de fonctionnement d'un code FEC systématique. Dans cet exemple, 2 paquets sont perdus pendant la transmission mais les 6 paquets sources sont reconstruits intégralement.

La figure 2.2 illustre le fonctionnement d'un code FEC systématique. Dans cet exemple, le nombre de paquets perdus (deux) est égal à la capacité de correction du code ce qui permet la reconstruction intégrale des paquets sources lors du décodage canal.

Codes correcteurs pour contenus multimédias

Un des codes FEC populaires pour la transmission de contenus multimédias sur des canaux non fiables est le code Reed-Solomon (RS). Ce code MDS, proposé par Irving S. Reed et Gustave Solomon en 1960 [Ree60], est utilisé dans des applications comme le stockage et la transmission de données sur des canaux à effacement de paquets. Par exemple, le code RS(204,188) est typiquement utilisé dans la norme DVB (*Digital Video Broadcasting*) [ets09]. La complexité des codes RS est assez élevée même si des optimisations pour la protection de paquets de données ont été proposées (par exemple dans [Riz97]). La complexité du codage est linéaire avec le nombre de paquets entrants k alors qu'elle est de $\mathcal{O}(klog^2k)$ pour le décodage.

Des codes correcteurs moins complexes mais aussi moins performants (car probabilistes) ont été proposés dans la littérature. Ces codes sont proches des codes MDS et sont appelés $(1+\epsilon)$MDS [Alo96] car ils nécessitent $(k+\epsilon)$ paquets parmi n pour reconstruire le message source. Par exemple, les codes Tornado [Lub97] ont pour de grandes tailles de données un temps de décodage 10000 fois moins important que celui des codes RS pour un surcoût ϵ de $3,2\%$.

D'autre part, la famille de codes Fountain comprend des codes à débit infini (*rateless*) qui peuvent être appliqués sur de très larges séquences de données tout en gardant une complexité acceptable. Un code Fountain implique que les données sources peuvent être reconstruites intégralement à partir de n'importe quel sous-ensemble de paquets ayant une longueur légèrement supérieure à la longueur des données sources. Parmi ces codes, nous retrouvons les codes

LT [Lub02] et leur extension, les codes Raptor [Sho06]. Les codes Raptor ont une complexité de décodage linéaire avec le nombre de messages à décoder. L'application des codes Raptor à la protection de services vidéos sur les réseaux sans fil a été proposée dans [Say07]. Pour un taux de pertes de paquets de 10% et un taux de redondance de 16%, une amélioration de qualité de 15 dB est notée par rapport à une approche sans protection.

2.2.1.2 Codes correcteurs d'erreurs avec protection inégale

Les paquets de données transmis sur un réseau IP *best effort* sont exposés de manière égale aux pertes alors que leur importance intrinsèque n'est pas la même. L'importance d'un paquet dépend de son contenu qui peut suivre une hiérarchisation lors du codage source. L'idée d'une protection inégale de contenus multimédias à travers des codes correcteurs d'erreurs est avancée par Albanese *et al.* dans un rapport technique en 1994 [Alb94] puis publiée dans *Information Theory* en 1996 [Alb96]. Dans un contexte de transmission de contenus multimédias sur un réseau non fiable, les auteurs proposent un système d'allocation de protection basé sur la priorité des segments de données. Cette priorité est fixée par l'utilisateur et sa valeur détermine le nombre minimal de paquets qui doivent être reçus pour garantir le décodage du segment concerné. Leicher applique ce système appelé PET (*Priority Encoding Transmission*) sur des vidéos codées en MPEG-1 dans [Lei94] et assigne les priorités selon le type de l'image à protéger, les images I et B recevant respectivement les priorités les plus et les moins élevées.

L'intérêt d'une protection inégale de l'information multimédia est mis en relief dans plusieurs travaux de la littérature. Girod *et al.* [Gir99] utilisent un code Reed-Solomon pour appliquer une protection inégale à des vidéos codées en H.263. Le codage effectué est un codage graduable (*scalable*) qui permet d'obtenir un flux hiérarchique à deux couches : la couche de base et la couche de raffinement. Un modèle de pertes Markovien à deux états (canal de Gilbert-Elliott) est classiquement utilisé pour simuler les pertes de paquets. Les performances de ce système de protection inégale sont comparées à un système de protection égale et à un autre où le flux de référence n'est pas protégé. Les résultats présentés montrent une dégradation de qualité graduelle pour le système à protection inégale à mesure que le taux de pertes augmente. Les deux autres systèmes sous test ne suivent pas cette allure mais enregistrent des chutes de qualité brusques à partir d'un certain taux de pertes. Il faut noter aussi que le flux non protégé a une meilleure qualité (32 contre 30 dB) dans le cas sans pertes. Ceci est dû au fait que les deux systèmes de protection doivent allouer une partie du débit disponible au codage canal. Si aucune protection n'est appliquée, la totalité du débit disponible est allouée au codage source ce qui permet une représentation de meilleure qualité de la source.

Mohr *et al.* proposent dans [Moh00] un système de protection inégale d'images dans lequel le codage est conjoint source-canal. Le codage source utilisé est un codage progressif SPIHT (*Set Partitioning in Hierarchical Trees*) dont le résultat est également un flux hiérarchique. Un

algorithme d'allocation de protection basé sur la maximation de la qualité de l'image reçue en termes de PSNR est proposé. Les tests menés sur deux images (*Lena* et une image médicale) démontrent la gradualité de la diminution de qualité avec l'augmentation des pertes de paquets. Par rapport à une protection égale de l'image, le système proposé présente une amélioration de qualité de 1 dB pour un faible taux de pertes. Cette amélioration s'élève à plus de 6 dB (de 18 à 24 dB) en présence d'un taux élevé de pertes.

Dans le même esprit, Bouazizi et Günes [Bou04] calculent les distorsions causées par la perte individuelle de la totalité ou d'un ensemble de paquets pour allouer la redondance de manière inégale. Ils mesurent, à l'aide du PSNR, la dégradation de la qualité d'une vidéo par simulation du processus de compensation de pertes au niveau de l'émetteur. De plus, ils considèrent l'effet de propagation temporelle de la perte due à la prédiction inter-images en calculant les distorsions de toutes les images jusqu'à l'image I suivante. Un facteur d'atténuation, qui varie en fonction de la distance entre l'image perdue et l'image dont on calcule la distorsion, est aussi intégré au calcul. L'approche testée fait preuve de robustesse pour les codes Reed-Solomon généralement utilisés pour la vidéo (taux de codage $R \geq 0,8$). Une approche similaire est présentée dans [Mar04] où la distorsion est calculée sur deux partitions de la vidéo contenant respectivement les composantes continues et hautes fréquences de la transformée DCT (*Discrete Cosine Transform*). Les résultats obtenus sur un canal à bruit blanc Gaussien additif montrent une amélioration de qualité atteignant 8 dB (de 17 à 25 dB) par rapport à une protection égale.

Plus récemment, une approche originale d'application des codes correcteurs d'erreurs sur une vidéo a été proposée par Huang et Apostolopoulos [Hua07]. Dans un contexte de *streaming*, il est proposé de protéger uniquement les paquets les plus importants. L'importance d'un paquet est déterminée en mesurant l'erreur induite si le décodage se fait sans ce paquet. De plus, pour pouvoir appliquer une forte protection aux paquets les plus importants, les paquets les moins importants sont délibérément éliminés au codage. Cette élimination permet d'augmenter la protection et de la concentrer sur les paquets qui contribuent le plus à la qualité finale. Les résultats présentés montrent cependant une légère amélioration de la qualité ($\Delta PSNR = 1,5$ dB) pour un taux de pertes égal à 15% par rapport à la simple protection des paquets les plus importants.

La protection inégale peut être guidée par la position temporelle de l'image. Dans ce cas, la protection d'une image diminue à mesure qu'elle est située plus loin de l'image de référence précédente. Cette approche est utilisée dans [Fan05] pour allouer plus de redondance aux images B et P les plus proches de l'image I du GOP. Le code FEC utilisé dans ce travail est le code RS. Le but visé est l'atténuation de l'effet de la propagation temporelle des pertes en protégeant les images susceptibles d'être les plus utilisées comme références des prédictions inter. Les résultats présentés justifient l'usage d'une telle approche sur des images P, montrant une amélioration de

qualité de 2 dB (de 27 à 29 dB) par rapport à une protection égale des images pour des taux de pertes de paquets de 5% et 10%.

L'association d'un codage graduable à une protection inégale semble être une bonne solution face aux pertes de paquets. L'intérêt du codage graduable est de créer un flux multimédia facilement adaptable au canal de transmission, à la variation de ses états et au terminal de réception. Dans SVC (*Scalable Video Coding*), l'extension graduable de H.264/AVC [Sch07], la graduabilité peut se situer à trois niveaux : la qualité, la résolution spatiale et la résolution temporelle. Pour ces niveaux, une couche de qualité de base est toujours codée avec éventuellement une ou plusieurs couches de raffinement. Dans [Ngu08], Nguyen *et al.* proposent de différencier la quantité de redondance allouée à chaque couche du flux en fonction de sa priorité. La vidéo est codée de manière à obtenir deux couches spatiales (QCIF - *Quarter Common Intermediate Format* et CIF - *Common Intermediate Format*), quatre couches temporelles et deux couches de qualité. Un code RS(255,239) est utilisé pour la protection des couches du flux. Les NALU de la couche de base obtiennent une protection égale au double de celle obtenue par les NALU de la première couche temporelle. Les tests effectués montrent une nette amélioration de qualité ($\Delta PSNR = 6$ dB de 20 à 26 dB) comparée à la qualité d'un flux sans protection au prix d'un surcoût de codage et de signalisation égal à 9%.

Dans [Ram06], Ramon *et al.* proposent d'associer un codage graduable Wyner-Ziv à une protection par code Reed-Solomon pour la transmission de vidéos. Le flux Wyner-Ziv est formé d'une version de la vidéo ayant une résolution spatiale réduite. Un code RS systématique est appliqué à cette version et seules les données de redondance sont transmises avec la vidéo codée à sa résolution initiale. En présence d'une perte de paquets, les slices perdues sont compensées à partir de la représentation de résolution inférieure. Les résultats expérimentaux montrent que ce système de protection est plus efficace que l'utilisation exclusive de codes FEC pour des taux de pertes de paquets élevés (supérieurs à 10%).

2.2.2 Codage et décodage robustes

Il existe des techniques de codage qui confèrent au flux multimédia codé une certaine robustesse face aux éventuelles erreurs subies sur le réseau. De même, au moment du décodage, le décodeur est capable d'appliquer une stratégie de compensation des parties du flux perdues ou corrompues. Ces deux approches sont abordées dans cette sous-section.

2.2.2.1 Codage vidéo H.264/AVC robuste

La norme H.264/AVC décrite à la section 1.3 du chapitre 1 propose des méthodes pour augmenter la robustesse de la vidéo codée face aux pertes de paquets. Ces méthodes se basent généralement sur une hiérarchisation des données sources en vue de protéger les parties les plus importantes de la vidéo.

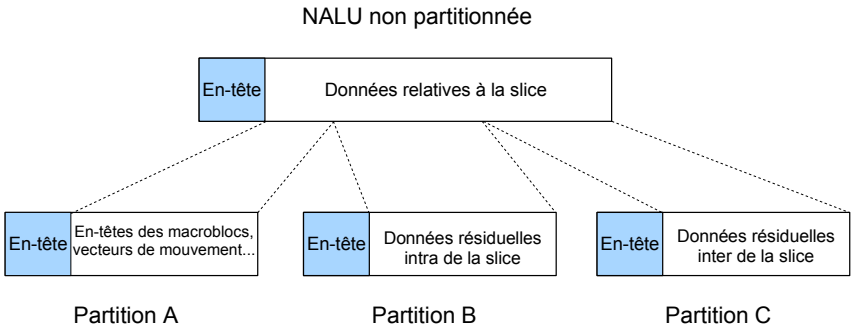


FIG. 2.3 – Le partitionnement de données DP de H.264/AVC.

Partitionnement de données

Le partitionnement de données DP (*Data Partitioning*) est une caractéristique de codage robuste disponible dans un seul profil H.264/AVC (*Extended Profile*). Conçu pour être utilisé sur les réseaux mobiles à bas débit, il consiste à diviser les données contenues dans une slice en trois partitions d'importance décroissante : la première, nommée A, contient les données les plus importantes que sont les en-têtes des macroblocs, les vecteurs de mouvement et les paramètres de quantification. Les partitions B et C incluent respectivement les coefficients de la transformée entière des macroblocs codés en intra et inter. Chacune des trois partitions est encapsulée dans une NALU déterminant ainsi son type. La perte d'une NALU B ou C n'interdit pas le décodage de la slice concernée et permet la reconstruction d'une représentation plus ou moins bonne du contenu perdu. Par contre, si la NALU A est perdue, le décodage est bloqué car les informations essentielles au démarrage du décodage manqueront. Le processus de partitionnement d'une slice est illustré figure 2.3.

Ghandi et Ghanbari étudient dans [Gha06] le surcoût de codage introduit par l'utilisation de DP et son efficacité dans les environnements sans fil. Ils concluent tout d'abord que les séquences codées avec l'outil DP activé ont une meilleure qualité (mesurée en termes de PSNR) que celles codées sans aucune forme de robustesse en présence d'un taux d'erreur binaires élevé (10^{-3}). Ils montrent aussi que le surcoût lié à l'utilisation de DP est négligeable à un taux d'erreur faible.

Dans [Wen03], Wenger utilise des traces de pertes réelles survenues sur des portions de l'Internet pour tester la robustesse de DP. Les pourcentages de perte utilisés sont 3%, 5%, 10% et 20%. Les résultats de l'étude sur deux séquences vidéos de résolutions spatiales QCIF et CIF montrent que les qualités subjectives et objectives sont satisfaisantes pour les quatre taux de pertes. Par exemple, pour la séquence *Paris*, une amélioration de qualité supérieure à 9 dB (de 20 à 29, 5 dB) par rapport à la même séquence codée sans outils de robustesse est obtenue pour

5% de pertes de paquets.

Slices redondantes

Un second outil de robustesse intrinsèque à H.264/AVC est la redondance de slices (*Redundant Slices*). Pour chaque slice de données dans le flux, il est possible d'en coder une représentation redondante qui pourra être utilisée si la slice principale (appelée primaire) est perdue. La slice redondante peut avoir des références de prédiction différentes de celles de la slice primaire au cas où des images de référence subissent des pertes ou des dégradations. Elle peut aussi être codée avec des paramètres de quantification plus sévères pour diminuer le surcoût de codage et en même temps réussir à obtenir une représentation grossière du contenu de la slice primaire perdue.

Tillo *et al.* présentent dans [Til08] une nouvelle méthode de génération de descriptions multiples basée sur les slices redondantes. La description multiple de l'information consiste à représenter des données sources par plusieurs descriptions indépendamment décodables. En partant d'une slice redondante pour toute slice primaire, deux flux binaires sont formés avec les deux types d'une même slice entrelacés dans chaque flux. La réception de ces deux descriptions permet le décodage intégral de la vidéo tandis que la perte d'une description engendre une reconstruction de moindre qualité. La redondance allouée aux descriptions est calculée à l'aide d'un modèle mathématique qui prend en compte la dégradation de qualité prévue au décodage. Cette dégradation dépend du contenu de la vidéo, du mode de codage et de l'état du réseau. Les résultats de simulation montrent une amélioration de la qualité atteignant 5 dB (de 29 à 34 dB) par rapport à d'autres approches de description multiple compatibles avec H.264/AVC présents dans la littérature.

Organisation flexible de macroblocs

Dans un système de codage classique, les macroblocs codés sont contenus dans une slice par ordre de balayage, de gauche à droite et de haut en bas. La perte d'une slice affecte alors une série de macroblocs adjacents couvrant la totalité ou une grande partie de l'image. La norme H.264/AVC propose un outil de dispersion des macroblocs appartenant à une même slice appelé FMO (*Flexible Macroblock Ordering*). Le but de l'utilisation de cet outil est de répartir spatialement les conséquences d'une perte pour qu'elles soient moins visibles. En effet, selon leur mode d'opération, les algorithmes de compensation de pertes (décrits à la sous-section 2.2.2.2) donnent de meilleurs résultats si les macroblocs à remplacer ont des voisins correctement décodés. De plus, l'assemblage de macroblocs ayant un même ordre d'importance dans une slice séparée apporte une flexibilité pour l'application d'une protection inégale au niveau du canal de transmission. Ainsi, la plus grande partie de la protection serait attribuée aux slices les plus importantes ce qui atténue l'effet des pertes sur la qualité finale de la vidéo.

Il existe sept types de configurations de FMO possibles selon le standard H.264/AVC. Les six premiers types (numérotés de 0 à 5) sont illustrés figure 2.4. La notion de "groupe de slices" est introduite pour décrire le partitionnement de l'image en slices qui contiennent des macroblocs non nécessairement adjacents. Le type 0 est le type "entrelacé" qui impose aux macroblocs d'être inclus (par ordre de balayage) dans des slices entrelacées. Le type 1 ("dispersé") permet de séparer les macroblocs selon le nombre de slices dans l'image. Ainsi, dans l'exemple de la figure 2.4, la séparation est faite sur la base de la parité du numéro du macrobloc dans l'image. Ce type de FMO est efficace dans le cas de la perte d'une des deux slices car il garantit, pour la compensation de pertes, la disponibilité d'échantillons couvrant toute l'image. Le type 2 appelé "premier et arrière-plan" décompose l'image selon des régions d'intérêts rectangulaires où les macroblocs importants sont classés dans les slices de premier plan (groupes 0 et 1). Les macroblocs restants sont disposés dans une dernière slice qui n'est généralement pas rectangulaire. Ce type peut être utile pour la séparation des objets d'une scène en vue de l'application d'une protection inégale en fonction de leur importance.

Les types 3, 4 et 5 de FMO sont des types dynamiques qui incluent les macroblocs en fonction de paramètres spécifiés au codage. Ces paramètres sont le nombre de groupes de slices, le nombre de macroblocs par slice et le sens de balayage. Le balayage est respectivement en spirale, horizontal et vertical pour les types 3, 4 et 5. En pratique, ces types sont rarement utilisés. Le type 6 est le mode FMO le plus explicite dans lequel l'appartenance aux slices est spécifiée au codage pour chaque macrobloc.

Une étude approfondie du surcoût de codage induit par l'utilisation de différents types de FMO est présentée dans [Lam06]. Il est ainsi montré que l'application du type 2 sur une image de résolution spatiale CIF 352×288 résulte en un surcoût moyen de $1,7\%$ pour plusieurs structures de GOP (I seulement, IBP... et IBBP...). Ce surcoût augmente avec la diminution du débit du fait que moins de bits seront disponibles tandis que le coût de signalisation reste le même. Des tests similaires sont effectués pour évaluer les performances du type 1. Les résultats rapportés montrent que ce type de FMO est assez pénalisant car il réduit les possibilités de codage intra et remet à zéro le contexte du codage entropique (à l'instar des tuiles en image fixe). Des surcoûts entre 10% et 20% sont ainsi notés avec des pics à plus de 35%.

Une comparaison entre le type 1 de FMO et les types 3, 4 et 5 dans des environnements à effacement de paquets montre la supériorité du premier [Maz09]. Pour un taux de pertes de 3%, la différence de qualité objective entre les vidéos décodées est de 2 dB (38 contre 36 dB). Cette différence augmente avec le taux de pertes pour atteindre quasiment 4 dB (31 contre 27 dB) à 10% de pertes.

Wenger compare dans [Wen03] la qualité de deux séquences vidéos codées avec activation de FMO par rapport à l'utilisation d'autres outils de robustesse offerts par H.264/AVC. Il déduit

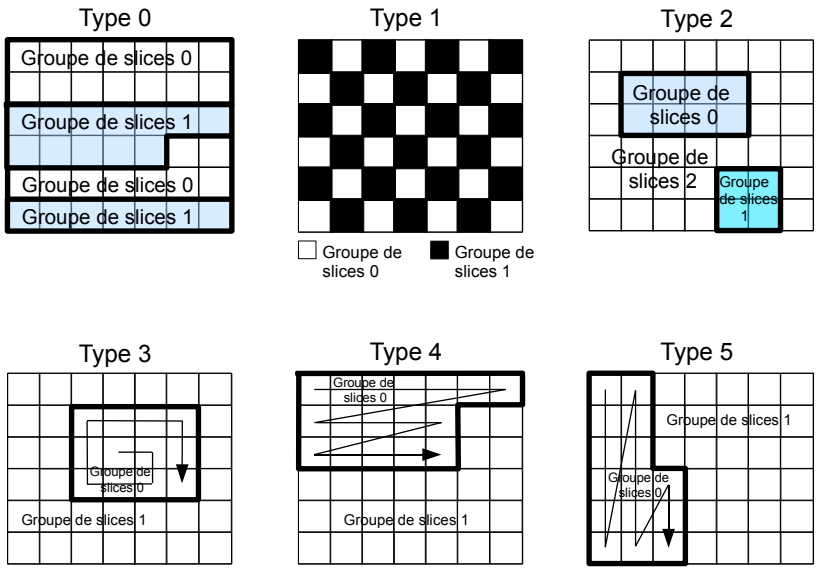


FIG. 2.4 – Les différentes configurations spatiales de FMO.

que pour des taux de pertes de paquets élevés (entre 10% et 15%), la qualité est le mieux préservée si FMO est activé. Ceci est largement dû à l'algorithme de compensation de pertes utilisé et à sa capacité à profiter de la présence de macroblocs non touchés par les pertes.

De même, Lambert *et al.* montrent dans [Lam06] que l'utilisation de l'outil FMO confère au flux vidéo une robustesse dont l'efficacité augmente à mesure que le taux de pertes de paquets devient plus élevé. Ceci est traduit par la différence de qualité qui se marque clairement entre la vidéo codée avec FMO et la vidéo codée sans activation d'aucun outil de robustesse. La simulation est effectuée pour deux niveaux de qualité initiaux : le premier faible avec un $PSNR<$ 30 dB et le second représentant une bonne qualité ($PSNR > 30$ dB). La décroissance de qualité avec l'augmentation des pertes est moins sévère dans le cas FMO que dans l'autre cas. Les différences de qualité liées à l'utilisation ou non de FMO sont plus grandes pour la vidéo ayant une bonne qualité de départ. Pour appuyer les notes de qualité obtenues à l'aide du PSNR, les auteurs demandent à quelques personnes d'évaluer visuellement la qualité des vidéos. Les observateurs notent alors que la taille de l'étendue spatiale des effets des pertes (dans le cas où FMO n'est pas utilisé) est la cause majeure de la dégradation rapide de qualité.

Dans [Kat07], Katz *et al.* proposent une nouvelle configuration de FMO pour améliorer l'efficacité de l'algorithme de compensation des pertes de paquets. Cette configuration est basée sur la combinaison des types 1 (configuration plateau d'échecs) et 3 (configuration spirale) pour favoriser la présence de macroblocs adjacents aux macroblocs éventuellement perdus. Son principe est le suivant : en partant du macrobloc d'indice 0 situé au coin gauche en haut de

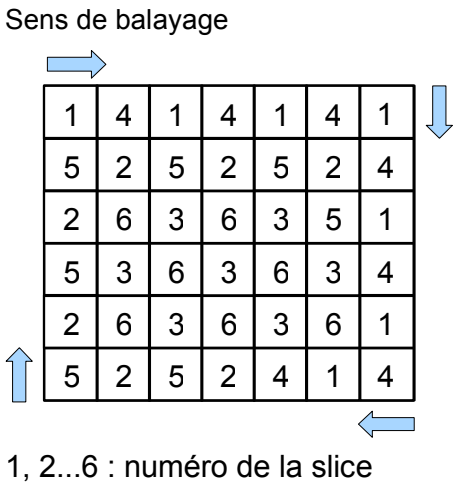


FIG. 2.5 – La configuration FMO combinant les types 1 et 3 [Kat07]. Dans cet exemple, l'image est formée de six slices chacune composée de sept macroblocs.

l'image, un balayage en spirale des macroblocs ayant des indices pairs est effectué. Quand le macrobloc central de l'image est assigné à la slice correspondante, le balayage reprend à partir du macrobloc d'indice 1 pour parcourir tous les macroblocs d'indices impairs. Une illustration de cette configuration de FMO est donnée figure 2.5 où chacune des six slices contient sept macroblocs. Cette configuration est mise en œuvre à l'aide du type 6 de FMO. Les tests effectués pour des séquences de résolution QCIF et pour des pertes allant de 10% à 50% de l'image montrent la supériorité de l'approche proposée par rapport aux types classiques de FMO. En effet, en termes de macroblocs correctement décodés et utiles à la compensation des pertes, une amélioration de 10% est notée quand 10% des macroblocs de l'image sont perdus.

D'autres outils de robustesse moins souvent utilisés sont proposés dans le standard H.264/AVC. Parmi ces outils, nous retrouvons l'actualisation par insertion de macroblocs codés en intra. Cette insertion se fait dans les slices B ou P pour réduire les dépendances temporelles entre les images successives et ainsi atténuer la propagation des pertes. Les résultats présentés dans [Hal04] montrent que l'insertion de macroblocs intra dans les régions contenant de grandes quantités de mouvement permet, par rapport à un codage classique, une dégradation graduelle de la qualité avec l'augmentation du taux de pertes. De même, Liang *et al.* proposent d'adapter le taux de macroblocs intra insérés au contenu de la vidéo et aux conditions de transmission [Lia06]. Ce taux est alors augmenté pour de fortes probabilités de pertes sur le canal de transmission et diminué si la vidéo est fortement texturée. En effet, dans ce dernier cas, le codage intra s'avère

coûteux en débit par rapport à un codage inter. Pour deux séquences vidéos à la résolution QCIF, la technique de robustesse proposée affiche les meilleures performances comparée à un codage sans actualisation intra ($\Delta PSNR > 4$ dB pour 1% de pertes). De même, une amélioration de qualité de 2 dB (de $24,5$ à $26,5$ dB) est notée par rapport à un codage avec un taux fixe de macroblocs intra pour un taux de pertes de 6%.

Enfin, la syntaxe du flux binaire H.264/AVC d'une part et sa décomposition en unités de transport NALU de priorités différenciées d'autre part contribuent aussi à sa protection. Ainsi, l'utilisation d'images IDR (*Instantaneous Decoding Refresh*) au lieu d'images I dans les GOP permet de remettre à zéro le contexte de codage. Ceci peut être utile pour arrêter la propagation temporelle des pertes d'un GOP au suivant. Par ailleurs, les NALU PPS ou SPS sont d'une telle importance au décodage qu'elles sont généralement mieux protégées au niveau du canal de transmission. Elles sont même parfois transmises sur un canal fiable pour en garantir la bonne réception.

Nous notons par ailleurs que la plupart des outils de robustesse évoqués dans cette section sont le plus souvent combinés à une forme de protection inégale comme nous allons le voir dans la sous-section 2.2.3 de ce chapitre.

Limitations du codage robuste L'inconvénient des techniques de codage robuste est l'introduction d'un surcoût de débit lié au découpage intra-image (qui réduit l'efficacité du codage) et à la signalisation. En effet, la hiérarchisation des données vidéos nécessite leur partitionnement en plusieurs entités indépendamment décodables ce qui limite l'espace de recherche des meilleures prédictions spatio-temporelles.

2.2.2.2 Méthodes de compensation

Un décodeur recevant un flux binaire de vidéo ayant subi la perte d'une ou plusieurs de ses unités de transport (les NALU dans le cas de H.264/AVC) peut soit remplacer les blocs d'image perdus soit choisir de ne pas le faire. Le remplacement de l'information perdue se fait en se basant sur les parties du flux correctement reçues.

Plusieurs méthodes de compensation de pertes existent dans la littérature. En général, elles sont divisées en deux groupes : les méthodes de compensation spatiale et celles de compensation temporelle. Ces méthodes peuvent agir sur une partie de l'image ou sur l'image entière, en fonction de la taille de la perte. Aucune méthode de compensation ne fait partie de la norme H.264/AVC même si le codeur de référence JM[3] (*Joint Model*) contient une implantation d'algorithmes particuliers.

Le mode d'action des méthodes de compensation du JM est déterminé par deux critères. Le premier critère est la taille de la perte et plus précisément si l'image entière est perdue ou une partie seulement. Le second critère est le mode de codage de l'image à compenser (intra ou inter).

[3]Disponible sur `http://iphome.hhi.de/suehring/tml/download/`.

Le document JVT-I049 du JVT [Sul03] spécifie le comportement du décodeur dans le cas d'une perte partielle de l'image. Pour ce type de pertes ayant lieu dans une image I, l'approche spatiale présentée dans [Sal98] est adoptée. Elle consiste à calculer, pour chaque pixel appartenant à un macrobloc perdu, une moyenne pondérée des pixels adjacents les plus proches du bloc contenant le pixel à compenser selon la formule de l'équation (2.7).

$$P(x,y) = \frac{\sum_{i=1}^{n} a_i * (B - d_i)}{\sum_{i=1}^{n} (B - d_i)} \tag{2.7}$$

où $P(x,y)$ est la valeur de l'échantillon (x,y) à compenser, a_i est la valeur de l'échantillon adjacent i, B est la largeur ou la hauteur du bloc dans lequel se situe l'échantillon à compenser et d_i est la distance entre l'échantillon à compenser et l'échantillon adjacent i. Généralement, le bloc à compenser est un macrobloc dont les dimensions sont 16×16 pixels. Les échantillons adjacents peuvent être au nombre maximal de 4. Il existe parfois des situations où les échantillons disponibles sont moins que 4, par exemple pour un macrobloc au bord de l'image ou un macrobloc ayant des macroblocs voisins eux-mêmes compensés. En effet, les échantillons appartenant uniquement à des macroblocs correctement décodés sont pris en compte. Mais si le nombre de ces derniers est inférieur à 2 alors les échantillons adjacents appartenant à des macroblocs compensés sont considérés. Le schéma de la figure 2.6 montre un échantillon perdu (marqué d'un x) et les échantillons correspondants qui sont utilisés dans le processus de compensation. Dans ce cas, nous supposons que les quatre macroblocs voisins du macrobloc perdu (de dimensions 16×16) ont été décodés correctement donc a_1, a_2, a_3 et a_4 peuvent être utilisés dans le calcul de la valeur moyenne de l'échantillon perdu. D'après l'équation (2.7), les coefficients les plus élevés sont attribués à a_1 et a_4 : $B - d_1 = B - d_4 = 16 - 2 = 14$.

Pour les pertes partielles dans une image codée en mode inter, le JVT propose dans [Sul03] un algorithme de compensation basé sur le travail de Lam *et al.* [Lam93]. Tout d'abord, la moyenne des vecteurs de mouvement des macroblocs de l'image correctement reçus est calculée. Sa valeur donne une indication de l'activité du contenu de la vidéo. Si cette moyenne est inférieure à un seuil prédéfini, chaque macrobloc perdu est simplement remplacé par celui situé à la même position dans l'image de référence. Cette méthode suppose que le mouvement est si faible entre les deux images qu'il est possible de confondre leurs macroblocs.

Si la moyenne calculée est supérieure au seuil, il est proposé d'utiliser les vecteurs de mouvement des blocs voisins pour remplacer le macrobloc perdu. Cette approche est justifiée par le fait que les mouvements de macroblocs voisins sont souvent corrélés. Le choix du vecteur de mouvement du macrobloc à compenser est basé sur la minimisation d'une mesure de distorsion calculée aux bordures des macroblocs. Le mécanisme d'opération est comme suit : pour tout vecteur de mouvement "voisin", la compensation de mouvement est effectuée pour obtenir les valeurs des échantillons à compenser. Ensuite, la somme des valeurs absolues des différences

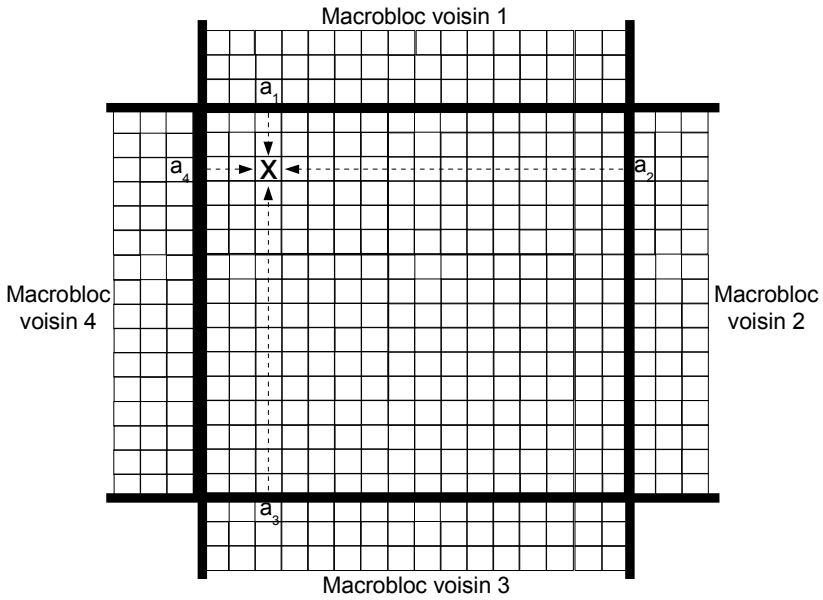


FIG. 2.6 – Compensation spatiale par moyenne pondérée. L'échantillon perdu est remplacé par la moyenne des valeurs des quatres échantillons voisins pondérée par leurs distances respectives.

entre les valeurs de luminance des échantillons "frontaliers" est calculée par vecteur de mouvement. Les échantillons frontaliers sont ceux qui se trouvent dans les colonnes (ou lignes) de part et d'autre des bordures du macrobloc comme illustré figure 2.7. La mesure de distorsion est calulée uniquement pour les macroblocs correctement décodés s'il en existe au moins un. Dans le cas contraire, les macroblocs compensés sont aussi considérés. Le vecteur de mouvement qui donne la distorsion la plus faible est alors choisi. Il faut aussi noter que le vecteur de mouvement nul, qui implique la copie du macrobloc situé à la même position dans l'image de référence, est toujours considéré comme candidat.

Halbach et Olsen ont comparé dans [Hal04] les performances de la compensation par remplacement direct des macroblocs perdus à celles de la compensation par calcul du meilleur vecteur de mouvement. La mesure des performances se fait ici à l'aide d'une métrique objective de qualité : SSIM (*Structural Similarity Image Metric*) [Wan04]. Cette métrique calcule la différence de luminance, de contraste et de structure entre les vidéos pour obtenir la note de qualité. Les auteurs concluent que la seconde méthode est aussi bonne que la première pour des séquences à faible quantités de mouvement alors qu'elle est légèrement plus performante pour les séquences où il y a beacoup de mouvement (type *Foreman*).

En ce qui concerne la perte d'une image entière, deux approches de compensation sont pro-

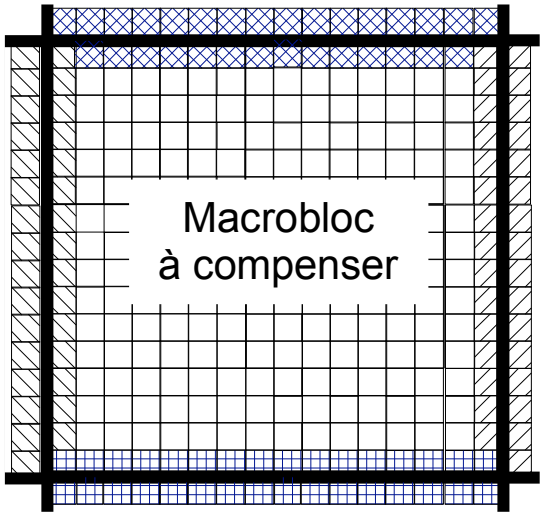


FIG. 2.7 – Échantillons de part et d'autre de la bordure d'un macrobloc à compenser. La mesure de distorsion est la somme des différences de luminance entre échantillons hachurés par le même motif.

posées dans le document JVT-P072 [Ban05]. La première consiste à remplacer tous les pixels de l'image perdue par les pixels correspondants de la dernière image de référence correctement décodée. Cette approche peut être considérée comme une technique de gel d'image où la dernière image correctement décodée est répétée pour toutes les images consécutives perdues. La seconde approche proposée prend en compte le mouvement supposé de l'image perdue en l'assimilant au mouvement de la dernière image de référence correctement décodée. Les vecteurs de mouvement et les indices de référence de cette dernière sont alors copiés et adaptés et la compensation de mouvement effectuée. La phase d'adaptation consiste à changer l'amplitude des vecteurs de mouvement en fonction de la distance de l'image à compenser par rapport à l'image de référence. Les résultats rapportés dans [Ban06] montrent que la méthode de copie de vecteurs de mouvements a de meilleures performances que celle de copie d'image, surtout pour les séquences contenant une grande quantité de mouvement. Par exemple, pour la séquence *Mobile* et à un taux de pertes égal à 3%, une différence de qualité égale à 3 dB est notée entre les deux approches.

Les méthodes de compensation implantées dans le codeur de référence ne sont pas les seules méthodes de la littérature. En effet, des méthodes de compensation spatiale, temporelle et spatio-temporelle existent. Les méthodes de compensation spatiale tentent de remplacer les macroblocs perdus par leurs voisins spatiaux appartenant à la même image. L'interpolation directionnelle est un exemple de compensation spatiale qui permet de préserver les contours de l'image. Elle consiste à détecter l'orientation des contours puis à appliquer une interpolation suivant cette

direction pour remplacer les pixels manquants. Les résultats rapportés dans [Nem05] montrent que la qualité des séquences décodées est légèrement améliorée par rapport à l'interpolation bilinéaire implantée dans le décodeur de référence. Une interpolation tri-directionnelle pondérée est proposée dans [Hua08] pour améliorer les performances de la compensation spatiale.

La compensation temporelle se base généralement sur les vecteurs de mouvement des macroblocs adjacents au macrobloc perdu. Kim *et al.* proposent dans [Kim05] d'utiliser en plus les modes de prédiction des macroblocs adjacents pour raffiner le processus de compensation. Dans [Han07], le vecteur de mouvement du macrobloc perdu est calculé par interpolation Lagrangienne (dans la direction verticale ou horizontale) des vecteurs de mouvement des blocs 8×8 adjacents. Les améliorations de qualité des deux méthodes précédentes ne sont pas significatives par rapport à la compensation temporelle implantée dans le codeur de référence. D'autre part, l'idée de l'utilisation des vecteurs de mouvement du macrobloc situé à la même position dans l'image de référence et de ses voisins est avancée dans [Xu08]. L'efficacité de cette approche est surtout notée dans les cas où les macroblocs adjacents au macrobloc perdu ne sont pas correctement décodés.

Les méthodes de compensation spatio-temporelle utilisent une des deux dimensions de compensation ou bien une combinaison des deux pour trouver le meilleur macrobloc qui remplacerait celui qui est perdu. Un algorithme de compensation spatio-temporelle est proposé dans [Bel03]. Cet algorithme effectue une compensation temporelle quasi-identique à celle du codeur de référence. La compensation spatiale prend en compte les contours et la texture de l'image et effectue un raffinement jusqu'au niveau du sous-macrobloc pour minimiser les distorsions dues aux effets de blocs. Le choix du mode de compensation se fait en comparant la valeur de distorsion engendrée par le macrobloc correspondant à chaque méthode de compensation à deux valeurs seuils. Les simulations sur deux séquences vidéos à la résolution QCIF montrent que l'amélioration de qualité obtenue avec l'algorithme proposé varie entre 1 et 6 dB ($PSNR < 30$ dB) pour des taux de pertes de 3% et 12%. Une approche différente pour choisir le mode de compensation est proposée dans [Agr06] où les modes de prédiction des macroblocs voisins du macrobloc à compenser sont considérés. Ainsi, la compensation spatiale est utilisée dans les images P si la majorité des macroblocs voisins sont codés en intra car ceci indique que l'image actuelle n'est pas assez corrélée avec l'image précédente.

2.2.3 Approches hybrides de protection de contenus vidéos

Dans la littérature, nous retrouvons des approches qui combinent les outils de codage source robuste à une protection inégale par codes correcteurs tout en assurant des mécanismes de compensation de pertes au décodage. Ainsi, Stockhammer et Bystrom proposent dans [Sto04] d'utiliser le partitionnement de données de H.264/AVC pour hiérarchiser le flux vidéo sur un réseau mobile. Ensuite, ils proposent d'utiliser un code convolutionnel pour ajouter de la redondance aux différentes partitions de données en veillant à ce que celle allouée à la partition

A soit supérieure à celles des deux autres partitions. Enfin, ils imposent une politique de compensation de pertes au décodeur en fonction de la partition reçue. Les résultats expérimentaux montrent que ce système de protection offre une qualité moyenne légèrement inférieure à celle d'un système de protection égale. Cependant, ce dernier est susceptible d'enregistrer des chutes de qualité brusques plus fréquemment. Une approche similaire de protection inégale dans ce même contexte de communications vidéos à bas débit est présentée dans [Har04]. Les performances en termes de qualité au décodage sont légèrement supérieures à celles du même système avec allocation égale de la protection ($\Delta PSNR = 0,05$ dB).

D'autre part, Thomos *et al.* [Tho06] étudient l'association de l'outil FMO de H.264/AVC avec les codes Reed-Solomon pour une protection inégale des macroblocs de la vidéo. Selon leur importance déterminée par une mesure de distorsion entre le signal original et le signal reconstruit, les macroblocs d'une image sont groupés dans trois slices d'importances faible, moyenne et élevée. Ensuite, des codes RS à différents taux de codage sont appliqués aux slices. Les résultats des simulations montrent que pour des taux de pertes de 10%, cette approche est plus robuste que celle proposée dans [Har04] enregistrant une différence de qualité de 2 dB (de 36 à 38 dB). Par contre, dans le cas sans pertes, l'utilisation de FMO influence négativement l'efficacité du codage et donc la qualité mesurée est légèrement inférieure à celle de [Har04].

Zhang *et al.* proposent dans [Zha06] de combiner le partitionnement de données DP à une optimisation débit-distorsion du choix du mode de codage. Les paramètres déterminant ce dernier sont la taille de macrobloc et son type de prédiction. De plus, ils appliquent une protection inégale en fonction du type de NALU transmise. Les résultats présentés montrent une amélioration de qualité de 3 dB (de 24 à 27 dB) pour un taux de pertes égal à 10% par rapport à l'utilisation de DP uniquement.

L'utilisation de DP avec un code Reed-Solomon pour l'augmentation de la robustesse d'un flux vidéo hiérarchisé est étudiée dans [Zha09]. La hiérarchisation dépend de trois critères : la distance par rapport à la première image du GOP, le débit source alloué à l'image et le type de partition de la NALU. Ainsi, une image proche du début du GOP sert de référence à un grand nombre d'images ce qui la rend plus importante qu'une image située vers la fin du GOP. Le débit et l'importance de l'image sont généralement corrélés du fait qu'une image I possède un débit supérieur à celui d'une image P ou B. La variation du débit canal alloué à chacun des paquets en fonction de son importance donne de bon résultats pour des taux de pertes entre 1% et 20%. La comparaison à un mécanisme de protection égale montre que l'approche proposée utilise moins de redondance et préserve un niveau de qualité supérieur (28 contre 32 dB pour un taux de pertes de 13%).

L'association de l'insertion périodique de slices codées en intra avec un code correcteur d'erreurs est proposée dans [Qu03] pour des communications sur réseau mobile. De plus, l'algorithme de compensation de pertes implanté dans le décodeur de référence est activé. Pour une séquence contenant une grande quantité de mouvement et à un taux de pertes de paquets de 7%, les

auteurs montrent que l'utilisation de la compensation de pertes uniquement n'est pas efficace. En la couplant à l'insertion périodique de slices intra ou à un code Reed-Solomon, le gain en qualité est de 6 dB (passant de 28 à 34 dB). En revanche, pour des séquences avec moins de mouvement, la compensation de pertes et l'actualisation intra améliorent à eux seuls la qualité de la vidéo décodée.

L'association de plusieurs méthodes de protection peut augmenter la complexité du système d'une manière significative ce qui pose problème dans les environnements ayant des contraintes strictes de délai. En effet, les codes Reed-Solomon sont essentiellement utilisés dans les travaux de la littérature mentionnés ci-dessus. Le décodage de ces codes implique au mieux une complexité quadratique. Il serait donc souhaitable d'utiliser des codes de moindre complexité. Cette approche sera étudiée dans le chapitre 4.

2.3 Discussion générale

En plus des mécanismes évoqués dans la section 2.2, d'autres méthodes sont proposées dans la littérature pour améliorer la qualité d'usage des services multimédias sur les réseaux IP. Parmi elles, nous retrouvons la description multiple de l'information [Goy01] qui consiste à coder l'information de manière à en obtenir plusieurs descriptions. En combinant ces descriptions au décodage, l'information intégrale est reconstruite. Si une ou plusieurs descriptions sont perdues, une représentation de qualité acceptable des données sources est toujours possible. La description multiple est utile dans le cas où plusieurs canaux de transmission de fiabilités différentes sont disponibles. Son application à des services multimédias a aussi été étudiée [Apo02, Wan05] et les résultats montrent une augmentation de la robustesse des flux vidéos.

Un récapitulatif des mécanismes évoqués dans ce chapitre avec des exemples de contributions est donné figure 2.8. Deux classifications de ces mécanismes sont considérées : la première par rapport au volet de la qualité visée (QdS ou QdU) et la seconde par rapport au niveau de la chaîne de transmission où le mécanisme agit (source, canal, réseau ou récepteur). En contrepartie de l'amélioration apportée par l'utilisation de ces mécanismes, quelques limitations sont à noter.

La classification de trafic est une solution générique au problème de la fluctuation de la qualité de service. Elle nécessite la création de domaines DiffServ où tous les routeurs implantent des stratégies de classification compatibles. Ceci pose problème à la mise en œuvre d'une telle solution à grande échelle. De même, le codage réseau suppose que tous les nœuds du réseau soient capables de combiner des paquets et de lire les vecteurs de coefficients transportés. Cependant, le codage réseau pourrait être une bonne solution s'il est utilisé au niveau des réseaux d'accès. À ce titre, le chapitre suivant propose un nouvel opérateur de codage réseau pour services multimédias.

En utilisant les codes correcteurs, un niveau satisfaisant de qualité est atteint surtout quand

la protection est réalisée d'une façon inégale. Mais la complexité des codes et les délais introduits peuvent parfois être un obstacle à leur utilisation. La deuxième partie propose un outil qui permet de réduire cette complexité.

L'emploi de techniques de codage robuste est une solution raisonnable quand la norme de codage offre de telles techniques. Cette solution permet d'atteindre une bonne qualité d'usage surtout avec l'utilisation de méthodes de compensation de pertes adéquates au décodage du flux multimédia. Par ailleurs, l'efficacité d'un codage source robuste peut être augmentée si la hiérarchie de la source est prise en compte ou si la perception humaine des dégradations est modélisée à l'avance. Ces deux aspects sont étudiés en détail dans les parties 3 et 4 de ce mémoire.

Conclusion

Nous avons vu dans ce chapitre qu'un ensemble d'approches peuvent être adoptées pour garantir d'une part la QdS et d'autre part la QdU dans un contexte de services multimédias sur IP. Les protocoles de différenciation de flux permettent d'associer un traitement particulier à ces services dans le réseau en vue de l'amélioration de la qualité de service. D'autre part, le codage réseau permet de son côté une utilisation optimale de la bande passante en combinant les paquets transmis pour réduire le nombre de transmissions nécessaires. L'amélioration de la qualité de service aboutit à une amélioration de la qualité d'usage.

L'amélioration de la qualité d'usage peut se faire aussi à l'aide de mécanismes opérant au niveau du codage source, du codage canal ou du décodage. Ces mécanismes sont préventifs ou correctifs auquel cas ils tentent d'atténuer les dégradations causées par les erreurs de transmission. Nous avons ainsi présenté les codes correcteurs d'erreurs FEC en mettant en avant leur application à des sources vidéos. Ces codes permettent dans certains cas d'adapter la protection à la hiérarchie de la source.

Nous avons également décrit les principales techniques de codage robuste présentes dans la norme H.264/AVC. Même si l'utilisation de ces techniques réduit l'efficacité de codage, elles se situent au plus près de la source à protéger et offrent donc une flexibilité dans l'application de la protection.

Enfin, nous avons donné des exemples de systèmes qui associent plusieurs mécanismes d'amélioration de la qualité d'usage. Une approche raisonnable consisterait à appliquer une protection en prévention des pertes puis compenser les pertes éventuelles au décodage.

Dans la suite de ce mémoire, nous proposons deux méthodes visant l'amélioration de la QdS (chapitre 3) et la QdU (chapitre 4). Ces méthodes sont fondées sur l'utilisation d'un outil simple : la transformation Mojette.

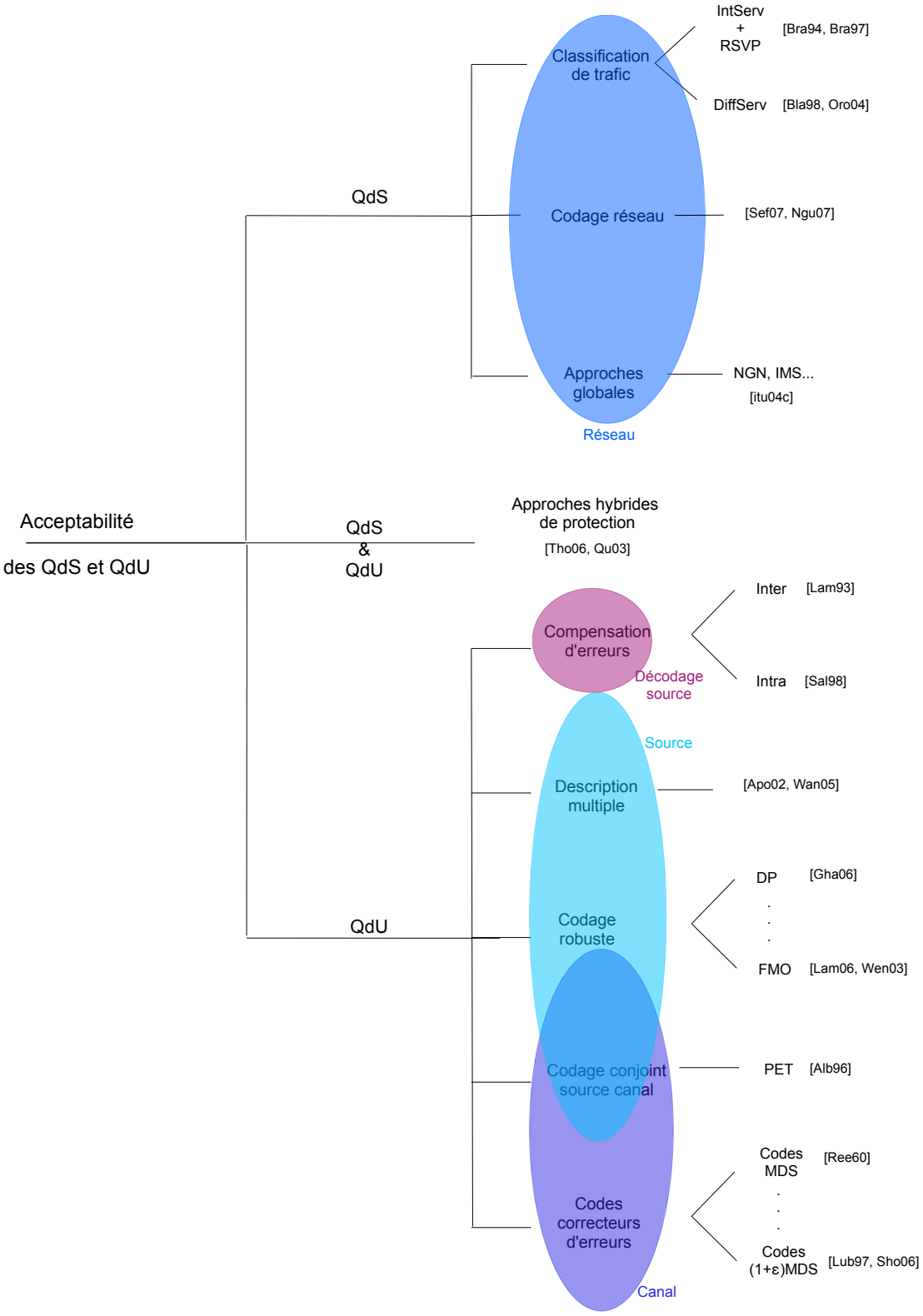


FIG. 2.8 – Classification générale des mécanismes d'amélioration de QdS et QdU évoqués dans ce chapitre.

Deuxième partie

Amélioration des qualités de service et d'usage par transformation Mojette

Chapitre 3

La transformation Mojette comme opérateur de codage réseau

Introduction

Comme nous l'avons vu dans la section 2.1.2 du chapitre 2, le codage réseau (*network coding*) se présente comme une méthode d'amélioration de la qualité de service. Cette méthode permet de réduire le nombre d'échanges de paquets nécessaires à la transmission d'une information et donc optimise la bande passante. La mise en œuvre d'une telle solution suppose que les nœuds médians du réseau sont capables d'effectuer des opérations arithmétiques simples sur les paquets avant de les retransmettre. Les paquets sont traités de façon équivalente sans priviligier l'un sur l'autre en fonction de son contenu.

D'autre part, la transformation Mojette est une transformation de Radon discrète qui permet de représenter un support de données 2D par des ensembles de données dits projections. Cette représentation redondante de l'information est effectuée à l'aide d'opérations simples avec une faible complexité. Cette propriété motive l'utilisation de la transformation Mojette en tant qu'opérateur de codage réseau.

Nous commençons ce chapitre par une description du mode de fonctionnement du codage réseau. Nous expliquons ainsi les processus de codage et de décodage puis nous rappelons quelques travaux de la littérature qui utilisent le codage réseau pour améliorer la qualité de service, en allant plus loin que le chapitre 2. Ensuite, nous présentons le formalisme mathématique de la transformation Mojette directe et de la transformation inverse qui sert à reconstruire le support original de données à partir des projections. Enfin, nous expliquons l'application de cette trans-

formation dans un contexte de codage réseau et discutons des avantages et des inconvénients de notre approche.

3.1 Le mode d'opération du codage réseau

Le codage réseau a été présenté brièvement dans le chapitre 2 et son utilité a été illustrée dans le cas de deux nœuds communiquant à travers un nœud relais (cf. figure 2.1). Prenons maintenant l'exemple d'un réseau "en papillon" comme celui de la figure 3.1. S_1 et S_2 sont deux sources émettant respectivement les symboles (bits) b_1 et b_2. R_1 et R_2 sont les deux nœuds récepteurs et A, B et C sont les nœuds intermédiaires. b_1 et b_2 doivent être reçus de R_1 et de R_2. Dans un contexte de transmission classique, R_1 reçoit en premier lieu le symbole b_1 du nœud A tandis que R_2 reçoit b_2 de B. Ensuite, le nœud C diffuse en un premier temps b_1 à R_2 et en un second temps b_2 à R_1. Il est évident que pour chacun des récepteurs, une diffusion parmi les deux était redondante. Si C était capable d'effectuer du codage réseau, par exemple au moyen d'une addition *modulo* 2 (XOR), il aurait eu besoin de diffuser uniquement $b1 \oplus b2$. Chaque récepteur ayant une partie de l'information, il peut obtenir l'autre partie qui lui manque. Dans cet exemple, trois transmissions sont effectuées au lieu de quatre donc le gain en efficacité est de 25%.

3.1.1 Codage et décodage

Les nœuds qui effectuent les opérations arithmétiques de codage réseau forment des combinaisons linéaires des paquets reçus. Ces opérations sont généralement des additions et des multiplications dans un corps de Galois $GF(2^p)$. Ainsi, pour N paquets sources, le nœud récepteur a besoin d'au moins N combinaisons linéaires de ces paquets pour décoder et reconstruire les paquets originaux. Les processus de codage et de décodage se déroulent comme suit : les paquets présents à un nœud sont combinés en leur attribuant des coefficients fixes ou variables (*cf.* ci-après). Les paquets combinés peuvent être eux-mêmes des combinaisons d'autres paquets. À la réception, les nœuds résolvent le système linéaire d'équations pour reconstruire l'intégralité des paquets sources.

Les paquets acheminés sur le réseau contiennent les données d'origine et le vecteur de codage (figure 3.2). Ce vecteur inclut l'ensemble des coefficients utilisés pour combiner les paquets sources. Les coefficients sont mis à jour par chaque nœud de codage en les multipliant par les nouveaux coefficients choisis par ce nœud. L'avantage principal de l'inclusion du vecteur de codage dans les paquets transmis est d'obtenir un système de décodage décentralisé. En effet, n'importe quel nœud recevant un nombre suffisant de combinaisons peut procéder à la reconstruction des paquets sources.

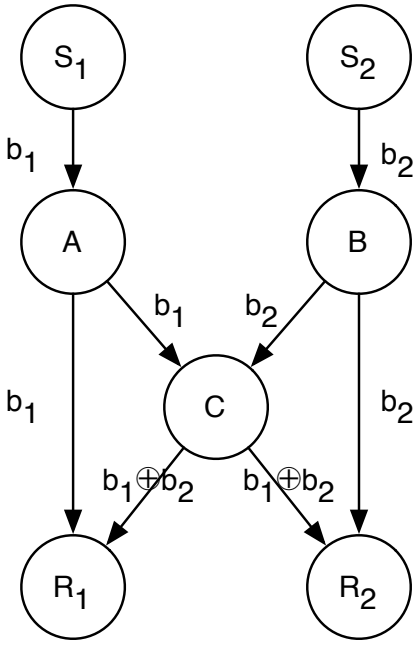


FIG. 3.1 – Codage réseau dans un réseau "en papillon" mettant en œuvre deux sources S_1 et S_2, deux récepteurs R_1 et R_2 et trois nœuds médians A, B et C.

Chou *et al.* montrent dans [Cho03] que la transmission du vecteur de codage dans les paquets n'induit pas un surcoût de débit significatif. Ils considèrent un réseau pour lequel la taille maximale d'un paquet est égale à 1400 octets en excluant les en-têtes. Pour la transmission de 50 symboles supplémentaires relatifs aux coefficients du vecteur de codage par exemple, le surcoût est égal à $\frac{50}{1400} = 3,6\%$ si chaque symbole est codé sur un octet.

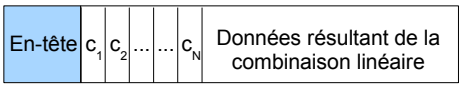


FIG. 3.2 – Un paquet contenant le résultat de la combinaison linéaire de N paquets et le vecteur des N coefficients utilisés.

Les coefficients utilisés pour former des combinaisons linéaires des paquets sont choisis dans un corps de Galois. Ils peuvent être choisis d'une façon aléatoire ou suivant une approche déterministe [Fra06]. Dans le premier cas, les coefficients sont choisis aléatoirement de façon à former des combinaisons linéairement indépendantes. La probabilité d'obtenir des combinaisons dépendantes tend vers zéro même pour de petits corps finis [Cho03]. Dans le second cas, chaque nœud

applique les mêmes coefficients à tous les paquets codés. Cette approche est moins flexible que l'approche aléatoire mais permet de ne pas transporter les vecteurs de codage dans les paquets.

3.1.2 Applications du codage réseau

Nous avons présenté dans le chapitre 2 des travaux utilisant le codage réseau dans un contexte d'amélioration de la qualité de service (par optimisation de l'utilisation de la bande passante). Le codage réseau trouve son application également dans la supervision de réseau [Ho05], les réseaux cellulaires [Lar06] et surtout les réseaux IP sans fil [Kat08]. Pour ce dernier type de réseaux, deux facteurs essentiels favorisent l'utilisation du codage réseau : la diffusion (*broadcast*) et la consommation d'énergie. La diffusion, de par sa nature physique (propagation d'ondes), est un moyen priviligié de transmission d'information dans les réseaux sans fil.

D'autre part, l'autonomie énergétique des nœuds est importante et une réduction du nombre de transmissions contribue à l'augmentation de la durée de vie du réseau. Les résultats rapportés dans [Deb05] montrent que le codage réseau réduit la consommation d'énergie moyenne des nœuds de 13% à 49% par rapport au multicast sur un réseau sans fil.

Le codage réseau a aussi été utilisé pour améliorer la robustesse des réseaux pair-à-pair contre les arrivées et les départs imprévisibles des pairs [Gka05]. Dans un réseau de distribution de contenus à large échelle CDN (*Content Distribution Network*), le nombre de pairs qui arrivent à finir le téléchargement d'un fichier quand le codage réseau est utilisé est 4 à 5 fois supérieur par rapport à une approche sans codage réseau.

La robustesse du codage réseau pour des transmissions en *multicast* est comparée à celle des codes correcteurs d'erreurs dans [Wan07]. Pour ce faire, une protection inégale est appliquée à une séquence vidéo codée en SVC à l'aide des codes correcteurs. D'autre part, un codage réseau aléatoire est utilisé. Les simulations effectuées pour des taux de pertes entre $0,05\%$ et $0,35\%$ montrent que la qualité de la vidéo décodée est meilleure dans le cas du codage réseau ($\Delta PSNR = 5$ dB). De plus, cette dernière approche évite les fluctuations de qualité observées quand les codes correcteurs sont utilisés.

3.2 La transformation Mojette

La transformation Mojette est une transformation de Radon discrète exacte développée au sein de l'équipe IVC (Image Vidéo Communication) du laboratoire IRCCyN (Institut de Recherche en Communications et Cybernétique de Nantes). Inventée par Jeanpierre Guédon au début des années 1990, cette transformation est essentiellement un outil de représentation redondante de l'information. Elle utilise la géométrie discrète pour transformer un ensemble de données 2D en groupes de données unidimensionnels qui sont des projections. Formant des combinaisons linéaires des données entrantes, la transformation Mojette est un outil potentiel de codage réseau.

La transformation Mojette trouve par ailleurs son application dans plusieurs domaines dont la tomographie [Ser05], la cryptographie d'image [Kin09], le stockage distribué [Gué01] ou la description multiple de l'information [Par01].

3.2.1 Présentation

Chaque angle de projection θ est défini par un couple d'entiers (p, q) premiers entre eux où $\tan\theta = \frac{q}{p}$. La transformation Mojette directe d'une image $f(k, l)$, notée $\mathcal{M}f$, représente un ensemble de N projections $\mathcal{M}_{p,q}f$ telles que

$$\mathcal{M}f = \{\mathcal{M}f_{p_i,q_i}, i = 1, 2...N\} \tag{3.1}$$

Pour obtenir chacune des projections, de simples additions et soustractions définies par les couples d'entiers (p, q) sont effectuées. Les projections ainsi calculées sont composées d'éléments appelés "bins". La valeur d'un bin de projection est égale à la somme des valeurs des pixels situés sur la ligne discrète déterminée par l'équation $m = -kq + pl$. La transformation Mojette est définie par les équations suivantes :

$$\begin{cases} [\mathcal{M}_{p,q}f](m) = proj(p, q, m) \\ [\mathcal{M}_{p,q}f](m) = \sum_k \sum_l f(k, l)\Delta(m + kq - pl) \end{cases} \tag{3.2}$$

avec Δ la fonction de Kronecker donnée dans l'équation (3.3) :

$$\Delta(m) = \begin{cases} 1 & si \quad m = 0 \\ 0 & si \quad m \neq 0 \end{cases} \tag{3.3}$$

Un exemple de projections Mojette est donné figure 3.3 où nous représentons un support de données de largeur 5 et de hauteur 3. La projection ayant un angle $(0, 1)$ se réduit à une addition verticale qui produit un vecteur contenant un nombre d'éléments égal à la largeur du support. La projection $(1, 0)$ est une addition horizontale qui résulte en un vecteur de trois bins (hauteur du support). La projection $(-2, 1)$ s'effectue en commençant par l'élément situé au coin gauche inférieur du support (de valeur 5).

Comme chaque pixel contribue nécessairement à un bin, la complexité d'une projection Mojette est $\mathcal{O}(PQ = I)$ avec P la largeur du support de données, Q sa hauteur et I le nombre total de pixels. Pour un ensemble de N projections, la complexité calculatoire est linéaire avec le nombre de pixels et le nombre de projections et donc devient $\mathcal{O}(IN)$.

Le nombre de bins d'une projection (p, q) effectuée sur un support rectangulaire de dimensions (P, Q) est donné par l'équation suivante :

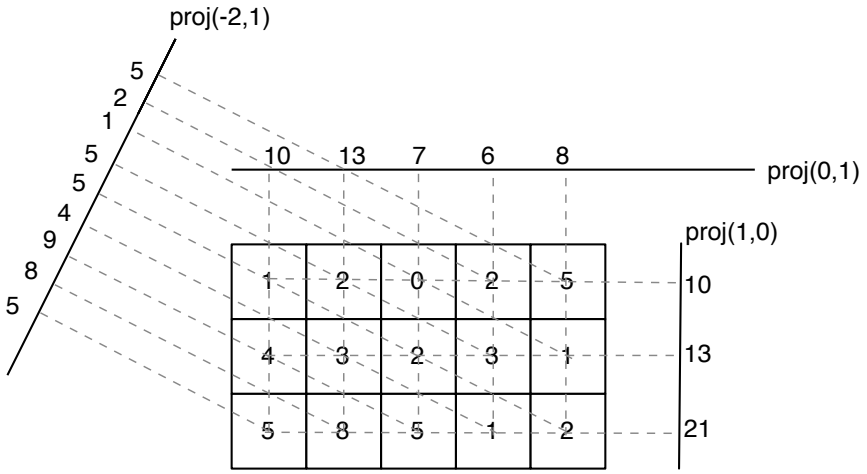


FIG. 3.3 – Trois projections Mojette d'un support de données 5×3 : $(-2, 1)$, $(0, 1)$ et $(1, 0)$. Cet ensemble de projections ne suffit pas à la reconstruction du support 2D à partir des projections 1D car ne satisfaisant pas au critère de Katz.

$$\#bins(P, Q, p, q) = |p|(Q-1) + |q|(P-1) + 1 \tag{3.4}$$

La transformation Mojette peut être également exprimée sous la forme du produit de la matrice de transformation M par le vecteur F des valeurs de pixels. Calculer la transformation Mojette inverse revient à résoudre le système linéaire $M.F = B$ où le vecteur B contient les valeurs des bins. L'algorithme de reconstruction Mojette est basé sur le fait que les bins ne correspondent pas au même nombre de pixels projetés. En particulier, il existe toujours un bin obtenu à partir d'un pixel singulier. La reconstruction commence alors en copiant la valeur de ce bin dans le pixel correspondant et en soustrayant la valeur de ce dernier de toutes les projections dans lesquelles il a été projeté. Cette opération fait apparaître de nouvelles correspondances unitaires.

La reconstruction d'une image à partir d'un ensemble de projections est donc un processus itératif qui :

1. cherche les correspondances unitaires pixel-bin ;
2. remplace le pixel par la valeur du bin (rétroprojection) ;
3. met à jour les projections.

L'itération s'arrête quand il n'y a plus de correspondances univoques ou si la reconstruction est terminée (les projections utilisées ont alors tous leurs bins à la valeur 0). La condition nécessaire et suffisante à la reconstruction intégrale d'une image rectangulaire de dimensions $P \times Q$ à partir de projections dans les directions $\{(p_i, q_i)\}$ est formulée dans le critère de Katz [Kat78] :

$$\begin{cases} \sum_i |p_i| \geq P \\ \text{ou} \\ \sum_i |q_i| \geq Q \end{cases} \tag{3.5}$$

Lors du calcul d'un ensemble de projections, nous choisissons de satisfaire une des deux conditions du critère de Katz de manière à limiter la redondance. Par exemple, si nous imposons une valeur de q_i égale à 1 pour toutes les projections, une image $P \times Q$ aura besoin d'exactement Q projections pour qu'elle soit totalement reconstruite, quelque soit la valeur de P. Dans un but de redondance, nous pouvons effectuer $N - Q$ projections supplémentaires ; dans ce cas précis, nous avons besoin de n'importe quel sous-ensemble de Q projections parmi les N calculées pour reconstruire intégralement l'image de départ. À q ou p constant, les projections possèdent un pouvoir de reconstruction équivalent.

Dans l'exemple de la figure 3.3, en ajoutant la projection $(1, 2)$ aux projections déjà effectuées, nous pouvons reconstruire l'intégralité des données car la somme des angles q est supérieure à la hauteur du support Q. Si nous voulons limiter la redondance à son minimum en retenant la capacité à reconstruire, nous pouvons garder les projections $(0, 1)$, $(1, 0)$ et $(1, 2)$ uniquement.

De plus amples détails sur la transformation Mojette et ses différentes applications sont fournis dans [Gué09]. L'utilisation de la transformation Mojette pour une protection inégale de l'information est détaillée au chapitre 4.

3.2.2 Application de la transformation Mojette au codage réseau

Les projections dans les différentes directions de l'espace peuvent être considérées comme des combinaisons linéaires des paquets entrants. La capacité de chaque projection à reconstruire les données sources est déterminée par l'angle de projection et par la forme du support à projeter. Ce support est apparenté à une mémoire tampon géométrique dont la largeur dépend de la taille des paquets entrants.

La figure 3.4 détaille le codage de trois paquets entrants à l'aide de l'opérateur XOR. Les trois projections qui sont produites sont suffisantes pour reconstruire les trois paquets. S'agissant d'une forme non rectangulaire mais $pq-$convexe, les conditions de reconstruction sont définies par la morphologie mathématique [Nor97].

Sur cet exemple, il apparaît que le nombre d'équations (les éléments des projections) est supérieur au nombre de symboles sources. Ceci présente un surcoût vis-à-vis de l'approche algébrique. Néanmoins, ce nombre d'équations redondantes est constant quelle que soit la taille du support pour un nombre de projections constant. Le rapport avec le nombre de symboles sources devient donc rapidement négligeable (pour un transfert de 36 kbits, il est déjà inférieur à $0,1\%$ dans cet exemple). En outre, leur formulation dans un contexte géométrique autorise une reconstruction progressive quel que soit le désordonnancement des paquets et permet d'atteindre

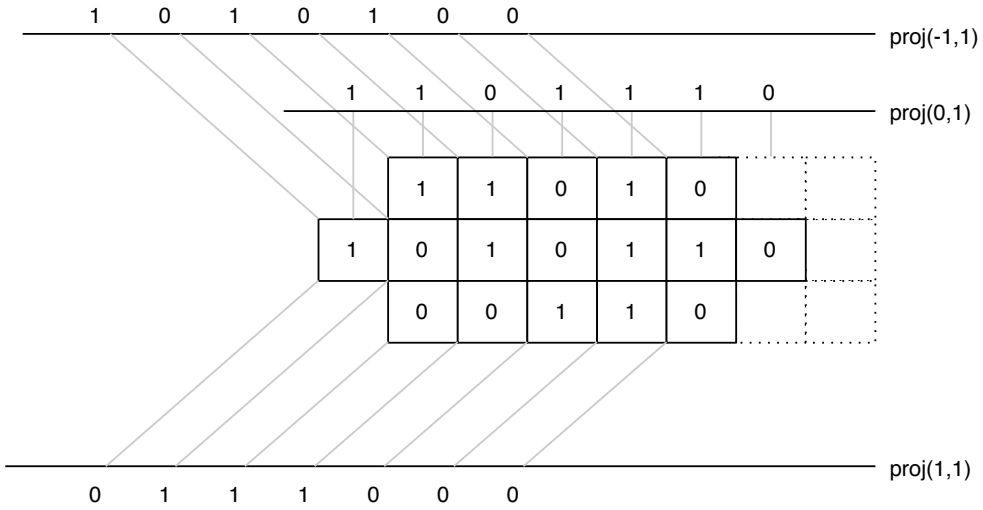


FIG. 3.4 – Un exemple d'ensemble suffisant de trois projections Mojette appliquées sur un support semi-infini de type hexagonal. L'usage de ce support permet d'obtenir des projections de taille constante (dans cet exemple 7 éléments).

une complexité linéaire avec écriture d'algorithmes d'inversion rapide [Nor06].

Nous illustrons sur la figure 3.5 une application pratique de la transformation Mojette. Cet exemple, inspiré de [Kat05], montre trois nœuds (circulaires) émettant et recevant des paquets par l'intermédiaire d'un nœud relais (carré). Nous supposons qu'un nœud a besoin de trois paquets différents pour reconstruire le message d'origine. La transmission du paquet résultant du XOR des trois paquets leur permet, en une seule transmission, de récupérer leur paquet manquant (partie gauche de la figure 3.5).

Dans le cas où la transformation Mojette est utilisée (partie droite de la figure 3.5), chaque paquet est une projection dans un angle discret déterminé. En appliquant la transformation Mojette inverse, le nœud relais reconstruit le message d'origine puis effectue une projection dans une direction bien précise : un quatrième paquet est généré. Il est alors transmis aux nœuds. La direction de projection doit être de la forme $(p, 1)$ pour permettre la reconstruction intégrale du message d'origine dans les nœuds (par transformation Mojette inverse). L'avantage de cette approche est de rendre la complexité des opérations effectuées dans tous les nœuds linéaire avec le nombre de paquets entrants. Cette complexité est au mieux $\mathcal{O}(n^2)$ pour les méthodes de résolution de systèmes d'équations linéaires.

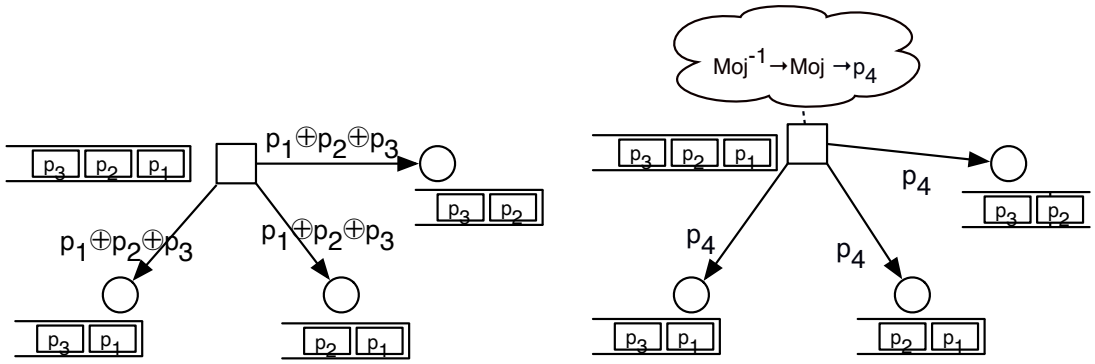


FIG. 3.5 – Codage réseau effectué par une opération XOR (gauche) et par la transformation Mojette (droite).

3.2.3 Performances et discussion

L'objectif principal du codage réseau est l'utilisation optimale de la bande passante. La possibilité de recouvrir à la totalité des messages suppose des mots de codes linéairement indépendants. Une approche optimale de codage réseau est capable de reconstruire k paquets sources à partir de n'importe quel sous-ensemble de k équations linéairement indépendantes.

La transformation Mojette est quasi-optimale en termes de capacité de reconstruction : elle nécessite $(k + \epsilon)$ projections pour la reconstruction de k paquets sources. Le paramètre ϵ est égal au rapport du nombre de bins de l'ensemble de projections avec le nombre total d'éléments sources. Il est donné par l'équation suivante :

$$\epsilon = \frac{\#nombre\ suffisant\ de\ bins}{I} - 1 \tag{3.6}$$

Le nombre suffisant de bins est égal au nombre de bins d'un ensemble de projections suffisantes. Pour $q = 1$, la taille d'une projection formulée dans l'équation (3.4) devient :

$$\#bins(P, Q, p, 1) = |p|(Q - 1) + P \tag{3.7}$$

Pour Q projections suffisantes, l'équation (3.6) devient alors :

$$\epsilon = \frac{(Q - 1)\sum_{i=1}^{Q} |p_i|}{I} \tag{3.8}$$

Ceci signifie que dans un ensemble suffisant de bins, les bins redondants sont constants pour un ensemble donné de projections. De plus, le surcoût est inversement proportionnel au nombre d'éléments sources I pour le même nombre de projections ce qui est favorable à l'envoi d'une grande quantité de données. Le tableau 3.1 montre des applications telles que la transmission d'images et de vidéos et leurs surcoûts correspondants. Les dimensions des mémoires tampons

TAB. 3.1 – Surcoûts introduits par la transformation Mojette pour une taille de projections de 1500 octets à différents débits. I représente la capacité de la mémoire tampon geométrique et Q le nombre de projections nécessaires pour la reconstruction de la source.

Source	I (kbits)	Q	Surcoût (%)
Image JPEG 2000	131	13	0,32
Vidéo QCIF (tampon : 12s)	322	29	1,72
Vidéo CIF (tampon : 3s)	444	39	3,18
Image médicale	617	55	6,59

géométriques nécessaires sont aussi données.

Toutes les projections effectuées sont de la forme $(p, 1)$ avec $\llcorner \frac{-(Q-1)}{2} \lrcorner \leq p < \ulcorner \frac{(Q-1)}{2} \urcorner$. Ceci implique que le nombre de projections nécessaires est égal à la hauteur du support. De plus, dans un contexte d'utilisation optimale de la bande passante, ce nombre est égal au nombre suffisant de projections. Dans le tableau 3.1, les tailles de projections sont fixées à 12000 bits (1500 octets) qui est la taille maximale d'unité de transfert MTU (*Maximum Transmission Unit*) sur un lien Ethernet. L'image JPEG 2000 (512×512) est codée avec pertes. L'image médicale concerne un codage sans pertes de type LAR [Bab03] pour une résolution de 350×350 pixels en niveaux de gris. Les séquences vidéos sont codées en H.264/AVC (*Main Profile*) aux résolutions QCIF et CIF . La durée en secondes est la durée du tampon géométrique qui représente le support de données avant de projeter.

Nous remarquons dans ce tableau que les surcoûts sont acceptables à l'exception de celui introduit par l'image médicale. Ce surcoût augmente avec le nombre de projections. En effet, le paramètre ϵ est une fonction quadratique du nombre de projections nécessaires selon l'équation (3.8). Ce nombre de projections caractérise la dispersion de la source dans les unités de transport. La multiplication des unités de transport liées à l'envoi d'un grand volume de données peut induire un coût prohibitif de cette dispersion. Cependant, la plupart des protocoles de transport autorisent et supportent la fragmentation au sens réseau. Dans ce cas, la contrainte sur la taille de projection peut être relâchée et P est augmenté de telle manière à faire diminuer ϵ. Par exemple, dans le cas de l'image médicale, pour une taille de projection de 18000 bits (soit 2250 octets), 35 projections sont nécessaires et par conséquent le surcoût est ramené à $1,7\%$.

Conclusion

Nous avons décrit dans ce chapitre la mise en œuvre de la transformation Mojette dans le cadre du codage réseau. Pour cela, nous avons utilisé une forme semi-hexagonale de support de données. Pour plusieurs types de sources multimédias, nous avons évalué le surcoût introduit par notre approche. Ce surcoût diminue à mesure que la taille du support de données augmente. En plus de sa faible complexité tant au codage qu'au décodage, cette transformation autorise la

reconstruction progressive de la source.

Le codage réseau traite tous les paquets de données d'une manière égale. Cette approche peut être suffisante dans un contexte d'amélioration de la qualité de service car elle réduit le nombre d'échanges de paquets par rapport aux paradigmes classiques de routage. Cependant, pour améliorer la qualité d'usage, il est important de prendre en compte le contenu de la source. Plus particulièrement, si cette source est hiérarchisée, il faut appliquer une protection en fonction de l'importance de chacune de ses parties. Le chapitre suivant s'inscrit dans cette direction.

Chapitre 4

Protection inégale perceptuelle de flux hiérarchiques par transformation Mojette

Introduction

L'efficacité de la protection inégale de contenus multimédias par codes correcteurs d'erreurs a été mise en avant dans plusieurs travaux [Alb96,Moh00] présentés au chapitre 2. L'application de cette protection nécessite toutefois une hiérarchisation préalable de la source. La norme JPEG 2000 étend dans sa partie 11 le processus de compression d'images aux aspects de transmission sans fil en proposant une protection inégale de l'information. Cette protection est guidée par des informations sémantiques renseignant sur la sensibilité de chaque portion du flux binaire aux erreurs de transmission. Dans cette partie de la norme, la sensibilité aux erreurs est déterminée à l'aide du PSNR et la protection effectuée au niveau des symboles.

Cependant, le PSNR n'est pas la meilleure métrique de qualité d'images. De plus, la protection au niveau des symboles est éloignée des canaux réels sans fil comme par exemple les réseaux basés sur le standard IEEE 802.11. En effet, JPEG 2000 Wireless (JPWL) ne considère pas l'efficacité des mécanismes opérant au niveau de la couche physique pour garantir l'intégrité des symboles transmis. L'approche proposée dans ce chapitre essaie de dépasser ces limitations en concentrant la protection au niveau des paquets. Cela suppose que la couche physique arrive quasiment toujours à préserver les symboles transmis de la corruption. Ainsi, un paquet est ou bien reçu ou bien perdu.

Dans ce chapitre, nous présentons un schéma d'allocation optimale de redondance applicable à une source hiérarchisée. En particulier, nous l'appliquons à des images codées avec JPEG 2000.

Ce schéma considère l'importance de chaque portion de l'information source et sa contribution à la qualité finale. Les incréments de qualité sont mesurés à l'aide du critère objectif perceptuel de qualité d'images C4 qui est décrit dans ce chapitre.

Nous proposons ensuite une nouvelle technique de protection inégale basée sur la transformation Mojette. Nous évaluons les performances de notre approche en la comparant à un schéma de protection basé sur les codes Reed-Solomon. Enfin, une extension de notre proposition à la vidéo clôt ce chapitre.

4.1 Graduabilité du flux JPEG 2000

La norme JPEG 2000 [jpe04] est une norme de compression d'images basée sur la transformée en ondelettes. JPEG 2000 permet l'obtention d'un flux graduable possédant plusieurs modes de progression dont les deux les plus importants sont la qualité et la résolution spatiale. Cette graduabilité est utile si les terminaux de réception ou les canaux de transmission du contenu multimédia sont hétérogènes. Une image est ainsi codée une seule fois et transmise dans un flux composé de sous-flux qui peuvent être décodés selon les capacités du terminal. De plus, la graduabilité améliore potentiellement la robustesse du flux contre les erreurs de transmission en permettant la transmission des couches les plus importantes sur un canal plus fiable que ceux transportant les couches de raffinement.

La graduabilité spatiale est obtenue dans JPEG 2000 grâce à l'analyse multirésolution inhérente à la transformée en ondelettes. La graduabilité en qualité est obtenue lors de la phase Tier2 du codage entropique EBCOT (*Embedded Block Coding with Optimal Truncation*) des coefficients de la transformée en ondelettes [Tau00]. Le codage EBCOT débute par la phase Tier1 pendant laquelle le train binaire est créé. Le codage entropique se fait par plans de bits en commençant par les bits les plus significatifs MSB (*Most Significant Bit*). EBCOT divise chaque sous-bande en petits blocs carrés appelés *code-block* (typiquement de taille 32×32 ou 64×64) qui sont codés de manière indépendante. La distribution de ces blocs sur plusieurs couches de qualité dans le flux binaire de l'image est effectuée par le processus d'optimisation débit-distorsion PCRD (*Post-Compression Rate-Distortion*). À l'issue de cette étape, la distorsion minimale (au sens de l'erreur quadratique moyenne) est obtenue pour une longueur de flux codé donnée. L'algorithme PCRD délivre les longueurs de chacun des *code-block* au sein des différentes couches de qualité de manière à minimiser la distorsion globale.

La figure 4.1 illustre un exemple de répartition des longueurs binaires de sept *code-block* sur cinq couches de qualité. Cette graduabilité en qualité est exploitée dans ce chapitre.

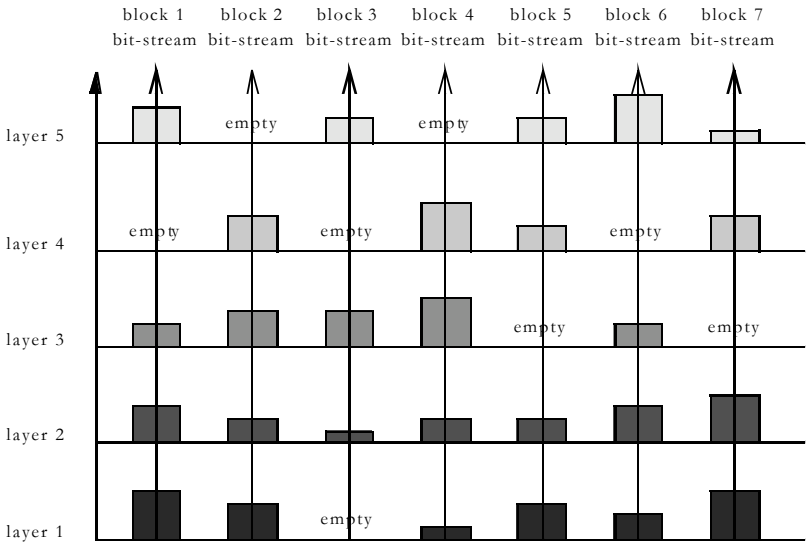


FIG. 4.1 – Un exemple de répartition des longueurs binaires de sept *code-block* d'une image JPEG 2000 et de leurs contributions à chacune des cinq couches de qualité. (extrait de [Chr00])

4.2 Allocation optimale de redondance

Pour qu'une allocation de redondance soit optimale, il faut qu'elle varie avec deux paramètres essentiels : l'importance des données à protéger et la probabilité de réception de ces données sans perte. Dans notre cas, pour un flux JPEG 2000 graduable, la redondance est allouée aux sous-flux (*substream*) en fonction de leur contribution à la qualité de la source et de la probabilité de recevoir une projection Mojette. En reprenant le formalisme de [Alb96], l'optimisation est effectuée à l'aide de la fonction de priorité ρ qui assigne à chaque sous-flux le nombre de projections nécessaires à sa reconstruction.

Soit s l'indice de sous-flux variant entre 1 et L où L est le nombre de résolutions ou couches de qualité. Comme la reconstruction de la source est progressive, reconstruire le sous-flux s est possible uniquement si le sous-flux $s-1$ est reconstruit. Il en résulte que la protection du sous-flux $s-1$ doit être supérieure à celle du sous-flux s. La fonction ρ est alors strictement croissante :

$$\rho_0 = 0 \leq \rho_1 \leq \rho_2 ... \leq \rho_L \tag{4.1}$$

Soit Q_s la mesure de qualité de l'image reconstruite à partir des sous-flux 1 à s. Pour une source graduable, nous supposons que $Q_s \geq Q_{s-1}$ et nous définissons alors l'incrément de qualité $\Delta Q_s \geq 0$ tel que :

$$\Delta Q_s = Q_s - Q_{s-1} \tag{4.2}$$

avec $\Delta Q_0 = Q_0$. Soit X une variable aléatoire qui représente le nombre de projections reçues pour chaque sous-flux. La fonction ρ doit maximiser la qualité au décodage représentée par l'espérance mathématique de la qualité $E[Q]$ selon l'équation (4.3) :

$$E[Q] = \sum_{s=0}^{L-1} Q_s Prob[\rho_s \leq X < \rho_{s+1}] \tag{4.3}$$

En posant $Q_s = \sum_{j=0}^{s} \Delta Q_j$, nous obtenons :

$$\begin{aligned} E[Q] &= \sum_{s=0}^{L} \left(\sum_{j=0}^{s} \Delta Q_j \right) Prob[\rho_s \leq X < \rho_{s+1}] \\ &= \sum_{j=0}^{L} \sum_{s=j}^{L} \Delta Q_j Prob[\rho_s \leq X < \rho_{s+1}] \\ &= \sum_{j=0}^{L} \Delta Q_j \sum_{s=j}^{L} Prob[\rho_s \leq X < \rho_{s+1}] \\ &= \sum_{j=0}^{L} \Delta Q_j Prob[X \geq \rho_j] \end{aligned} \tag{4.4}$$

L'équation (4.4) considère l'espérance mathématique de la qualité comme la somme des incréments de qualité ΔQ_j pondérés par la probabilité de recevoir les flux correspondants. Cette même équation considère conjointement les propriétés de la source (les incréments de qualité et donc sa graduabilité) et celles du canal (la probabilité de réception). Le système d'allocation de redondance utilisé est donc une optimisation conjointe des conditions des codages source et canal [Par01].

4.3 Contribution perceptuelle des incréments de qualité

Le PSNR est le critère objectif de qualité le plus utilisé à cause de sa faible complexité et la facilité de son implantation. Huynh-Thu et Ghanbari [HT08] ont montré que ses performances sont généralement cohérentes pour le même contenu et le même codec mais se dégradent pour des contenus différents. De plus, la fiabilité du PSNR a été mise en cause dans plusieurs contextes de codage [Wan03, Car04].

Nous choisissons dans ce travail d'utiliser la métrique objective de qualité perceptuelle C4 proposée dans [Car05] pour évaluer la contribution à la qualité finale de chacun des sous-flux

de l'image. Notre choix est motivé par les bonnes performances de cette métrique à référence réduite (coefficient de corrélation linéaire égal à $0,95$) sur une base de données de 100 images codées en JPEG 2000[1]. La métrique C4 opère comme suit : une description réduite de chacune des images (de test et de référence) est formée puis les descriptions sont comparées pour obtenir la note de qualité. Pour élaborer les descriptions, l'image est d'abord transformée dans un espace perceptuel de couleurs, l'espace de Krauskopf.

Ensuite, la sensibilité du SVH au contraste et les effets de masquage sont modélisés. La sensibilité au contraste du SVH varie avec la fréquence spatiale du stimulus. Elle est élevée pour les basses fréquences de l'image et diminue au fur et à mesure que les détails de l'image deviennent plus fins (hautes fréquences spatiales). Les effets de masquage ont lieu quand la visibilité d'un signal est diminuée ou augmentée par la présence d'un autre signal. Par exemple, l'amplitude d'une distorsion par effets de blocs dans une zone fortement texturée est inférieure à l'amplitude de cette même distorsion dans une zone uniforme de l'image.

Enfin, les traits caractéristiques sont extraits à partir de positions spatiales précises dans l'image appelés points caractéristiques. Ces points correspondent aux zones de l'image attirant l'attention du SVH. Les descriptions ainsi obtenues sont génériques et ne dépendent pas du système de dégradation de l'image. Le diagramme de blocs de C4 est illustré figure 4.2. Les notes de qualité produites sont comprises entre 1 (mauvaise qualité) et 5 (qualité excellente).

4.4 Protection inégale par transformation Mojette

La transformation Mojette peut décrire une information source d'une façon redondante en calculant un nombre de projections supérieur au nombre suffisant nécessaire à la reconstruction de cette information. Cette redondance est utile pour la protection de l'information contre d'éventuelles pertes de projections sur le réseau.

Le processus d'optimisation de l'allocation de redondance indique, pour chaque sous-flux de la source, le nombre de projections nécessaires à sa reconstruction. Cette reconstruction à la réception est possible si et seulement si le critère de Katz est vérifié (*cf.* équation (3.5) du chapitre 2). Les angles de projection sont choisis de telles façons à obtenir des projections équivalentes pour la reconstruction et à satisfaire la direction verticale du critère de Katz. Ils sont donc de la forme $(p_i, 1)$ pour $i = 1, 2...N$. De cette manière, Q_s projections parmi N sont nécessaires et suffisantes pour resonstruire le sous-flux s. Connaissant Q_s et la taille des sous-flux, des supports 2D sont constructibles pour chaque sous-flux s. Ces supports agissent comme des mémoires tampons géométriques dont la capacité est égale aux sous-flux qu'elles protègent. La protection inégale conférée par ce système vient du fait que ces mémoires tampons ont des

[1]H. R. Sheikh, Z. Wang, L. Cormack and A. C. Bovik, "LIVE Image Quality Assessment Database Release 1", `http://live.ece.utexas.edu/research/quality`.

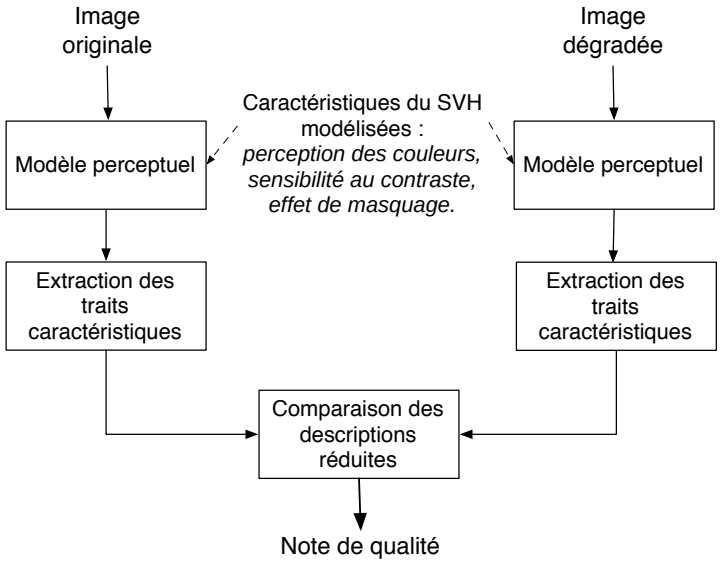


FIG. 4.2 – Le diagramme de blocs de la métrique de qualité C4.

niveaux de reconstruction différents en fonction des priorités des sous-flux. En effet, la protection diminue quand la hauteur de la mémoire tampon augmente.

La figure 4.3 illustre un système de protection inégale par transformations Mojette. Dans cet exemple, trois sous-flux de différentes priorités sont assignés à trois mémoires tampons de tailles différentes. Pour chacun des sous-flux, quatre projections de la forme $(p_i, 1)$ sont effectuées et leur ensemble $\{proj_{-2,1}, proj_{-1,1}, proj_{0,1}, proj_{1,1}\}$ est transmis. À la reconstruction, les sous-flux ont besoin de nombres de projections différents pour leur reconstruction. Ainsi, le sous-flux 1 bénéficie de deux projections redondantes tandis que les sous-flux 2 et 3 bénéficient respectivement d'une et d'aucune projection redondante.

La mise en paquets consiste à concaténer des projections appartenant à différentes mémoires tampons. Un paquet pourra contenir par exemple les trois projections $proj_{-2,1}$ ou $proj_{-1,1}$. De manière générale, on pourra chercher à optimiser la taille des paquets obtenus en mixant les angles de projection. La perte d'un paquet influencera alors différemment la reconstruction de chacun des sous-flux selon que le nombre suffisant de projections reçues est atteint ou non.

4.5 Évaluation de performances

Nous testons notre système d'allocation optimale de redondance sur une image *Lena* en niveaux de gris de dimensions 512×512. Nous simulons un environnement de transmission sans

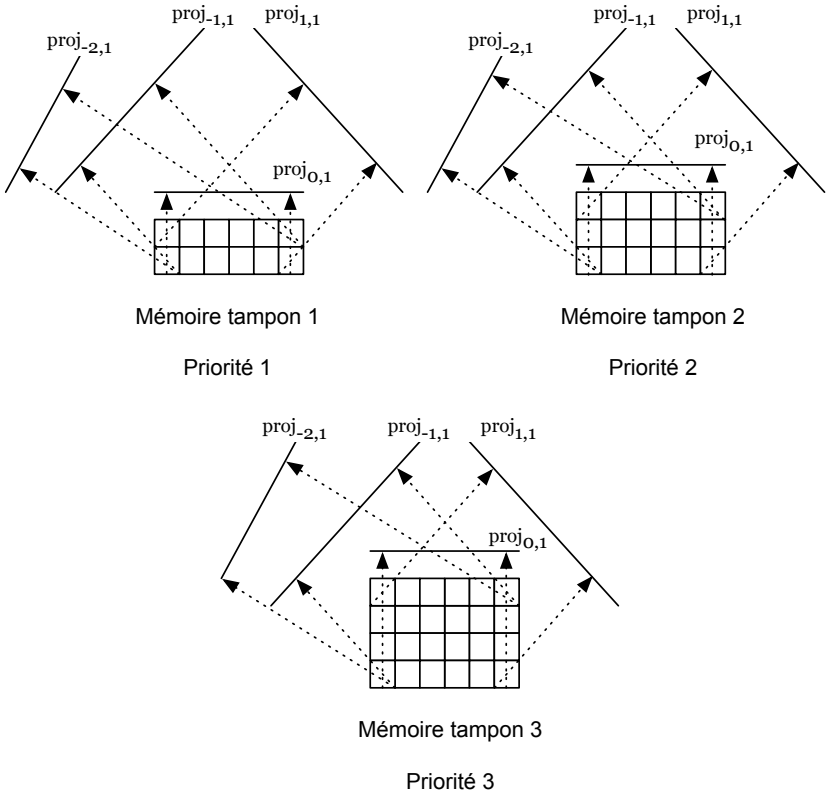


FIG. 4.3 – Système de protection inégale par transformation Mojette. Dans cet exemple, chaque mémoire tampon est dédiée à des données ayant une même priorité. Deux, trois et quatre projections parmi quatre sont respectivement nécessaires à la reconstruction des mémoires tampons 1, 2 et 3.

fil avec un modèle de pertes de paquets exponentiel reproduisant des pertes en rafale. L'image est codée en JPEG 2000 à un débit cible de $0,5$ bpp avec 5 niveaux de décomposition ondelette et 2 couches de qualité. Un paquet JPEG 2000 correspond à une seule couche de qualité d'un niveau de décomposition. Nous obtenons ainsi six paquets JPEG 2000 (incluant la sous-bande LL) par couche de qualité et par suite 12 paquets pour les deux couches de qualité considérées. Chacun de ces paquets a une note de qualité donnée par la métrique objective de qualité C4. Cette note, comprise entre 1 et 5, correspond à la qualité de l'image décodée à partir de ce paquet et de ses prédécesseurs.

Pour obtenir des sous-flux hiérarchisés par leur note objective de qualité, nous agrégeons les paquets n'apportant pas un incrément de qualité significatif en un seul sous-flux. De cette façon, nous avons six sous-flux qui correspondent à l'agrégation des paquets 1, 2 et 3 ; au paquet 4 ; au paquet 5 ; à l'agrégation des paquets 6, 7 et 8 ; au paquet 9 ; au paquet 10 et enfin à l'agrégation

TAB. 4.1 – L'agrégation des paquets JPEG 2000 et les sous-flux obtenus.

#paquet	1	2	3	4	5	6	7	8	9	10	11	12
taille (en octets)	247	542	1266	2303	2615	991	1	113	472	1276	2741	3538
qualité	1,00	1,00	1,10	3,41	4,21	4,50	4,50	4,50	4,63	4,68	4,88	4,89
#sous-flux	1			2	3	4			5		6	
débit source (en bpp)	0,08			0,15	0,23	0,26			0,32		0,51	

des paquets 11 et 12. Le tableau 4.1 résume la constitution des six sous-flux sources. Dans ce tableau, la taille de chaque paquet, sa contribution à la qualité de l'image et le débit source correspondant sont donnés.

À partir de cette description grabuable de la source, nous appliquons l'allocation optimale de redondance en considérant les incréments de qualité et le profil exponentiel de pertes de paquets. La protection inégale s'applique alors sur l'ensemble du flux (en-tête et données images). Elle se caractérise dans le cas d'un code Mojette par un nombre de projections suffisantes parmi un nombre N de projections envoyées. Nous choisissons ici $N = 16$ dans le même ordre de grandeur du nombre de paquets JPEG 2000 et permettant d'obtenir des projections de taille inférieure à la MTU sur Ethernet (1500 octets).

La courbe de la figure 4.4 représente les valeurs du couple (espérance de la note objective de qualité, débit global) pour un taux de pertes de paquets moyen égal à 10%. Les trois courbes correspondent respectivement aux codes MDS (la référence en termes de rapport débit/qualité) et à deux types de transmission prioritaire par transformation Mojette. Ces deux types diffèrent selon que les sous-flux ayant le même niveau de protection sont contenus dans la même mémoire tampon géométrique (courbe *opt1*) ou non. L'intérêt de cette opération est de réduire le surcoût de débit dû aux projections Mojette.

Les allures des courbes sont le résultat d'une recherche exhaustive de tous les motifs de protection respectant la monotonie de la fonction ρ. Cette recherche est effectuée sans aucune contrainte de débit ou d'espérance de qualité pour obtenir le plus grand nombre de motifs. Les différentes singularités (au nombre de six) correspondent à la décision du système de protection de transmettre un sous-flux supplémentaire pour atteindre l'objectif de qualité.

En choisissant la qualité d'image désirée à la reconstruction sur l'axe des ordonnées, nous pouvons déduire le motif de protection inégale à appliquer. Par exemple, une note de qualité de $4,51$ (pour un débit global source-canal de $0,41$ bpp) peut être obtenue par une protection de type 11-11-12-12-13-17 pour les six sous-flux. Le nombre 13 représente les projections nécessaires à la reconstruction du sous-flux 5. Il faut noter que dans ce cas de figure, 17 projections sont

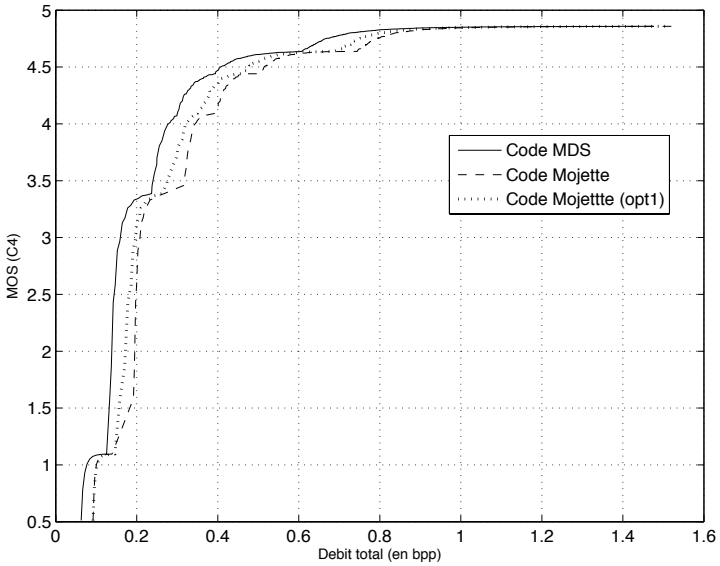


FIG. 4.4 – Résultats de l'optimisation débit global/qualité espérée pour les codes MDS et Mojette. Le débit global comprend le débit alloué à la protection. Les transmissions prioritaires par transformation Mojette diffèrent selon que les paquets sont agrégés en sous-flux auxquels est appliquée une même protection (courbe *opt1*) ou non.

nécessaires à la reconstruction du sous-flux 6. Ce dernier n'est donc pas transmis pour atteindre le niveau de qualité désiré. La taille des projections est de 843 octets.

De même, pour une valeur maximale de débit lue sur l'axe des abscisses, nous pouvons prédire la qualité obtenue à la reconstruction et le motif de protection inégale correspondant. Ainsi, si un taux de redondance de 5% est accepté pour la protection d'une source JPEG 2000 (ce qui fait monter le débit global à $0,5 + 0,5 * \frac{5}{100} = 0,525$ bpp), une qualité maximale égale à $4,62$ est attendue quand une protection 8-8-10-10-10-17 est appliquée aux six sous-flux. Dans ce cas, les sous-flux 1 et 2 peuvent être reconstruits même si la moitié de leurs projections est perdue. Contrairement aux approches par codes correcteurs classiques [Moh00], la protection appliquée au niveau des paquets JPEG 2000 nous autorise une fine granularité de protection. Ces deux exemples illustrent l'usage de la recherche exhaustive des motifs de protection une seule fois pour un contenu.

Comparé à un système de protection inégale à base de codes MDS, notre système fondé sur les projections Mojette a un surcoût global moyen au codage de $8,18\%$ dans le cas de l'image *Lena*. L'application du système de protection sur une image médicale de taille 512×512 codée sans pertes montre que le surcoût est réduit à $2,5\%$ et ce indépendamment du taux moyen de pertes de paquets. Il est aussi intéressant de noter que la réduction du surcoût est en moyenne de 4% quand les protections sont agrégées (courbe *opt1*).

Néanmoins, la complexité au décodage est inférieure dans le cas d'un code Mojette par rapport aux codes MDS. En effet, le décodage Mojette a une complexité en $\mathcal{O}(IN)$ linéaire avec le nombre de symboles d'information I et celui de projections N. Le décodage d'un code MDS a une complexité de la forme $\mathcal{O}(Ilog^2I)$ ou $\mathcal{O}(log^2I)$ dans le meilleur des cas [Lac09].

D'autre part, le code Mojette a les propriétés d'un code $(1+\epsilon)$MDS où ϵ représente le surcoût au décodage d'un nombre suffisant de projections. D'après l'équation (3.8) du chapitre 3, ce surcoût diminue quand le nombre total d'éléments d'information à transmettre augmente (dénominateur du rapport). Ceci motive l'utilisation de notre système pour des données volumineuses (vidéo, imagerie médicale, *etc.*).

4.6 Extension à la vidéo

Ce constat d'optimalité pour des données volumineuses nous motive à proposer la protection inégale par transformation Mojette pour la vidéo. Dans le contexte de la norme H.264/AVC, une protection inégale peut être appliquée aux NALU qui contiennent les données relatives au codage, les NALU VCL. Les NALU non-VCL sont supposées être reçues intégralement et sans aucune corruption car elles sont essentielles au démarrage du décodage. En particulier, nous proposons d'appliquer la protection inégale sur les NALU ayant subi le partitionnement de données DP. Selon leur type, les NALU sont affectées à des mémoires tampons géométriques qui déterminent le niveau de protection dont elles jouissent. Les NALU de type DP A sont traitées dans la mémoire tampon T_1 tandis que les NALU de types DP B et C sont mises dans la mémoire tampon T_2 où $hauteur(T_1) < hauteur(T_2)$. La figure 4.5 illustre ce processus. Ainsi, T_1 and T_2 contiennent respectivement les couches de base et de raffinement.

Au décodage, le nombre de projections nécessaires à la reconstruction des NALU A est inférieur à celui nécessaire pour reconstruire les NALU B et C. Donc un nombre plus grand de paquets perdus peut être toléré dans le cas des NALU A.

Cette approche est justifiée par le fait que les NALU B et C dépendent fortement de la NALU A qui, perdue, bloque le décodage de la slice entière contenue dans les trois types de NALU. De plus, si les NALU B et C sont corrompues ou perdues, les en-têtes de macroblocs et les vecteurs de mouvement contenus dans les NALU A peuvent toujours être utilisés pour obtenir une représentation grossière de l'image [Sto04].

Conclusion

Nous avons présenté dans ce chapitre un système de protection inégale d'images basé sur un algorithme d'allocation optimale de redondance visant à maximiser l'espérance mathématique de la qualité. L'allocation se fait en tenant compte des propriétés de la source (la hiérarchie

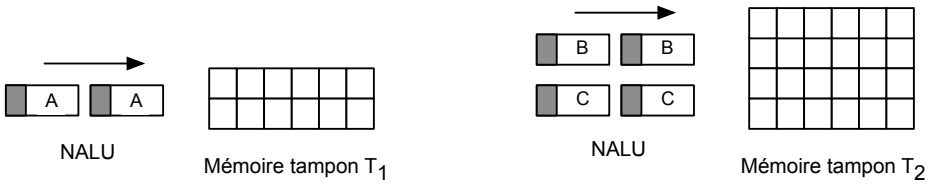


FIG. 4.5 – Protection inégale des partitions A, B et C. La mémoire tampon T_1 génère plus de redondance que la mémoire tampon T_2 pour un même nombre de projections.

qualitative) et des conditions du canal (la probabilité de réception). L'originalité de l'approche réside dans le codage redondant utilisé qui est une transformation de Radon discrète et exacte. De plus, l'importance des paquets à protéger est déterminée par C4 qui est une métrique objective de qualité modélisant des mécanismes du SVH. Nous avons dans ce cadre évoqué l'importance de l'intégration de caractéristiques du processus cognitif humain dans l'évaluation de la qualité visuelle. Enfin, nous avons proposé une extension de la protection inégale à la vidéo et plus précisément aux NALU constituant le flux binaire.

Ce travail est une première tentative d'inclure les paramètres liés à la perception humaine dans une solution au problème des pertes de paquets. Cette voie sera exploitée davantage dans les chapitres suivants.

Troisième partie

Vers une hiérarchie perceptuelle guidée par le contenu de sources vidéos

Chapitre 5

Caractérisation perceptuelle des effets de pertes de paquets

Introduction

La perte de paquets d'un flux vidéo conduit souvent à des dégradations de la qualité visuelle. Les causes des pertes et des exemples de leurs conséquences ont été présentés au chapitre 1. Dans ce chapitre, nous voulons apprécier la relation entre les pertes qui ont lieu au niveau du réseau et la qualité visuelle de vidéos codées en H.264/AVC.

L'impact perceptuel des pertes de paquets varie en fonction de plusieurs facteurs. Le facteur le plus intuitif est le nombre de paquets perdus ou l'amplitude de la perte. La qualité d'une vidéo ayant perdu plus de 1% de ses paquets peut être considérée comme mauvaise [Iva06]. Mohamed et Rubino [Moh02] montrent quant à eux que ce facteur n'est pas un bon indicateur de la qualité perceptuelle. En effet, la nature des données contenues dans le(s) paquet(s) perdu(s) ou la fréquence des pertes influencent également la perception des dégradations par l'utilisateur.

Pour évaluer l'impact perceptuel des pertes de paquets, nous adoptons une approche fondée sur les tests subjectifs d'évaluation de la qualité de vidéos. Durant ces tests, des vidéos dégradées par plusieurs motifs de pertes sont présentées aux observateurs qui doivent leur attribuer des notes de qualité. Les motifs de pertes utilisés sont introduits à des positions spatio-temporelles particulières dans le flux vidéo.

Nous abordons ce chapitre par une étude visant à identifier les paramètres qui peuvent atténuer ou amplifier l'effet perceptuel d'une perte de paquets. Pour ce faire, nous passons en revue les travaux de la littérature qui ont traité cette problématique et nous dressons le bilan des contributions principales. Ensuite, nous décrivons la procédure de simulation des pertes de paquets que nous utilisons et qui nous permet de contrôler de près l'introduction des pertes. Nous expliquons aussi la méthodologie générale des tests subjectifs d'évaluation de qualité de

vidéos et détaillons la mise en œuvre des deux tests (sections 5.4 et 5.5) conduits selon la recommandation BT.500-11 [itu02] de l'ITU. Enfin, nous analysons les résultats des tests et dégageons les principales conclusions.

5.1 Impact perceptuel d'une perte de paquets

Les conséquences de la perte d'un paquet dépendent de plusieurs facteurs liés aux différentes étapes de la chaîne de transmission. Selon Reibman et Poole [Rei07], quatre facteurs sont importants : l'algorithme et la configuration de codage, la mise en paquets, le comportement du décodeur face aux pertes et le contenu de la vidéo. En effet, les paramètres de codage déterminent la qualité de départ de la vidéo et sa robustesse contre les erreurs de transmission. La mise en paquets donne une indication sur le contenu d'un paquet et donc sur l'impact de sa perte. À la suite d'une perte, le décodeur utilise généralement un algorithme de compensation de pertes qui peut, suivant son efficacité, réduire l'influence de la perte sur la qualité. Il est enfin important de noter que les effets des pertes peuvent être amplifiés ou masqués en fonction du contenu de la séquence vidéo.

Dans ce chapitre, nous nous intéressons à l'étude de l'impact du motif de pertes et du contenu de la vidéo sur la qualité visuelle.

5.1.1 Travaux antérieurs

Plusieurs travaux de recherche se sont intéressés à l'étude des effets des pertes de paquets sur la qualité visuelle. Nous classons ces travaux en fonction de l'aspect étudié : la visibilité des pertes, la relation entre le motif de pertes et la qualité et l'impact perceptuel du gel d'images (à l'affichage) dû à la perte d'images entières.

Visibilité des pertes

Kanumuri *et al.* examinent le seuil de visibilité des dégradations liées aux pertes de paquets dans [Kan04, Kan06]. En vue du développement d'un modèle objectif linéaire pour le suivi de la qualité de vidéos sur un réseau IP, des tests subjectifs sont conduits pour quantifier l'impact des pertes sur la qualité visuelle. Durant ces tests, les observateurs doivent réagir quand ils perçoivent une dégradation de qualité dans des séquences vidéos codées en MPEG-2 [Kan04]. Il est ainsi montré que des facteurs comme la quantité de mouvement présente dans les paquets perdus, le nombre de paquets perdus et le type des images contenues dans ces paquets influencent nettement la visibilité des dégradations. Des résultats similaires sont obtenus dans [Kan06] pour des vidéos codées en H.264/AVC. Une différence est toutefois notée : la quantité de mouvement relative au contenu perdu n'est plus à considérer dès lors qu'une compensation de pertes basée mouvement est appliquée.

Dans [Rei07], les effets des caractéristiques du contenu des séquences sur la visibilité des dégradations sont étudiés. Les auteurs concluent que cette visibilité augmente si la perte a lieu au changement de scène et diminue si elle est située avant (jusqu'à 350 ms) ou après (quelques images). Ils montrent aussi que le mouvement de la caméra (zoom par exemple) peut augmenter la visibilité des pertes.

Impact perceptuel de la distribution des pertes

D'autre part, la relation entre la longueur de la séquence de paquets perdus et la qualité de la vidéo décodée a aussi été examinée. Dans [Lia03], Liang *et al.* montrent que la dégradation résultante d'une perte de paquets dépend fortement du modèle de pertes ; ainsi, la perte d'un nombre de paquets consécutifs produit une dégradation plus importante que celle produite par le même nombre de paquets perdus isolément.

Dans [Rah06], des tests subjectifs comprenant des vidéos à la résolution SD (*Standard Definition*) pour des applications de TV sur IP montrent que pour des pertes de moins de 200 ms, la qualité d'usage est plus influencée par le nombre de pertes que par leur longueur. Ceci implique que pour un même nombre de paquets perdus, la qualité est meilleure dans le cas d'une perte de plusieurs paquets consécutifs que dans le cas de plusieurs pertes singulières. L'importance de la distribution des pertes est soulignée dans [Moh02] où il est montré que la qualité visuelle est fortement influencée par le nombre de paquets consécutifs perdus. Plus précisément, les résultats de tests subjectifs montrent qu'à basse qualité (faible débit ou taux de pertes élevé), l'augmentation de la longueur de la séquence de pertes en conservant le même taux de pertes conduit à une meilleure qualité.

Impact perceptuel du gel d'images

Une étude approfondie de l'effet des pertes d'images entières sur la qualité perçue par l'utilisateur final est conduite par Pastrana-Vidal *et al.* [PV04b, PV04a, PV04c]. En particulier, les auteurs s'intéressent à l'élimination d'images. Cette élimination peut avoir lieu dans un contexte de communications multimédias interactives à délais courts. Dans ce contexte, des limites strictes sont imposées par le décodeur sur les délais d'arrivée des images. Ceci le mène à considérer une image comme étant perdue si elle accuse un délai supérieur au seuil fixé.

D'autre part, comme ces applications fonctionnent généralement à bas débit, le codeur peut aussi décider de ne pas coder une image pour économiser le débit. L'élimination d'images conduit à une rupture de fluidité qui représente un changement brusque de mouvement dans une séquence causant une discontinuité du mouvement apparent. Les images perdues sont remplacées par la dernière image correctement décodée provoquant ainsi un phénomène de "gel d'images".

Dans [PV04b], des tests subjectifs d'évaluation de qualité sont effectués pour déterminer le seuil de détection des discontinuités temporelles qui résultent de la perte d'images de la vidéo. L'influence de la durée de la perte, sa position temporelle et sa distribution dans le

temps ont été aussi étudiées. Une des conclusions de ce travail est que les séquences contenant des mouvements de personnages humains sont les plus sévèrement pénalisées par la rupture de fluidité. Par exemple, une discontinuité d'un mouvement de mains ou de corps est détectée facilement et cause une gêne pour l'observateur. Ceci est expliqué par le fait que notre perception est très sensible aux dégradations impliquant des humains.

Dans [PV04a], un modèle objectif d'évaluation de la qualité de séquences affectées par la perte d'images est proposé. Il est aussi montré que l'élimination régulière d'images est moins gênante pour les observateurs que les pertes irrégulières. Dans [PV04c], l'effet de masquage des changements de scènes est quantifié : les discontinuités temporelles sont généralement masquées si elles ont lieu jusqu'à 200 ms avant le changement de scène. Les auteurs insistent aussi sur la forte dépendance entre le contenu de la séquence vidéo et la note de qualité, même si la variation des valeurs de MOS avec le contenu n'excède pas les 10%. Cette dépendance est aussi notée dans [Feg05] quand la fréquence d'images varie.

5.1.2 Discussion

Les travaux de recherche présentés dans la sous-section précédente offrent des éléments de réponse à la question sur la relation qui lie les pertes de paquets à leurs effets perceptuels. De plus, ils adoptent des approches d'introduction de pertes qui est différente selon le contexte de chaque travail.

5.1.2.1 Sur la caractérisation de l'effet perceptuel des pertes de paquets

Les résultats rapportés dans [Kan04], [Kan06] et [Rei07] sont limités à la visibilité des dégradations causées par les pertes de paquets et non pas à la qualité perçue de la vidéo. D'autre part, un codage H.263 à bas débit est considéré dans [Moh02] tandis que le contexte de notre travail est le codage H.264/AVC à débits plus élevés (services type VOD).

Concernant l'impact perceptuel de la distribution des pertes, les résultats de [Lia03] sont contradictoires avec ceux de [Rah06]. Ceci nous mène à étudier de plus près cet aspect en faisant varier les motifs de pertes utilisés.

Enfin, les travaux de Pastrana-Vidal *et al.* concernent un contexte bien particulier, celui du service de visioconférence. Pour ce type de service, les vidéos sont codées à bas débit et ont une faible résolution. Ceci implique qu'une image peut être contenue dans un nombre limité de paquets voire un seul. Dans cette thèse, nous somme intéressés par des motifs de pertes différents. En effet, la perte d'une image entière à la résolution SD et aux débits utilisés implique la perte d'un très grand nombre de paquets. Nous avons jugé ce scénario assez rare et donc nous avons exclu la possibilité de perdre une image entière.

Les paramètres dégagés de cette étude bibliographique et qui sont d'intérêt pour notre travail sont le changement de scène et le motif de pertes. Le premier joue un rôle important de par sa

position et celle de la perte dans le flux vidéo. La distribution des pertes reflète la propagation temporelle de l'erreur qui a un effet non négligeable sur la qualité.

Par ailleurs, nous savons qu'il existe en général une région d'intérêt dans l'image sur laquelle se concentre l'observateur quand il regarde un contenu vidéo. Nous étudions donc l'influence de la position spatiale de la perte dans l'image en supposant dans un premier temps que la région d'intérêt se situe au centre de l'image. Nous examinons également l'effet perceptuel de la perte de différents pourcentages de paquets d'une image I. Ceci nous permet de quantifier l'impact de la perte d'une image I entière ou d'une partie de cette image sur les images suivantes et sur la qualité visuelle.

5.1.2.2 Sur l'introduction des pertes de paquets

La quantité de pertes de paquets introduite dans [Moh02] dépend du taux de pertes considéré. Cependant, la distribution temporelle des pertes dans le flux est aléatoire. Dans [Rah06], plusieurs motifs de pertes sont définis ce qui permet de contrôler les positions temporelles des pertes. Mais les types d'images dont les paquets sont perdus ne sont pas considérés. La propagation temporelle des pertes n'est donc pas rigoureusement étudiée.

Dans les travaux de Pastrana-Vidal *et al.*, un comportement particulier du décodeur vidéo est considéré : le gel d'images en cas de pertes de paquets. Cette approche, raisonnable dans un contexte de services de visioconférence, ne permet pas de généraliser les conclusions tirées quant à l'impact perceptuel de la perte d'images entières sur d'autres contextes.

Dans [Kan04, Kan06, Rei07], les auteurs adoptent une approche intéressante de simulation de pertes. Les pertes sont en effet introduites d'une manière spécifique, à des positions spatio-temporelles déterminées en fonction de la condition à tester (par exemple la position de la perte par rapport au changement de scène). L'intérêt d'une telle approche est qu'elle permet de connaître *a priori* le type d'informations perdues pour chaque motif de pertes ce qui aide à comprendre la relation entre les pertes et la qualité perçue.

Pour dégager des conclusions crédibles de notre étude, il est donc impératif de contrôler finement les pertes de paquets pour caractériser exactement chaque effet perceptuel. Ainsi, nous devons connaître le nombre de pertes introduites, leur distribution temporelle et le type d'images dans lesquelles les pertes ont lieu. De plus, nous souhaitons faire varier la position spatiale de la perte dans l'image.

Approche méthodologique pour l'étude des effets perceptuels des pertes de paquets

Le cadre général du travail présenté dans ce chapitre est donné figure 5.1. La première étape est le codage des séquences en tenant compte de la taille maximale d'une slice. Ensuite, le simulateur applique les motifs de pertes au flux vidéo et donne en sortie le flux dégradé. Ce flux est alors décodé par le décodeur et la vidéo obtenue est ainsi prête à être intégrée au test

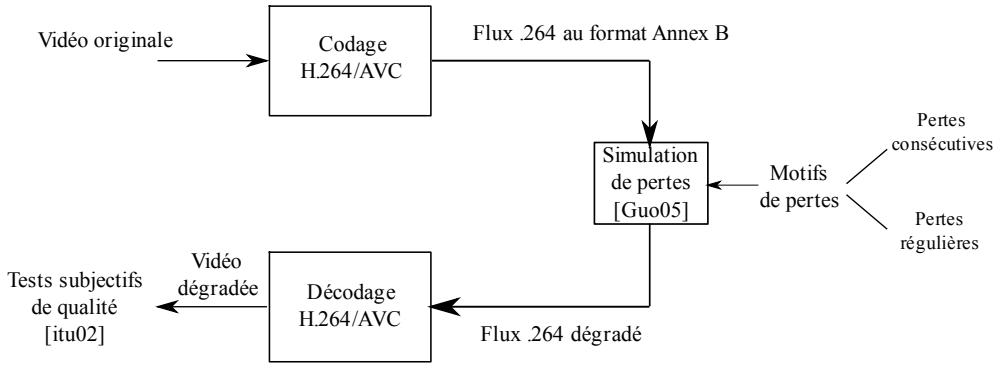


FIG. 5.1 – Le cadre général du travail présenté dans ce chapitre.

subjectif. Nous nous intéressons dans les deux sections suivantes à la simulation des pertes de paquets et à la méthodologie des tests subjectifs de qualité.

5.2 Simulation contrôlée de pertes de paquets

Tout en utilisant des motifs de pertes réalistes, les pertes doivent être introduites de manière à caractériser les effets perceptuels résultant de chaque motif.

5.2.1 Caractéristiques du simulateur

Nous choisissons de simuler les pertes de paquets au niveau des unités de transport du flux vidéo, les NALU. Cette approche nous permet de nous placer au plus près du réseau dans le contexte d'une transmission réelle sur un réseau de paquets.

Comme nous l'avons vu sur la figure 1.8 du chapitre 1, chaque slice de données codées en H.264/AVC est encapsulée dans une unité NALU. L'introduction des pertes au niveau NAL offre la possibilité de contrôler les positions temporelles et spatiales de la perte.

La position temporelle de la perte est déterminée par la position de la slice perdue dans le flux binaire. Plus précisément, elle dépend du numéro de séquence de la NALU dans la séquence de NALUs ordonnées par le codage. La position spatiale de la perte représente sa localisation dans l'image et elle est approximativement déterminée au niveau d'une slice car cette dernière peut contenir une ou plusieurs rangées de macroblocs.

Un exemple de variation de la position spatiale des NALU dans une même image est donné figure 5.2 qui illustre deux flux binaires contenant des NALU appartenant à des images I, P et B. Nous supposons dans cet exemple que chaque slice contient une rangée de macroblocs et que

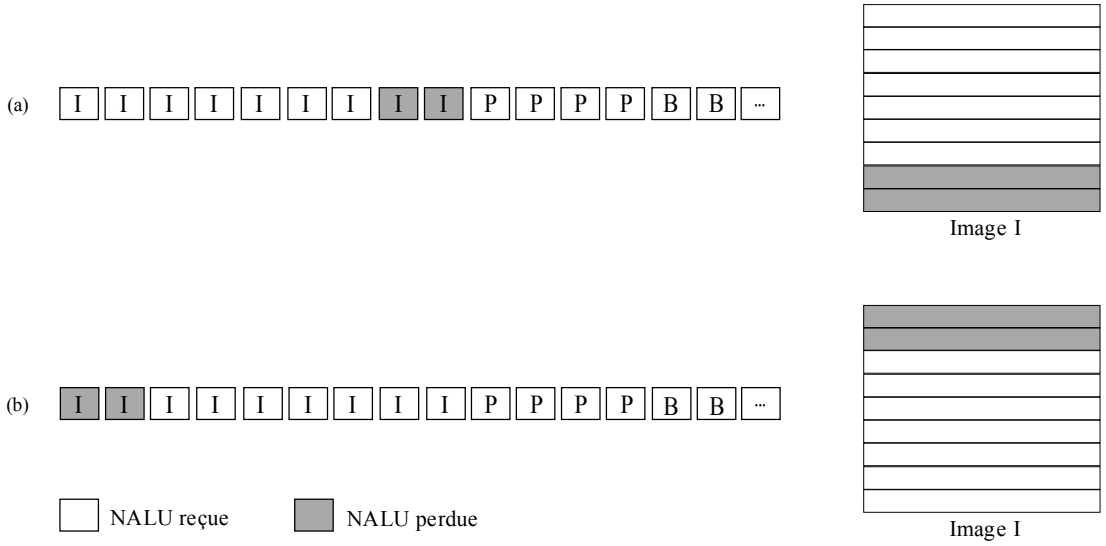


FIG. 5.2 – Représentation schématique de la variation de la position spatiale de la perte de deux slices au niveau du flux binaire (gauche) et sa représentation dans l'image (droite). Les parties inférieure et supérieure de l'image I sont respectivement perdues dans (a) et (b).

l'ordre de codage est de gauche à droite et de haut en bas. Comme chaque slice est encapsulée dans une NALU alors les NALU représentent des régions de même taille dans l'image. Les NALU perdues sont ici grisées. Sur la figure 5.2.a, les NALU perdues sont les dernières de l'image I donc elles se situent dans la partie inférieure de l'image. Par contre, les NALU du flux de la figure 5.2.b qui sont perdues représentent la partie supérieure de l'image.

Pour rendre notre simulation réaliste, nous fixons la taille maximale d'une slice à 1450 octets, en deçà de la taille maximale d'une unité de transport sur Ethernet (1500 octets). Les octets restants (au nombre de 50) servent à l'en-tête RTP/UDP/IP (40 octets) et aux éventuels octets additionnels utilisés par le codeur au-delà de la limite fixée. Les slices ne contiennent donc pas nécessairement le même nombre de macroblocs qui peut être supérieur ou inférieur au nombre de macroblocs d'une rangée de l'image. L'imposition de la taille maximale d'une slice est possible au moment du codage à l'aide du fichier de configuration du codeur. Ainsi, chaque NALU est encapsulée dans un paquet IP ce qui permet que le taux de pertes au niveau NAL corresponde au taux de pertes au niveau application (RTP par exemple). Dans la suite de ce chapitre, les termes "NALU" et "paquet" seront donc confondus.

5.2.2 Mise en œuvre du simulateur de pertes

Le simulateur de pertes utilisé dans notre travail est basé sur une version améliorée du simulateur du JVT [Guo05]. Initialement, ce simulateur simple prend en entrée un flux binaire H.264/AVC au format "brut" (appelé aussi Annex B) et un fichier de traces de pertes. Par format "brut" nous signifions un flux constitué de NALU disposées les unes à la suite des autres (une autre possibilité étant l'encapsulation des NALU dans des paquets RTP). Les motifs de pertes peuvent provenir de fichiers de traces de pertes réelles comme ceux fournis dans le document JVT-Q069 [Guo05]. Ces derniers sont le résultat d'une série de mesures effectuées sur le *backbone* de l'Internet en 1999 [Wen99]. Ils représentent des pertes sporadiques ou en courtes rafales (2 à 3 paquets consécutifs) accusant des taux de pertes de 3%, 5%, 10% et 20% en moyenne. Nous n'utilisons pas ces fichiers pour deux raisons. Premièrement, les pourcentages de pertes sont grands et ne couvrent pas l'intervalle $0,1\% - 2\%$ qui correspond mieux à notre domaine d'étude. Ensuite, les pertes s'étalent sur la totalité de la durée des séquences vidéos ce qui ne permet pas d'isoler les pertes pour étudier de plus près leurs effets.

En sortie du simulateur, nous obtenons le flux binaire dégradé. Ce flux contient nécessairement moins de NALU que le flux d'entrée car des NALU ont été éliminées en fonction du fichier de pertes. De plus, le simulateur délivre des informations sur les flux originaux et dégradés et sur les pertes introduites (taille des paquets, quantité d'information perdue, positions spatiale et temporelle de la perte, *etc.*). La version du simulateur que nous avons développée conserve la simplicité initiale tout en ajoutant de nouveaux modes d'introduction de pertes (spécification des NALU à éliminer, du type des images concernées par les pertes ou de la taille de rafale, *etc.*) et en délivrant des informations plus détaillées sur les flux d'entrée et de sortie (types des NALU, types des images ayant subies des pertes).

5.3 Méthodologie générale des tests subjectifs

Pour une évaluation fiable des effets perceptuels des pertes de paquets, nous adoptons une méthodologie rigoureuse de tests décrite dans cette section.

5.3.1 Les observateurs

Un test subjectif commence par la sélection des observateurs. Généralement, les personnes passant le test doivent avoir une acuité visuelle correcte et une bonne distinction des couleurs. Cette vérification est effectuée avant le début des tests au travers de tests de vision. Les personnes choisies pour un test doivent aussi représenter la population d'intérêt. Par exemple, pour des applications visant le grand public, il est nécessaire que le groupe d'observateurs soit composé de personnes n'ayant pas une expérience significative en traitement de l'image.

5.3.2 L'environnement de test

L'environnement dans lequel se déroule le test doit être conforme à des spécifications normalisées. Le lieu de test doit être une salle fermée où l'observateur n'est pas distrait par des facteurs extérieurs comme le son ou le bruit des machines. La distance à laquelle s'assoit la personne est fixée à l'avance. Par exemple, cette distance est égale à six fois la hauteur de l'écran pour des vidéos en SD. Des contraintes sont aussi imposées sur la luminosité de la salle et la luminance et le contraste des écrans d'affichage. Les séquences sont généralement présentées dans un ordre aléatoire.

Quand toutes les préparations techniques sont terminées, les consignes sont données à l'observateur et une explication du processus de vote est fournie. Un test préliminaire avec un échantillon réduit de séquences est aussi effectué. Le but de cette étape est de vérifier que l'observateur a bien compris la tâche à réaliser. De plus, elle permet de le familiariser avec les conditions du test et le type de dégradations de contenus qu'il peut rencontrer par la suite.

5.3.3 Traitement des résultats

Les notes recueillies à l'issue du test sont traitées d'abord pour valider ou non la participation de l'observateur. En effet, si les notes d'une personne semblent inconsistantes avec les notes attribuées par la moyenne des autres observateurs alors l'observateur en question est rejeté et ses notes ne sont pas incluses dans le calcul des MOS. Les détails de l'analyse des notes subjectives peuvent être trouvés dans [itu02]. Le but d'un test subjectif étant d'obtenir une note par séquence (MOS), la moyenne des notes attribuées par tous les observateurs retenus à une séquence donnée est calculée.

5.3.4 Classes de méthodologies

Parmi les méthodologies d'évaluation de la qualité subjective de vidéos, nous distinguons trois classes : la classe des méthodes à simple stimulus, celle des méthodes à double stimuli et celle des méthodes à stimuli multiples.

Les méthodes appartenant à la première classe consistent à présenter à l'observateur uniquement la séquence à évaluer.

Dans le cas des méthodes à double stimuli, la séquence de référence (originale) est explicitement présentée à l'observateur suivie ou précédée de la séquence dégradée. Cette méthode permet à l'observateur de baser son jugement de qualité sur une référence visualisée.

La dernière classe de méthodologies donne la liberté aux observateurs de visualiser les séquences autant de fois qu'ils le souhaitent. Le principal inconvénient de cette méthode est la longue durée qu'un test peut nécessiter ce qui réduit le nombre de contenus différents pouvant être évalués. Une explication détaillée de méthodologies ainsi qu'une comparaison entre elles sont fournies dans [Péc08].

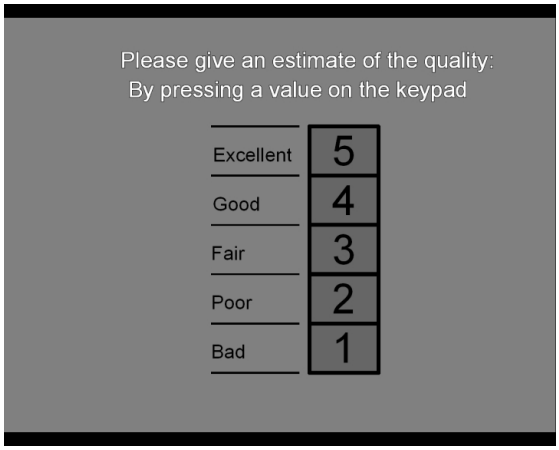


FIG. 5.3 – L'interface de vote illustrant l'échelle à cinq catégories utilisée dans le premier test subjectif.

5.3.4.1 La méthodologie choisie pour nos travaux

Nous choisissons pour notre travail la méthode d'évaluation par catégories absolues ACR (*Absolute Category Rating*) qui appartient à la première classe de méthodologies. Cette méthode, standardisée dans [itu99], consiste à évaluer les séquences l'une à la suite de l'autre à l'aide d'une échelle catégorielle à cinq niveaux. La figure 5.3 représente l'échelle de cinq catégories utilisée dans nos tests.

Pour chaque contenu, la séquence originale est incluse une fois dans le test sans que l'observateur ne le sache. Les raisons du choix de la méthode ACR sont les suivantes :

1. elle a été utilisée dans le *Multimedia Test Plan*[1] de VQEG (*Video Quality Experts Group*), le groupe international en charge de la formalisation des aspects de la qualité. Ses performances ont été validées par plus de dix laboratoires internationaux ;
2. elle permet de tester un grand nombre de séquences en un temps raisonnable du fait que la durée de vote est fixe et que la séquence de référence n'est pas affichée pour chaque séquence à évaluer ;
3. elle s'apparente à un scénario de services de vidéos sur IP où l'utilisateur évalue la qualité des vidéos dans l'absolu sans se baser sur une version "idéale".

[1] Disponible sur `ftp://vqeg.its.bldrdoc.gov/Documents/Projects/multimedia/MM_Final_Report/`.

TAB. 5.1 – Les caractéristiques des séquences vidéos utilisées dans le premier test subjectif.

Séquence	Description	Caractéristiques
Athletism	Des athlètes font une course	Beaucoup de mouvement, régions uniformes
Bluepeter	Discours prononcé lors d'une cérémonie officielle	Sombre, changement de scène
Football	Match de football, *zoom in*	Beaucoup de mouvement, régions texturées
Holiday_1	Bateau à voile en mer	Mouvement lent, régions texturées
Holiday_2	Paysages naturels avec ciel bleu	Images fixes, changement de scène (effet de dissolution)
Labourparty	Discussion dans un bureau	Tête-et-épaules, régions texturées
News	Présentation du journal d'informations	Tête-et-épaules, régions uniformes
Pool	Joueur de billard en train de tirer	Changement de scène, régions texturées

5.4 Première caractérisation des effets perceptuels des pertes de paquets

Nous décrivons dans cette section le premier des deux tests subjectifs mis en place[2]. Les séquences vidéos sont affichées sur un écran CRT à la fréquence de 25 images par seconde. Seize observateurs non experts ont participé au test dont la durée moyenne était de 35 minutes.

Les observateurs, assis à une distance égale à six fois la hauteur de l'écran, devaient noter la qualité des vidéos à l'aide d'un clavier numérique Les tests ont tous commencé par une phase d'entraînement comprenant quatre séquences vidéos. Les qualités de ces séquences sont choisies de façon à couvrir toute l'échelle de qualité.

5.4.1 Les séquences de test

Le premier test subjectif comprend huit séquences vidéos à la résolution SD (720×576) de dix secondes chacune. Ces séquences, propriété de BT, ont des contenus à quantités de mouvement et niveaux de texture variés. Le tableau 5.1 décrit le contenu et les caractéristiques de chacune des séquences.

5.4.2 Création des séquences dégradées

Pour une séquence vidéo source SRC (*Source Reference Circuit*), une séquence de test PVS (*Processed Video Sequence*) est créée à partir d'une certaine condition de test HRC (*Hypothetical*

[2]Les tests décrits dans ce chapitre ont été effectués au sein du centre de recherche de BT plc. à Ipswich au Royaume-Uni entre novembre 2007 et mars 2008.

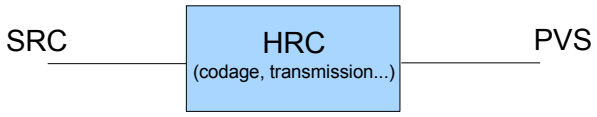


FIG. 5.4 – Désignation des séquences et des conditions de test : création d'une séquence PVS (*Processed Video Sequence*) à partir d'une source SRC (*Source Reference Circuit*) et d'une condition de test HRC (*Hypothetical Reference Circuit*).

Reference Circuit) comme illustré figure 5.4. Les HRC représentent tout traitement ou processus qu'une vidéo peut subir (codage, redimensionnement, transmission, post-traitement, *etc.*). Un HRC peut combiner plusieurs conditions de test simultanément.

Dans le contexte de ce travail, les HRC sont formés à partir d'un débit de codage donné et dans la majorité des cas d'un motif de pertes introduit.

5.4.2.1 Codage et décodage

Le codage et le décodage se font à l'aide du codec de BT. La structure du GOP est de la forme IBBPBBP... et sa longueur est de 24 images. Le profil de codage H.264/AVC utilisé est le *High Profile*. Le codage se fait à débit fixe. Deux débits de codage sont choisis de façon à représenter en même temps deux niveaux de qualité distincts et des débits réalistes pour des services de TV sur IP (1,5 et 4 Mbit/s). En considérant la contrainte imposée sur la taille de la slice, les images I des séquences codées à 1,5 Mbit/s sont composées d'un nombre de slices variant entre 20 et 50 selon le contenu de la séquence. À un débit de 4 Mbit/s, le nombre de slices par image I est compris entre 55 et 100 approximativement.

Aucune fonctionnalité de codage robuste offerte par la norme H.264/AVC (FMO, DP, *etc.*) n'est utilisée. Ce choix est justifié par la volonté d'évaluer la qualité de séquences vidéos ayant subi un codage classique. Un algorithme de compensation de pertes non standard est cependant implanté dans le décodeur. Cet algorithme agit dès que la perte d'une partie du flux est détectée. La stratégie de compensation est basée sur une compensation temporelle sauf dans le cas d'une image IDR (*Instantaneous Decoding Refresh*) pour laquelle une compensation spatiale est effectuée.

La compensation temporelle agit comme suit : pour chaque macrobloc perdu, un vecteur de mouvement pour chaque bloc 8×8 est calculé comme étant le vecteur moyen des blocs voisins (trois au maximum). La référence utilisée pour l'application du vecteur de mouvement est la dernière image de référence décodée. Si une image entière est perdue, elle est remplacée par l'image de référence la plus récente. Comme les images IDR remettent à zéro le contexte

TAB. 5.2 – Les HRC du premier test subjectif.

HRC	Débit (Mbit/s)	Fréquence des pertes (par séquence)	Taille de la rafale (en paquets)| Pourcentage de paquets perdus
0	1,5 ou 4	0	0
1	1,5	1	4 | 0,2%
2			10 | 0,5%
3			20 | 1,1%
4			40 | 2,2%
5		4	4 | 0,9%
6			10 | 2,2%
7			20 | 4,5%
8		8	20 | 8,9%
9	4	1	4 | 0,1%
10			10 | 0,2%
11			20 | 0,4%
12		4	4 | 0,4%
13			10 | 0,9%
14			20 | 1,8%

de codage et plus précisément causent le vidage des mémoires tampons du décodeur de leurs images de référence, une compensation spatiale s'impose. Cette compensation est basée sur une moyenne pondérée des échantillons voisins semblable à celle décrite dans la sous-section 2.2.2.2 du chapitre 2.

5.4.2.2 Motifs de pertes

Nous choisissons d'utiliser nos propres motifs de pertes où la longueur de la rafale et le nombre de pertes sont variés. Ces motifs sont réalistes mais aussi contrôlés dans un souci d'analyse de leurs conséquences. Les taux de pertes correspondants sont diversifiés et se situent entre $0,1\%$ et $8,9\%$. Un résumé de tous les motifs de pertes utilisés dans le premier test subjectif est donné dans le tableau 5.2. Les pertes sont généralement introduites dans les images I et un test d'inspection visuelle préliminaire avec experts vérifie que les séquences obtenues couvrent l'intégralité de l'échelle de qualité.

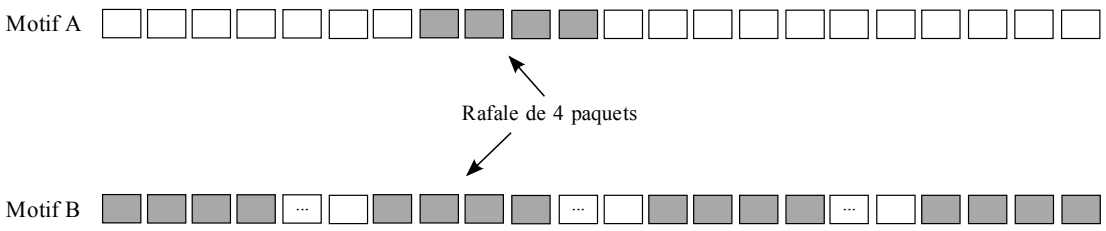


FIG. 5.5 – Exemples de motifs de pertes représentant la perte de quatre et seize paquets selon une distribution différente.

Dans le tableau 5.2, la fréquence des pertes indique leurs occurrences toutes tailles de rafale confondues. La taille d'une rafale (dernière colonne du tableau) représente le nombre de paquets consécutifs perdus. Les notions de rafales et de pertes sont illustrées sur la figure 5.5. Pour le motif A (HRC 1 et 9), nous notons la perte de quatre paquets consécutifs alors que le motif B (HRC 5 et 12) est constitué de quatre rafales de quatre paquets chacune. Le pourcentage de paquets perdus est équivalent au rapport (multiplié par 100) du nombre de paquets perdus de la séquence sur le nombre total de paquets dans la séquence. Par exemple, pour la séquence *Football* de dix secondes codée à $1,5$ Mbit/s, 4 pertes de paquets de longueur 20 paquets consécutifs chacune engendrent une perte de $\frac{80}{1789} * 100 = 4,5\%$ du nombre total de paquets dans la séquence. Les séquences ayant un nombre total de NALU très proche, les pourcentages de pertes sont les mêmes pour chaque motif de pertes. Nous appelons les motifs de pertes qui se produisent une seule fois "motifs singuliers" même s'ils concernent plusieurs paquets consécutifs et les motifs de pertes répétés "motifs réguliers".

5.4.2.3 Choix de la position temporelle des pertes

La position temporelle des dégradations de qualité dans la vidéo influence le jugement des observateurs. En effet, les dégradations qui apparaissent à la fin de la séquence sont susceptibles de marquer l'observateur plus que les dégradations ayant lieu à son début. Cet effet de récence (*recency*) a été mis en évidence par Hands et Avons [Han01] à travers une série de tests subjectifs de qualité sur des vidéos de trente secondes.

Pour éviter que la position temporelle de la perte n'introduise un biais sur le jugement de qualité des observateurs, nous conservons les mêmes positions temporelles pour toutes les séquences. De plus, nous choisissons les pertes de façon à ce qu'elles soient situées au milieu de la vidéo. Ainsi, les pertes uniques ont toujours eu lieu à la sixième seconde de la séquence vidéo, indépendamment de la taille de la rafale. Pour les motifs de pertes à plusieurs occurrences (4 et 8), la séquence temporelle pendant laquelle les pertes sont survenues se situe entre la quatrième et la septième seconde. L'intervalle de temps séparant les quatre et huit pertes a été

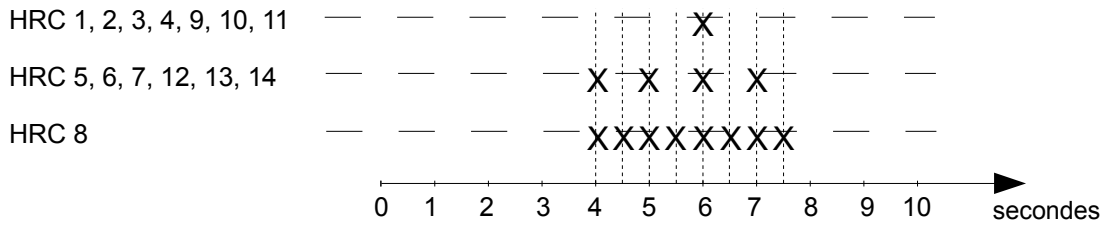


FIG. 5.6 – Exemples de motifs de pertes représentant un, quatre et huit événements de pertes de NALU en rafale (ensemble de 4, 10, 20 ou 40 paquets consécutifs)

fixé respectivement à une et une demi-seconde. La figure 5.6 illustre un motif singulier (HRC 1, 2, 3, 4, 9, 10 et 11) et deux motifs réguliers (l'un correspondant aux HRC 5, 6, 7, 12, 13 et 14 et l'autre au HRC 8). La perte est indiquée par une croix et elle peut concerner entre 4 et 40 paquets consécutifs.

Il est important de noter que dans toutes nos simulations, nous évitons de perdre les NALU PS (*Parameter Set*) qui ne contiennent pas des données vidéos mais qui guident le processus de décodage. En pratique, ces NALU peuvent être protégées avec des codes correcteurs à taux de redondance très élevés ou transmises sur des canaux fiables.

Nombre total de PVS

Les HRC 1 à 14 du tableau 5.2 ont été appliqués aux 8 séquences vidéos obtenant ainsi $14 \times 8 = 112$ séquences dégradées. Nous avons intégré les séquences codées aux deux débits et n'ayant pas subi de pertes de paquets (HRC 0) à l'ensemble des vidéos de test portant ainsi le total à 128 séquences.

5.4.3 Résultats expérimentaux

Les résultats obtenus caractérisent plusieurs aspects perceptuels des effets de pertes de paquets qui sont présentés dans cette section.

Les valeurs des MOS sont données avec leurs intervalles de confiance à 95% $[MOS_{jk} - e_{jk}; MOS_{jk} + e_{jk}]$ où MOS_{jk} correspond à la note moyenne de la séquence originale j ayant subi la dégradation k et e_{jk} à la grandeur calculée selon la formule suivante :

$$e_{jk} = 1,96 \cdot \frac{\sigma_{jk}}{\sqrt{N_{obs}}} \tag{5.1}$$

avec σ_{jk} l'écart-type de MOS_{jk} et N_{obs} le nombre d'observateurs retenus. L'intervalle de confiance à 95% permet d'apprécier la fiabilité des résultats des tests subjectifs.

5.4.3.1 Qualité visuelle sans pertes de paquets

L'analyse des notes de qualité des séquences codées et n'ayant pas subi de pertes de paquets (HRC 0) renseigne sur la difficulté de codage de chacune des séquences. Sur la figure 5.7, nous remarquons que la qualité de toutes les séquences codées à 4 Mbit/s est qualifiée entre "bonne" et "excellente" par les observateurs (MOS entre 4 et 5).

Par contre, les séquences codées à $1,5$ Mbit/s n'ont pas toutes une bonne qualité. La séquence *Athletism* contient une grande quantité de mouvement ce qui la rend difficile à coder (MOS de $3,4$). De même pour la séquence *Pool* qui contient des niveaux de détails très fins et qui obtient un MOS de $3,9$. À noter que les différences de qualité surprenantes entre les versions codées aux deux débits des séquences *Labourparty* et *News* ne sont pas significatives.

Nous pouvons donc dire à partir des notes de qualité données figure 5.7 que les séquences ont généralement une bonne qualité aux débits de codage utilisés. Ceci nous permettra de mieux évaluer la diminution de qualité engendrée par les pertes de paquets.

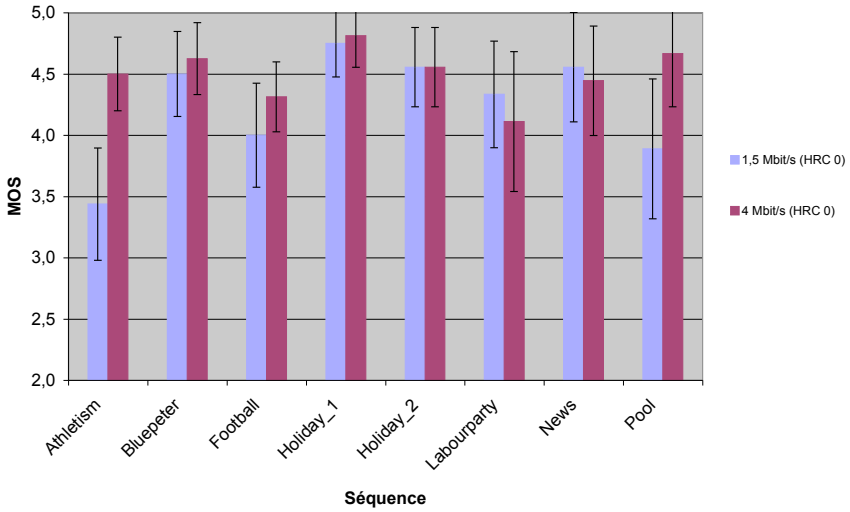


FIG. 5.7 – Qualité initiale des séquences codées aux débits $1,5$ et 4 Mbit/s n'ayant subies aucune perte de paquets (HRC 0).

5.4.3.2 Influence combinée du débit et des pertes sur la qualité

Le cas de la séquence *Athletism* est intéressant car les PVS présentent une différence d'une unité MOS entre les deux débits ce qui correspond à deux niveaux de qualité distincts. Ce cas de figure peut illustrer l'interaction entre le débit de codage (qui détermine la qualité initiale) et

les pertes de paquets (qui déterminent la qualité finale). Les courbes “avec pertes” de la figure 5.8 correspondent, pour cette séquence, aux HRC 1, 2 et 3 du tableau 5.2 pour le codage à 1,5 Mbit/s et aux HRC 9, 10 et 11 pour le codage à 4 Mbit/s. Les HRC 1 et 9 ; 2 et 10 ; 3 et 11 représentent respectivement la perte de 4, 10 et 20 paquets consécutifs. La diminution de qualité observée avec l'augmentation du nombre de paquets perdus a la même allure pour les deux séquences.

Pour quatre paquets consécutifs perdus, les qualités des deux versions de la séquence chutent d'une valeur MOS chacune. La qualité de la séquence codée à 4 Mbit/s devient égale à la qualité initiale de celle codée à débit plus faible. Ce cas illustre l'intérêt d'une approche de codage conjoint source-canal : une séquence codée à bas débit et bien protégée contre les pertes satisfait l'utilisateur de la même manière qu'une séquence codée à plus haut débit mais affectée par les pertes. À partir d'un certain seuil (10 paquets dans cet exemple), la qualité devient médiocre indépendamment du codage source.

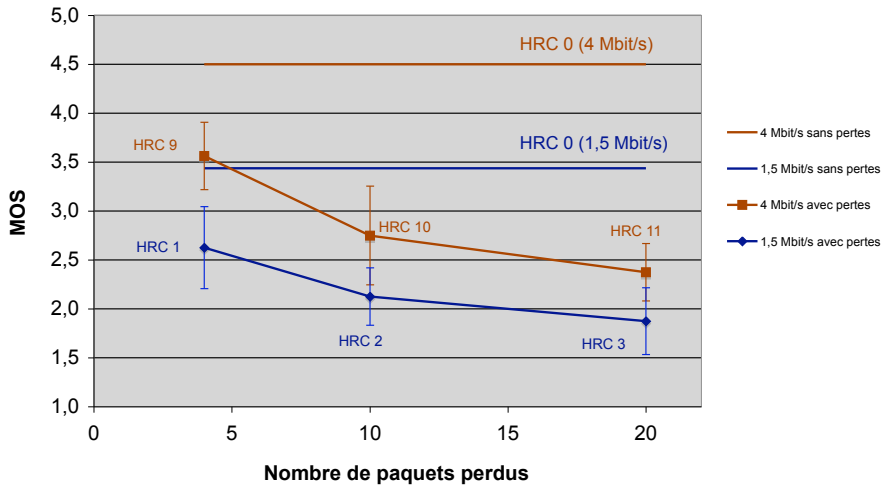


FIG. 5.8 – Comparaison de la séquence *Athletism* aux débits 1,5 et 4 Mbit/s pour des motifs de pertes singuliers (respectivement HRC 1 ; 2 ; 3 et 9 ; 10 ; 11).

5.4.3.3 Qualité moyenne par débit en fonction du nombre de paquets perdus

La qualité d'une séquence tend généralement à diminuer avec l'augmentation du nombre de paquets perdus. Nous présentons sur la figure 5.9 la qualité moyenne des huit séquences pour chaque débit et pour les motifs de pertes singuliers (HRC 1 ; 2 ; 3 et 9 ; 10 ; 11) et réguliers (HRC 5 ; 6 ; 7 et 12 ; 13 ; 14). L'allure générale des deux courbes montre que la qualité diminue en effet

à mesure que la quantité d'information perdue augmente. Cependant, deux remarques sont à noter sur cette figure.

Premièrement, les intervalles de confiance à 95% sont relativement larges et couvrent pour certains motifs de pertes plus d'une unité MOS. Ainsi, à un même débit, le même nombre de paquets perdus conduit à des niveaux de qualité distincts. Cette constatation indique une forte dépendance entre l'impact perceptuel d'une perte et les caractéristiques du contenu de la séquence.

Ensuite, nous observons une singularité au niveau de la courbe de qualité lorsque le nombre de paquets perdus passe de 16 à 20. La qualité ne diminue pas (elle augmente légèrement) entre les motifs réguliers HRC 5 et 12 et les motifs singuliers HRC 3 et 11. Ceci implique une certaine relation entre la distribution des pertes et leur impact perceptuel que nous étudions dans ce qui suit.

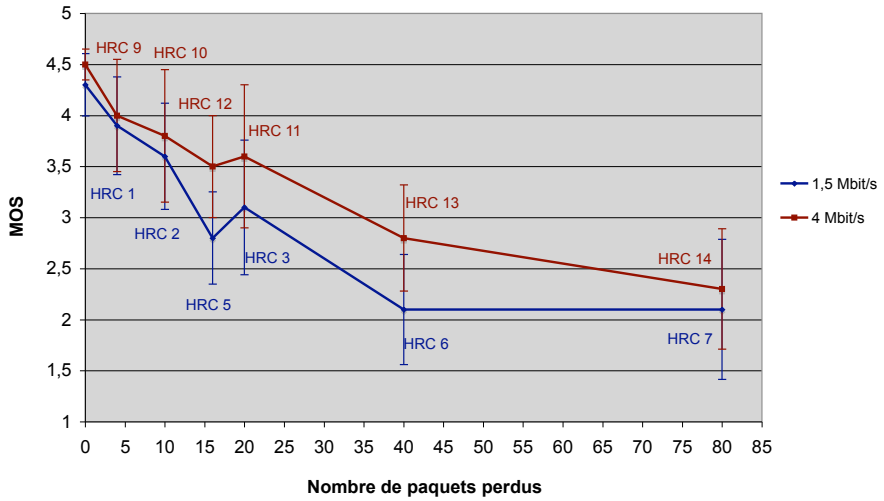


FIG. 5.9 – La qualité moyenne des huit séquences par débit et en fonction du nombre de paquets perdus.

5.4.3.4 Influence de la distribution des pertes sur la qualité visuelle

Nous étudions l'impact que peut avoir la distribution des pertes de paquets sur la note de qualité attribuée par un observateur à une séquence vidéo codée à $1,5$ Mbit/s. Nous avons ainsi comparé deux scénarios de pertes différents mais qui accusent un même taux de pertes global ($2,2\%$). Nous avons choisi les motifs suivants : perte unique de 40 paquets consecutifs (HRC 4)

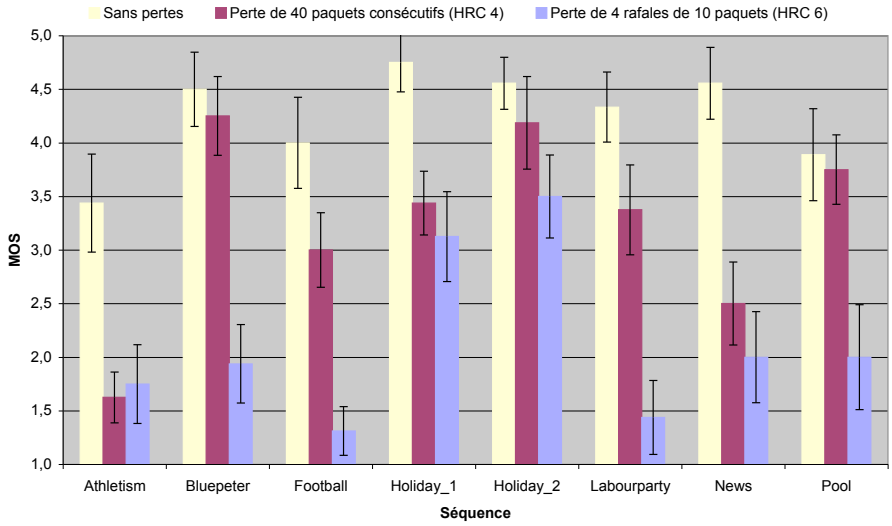


FIG. 5.10 – L'impact de la distribution des pertes sur la qualité perceptuelle.

et quatre pertes régulières de 10 paquets chacune (HRC 6). Les résultats sont présentés figure 5.10.

Nous pouvons déduire du diagramme en bâtons de la figure 5.10 que la qualité d'une vidéo est généralement meilleure dans le cas d'une seule perte importante que dans le cas de plusieurs pertes moins importantes. En particulier, la qualité des séquences *Football* et *Labourparty* diminue d'une unité MOS pour HRC 4 alors que cette diminution est de quasiment trois unités MOS pour HRC 6. Ceci est dû à la durée de propagation de l'erreur. En effet, pour le HRC 4, les 40 paquets perdus appartiennent à l'image I et aux images B et P suivantes. La propagation temporelle des dégradations qui a lieu reste cependant confinée à l'intérieur du GOP concerné.

Pour le HRC 6, les pertes ont lieu dans quatre images I consécutives. Comme le GOP a une durée proche d'une seconde (24 images à la fréquence 25 ips), l'effet de propagation de l'erreur peut durer jusqu'à quatre secondes (40% de la durée totale de la séquence). À noter que la qualité de la séquence *Athletism* qui contient une grande quantité de mouvement est mauvaise pour les deux motifs de pertes car l'algorithme de compensation de pertes n'a pas de bonnes performances à ce taux de pertes.

5.4.3.5 Impact perceptuel du changement de scène en présence de pertes de paquets

Nous présentons sur la figure 5.11 les valeurs des MOS des huit séquences codées à $1,5$ Mbit/s. Pour chaque séquence, nous présentons quatre versions correspondant au cas sans pertes

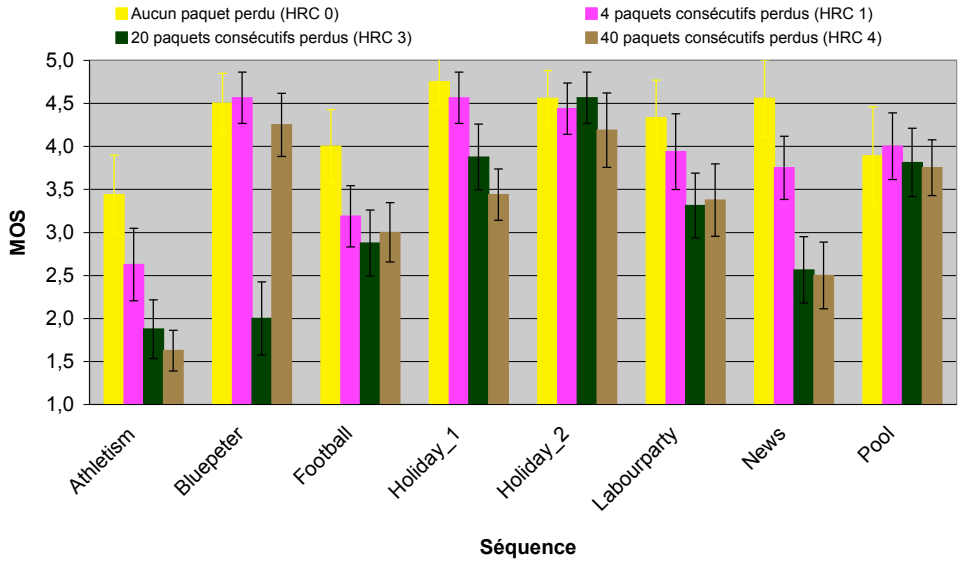


FIG. 5.11 – L'impact du changement de scène sur la qualité perceptuelle. Les pertes de 4, 20 et 40 paquets correspondent respectivement aux HRC 1, 3 et 4.

de paquets et aux trois HRC 1, 3 et 4 (impliquant respectivement 4, 20 et 40 paquets perdus).

Les séquences contenant un changement de scène sont *Bluepeter*, *Holiday_2* et *Pool*. Pour les séquences sans changement de scène, nous remarquons une diminution de la qualité à mesure que le nombre de paquets perdus augmente. Par contre, ces trois séquences n'ont pas ce comportement.

Généralement, la première image d'une nouvelle scène est codée en mode intra car c'est souvent la solution la moins coûteuse en termes de débit. En effet, un codage inter à partir d'images appartenant à l'ancienne scène n'est pas efficace. En observant les diagrammes en bâtons des séquences *Holiday_2* et *Pool*, nous pouvons constater que la variation de MOS entre les différents scénarios de pertes est faible sachant que le taux de pertes augmente de $0,22\%$ à $2,2\%$. Ceci est dû au fait que la propagation temporelle de l'erreur causée par le processus de prédiction inter est stoppée quand une image codée en intra est rencontrée. Dans les deux cas susmentionnés, les pertes ont lieu dans l'image située juste avant le changement de scène ce qui empêche l'erreur de se propager vers un nouveau GOP. Le codage en mode intra agit donc comme une forme de protection. La figure 5.12 illustre deux exemples de pertes où l'un concerne l'image juste avant le changement de scène (figure 5.12.a) et l'autre concerne la première image de la nouvelle scène (figure 5.12.b).

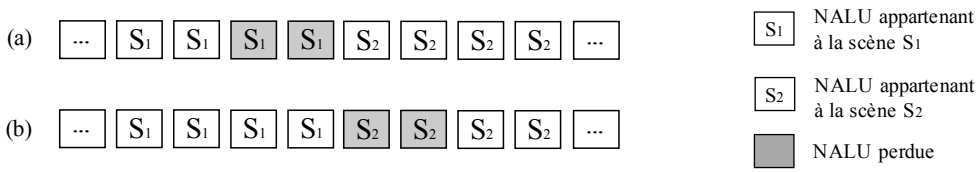


FIG. 5.12 – Un exemple de pertes ayant lieu (a) avant le changement de scène et (b) exactement au changement de scène.

Dans la séquence *Bluepeter*, nous remarquons le même comportement pour tous les scénarios de pertes sauf celui où les pertes ont une longueur de vingt paquets. En effet, dans ce dernier cas, la perte a lieu dans la première image d'une nouvelle scène qui est codée en mode intra. L'algorithme de compensation de pertes implanté dans le décodeur cherche alors à utiliser la dernière image de référence décodée pour compenser les paquets perdus. Or, la dernière image dans la mémoire tampon du décodeur appartient à la scène précédente ce qui conduit à des résultats désastreux comme ceux présentés sur la figure 5.13. Le mélange de contenus dû à la perte de cinq paquets dans la première image d'une nouvelle scène est un effet très gênant pour l'observateur. Pour les deux HRC 1 et 4, la qualité est quasiment inchangée du fait que les pertes ont lieu dans les images situées juste avant le changement de scène.

La figure 5.11 montre aussi que les versions de *Bluepeter*, *Holiday_2* et *Pool* avec pertes juste avant le changement de scène ont une diminution de moins de $0,5$ MOS comparées à la version sans pertes. Ceci permet de conclure sur une hiérarchisation du flux qui est intrinsèque au contenu. Dans une perspective de protection inégale, les paquets contenant les dernières images d'une scène peuvent être protégés beaucoup moins que la première image de la nouvelle scène.

5.4.3.6 Impact perceptuel de la position spatiale de la perte dans l'image

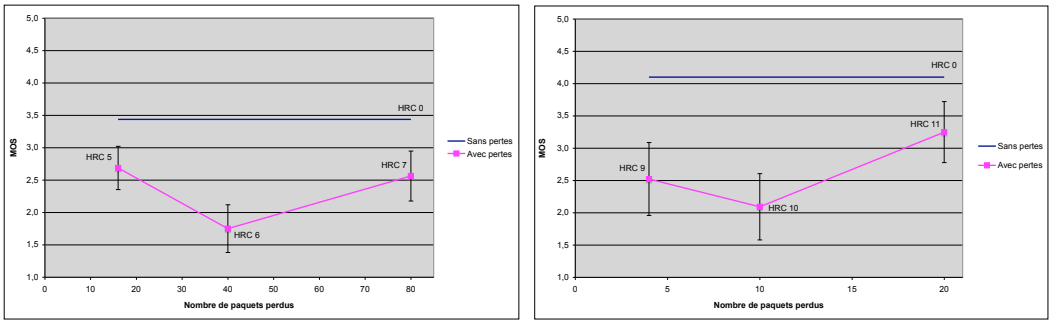
Généralement, pour un même motif de pertes, la qualité diminue quand le nombre de paquets perdus augmente. Mais les figures 5.14.a et 5.14.b montrent des allures contradictoires. Commençons par la figure 5.14.a qui montre les valeurs de MOS de la séquence *Athletism* codée à $1,5$ Mbit/s et affectée par un motif régulier de quatre rafales de pertes de 4, 10 et 20 paquets chacune (respectivement HRC 5, 6 et 7). Quand la taille de la rafale est augmentée de 10 à 20, la valeur du MOS augmente d'une unité à peu près. Ceci peut être en partie expliqué par le fait que les pertes dans le cas de 80 paquets perdus ne touchent pas les régions d'intérêt de l'image. En effet, les dégradations sont observables dans la région de l'image contenant les gradins qui ne représentent pas une zone d'intérêt pour l'observateur plutôt concentré sur l'action de la course. La variation de la zone d'impact de la perte est réalisée comme décrit à la section 5.2 (figure

(a) La dernière image de l'ancienne scène de *Fries*.

(b) La première image de la nouvelle scène de *Fries*.

(c) La première image de la nouvelle scène ayant perdu 3 paquets.

FIG. 5.13 – Un exemple de pertes dans la première image d'une nouvelle scène.



(a) La séquence *Athletism* à 1,5 Mbit/s pour des motifs de pertes réguliers.

(b) La séquence *Labourparty* à 4 Mbit/s pour des motifs de pertes singuliers.

FIG. 5.14 – L'impact de la position spatiale de la perte sur des contenus, des débits et des motifs de pertes différents.

5.2) de ce chapitre.

La figure 5.14.b représente la qualité de la séquence *Labourparty*, codée à 4 Mbit/s et affectée par un motif singulier de 4, 10 et 20 paquets (respectivement HRC 9, 10 et 11). Une augmentation de qualité est notée quand le nombre de paquets perdus passe de 10 à 20. Dans ce cas, les HRC 9 et 10 dégradent la région d'intérêt de l'image comprenant les personnes en discussion. Par contre, le HRC 11 impacte l'arrière-plan de l'image fortement texturé ce qui ne représente pas une source importante de gêne pour l'observateur.

Nous pouvons donc conclure que d'un point de vue perceptuel, la position spatiale de la perte influence plus la qualité que la quantité d'information perdue. De plus, il existe une hiérarchie des données au sein du flux vidéo.

5.5 Identification de la relation entre le taux de pertes et la qualité visuelle

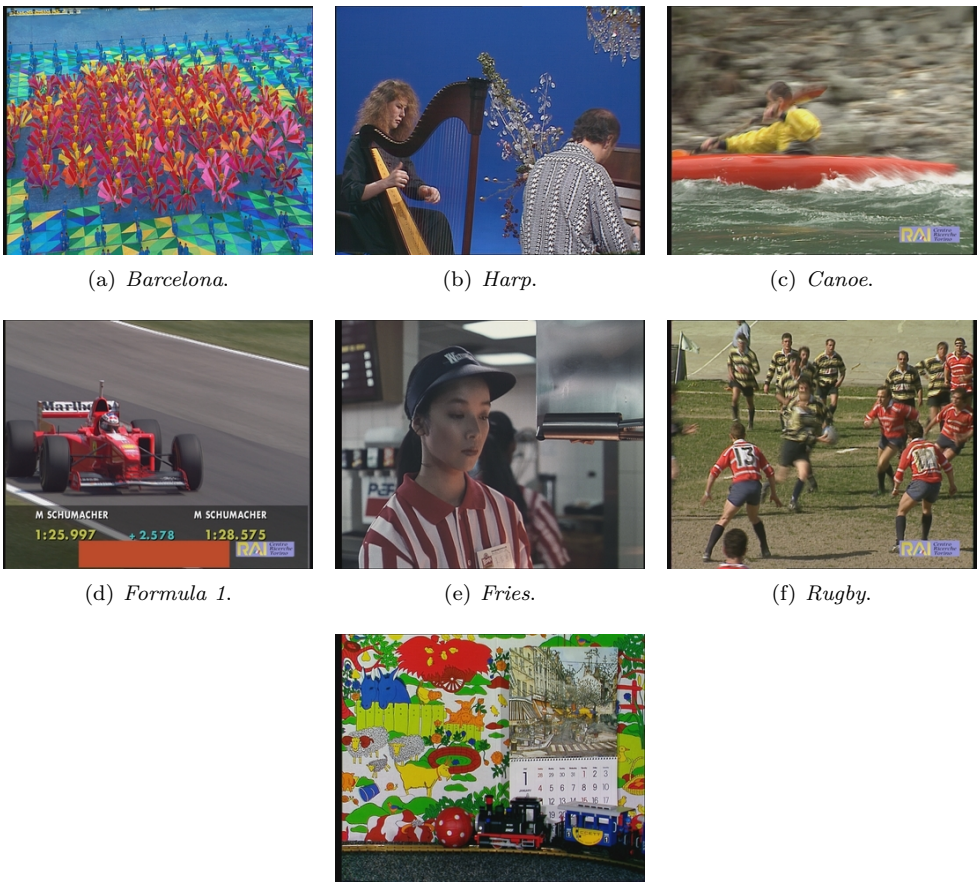
Dans le second test subjectif, nous utilisons le même dispositif expérimental mais nous changeons les motifs de pertes en fonction des effets perceptuels à étudier. Nous voulons en effet caractériser l'effet perceptuel de la perte totale ou partielle d'une seule image I. D'une part, ces motifs nous permettent de quantifier l'importance de l'impact de la propagation temporelle des dégradations constaté dans le premier test. D'autre part, nous essayons d'établir une relation entre la quantité d'information perdue dans l'image I et la diminution de qualité résultante.

5.5.1 Environnement de test

Nous utilisons le même environnement de test (écran, salle, échelle de qualité, *etc.*) que pour le premier test. Quinze observateurs non experts ont participé au test. Le codec de BT est utilisé et le débit source est fixé à 2 Mbit/s.

5.5.2 Séquences de test

Sept séquences vidéos de huit secondes chacune tirées de la phase 1 de VQEG FR-TV [vqe00] sont utilisées dans le test. Ces séquences, libres de droit, présentent des contenus diversifiés ayant des niveaux de texture et de mouvement assez variés. La première image de chacune des séquences est donnée figure 5.15.



(a) *Barcelona.* (b) *Harp.* (c) *Canoe.*

(d) *Formula 1.* (e) *Fries.* (f) *Rugby.*

(g) *Calendar.*

FIG. 5.15 – Les séquences utilisées dans le second test subjectif.

Pour quantifier les différences entre les contenus de ces séquences, nous présentons deux mesures standardisées par l'ITU [itu99] qui reflètent l'information spatiale et l'information temporelle d'une séquence vidéo. Cette caractérisation n'était pas possible auparavant car les séquences du premier test ne sont pas libres de droit.

5.5.2.1 Caractérisation de l'information spatiale

L'information spatiale perceptuelle SI (*Spatial Information*) est basée sur le filtrage de la composante de luminance des images de la séquence par deux filtres de Sobel, l'un horizontal et l'autre vertical. Pour chaque image n filtrée $I_f(n)$, l'écart-type des valeurs des pixels σ_{espace} est calculé. L'information spatiale est alors donnée par l'équation suivante :

$$SI = max_{temps}\{\sigma_{espace}[I_f(n)]\} \tag{5.2}$$

où max_{temps} est la valeur maximale de σ_{espace} sur toutes les images de la séquence. Une valeur élevée de SI indique généralement la présence de fines textures dans l'image.

5.5.2.2 Caractérisation de l'information temporelle

L'information temporelle perceptuelle TI (*Temporal Information*) est basée sur le calcul du mouvement effectué entre deux images successives. Ce mouvement est représenté par la différence $M_n(i,j)$ des valeurs de luminance des pixels situés à la même position spatiale dans les deux images consécutives n et $(n-1)$. L'écart-type des valeurs de tous les pixels (i,j) de $M_n(i,j)$ est ensuite calculé pour chaque deux images consécutives. L'information temporelle est alors donnée par l'équation suivante :

$$TI = max_{temps}\{\sigma_{espace}[M_n(i,j)]\} \tag{5.3}$$

où max_{temps} est la valeur maximale de σ_{espace} sur toutes les différences $M_n(i,j)$ de la séquence. Une grande quantité de mouvement entre images successives implique une valeur élevée de TI.

5.5.2.3 Classification des contenus des séquences

Nous avons calculé les valeurs de SI et de TI pour chacune des sept séquences utilisées dans ce test. Leurs valeurs sont données dans le tableau 5.3. Une première lecture de ce tableau par colonne nous permet de constater que la séquence *Calendar*, qui contient une grande quantité de texture, a la valeur de SI la plus grande $(147,7)$.

D'autre part, la séquence *Fries* a une valeur de TI égale à $64,8$ (la plus élevée), justifiée en partie par la présence d'un changement de scène. Un nouveau calcul de TI sans y inclure

TAB. 5.3 – Les valeurs de SI et de TI pour les sept séquences vidéos utilisées dans le second test subjectif.

Séquence	SI	TI
Barcelona	113,8	23,4
Harp	141,3	32,7
Canoe	81,8	33,5
Formula 1	105,7	46,7
Fries	101,1	64,8
Rugby	102,4	43,1
Calendar	147,7	38

le changement de scène donne un résultat très proche $(60, 9)$ ce qui implique que la séquence a effectivement une forte activité temporelle. Nous décidons donc de garder la valeur de TI initialement calculée.

Afin d'obtenir une classification des séquences selon l'information spatio-temporelle qu'elles contiennent, nous illustrons figure 5.16 les couples (SI, TI) des séquences. Nous pouvons ainsi distinguer cinq classes C de contenus numérotées de 1 à 5.

Les classes C_1, C_3 et C_5 représentent des contenus ayant une complexité temporelle moyenne et des complexités spatiales respectivement faible, moyenne et élevée. Nous pouvons donc grouper ces trois classes dans une métaclasse M_T de complexité temporelle T.

Les classes C_2, C_3 et C_4 représentent des contenus à complexité spatiale moyenne mais ayant des complexités temporelles respectivement faible, moyenne et élevée. De même, ces classes peuvent être groupées dans une métaclasse M_S de complexité spatiale S.

Les séquences *Formula 1* et *Rugby* d'une part et *Calendar* et *Harp* d'autre part ont été associées à une seule classe car elles ont des valeurs de SI et de TI très proches.

5.5.3 Motifs de pertes

Les taux de pertes utilisés sont 20%, 50% et 100% de l'image I du cinquième GOP de la séquence. Une perte de 100% dans l'image I signifie sa perte totale. Les pourcentages sont calculés comme étant le rapport du nombre de NALU perdues sur le nombre total de NALU dans l'image I. Les NALU perdues sont toujours consécutives. Le cinquième GOP est choisi car il se situe temporellement au milieu de la séquence.

Nous définissons le Taux de Pertes Global (TPG) comme étant le pourcentage que représente le nombre de NALU perdues dans l'image I sur le nombre total de NALU dans la séquence :

$$TPG = \frac{nombre\ de\ NALU\ perdues}{nombre\ total\ de\ NALU\ dans\ le\ flux} * 100 \tag{5.4}$$

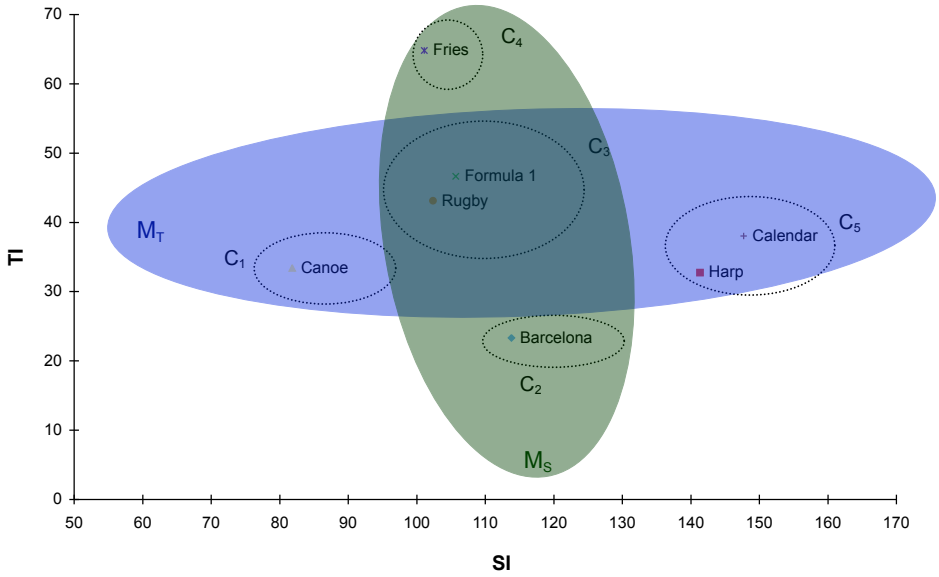


FIG. 5.16 – Les couples (SI, TI), les classes C et les métaclasses M correspondant aux sept séquences utilisées dans le second test subjectif.

Pour un même nombre de NALU perdues et à un même débit, TPG dépend du contenu de la vidéo. Les trois taux de pertes utilisés dans ce test correspondent à des valeurs de TPG comprises entre $0,2\%$ et $3,2\%$.

Nombre total de PVS Trois versions de chaque séquence correspondant aux HRC 1, 2 et 3 du tableau 5.4 sont incluses dans le test. Avec les sept séquences codées et n'ayant pas subi de pertes (HRC 0), le nombre total de séquences présentées aux observateurs devient 28.

TAB. 5.4 – Les HRC du second test subjectif.

HRC	Débit (Mbit/s)	Pourcentage de l'image I perdu
0	2	0%
1		20%
2		50%
3		100%

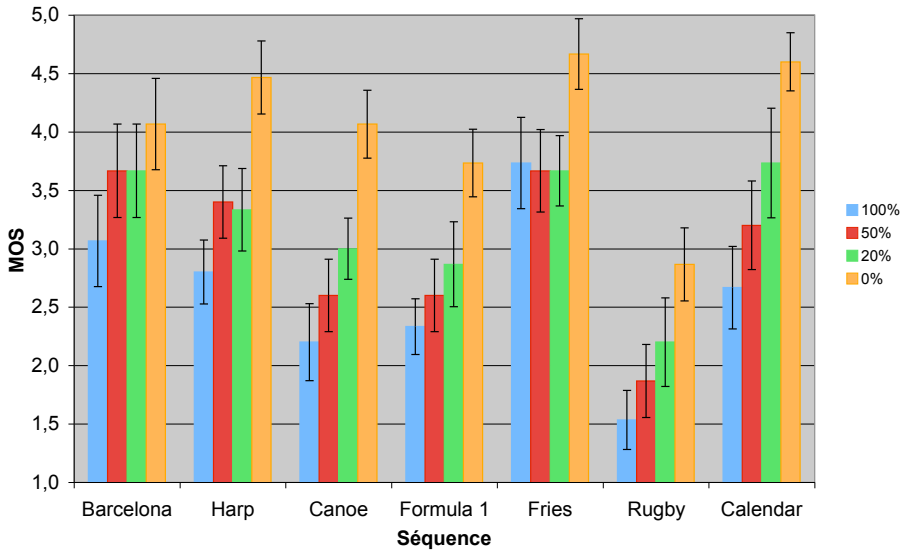


FIG. 5.17 – La qualité des sept séquences pour différents pourcentages d'image I perdus.

5.5.4 Résultats expérimentaux

Les résultats de ce second test subjectif renseignent sur l'impact perceptuel de la perte d'une partie ou de la totalité de l'image I.

5.5.4.1 Perte d'une partie de l'image I

La figure 5.17 montre que la qualité des séquences affectées par un taux de pertes de 20% reste acceptable ($MOS \simeq 3$) sauf pour la séquence *Rugby*. En effet, cette séquence n'est pas de bonne qualité quand elle est codée à 2 Mbit/s du fait de la grande quantité de mouvement et de la texture qu'elle contient.

Pour le taux de pertes de 50%, la qualité diminue de moins de $0,5$ MOS par rapport à un taux de 20% ce qui indique un impact perceptuel similaire pour ces deux motifs de pertes. De plus, la qualité descend sous la barre de la qualité acceptable ($MOS = 3$) pour les séquences *Canoe* et *Formula 1* uniquement (en excluant le cas particulier de la séquence *Rugby*).

Nous pouvons donc dire qu'une séquence, ayant une qualité au codage initialement bonne, conserve une qualité acceptable si moins de la moitié des paquets d'une image I sont perdus. Cette conclusion tient généralement pour tous les contenus.

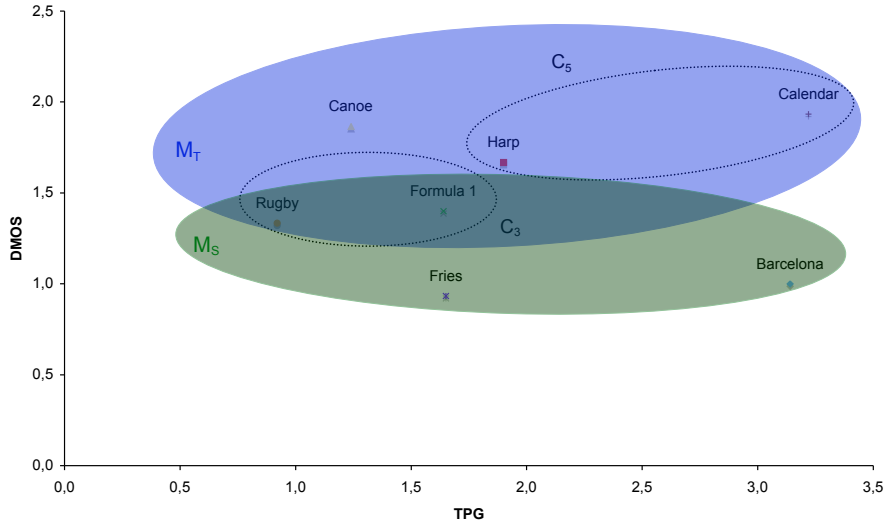


FIG. 5.18 – La diminution de qualité engendrée par la perte d'une image I entière (HRC 3) en fonction du TPG correspondant pour les sept séquences.

5.5.4.2 Perte d'une image I entière

La figure 5.18 présente, pour les sept séquences, les valeurs de la diminution de qualité engendrée par la perte de la totalité d'une image I. La valeur de $DMOS$ est calculée selon l'équation suivante :

$$DMOS = MOS_{TPG_0} - MOS_{TPG_{100}} \tag{5.5}$$

où MOS_{TPG_0} est la valeur du MOS de la séquence codée à 2 Mbit/s et n'ayant subi aucune perte de paquets ; $MOS_{TPG_{100}}$ est la valeur du MOS de la séquence codée ayant subi la perte de sa cinquième image I.

Relation entre le changement de scène et DMOS

Ce second test confirme l'importance de la position temporelle de la perte par rapport au changement de scène dans une séquence. Dans la séquence *Fries*, la perte de l'image I a lieu juste avant le changement de scène. La même qualité pour les trois pourcentages de pertes introduits est alors observée figure 5.17. De plus, la valeur de DMOS de la séquence *Fries* (0,9 MOS) est la plus faible parmi les valeurs de DMOS de toutes les séquences de la figure 5.18. L'effet d'atténuation des dégradations est encore une fois mis en évidence.

Relation entre TPG et DMOS

Nous constatons sur la figure 5.18 que la diminution de qualité ne corrèle pas avec le taux de paquets perdus dans le flux. Ainsi, les séquences *Barcelona* et *Calendar*, qui subissent des taux de pertes très proches (respectivement $3,14\%$ et $3,22\%$), ont des DMOS respectifs de 1 et $1,9$. De même, la séquence *Fries* a un DMOS de $0,9$ pour un TPG de $1,65\%$ tandis que la séquence *Barcelona* a un DMOS quasiment égal pour une valeur double de TPG.

Relation entre le contenu et DMOS

Trois conclusions concernant la relation entre la complexité spatio-temporelle des contenus des séquences et les valeurs de DMOS peuvent être dégagées à partir de la figure 5.18.

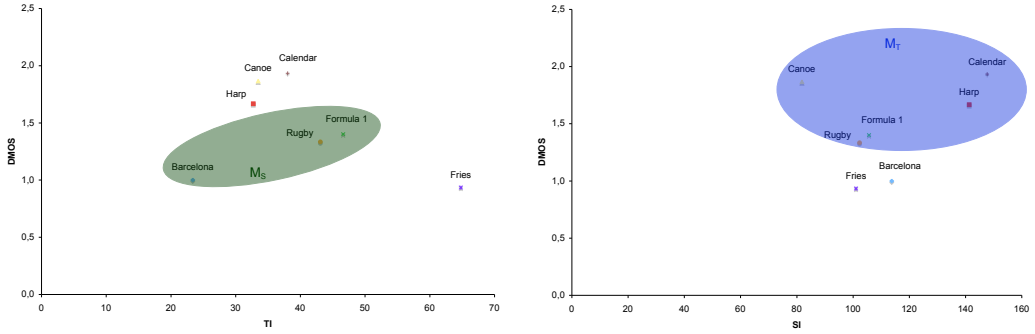
Tout d'abord, nous nous intéressons aux classes de contenus C_3 et C_5. Deux séquences appartenant à la même classe ont des DMOS très proches. Ainsi, *Rugby* et *Formula 1* ont des DMOS respectifs de $1,3$ et $1,4$ alors que leurs TPG sont sensiblement différents. La même allure est constatée pour *Harp* et *Calendar* qui ont des DMOS respectifs de $1,7$ et $1,9$.

Ensuite, nous discutons de la variation de DMOS au sein d'une même métaclasse. Dans la métaclasse M_S, en excluant la séquence *Fries* dont la valeur de DMOS a été justifiée précédemment, les séquences *Rugby*, *Formula 1* et *Barcelona* ont des DMOS respectifs de $1,3$; $1,4$ et 1. Pour mieux comprendre la variation de DMOS au sein d'une métaclasse, nous illustrons figure 5.19.a les valeurs de DMOS en fonction de TI. Sur cette figure, nous constatons pour les séquences de M_S que DMOS augmente à mesure que TI augmente. Autrement dit, pour des valeurs de SI similaires et pour la perte d'une image I entière, la diminution de la qualité d'une séquence dépend de sa complexité temporelle.

Dans la métaclasse M_T, les valeurs de DMOS des séquences *Canoe*, *Harp* et *Calendar* sont très proches et en même temps supérieures à celles de *Rugby* et *Formula 1*. Nous représentons DMOS en fonction de SI sur la figure 5.19.b pour observer la variation de DMOS au sein de la métaclasse M_T. La séquence *Canoe* a une valeur de SI inférieure à celles de *Rugby* et *Formula 1* alors que *Harp* et *Calendar* ont une complexité spatiale supérieure. Ce constat ne nous permet pas de tirer une conclusion quant à l'influence de l'augmentation de SI sur DMOS à complexité temporelle presque égale.

Discussion sur l'utilisation des grandeurs SI et TI

Il est important de noter que les valeurs de SI et TI considérées individuellement ne renseignent pas sur la vulnérabilité d'une séquence face à la perte d'une image I. En effet, les figures 5.19.a et 5.19.b ne montrent pas une corrélation entre DMOS et ces mesures d'information spatiale et temporelle. Ceci peut être dû à deux raisons.



(a) (b)

FIG. 5.19 – La variation de DMOS en fonction de (a) TI et (b) SI.

Premièrement, la complexité spatio-temporelle d'une séquence pourrait simplement ne pas avoir d'influence sur l'impact perceptuel d'une perte. Cette conclusion n'est cependant pas crédible du fait que la dépendance entre le contenu et la qualité perçue a été montrée dans des travaux de la littérature [PV04c, Feg05] et dans le cadre de ce travail (*cf.* sous-section 5.4.3.3).

La seconde raison est liée à la méthode de calcul de SI et TI. Ces métriques simples ont été introduites essentiellement pour caractériser les contenus des séquences utilisées pendant les tests subjectifs. SI et TI caractérisent une séquence par la valeur maximale d'un écart-type dans le temps ce qui peut parfois fausser les résultats. Par exemple, un contenu qui présente une complexité temporelle élevée pendant deux secondes sur dix est classé dans la catégorie des TI élevés alors qu'il ne doit pas l'être. Une moyenne temporelle peut être une approche plus raisonnable pour caractériser spatio-temporellement un contenu.

Fenimore *et al.* [Fen98] ont proposé une mesure de la difficulté de codage d'un contenu ou *scene criticality*. Cette mesure est fondée sur la moyenne temporelle du produit de SI et TI par image. Pour chaque image n, SI_n et TI_n sont calculées selon les formules suivantes :

$$SI_n = rms_{espace}[I_f(n)] \tag{5.6}$$

$$TI_n = rms_{espace}[M_n(i,j)] \tag{5.7}$$

où $I_f(n)$ et $M_n(i,j)$ sont respectivement l'image filtrée et la différence de luminance entre deux images consécutives (*cf.* équations (5.2) et (5.3)) ; rms_{expace} est la moyenne quadratique (*Root Mean Square*) des pixels de l'image.

La difficulté de codage $o(SI,TI)$ d'une séquence est alors donnée par :

$$o(SI,TI) = log_{10}(moy_{temps}[SI_n * TI_n]) \tag{5.8}$$

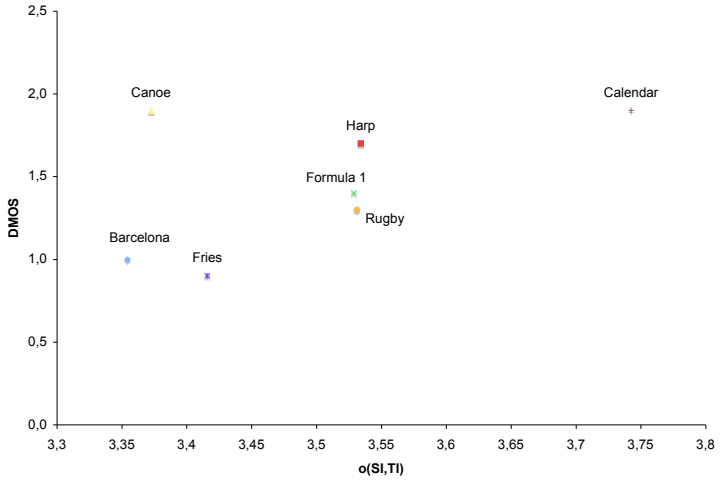


FIG. 5.20 – La variation de DMOS en fonction de $o(SI, TI)$.

Les résultats présentés dans [Fen98] montrent une corrélation de $0,82$ entre cette mesure et la difficulté de codage. La difficulté de codage est mesurée au travers de tests subjectifs de qualité pendant lesquels les observateurs jugent la qualité de séquences codées à différents débits.

Pour évaluer l'efficacité de l'utilisation de cette mesure comme indicateur de la vulnérabilité d'une séquence aux pertes de paquets, nous calculons la valeur de $o(SI, TI)$ pour chacune des sept séquences du second test. La variation de DMOS en fonction de $o(SI, TI)$ est illustrée figure 5.20.

L'allure générale des points de cette figure indique une augmentation de la valeur de DMOS à mesure que $o(SI, TI)$ augmente. Le coefficient de corrélation obtenu est de $0,5$. En excluant la séquence *Canoe* dont la valeur de DMOS est très élevée, le coefficient de corrélation devient égal à $0,9$. La singularité de *Canoe* est probablement due au fait que l'objet en mouvement occupe une grande partie de l'image (gros plan) ce qui implique une étendue spatiale importante des dégradations.

D'autre part, la forte corrélation obtenue, même si elle ne concerne qu'un nombre limité de points, nous permet de dégager une tendance qualitative entre la perte d'une image I entière et la complexité spatio-temporelle. Plus précisément, en présence de pertes de paquets, la qualité d'une séquence est d'autant plus faible que son contenu a une complexité spatio-temporelle élevée. Ce résultat étant prévisible, cette étude a montré que la mesure de difficulté de codage $o(SI, TI)$ peut être un bon indicateur de l'impact perceptuel des pertes de paquets.

Conclusion

Nous avons présenté dans ce chapitre deux tests subjectifs visant une compréhension plus profonde de l'impact perceptuel des pertes de paquets sur la qualité visuelle des vidéos. Plus particulièrement, nous avons identifié les principales vulnérabilités du flux codé H.264/AVC. Pour ce faire, nous avons décrit le processus de simulation de pertes au niveau du flux binaire de la vidéo. L'intérêt d'une simulation à ce niveau est de contrôler les positions temporelles et spatiales de la perte.

Les résultats obtenus peuvent être classés dans deux catégories : les résultats liés au contexte de transmission et ceux liés aux caractéristiques de la séquence vidéo. Dans la première catégorie, nous avons étudié l'impact du motif de pertes sur la qualité. Les résultats des tests subjectifs montrent que pour un taux de pertes équivalent, plusieurs pertes en rafales sont plus gênantes qu'une perte unique d'un grand nombre de paquets consécutifs. De plus, nous avons déduit que la perte d'une image I complète réduit sévèrement la qualité ($MOS < 3$). Cependant, la diminution de qualité dépend fortement du contenu et ainsi la mesure de la perte elle-même n'est pas suffisante pour prédire la qualité finale.

D'autre part, nous avons identifié l'importance de l'effet d'atténuation ou d'amplification de l'erreur que peut causer un changement de scène dans une séquence vidéo. En effet, les pertes (jusqu'à 40 paquets consécutifs) ayant lieu juste avant le changement de scène engendrent des dégradations qui ne se propagent pas à la nouvelle scène et n'ont donc pas une influence significative sur la qualité. Par contre, la perte de 20 paquets appartenant à la première image de la nouvelle scène pénalise gravement la qualité du fait de l'inefficacité de la compensation de pertes effectuée par le décodeur.

Une conclusion importante liée à la position spatiale de la perte est aussi dégagée. L'amplitude de la perte (exprimée par le nombre de paquets) devient un facteur quasiment négligeable quand la position spatiale de la perte dans l'image est variée. En particulier, une perte située dans la région d'intérêt de l'image gêne plus l'observateur qu'une perte d'amplitude double située loin de cette région. Une diminution de qualité d'approximativement deux unités MOS est parfois notée pour la perte de quelques paquets dans la région d'intérêt de l'image.

Sur la base de cette pré étude, nous nous intéressons dans le reste de ce mémoire à cet aspect perceptuel des pertes de paquets. Dans le but d'atténuer l'impact des pertes sur la qualité visuelle, nous étudions les phénomènes d'attention visuelle et proposons des méthodes pour protéger les régions les plus importantes de l'image contre les pertes de paquets.

Chapitre 6

Étude du déploiement de l'attention visuelle en visualisation de séquences vidéos

Introduction

Dans le chapitre précédent, nous avons montré qu'un des facteurs influençant l'effet perceptuel d'une perte de paquets sur la qualité est la position spatiale de la perte résultante dans l'image. On peut suspecter que les conséquences d'une perte n'ont donc pas le même impact selon qu'elles se situent dans une région importante ou non de l'image. En effet, un observateur qui regarde une séquence vidéo porte son attention sur des détails différents en fonction du contenu de la vidéo.

Dans ce chapitre, nous continuons le processus d'identification de la hiérarchie d'une vidéo en vue de l'application d'une protection adaptée. Cette hiérarchie est déterminée par l'attractivité visuelle des différentes régions de l'image. L'attention visuelle est mesurée pratiquement à l'aide de tests oculométriques consistant à suivre et enregistrer les mouvements des yeux d'un observateur. Des travaux de la littérature se sont également intéressés à la modélisation de l'attention visuelle.

Nous introduisons tout d'abord les mouvements oculaires et les mécanismes de sélection de l'attention visuelle. Ensuite, nous présentons le test oculométrique que nous avons réalisé en justifiant le choix d'une telle approche. Enfin, nous discutons les résultats de ce test. Bien que l'objectif principal du test soit la création d'une base de données de séquences vidéos avec leurs séquences de saillance, nous essayons aussi d'identifier l'impact des pertes de paquets sur l'attention visuelle. À travers ce dernier test, nous fournissons des pistes de réponse à la question

sur la relation entre le contenu de la séquence et l'attractivité de la perte.

6.1 L'attention visuelle

L'attention visuelle est un processus cognitif qui consiste à sélectionner les informations visuelles les plus pertinentes dans notre champ visuel. Ce processus implique les mouvements oculaires et les mécanismes de sélection décrits ci-dessous.

6.1.1 Les mouvements oculaires

L'œil humain est la source des informations visuelles envoyées au cerveau. La façon dont nous percevons une scène est déterminée par les mouvements de nos yeux. Les mouvements oculaires sont généralement divisés en deux grandes classes : les saccades et les fixations. D'autres types de mouvements oculaires secondaires comme les mouvements de poursuite sont aussi connus. Ces trois types de mouvement sont décrits ci-dessous.

- Les saccades sont des mouvements oculaires rapides de vitesse angulaire moyenne supérieure à 200 degrés d'angle visuel par secondes (°/s) avec des pics de vitesse pouvant atteindre 700 °/s [Sal99]. Pendant les saccades, aucun traitement d'information visuelle n'est effectué dans le SVH. Les saccades servent à diriger la fovéa de l'œil vers la zone d'intérêt. La fovéa est la partie de la rétine qui permet la visualisation des détails au niveau spatial et au niveau temporel. La vision fovéale est donc la composante essentielle de notre vision, l'autre composante étant la vision périphérique. Cette dernière composante ne permet pas une visualisation précise des détails mais confère une capacité de détection d'événements en périphérie [Car06].
- Les fixations sont des phases durant lesquelles un observateur fixe une région de l'image pour une certaine durée. Les fixations sont caractérisées par une forte activité de traitement perceptuel au niveau du SVH [vD99]. En effet, l'œil acquiert les informations et les transmet aux autres parties du SVH qui les codent et les transmettent à leur tour au cerveau. La région fixée est généralement définie comme étant centrée autour d'un point (pixel dans le cas d'une image par exemple). Le mouvement de l'œil entre deux fixations est assimilé à un mouvement de saccade.
- Les mouvements de poursuite représentent le suivi par l'œil des objets en mouvement. La vitesse relative d'un objet est la différence entre sa vitesse et celle du regard d'un observateur. Pendant un mouvement de poursuite, cette vitesse est nulle donc l'objet est stabilisé sur la rétine ce qui permet son examination par la fovéa. L'œil peut suivre des objets ayant une vitesse angulaire maximale de 60 °/s [Lan05].

6.1.2 Les mécanismes de sélection de l'attention visuelle

L'attention visuelle désigne le mécanisme de sélection des informations visuelles spatio-temporelles pertinentes du monde visible [Meu05]. Cette sélection est due au fait que toute l'information couverte par le champ visuel ne peut être traitée par le cerveau. Les mécanismes de sélection sont qualifiés de passifs s'ils sont inhérents à la physiologie même du SVH et actifs s'ils sont guidés par des facteurs externes. Des exemples de mécanismes passifs sont la sensibilité du SVH aux fréquences spatiales et son excitation par le contraste. En effet, le SVH est stimulé par une différence de luminance (contraste) et non pas par des valeurs absolues de luminance. Les mécanismes actifs de sélection de l'attention visuelle peuvent être exogènes ou endogènes [Pos80a].

Un mécanisme exogène (*bottom-up*) désigne l'attraction de l'attention de l'observateur uniquement par l'information visuelle présentée. C'est un mécanisme rapide et involontaire qui déplace l'attention vers les objets les plus saillants d'une scène. Par exemple, dans le cas d'une image d'un arbre dans le désert, notre regard est naturellement attiré par la forme qui se différencie de son voisinage. De même pour une tache de couleur sur une surface enneigée.

Un mécanisme endogène (*top-down*) désigne le déplacement de l'attention d'une manière volontaire suivant des critères cognitifs comme la reconnaissance des formes. Il module le mécanisme exogène dans le but de rechercher des éléments précis dans une scène. Les mécanismes endogènes peuvent aussi avoir lieu à la suite de l'affectation d'une tâche à l'observateur (par exemple la recherche d'une forme particulière dans une image).

6.2 Saillance ou régions d'intérêt ?

L'attention visuelle trouve son application dans plusieurs domaines comme le codage vidéo [Dho05], l'évaluation objective de qualité [Nin07] ou l'amélioration de la robustesse des vidéos contre les erreurs de transmission (travaux décrits dans le chapitre suivant). Ceci a motivé l'élaboration de modèles de l'attention visuelle pouvant être utilisés dans ce genre d'applications. Ces modèles sont généralement classés suivant le mécanisme d'attention (exogène ou endogène) mis en œuvre.

Un exemple de modèle adoptant l'approche *top-down* est celui d'Osberger et Maeder [Osb98] qui établissent une carte d'importance en se basant sur cinq facteurs calculés sur l'image segmentée en régions homogènes. Ces cinq facteurs sont le contraste de la région par rapport à son voisinage, sa taille, sa forme, sa position dans l'image et le plan auquel elle appartient (avant-plan ou arrière-plan). La combinaison de ces facteurs pour chaque région forme une carte d'importance dont la précision dépend fortement des performances de l'algorithme de segmentation utilisé. Pinneli et Chandler [Pin08] proposent une extension de ce travail vers la notion d'intérêt perceptuel des objets. La figure 6.1 montre une comparaison des performances des deux approches avec les évaluations subjectives.

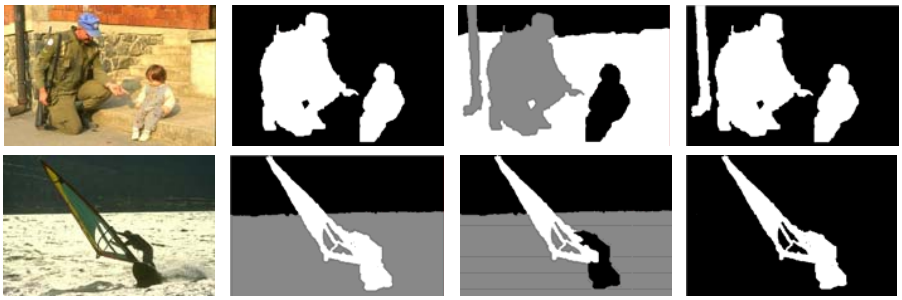


FIG. 6.1 – Deux exemples d'images avec leurs cartes d'importance (respectivement de gauche à droite) dégagées par les tests subjectifs, l'algorithme de Osberger *et al.* [Osb98] et l'algorithme de Pinneli et Chandler [Pin08]. La couleur d'une région indique son importance, le blanc représentant la région la plus importante et le noir la moins importante. (extrait de [Pin08])

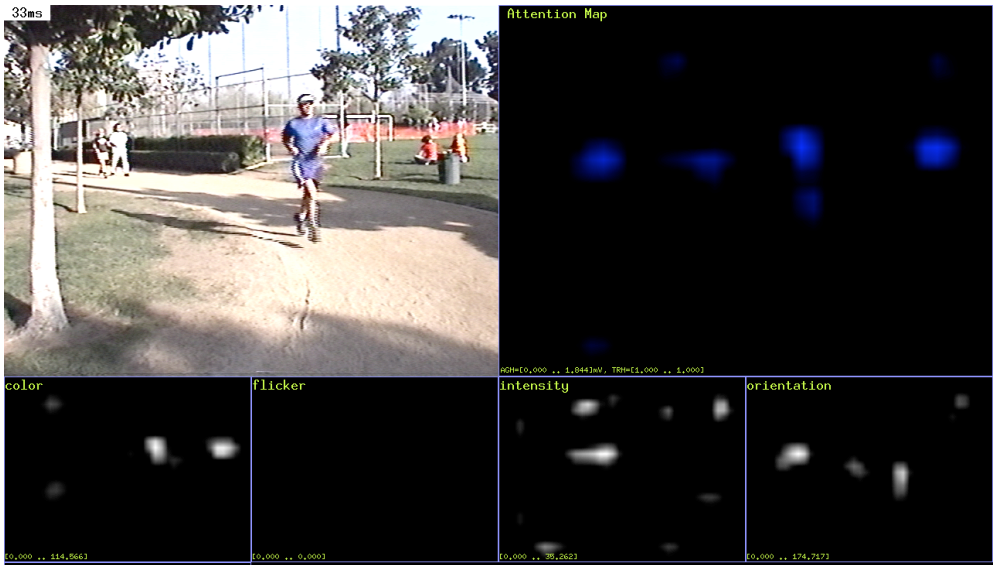
D'autre part, les modèles de détermination de la saillance adoptent une approche *bottom-up* de l'attention visuelle. Ils n'utilisent aucun mécanisme haut niveau de l'attention visuelle. Un modèle d'attention visuelle couramment utilisé est celui d'Itti *et al.* [Itt98]. Son principe de fonctionnement est le suivant : tout d'abord, des caractéristiques visuelles primitives (intensité, couleur et orientation) sont extraites à partir de l'image pour former des cartes de traits caractéristiques. Cette étape modélise essentiellement les étapes de la perception du SVH. Ensuite, les cartes sont normalisées par un filtrage à l'aide d'une différence de Gaussienne bidimensionnelle. Dans chaque carte, les régions les plus saillantes sont sélectionnées et enfin les cartes sont combinées par une moyenne pondérée pour obtenir la carte de saillance de l'image. Cette carte prend la forme d'une carte de niveaux de gris où les régions les plus claires correspondent aux régions les plus saillantes. Un exemple d'application du modèle d'Itti sur une image est donné figure 6.2.

Discussion

Les mécanismes de l'attention visuelle n'agissent pas en réalité d'une façon exclusive : la présence de l'un n'implique pas l'absence de l'autre. Conner *et al.* [Con04] déduisent, sur la base d'une étude bibliographique, que les mécanismes *bottom-up* sont déclenchés en premier puis sont suivis des mécanismes *top-down* lors d'expérimentations de recherche visuelle. Ces expérimentations consistent à trouver un objet cible enfoui parmi d'autres objets agissant comme des distracteurs. Cependant, cette conclusion n'est pas évidente pour les images de scènes naturelles. Parkhurst [Par02] montre, au travers d'expérimentations oculométriques, que les deux

(a) L'image originale.

(b) Les 5 points de fixation de l'image.



(c) L'image originale et ses cartes correspondantes.

FIG. 6.2 – Un exemple d'image fixe, ses points de fixation et ses cartes de saillance données par le modèle d'Itti. Dans (c), la carte de saillance (en haut à droite) combine les cartes de saillance d'intensité, de couleur et d'orientation (respectivement la première, la troisième et la quatrième de gauche en bas). Notons que la carte de vacillement (*flicker*) relative au mouvement est nulle pour une image fixe.

mécanismes d'attention visuelle ne peuvent être dissociés pour ce type d'image. Plus généralement, il suggère qu'il n'existe pas de situation naturelle pour laquelle l'attention visuelle dépende complètement d'un seul mécanisme.

Dans le cadre de notre travail de hiérarchisation du contenu d'une vidéo, l'utilisation d'un modèle objectif d'attention visuelle comme celui d'Itti [Itt98] ou de Pinneli [Pin08] n'est pas le meilleur choix car ces modèles considèrent un seul type de mécansime d'attention. Nous optons donc pour le choix des expérimentations oculométriques qui permettent d'obtenir la vérité terrain de l'attractivité du contenu d'une vidéo. Ces expérimentations doivent se dérouler dans un contexte *task-free* où les observateurs visualisent des séquences sans qu'aucune tâche ne leur soit assignée pour ne pas déclencher artificiellement les mécanismes *top-down*. Dans ce contexte, ces tests permettent le déclenchement naturel des mécanismes exogènes et endogènes qui vont déterminer les parties de l'image attirant le plus l'attention des observateurs.

6.3 Le test de suivi des mouvements oculaires

L'attention visuelle peut être mesurée au travers d'expérimentations oculométriques. Comme leur nom l'indique, ces expérimentations servent à quantifier les mouvements oculaires en suivant les déplacements de l'œil humain. S'ils sont menés rigoureusement, les tests d'*eye tracking* conduisent à l'obtention de la vérité terrain concernant une application particulière (navigation sur une page Web, étude de l'influence de la désynchronisation audio-vidéo sur l'attention visuelle [Gul07], *etc.*). Appliqués à une image ou une vidéo, ces tests indiquent les régions attirant l'attention de l'observateur au cours du temps.

Nous décrivons dans cette section l'approche expérimentale adoptée pour dégager les séquences de saillance de vidéos.

6.3.1 Objectifs du test

Deux objectifs principaux sont visés à travers ce test. Premièrement, nous voulons mesurer la saillance de vidéos dans le but d'appliquer une protection en fonction de l'importance perceptuelle des différentes parties du flux vidéo. Ces mesures de saillance permettent de répondre à la question suivante : qu'est-ce qui attire notre attention quand nous regardons une vidéo ? Nous utilisons donc un sous-ensemble de vidéos qui ne contiennent pas de dégradations (ni de transmission ni de codage) pour mesurer fidèlement la saillance des contenus.

Deuxièmement, nous voulons établir une relation entre les pertes de paquets et l'attention visuelle. Plus précisément, nous voulons comprendre comment les dégradations dues à ces pertes sont perçues par les observateurs. Pour ce faire, nous intégrons au test un second sous-ensemble de séquences dégradées par des pertes de paquets pour comparer leurs séquences de saillance aux séquences des vidéos n'ayant pas subi de pertes de paquets.

FIG. 6.3 – L'oculomètre utilisé dans nos expérimentations.

TAB. 6.1 – Les spécifications techniques de l'oculomètre utilisé.

Technique de mesure	Pupille et images de Purkinje
Fréquence d'échantillonnage	50 Hz
Résolution	$0,1°$
Précision	$0,25°$-$0,5°$
Excursion horizontale | verticale	$\pm 40°$ | $\pm 20°$
Mouvement de tête autorisé	± 10 mm

6.3.2 Le dispositif expérimental

Nous utilisons dans nos expérimentations l'oculomètre (*eye tracker*) de Cambridge Research Systems[1] illustré figure 6.3. L'oculomètre est fixé sur une table devant laquelle s'assoit l'observateur. Ce dernier appose son menton sur le support horizontal en réglant sa hauteur de façon à se rendre confortable. La sangle verticale sert aussi d'appui pour le front de l'observateur pour éviter tout mouvement de la tête pendant le test. Deux sources infrarouges illuminent l'œil de l'observateur et les reflets sur un miroir infrarouge sont filmés au moyen d'une caméra infrarouge. Le but de l'utilisation de deux sources d'illumination est de déterminer le positionnement précis de l'œil en illuminant la pupille et la cornée. Ceci est fait en comparant la position relative des rayons infrarouges reflétés sur la pupille par rapport à ceux réfléchis sur la cornée. Les spécifications techniques de l'oculomètre sont données dans le tableau 6.1.

Pour garantir une meilleure précision des mesures, une phase de calibrage est effectuée pour chaque observateur au début du test et répétée à plusieurs reprises pendant le test. Le calibrage

[1] `http://www.crsltd.com`.

vise à vérifier le réglage de l'appareil pour être sûr que l'œil est toujours bien illuminé par les sources infrarouges et que les mouvements de la pupille sont bien suivis. Cette étape consiste à demander aux observateurs de fixer un nombre de points qui apparaissent à l'écran et qui disparaissent dès que l'oculomètre valide la fixation. Un seul point est visible à chaque fois et la position des points dans l'image change à chaque apparition.

La figure 6.4 illustre ce processus. Dans 6.4.a, nous avons à gauche une image de l'œil d'un observateur où sont représentées les reflets des deux sources infrarouges. À droite, nous avons une capture de l'écran de contrôle sur laquelle le point à fixer est encadré. Sur la figure 6.4.b, nous présentons un calibrage réussi où les points de fixation de l'observateur se superposent sur les points affichés à l'écran. Par contre, le calibrage illustré sur la figure 6.4.c est à refaire puisque deux décalages importants (correspondants aux points encadrés) sont observés.

6.3.3 Les séquences vidéos

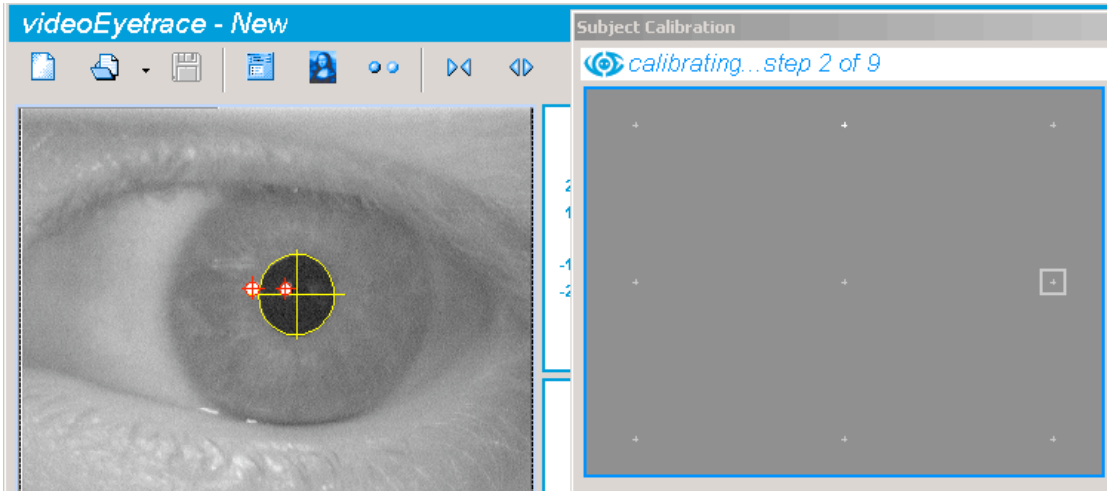
30 séquences vidéos à la résolution SD et 38 à la résolution HD, dont 23 contenus communs aux deux résolutions, sont utilisées dans ce test. Les séquences appartiennent à des bases de vidéos de VQEG[2] et sont toutes libres de droit. Elles ont des durées de 8 ou de 10 secondes. Les contenus communs sont initialement à la résolution HD. Le passage à la résolution SD nécessite la prise en compte du changement de format de l'image de 16/9 en HD à 4/3 en SD. Pour ce faire, nous rognons l'image centrale de la version HD située à 285 pixels à partir du bord gauche de l'image et ayant une largeur de 1350 pixels et une hauteur de 1080 pixels. À l'aide d'un filtre Lanczos [Duc79], nous redimensionnons l'image rognée et nous avons alors l'image de dimensions 720×576.

Les séquences sont affichées sur un écran LCD Haute Définition. Pour les séquences à la résolution SD, elles sont placées au centre d'une image grise de dimensions 1920×1080. La figure 6.5 illustre un exemple d'une image aux deux résolutions comme elle est affichée à l'écran. Les distances d'observation sont respectivement trois et six fois la hauteur de l'image pour les séquences HD et SD.

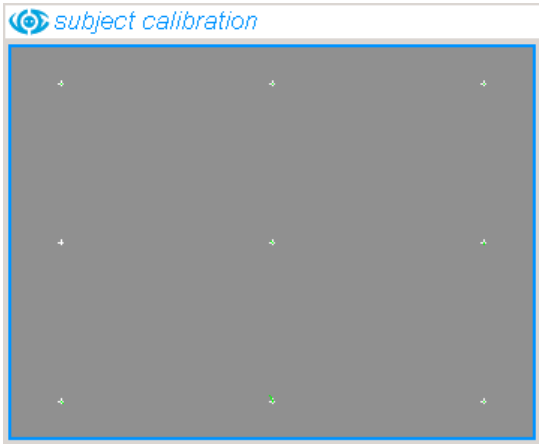
6.3.4 Codage et motifs de pertes

Pour les séquences ne comprenant pas de pertes de paquets, nous voulons mesurer fidèlement la saillance du contenu de la vidéo. Pour ce faire, nous devons coder les séquences de manière à n'avoir pas de dégradations liées au codage. Nous choisissons donc des débits de codage de 4 à

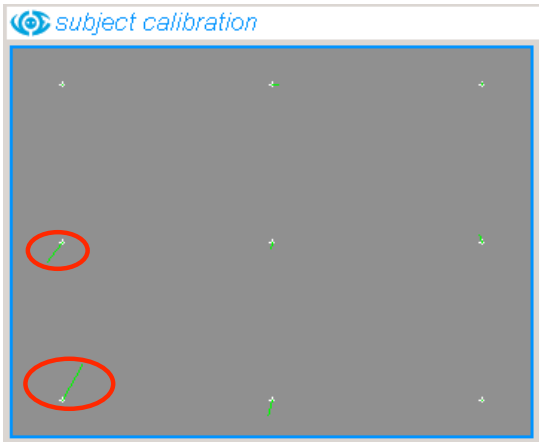
[2]Base NTIA : `ftp://vqeg.its.bldrdoc.gov/HDTV/NTIA_source/`
Base SVT : `ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat/`
Base Phase1 : `ftp://vqeg.its.bldrdoc.gov/SDTV/VQEG_PhaseI/TestSequences/Reference/`



(a) Une étape de la phase de calibrage.



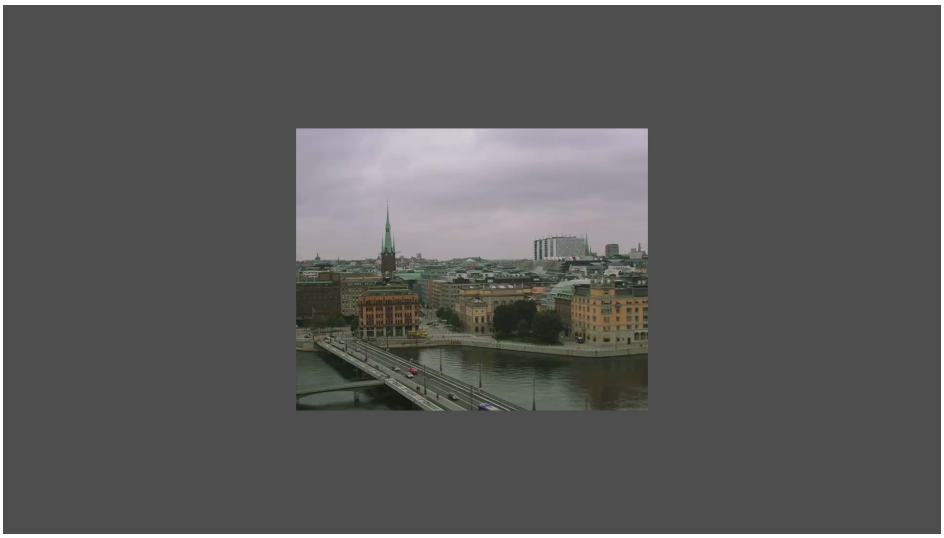
(b) Un calibrage réussi.



(c) Un calibrage non réussi.

FIG. 6.4 – La phase de calibrage et ses résultats possibles. La capture d'écran présentée dans (a) illustre une des neuf étapes du calibrage. Les neuf points de fixation de l'observateur correspondent bien dans (b) aux points affichés à l'écran. Ceci n'est pas le cas pour le calibrage présenté dans (c) où une disparité est clairement notée pour deux points.

(a) La séquence *Stockholm* à la résolution HD.



(b) La séquence *Stockholm* à la résolution SD.

FIG. 6.5 – La première image de la séquence *Stockholm* à la résolution (a) HD et (b) SD. Cette dernière est entourée d'un niveau de gris moyen pour permettre son affichage sur l'écran HD.

TAB. 6.2 – Les HRC du test de suivi des mouvements oculaires.

HRC	Débit (Mbit/s)		Slices perdues
	SD	HD	
0	4 à 6	12 à 16	0
1			5

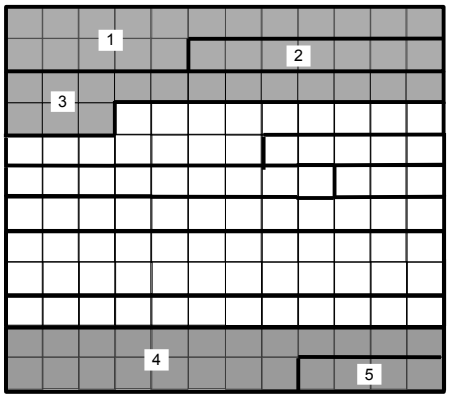


FIG. 6.6 – Illustration du motif de pertes utilisé pour un sous-ensemble de vidéos. Les cinq slices grisées correspondent aux cinq slices perdues de l'image I.

6 Mbit/s pour les séquence SD et de 12 à 16 Mbit/s pour les séquences HD. Le profil de codage H.264/AVC utilisé est le *High Profile* avec une structure de GOP IDRBBPBBP... de longueur 24 images. La version 14.0 du codeur de référence sert à coder et décoder les séquences.

Nous choisissons un seul motif de pertes permettant de vérifier à quel point les pertes attirent notre attention quand elles arrivent relativement loin de la région d'intérêt de l'image. Ce motif comprend la perte de cinq slices appartenant à la sixième image I de la vidéo (*cf.* tableau 6.2). Les cinq slices ne sont pas consécutives mais bien séparées comme l'illustre la figure 6.6. Trois d'entre elles se situent dans la partie supérieure de l'image tandis que les deux autres concernent la partie inférieure de l'image. Ces positions spatiales garantissent avec une forte probabilité que la région d'intérêt de l'image ne soit pas affectée par les pertes. Nous ciblons les images I pour obtenir une propagation maximale de l'erreur. Le pourcentage de slices perdues dans l'image I se situe dans les intervalles 5%-20% et 2%-5% pour les séquences SD et HD respectivement. Le taux de pertes global (TPG) est généralement inférieur à $0,5\%$ du nombre total de slices dans le flux.

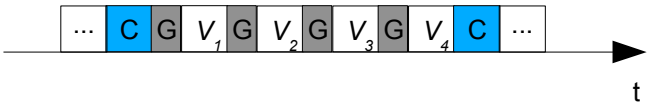


FIG. 6.7 – La structure de présentation des séquences dans le test. C désigne la phase de calibrage. V_n et G représentent respectivement la visualisation de la séquence vidéo n et des images grises de séparation.

Nombre total de PVS

Nous appliquons à 20 séquences SD et 12 séquences HD parmi les 68 séquences sources le même motif de pertes correspondant à HRC 1. Ces 32 séquences ajoutées aux 68 séquences n'ayant pas subies de pertes de paquets (HRC 0) forment donc un ensemble de 100 PVS.

6.3.5 Le déroulement de l'expérimentation

Pour un test en visualisation libre (*task-free*) comme celui que nous réalisons, seuls les mécanismes d'attention visuelle *bottom-up* doivent être mis en œuvre. Si une scène est visualisée plusieurs fois de suite dans un laps de temps court, ceci peut causer l'activation de mécanismes endogènes et par suite modifier la façon de regarder la séquence. Il faut donc éviter que deux séquences représentant le même contenu, l'une sans aucune perte de paquets et l'autre dégradée par le motif de pertes, ne soient visualisées par le même observateur durant le même test. Par contre, il est tolérable qu'un observateur visualise le même contenu en résolutions SD et HD dans le même test car les deux séquences ne sont pas identiques (conséquence du rognage).

Nous décidons de former deux ensembles de vidéos comprenant chacun 50 vidéos. Ces vidéos sont visualisées par les observateurs lors de deux tests différents. L'ordre de présentation des séquences dans un test est aléatoire et varie pour chaque observateur. Deux vidéos consécutives sont séparées par des images grises affichées sur l'écran pendant quatre secondes.

Le calibrage de l'oculomètre est effectué après chaque série de quatre vidéos consécutives selon l'illustration de la figure 6.7. La période est ainsi d'environ une minute en tenant compte des durées d'affichage des images grises. Nous utilisons neuf points pour le calibrage. Si son résultat est négatif, nous le répétons jusqu'à ce qu'il soit validé. La durée de chaque test est de 25 minutes environ.

37 observateurs non experts ayant une acuité visuelle normale ou parfaitement corrigée ont participé à l'expérimentation conduite dans un environnement normalisé suivant la recommendation BT.500-11 de l'ITU [itu02]. Nous ne leur avons assigné aucune tâche durant le test. Ils

devaient donc tout simplement regarder les séquences vidéos affichées à l'écran. Un test préliminaire (test de Porta [Rot02]) est réalisé au début de l'expérimentation pour déterminer l'œil directeur (*ocular dominance*) de l'observateur. L'œil directeur est l'œil dominant d'une personne qui présente une activité supérieure au second œil. La caméra de l'oculomètre est placée de façon à enregistrer les mouvements de l'œil directeur.

6.3.6 Des données oculométriques à la saillance

En sortie de l'oculomètre, nous avons, pour chaque observateur, les positions successives dans le référentiel de l'image du point d'intersection entre la direction du regard de l'observateur et le plan dans lequel est affichée l'image. L'oculomètre utilisé a une fréquence d'échantillonnage de 50 Hz, soit un point enregistré toutes les 20 ms. Pour une vidéo de 100 images affichées à la fréquence de 25 images par seconde (ips) par exemple, nous obtenons au plus deux points par image par observateur.

Les données brutes enregistrées ne sont pas significatives en elles-mêmes. Il faut en effet identifier les points correspondant aux fixations et aux poursuites et ceux correspondant aux saccades. Ces dernières ne sont pas utiles à la création d'une séquence de saillance puisqu'elles ne permettent pas l'acquisition d'informations visuelles pendant la durée du mouvement. La principale difficulté de cette procédure d'identification réside dans la distinction entre les mouvements de poursuite et de saccade, les premiers pouvant avoir des vitesses assez élevées. Il faut noter que ce problème n'est pas rencontré pour les images fixes car l'absence de la dimension temporelle implique l'absence de mouvements de poursuite.

L'algorithme de séparation des saccades et des fixations/poursuites que nous utilisons est inspiré du travail de Salvucci et Goldberg [Sal00]. Cet algorithme se base sur la mesure de la vitesse point-à-point pour décider si un point appartient à une fixation (ou poursuite) ou une saccade. Ce critère semble assez pertinent puisque les vitesses des mouvements de poursuite et de saccade se situent pratiquement dans des intervalles séparés (respectivement < 100 °/s et > 300 °/s selon Salvucci et Goldberg). Cependant, le seuil de vitesse utilisé pour la distinction entre les saccades et les poursuites doit être choisi avec précaution. Le choix d'un seuil discriminant est discuté à la section 6.3.6.2.

6.3.6.1 L'algorithme de création des séquences de saillance

Comme première étape, la vitesse de chaque point est calculée sur la base des données oculométriques : les coordonnées (x, y) des points et leurs instants d'échantillonnage. Après comparaison de la vitesse de chaque point au seuil fixé, les points correspondant à des saccades sont éliminés et les points de fixation consécutifs sont regroupés. La durée de chacun des groupes de fixations est alors comparée à un seuil pour valider ou non ce groupe comme fixation (ou

poursuite). Si la durée de la fixation est supérieure au seuil alors les points la constituant n'appartiennent pas à une saccade entre deux fixations mais bien à une fixation.

En résumé, les étapes de l'algorithme de séparation des saccades et des fixations sont les suivantes :

1. calculer la vitesse point à point de chaque échantillon ;
2. étiqueter chaque échantillon dont la vitesse point à point est inférieures au seuil comme fixation (ou poursuite) ; sinon, le considérer comme saccade ;
3. regrouper les échantillons consécutifs étiquetés comme fixation (ou poursuite) en groupe de fixations (ou poursuites) et supprimer les échantillons étiquetés comme saccade ;
4. supprimer les groupes de fixations (ou poursuites) dont la durée est inférieure au seuil pré-déterminé.

Nous pouvons maintenant construire la carte de fixations à partir des coordonnées des points obtenus. La carte de fixations (et de poursuites) $CS^{(k)}$ d'un observateur k dépend du nombre de fixations identifiées et de leurs durées. Elle est calculée selon la formule de l'équation suivante :

$$CS^{(k)}(x, y, t) = \sum_{j=1}^{N_F} \delta(x - x_j, y - y_j, t - t_j).d(x_j, y_j) \tag{6.1}$$

où N_F est le nombre de fixations identifiées, δ est le symbole de Kronecker, (x, y, t) représente les coordonnées de la fixation et son instant temporel et $d(x_j, y_j)$ est la durée de la fixation au point (x_j, y_j).

La carte de fixations moyennes des observateurs est calculée comme suit :

$$CS(x, y, t) = \frac{1}{N_{obs}} \sum_{k=1}^{N_{obs}} CS^{(k)}(x, y, t) \tag{6.2}$$

où N_{obs} est le nombre total d'observateurs.

La carte de fixations ainsi obtenue nécessite la prise en compte de deux facteurs supplémentaires pour qu'elle représente fidèlement la saillance de la vidéo. Premièrement, l'œil humain ne fixe pas un point précis de l'espace comme s'il était séparé de son voisinage. Plus précisément, notre vision n'est pas un processus discret qui concerne un point (le point fixé) sans affecter les points voisins. Il faut donc émuler la zone fovéale dans la construction de la carte de saillance. Deuxièmement, l'oculomètre est un appareil à précision limitée (comprise entre $0,25$ ° et $0,5$ ° selon le tableau 6.1). Pour ces deux raisons, la densité de saillance DS est calculée par convolution de la cartes de fixations moyenne avec une fonction Gaussienne bi-dimensionnelle :

$$DS(x, y) = CS(x, y) * g_\sigma(x, y) \tag{6.3}$$

où $g_\sigma(x, y)$ est la Gaussienne bi-dimensionnelle donnée par la formule suivante :

$$g_{\sigma_x,\sigma_y}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-(\frac{(x-\mu_x)^2}{2\sigma_x^2} + \frac{(y-\mu_y)^2}{2\sigma_y^2})} \tag{6.4}$$

avec $\sigma_x = \sigma_y$ les écarts-types dont la valeur est fixée à $0,5$ °. Cette valeur est du même ordre de grandeur que la précision de l'oculomètre d'une part et de la précision des fixations de l'œil humain d'autre part [Jor09].

Dans le reste du mémoire, nous confondons les termes "carte" et "densité" de saillance. Un exemple de carte de saillance obtenue avec l'algorithme décrit dans cette section est donné figure 6.8.

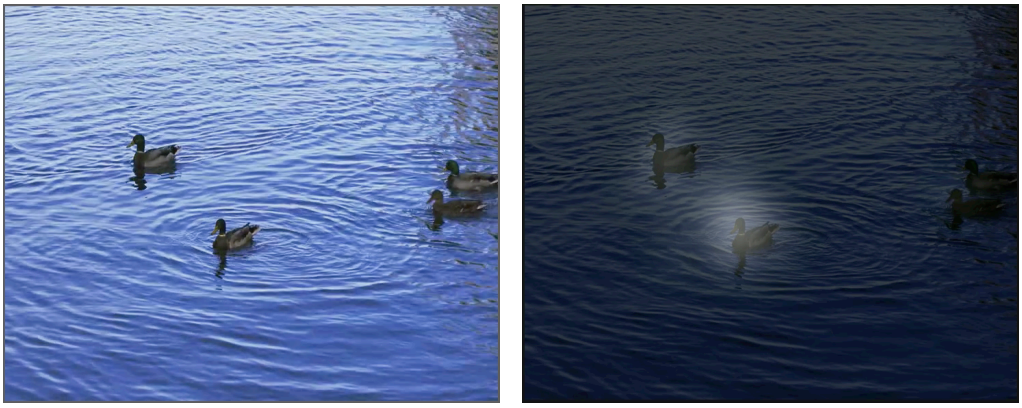


FIG. 6.8 – Un exemple d'une image de la séquence *Duck fly* (gauche) et sa carte de saillance (droite). Cette dernière est assombrie pour permettre une visualisation plus claire des régions saillantes de l'image.

6.3.6.2 Choix des valeurs des paramètres de l'algorithme

La spécification des valeurs seuils de la vitesse des mouvements des yeux et de la durée des fixations est expliquée ci-dessous.

– Seuil de la vitesse point-à-point : pour la séparation des fixations et des saccades, nous utilisons comme vitesse seuil la valeur 25 °/s. En effet, même si l'œil a une capacité de poursuite au-delà de cette valeur, Ninassi [Nin09] a montré que les objets les plus rapides d'une scène ont généralement une vitesse inférieure à 25 °/s. Cette étude, réalisée sur des séquences usuelles de test, est basée sur une observation des valeurs des vecteurs de mouvement des objets de l'image. Nous considérons que les résultats de Ninassi peuvent

être adoptés dans le cadre de cette thèse puisque la majorité des séquences utilisées dans les tests font partie des séquence étudiées dans [Nin09].

– Durée seuil des fixations : toute fixation n'est pas considérée comme telle si elle ne dure pas un certain temps. Cette durée dépend du contexte d'observation et de la tâche à effectuer. Ainsi, la lecture (de mots) nécessite une durée de fixation de 50 à 60 ms tandis que la compréhension d'une scène visuelle nécessite environ 150 ms [Ray09]. Manor et Gordon [Man03] estiment qu'une valeur de seuil égale à 100 ms discrimine efficacement les fixations des autres mouvements oculaires en observation libre (*task-free*). Ce seuil est aussi utilisé par Ninassi [Nin09]. Comme nos expérimentations sont conduites dans un contexte *task-free*, nous fixons ce seuil également à 100 ms.

6.4 Résultats et discussion

En plus de la création d'une base de données de vidéos avec leur séquences de saillance correspondantes, les expérimentations oculométriques fournissent des éléments de réponse à la question sur la relation entre les caractéristiques du contenu de la séquence et l'attention visuelle de l'observateur. La réponse à cette question n'étant pas l'objectif principal du test, l'analyse présentée ici sera limitée à un sous-ensemble des vidéos de la base.

6.4.1 L'impact d'une perte sur le déploiement de l'attention visuelle

L'impact d'une perte de paquets sur la qualité visuelle, étudié dans le chapitre précédent, dépend de facteurs comme la position spatiale dans l'image ou le contenu de la séquence vidéo. Quand la perte n'est pas dans la région d'intérêt, elle peut attirer l'attention de l'observateur et dévier son regard de la région fixée.

Quantification de l'attractivité d'une perte

Pour mesurer l'attractivité d'une perte dans une image, nous comparons, image par image, les cartes de saillance de la séquence n'ayant pas subi de pertes de paquets avec les cartes de saillance de la séquence affectée par le motif de pertes. Si la perte a attiré l'attention des observateurs, nous devons observer une différence entre les cartes de saillance des images ultérieures à l'image où la perte a lieu.

Nous choisissons d'utiliser la divergence de Kullback-Leibler [Kul51] qui est une mesure de dissimilarité entre deux ensembles. La divergence de Kullback-Leibler entre deux lois de probabilités p et q est donnée par l'équation suivante :

$$D_{KL}(p|q) = \sum_x p(x) log(\frac{p(x)}{q(x)}) \tag{6.5}$$

Dans cette section, p et q représentent respectivement la loi de probabilité de saillance provenant de la séquence sans perte de paquets et celle provenant de la séquence affectée par le motif de pertes. Une faible divergence reflète des cartes de saillance proches et par suite une faible attractivité de la perte (dans les images contenant des dégradations).

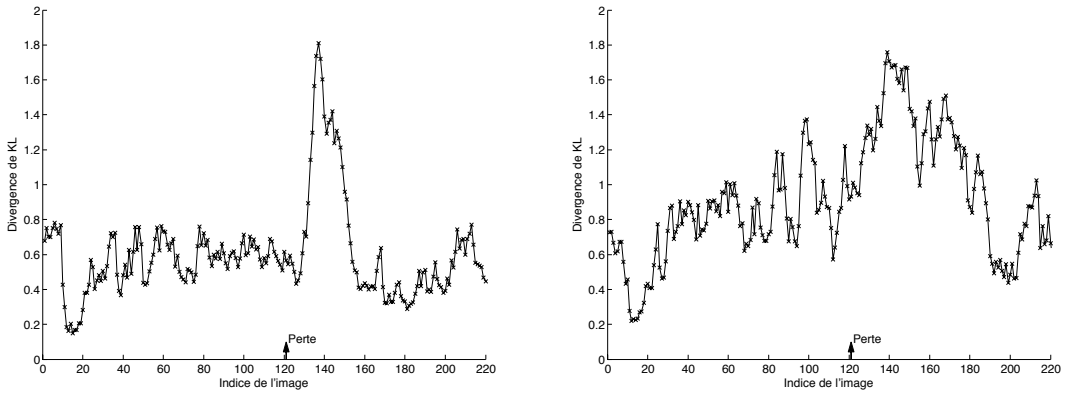
Les résultats du test effectué permettent de tirer les conclusions suivantes :

Le contenu de la séquence

Comme le nombre de slices perdues est le même pour toutes les séquences, l'amplitude des pertes est donc égale. Cependant, les conséquences de telles pertes au niveau du contenu de la séquence ne sont pas les mêmes. En effet, les slices étant d'une même taille approximativement (1450 octets), la surface de l'image qu'elles couvrent dépend de leur contenu. Une région uniforme est ainsi contenue dans un nombre de slices inférieur à celui utilisé par une région texturée. Par exemple, pour la séquence *Harp*, la perte de trois slices dans la partie supérieure de l'image impliquent la perte de $21,1\%$ du nombre total de macroblocs dans l'image. L'étendue spatiale de la perte peut évidemment influencer son attractivité.

Pour étudier l'impact d'une perte dans une région uniforme sur l'attention visuelle, nous prenons deux exemples de contenus : *Harp* et *Ulriksdals*. La figure 6.9 illustre la variation des valeurs de la divergence de Kullback–Leibler entre la carte de saillance de l'image i de la séquence n'ayant pas subi de pertes de paquets et la carte de saillance de la même image de la séquence affectée par des pertes de paquets.

Les courbes de variation de la divergence de KL présentent des pics qui font suite à la perte de paquets. Ces pics de divergence correspondent à la fixation de deux régions différentes dans les deux séquences : la région dégradée dans la séquence dégradée et la région d'intérêt dans la séquence originale. Les pertes de paquets, qui ont lieu à la 121ème image de la séquence, ne sont pas fixées instantanément par l'observateur. Ainsi, la divergence de KL augmente graduellement à partir de la 126ème image pour la séquence *Harp* et la 125ème pour *Ulriksdals*. Ce délai de quelque centaines de millisecondes entre l'apparition des dégradations et la fixation de la région dégradée par les observateurs est dû à deux facteurs. Premièrement, le SVH nécessite un temps de réponse pour réagir à l'apparition d'un stimulus dans la périphérie de sa vision fovéale. Deuxièmement, un mouvement de saccade est entrepris pour déplacer les yeux de la région d'intérêt à la région de la dégradation. Comme nous comparons des cartes de saillance reflétant les fixations des observateurs, la durée de la saccade n'est pas considérée ce qui justifie le fait que les fixations aient lieu quelques images après l'apparition des dégradations.



(a) La séquence *Harp*.

(b) La séquence *Ulriksdals*.

FIG. 6.9 – La variation de la divergence de KL pour les séquences *Harp* et *Ulriksdals* ayant subies la perte de cinq slices dans la 121ème image.

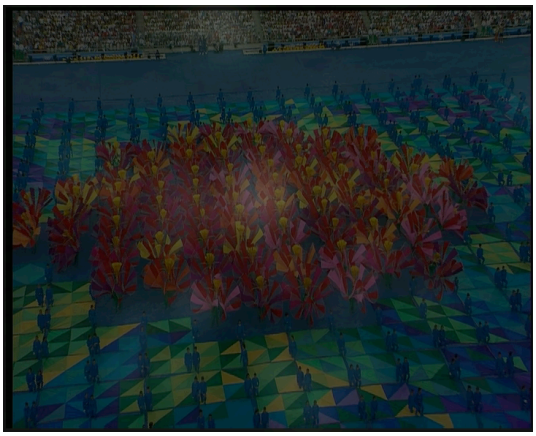
La longueur du GOP utilisé étant de 24 images, les dégradations peuvent durer jusqu'à l'image 144. Une diminution de la valeur de la divergence de KL est notée sur les figures 6.9.a et 6.9.b à mesure que les dégradations s'atténuent. L'atténuation des dégradations est due à la distance entre l'image dégradée et l'image I où la perte a lieu.

Sur l'exemple de la séquence *Harp*, la divergence de KL varie faiblement pour l'ensemble des images non dégradées par les pertes de paquets alors que cette variation est plus élevée dans le cas de la séquence *Ulriksdals*. Cette différence, due au contenu de la séquence, est expliquée dans le paragraphe suivant.

L'absence d'une région d'intérêt intrinsèque au contenu

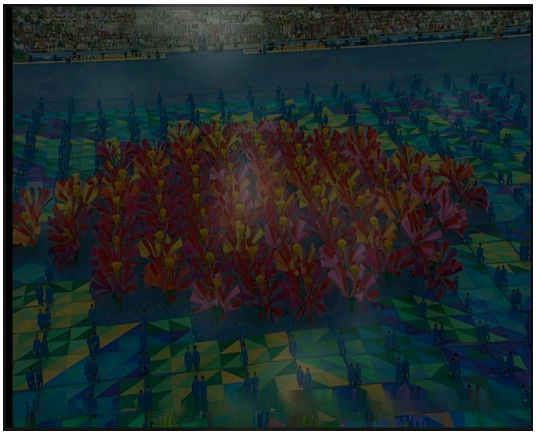
Dans les séquences utilisées, certaines ont une région d'intérêt qui attire naturellement le regard de l'observateur (ballon de rugby, visage d'une personne, *etc.*). Par contre, dans d'autres séquences, les observateurs ne sont pas unanimes quant à la région fixée pendant la durée de la vidéo. Cette situation a lieu surtout dans les séquences comprenant un grand nombre d'objets, fixes ou en mouvement, identiques ou distincts. Du fait de l'absence d'un point d'accroche dans la séquence, les observateurs sont facilement attirés par l'apparition d'une dégradation dans l'image. Ceci est mis en évidence pour les contenus de la figure 6.10 donnés avec leurs séquences de saillance. Sur cette figure, nous remarquons que les séquences de saillance changent en présence de pertes. Ainsi, sur la figure 6.10.c, une région saillante qui n'existait pas dans 6.10.a apparaît au niveau de l'arbre. De même pour la séquence *Barcelona* pour laquelle l'observateur moyen dévie son attention dans 6.10.d vers la région où la perte a lieu.

(a) La carte de saillance de *Above marathon* sans aucune perte.



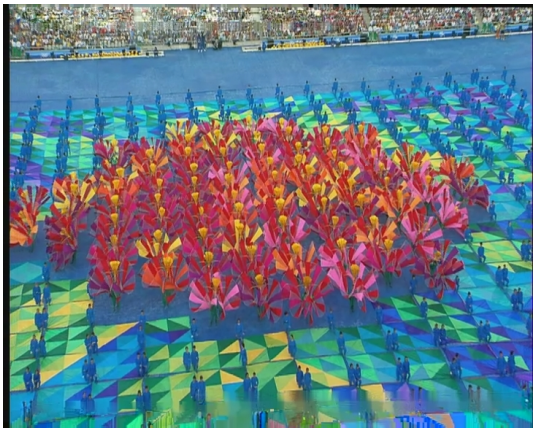
(b) La carte de saillance de *Barcelona* sans aucune perte.

(c) La carte de saillance de *Above marathon* affectée par la perte de cinq de ses slices.



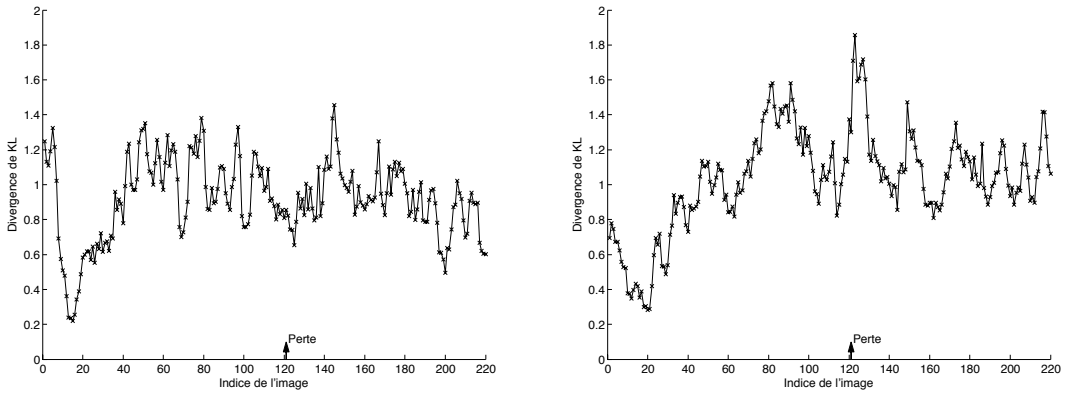
(d) La carte de saillance de *Barcelona* affectée par la perte de cinq de ses slices.

(e) La séquence *Above marathon* affectée par la perte de cinq de ses slices.



(f) La séquence *Barcelona* affectée par la perte de cinq de ses slices.

FIG. 6.10 – Les cartes de saillance de (a) l'image 135 de *Above marathon* sans aucune perte ; (b) l'image 142 de *Barcelona* sans aucune perte ; (c) l'image 135 de *Above marathon* avec pertes et (d) l'image 142 de *Barcelona* avec pertes. Les images 135 de *Above marathon* et 142 de *Barcelona* avec pertes sont respectivement données dans (e) et (f).



(a) La séquence *Above marathon*.

(b) La séquence *Barcelona*.

FIG. 6.11 – La variation de la divergence de KL pour les séquences *Above marathon* et *Barcelona* ayant subies la perte de cinq slices dans la 121ème image.

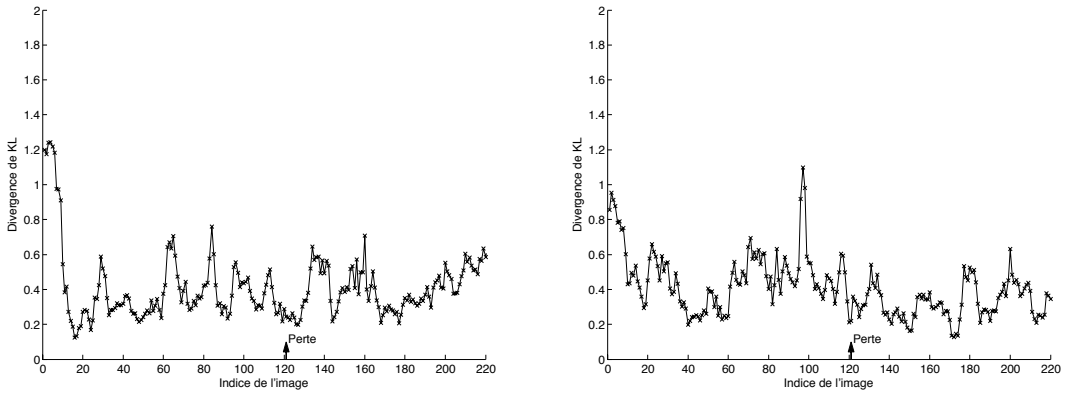
Nous illustrons sur la figure 6.11 la variation de la divergence de KL pour les deux séquences *Above marathon* et *Barcelona*. Un trait commun à ces deux figures est la forte variation de la divergence sur toute la durée de la séquence vidéo. En effet, l'observateur ne fixant pas un détail ou objet particulier dans la région d'intérêt, les point de fixation sont différents d'un observateur à l'autre.

L'augmentation de la divergence due à la perte dans le cas de la séquence *Barcelona* (figure 6.11.b) commence à partir de l'image 121 et le pic se situe à l'image 123 ce qui représente une fixation rapide de la région dégradée. Ceci peut être expliqué par l'absence d'une région d'intérêt intrinsèque au contenu. La séquence *Above marathon* (figure 6.11.a), dont le pic est situé à l'image 145, ne permet néanmoins pas de généraliser cette allure sur toutes les séquences ayant ce type de contenu.

Nous pouvons donc dire que l'apparition d'une dégradation due à une perte de paquets attire facilement l'attention visuelle en l'absence d'une région d'intérêt intrinsèque au contenu.

Le mouvement de la région d'intérêt

L'étude du contenu des séquences et du comportement correspondant de l'observateur moyen nous livre un constat intéressant : si la région d'intérêt intrinsèque au contenu est en mouvement, alors il est plus probable que la perte (ayant lieu à l'extérieur de cette région) n'attire pas l'attention de l'observateur. En effet, l'observateur qui suit une action voudrait la suivre jusqu'au bout, surtout s'il assimile le contexte et le but de l'action (par exemple l'athlète pendant une course ou le ballon dans un match de football).



(a) La séquence *Group disorder*.

(b) La séquence *Rugby*.

FIG. 6.12 – La variation de la divergence de KL pour les séquences *Group disorder* et *Rugby* ayant subies la perte de cinq slices dans la 121ème image.

Les séquences *Group disorder* et *Rugby* sont des exemples de séquences pour lesquelles l'observateur moyen n'a pas relativement dévié son attention de l'action d'intérêt de la scène en présence de dégradations. Ces séquences comprennent respectivement des mouvements rapides d'acteurs et de joueurs tenant un ballon. La figure 6.12 présente la variation de la divergence de KL en fonction du temps pour ces deux séquences.

Sur la figure 6.12.a, nous remarquons que la divergence varie irrégulièrement avec le temps. Cette irrégularité est due à l'entrée en scène d'acteurs en train de courir qui attirent l'attention de l'observateur. Quand la perte a lieu, nous remarquons une augmentation faible de la divergence qui reste inférieure aux autres pics de la courbe. Cette faible augmentation montre que les observateurs n'étaient pas tous attirés par les dégradations.

Sur la figure 6.12.b, une forte variation de la divergence est notée sur toute la durée de la séquence à cause de la rapidité de l'action. Le pic lié à la perte de paquets (à partir de l'image 120) ne se distingue presque pas des autres pics. Cependant, un pic surprenant est situé à l'image 97 qui ne contient aucune perte de paquets dans la séquence affectée par les pertes.

Ce pic est dû à un phénomène d'anticipation déjà identifié dans les travaux de Le Meur [Meu05]. En effet, lorsque le joueur lance le ballon, les observateurs anticipent la trajectoire du ballon et fixent la zone qu'ils estiment être le point d'arrivée du ballon. Comme cette estimation est très subjective, une forte divergence est notée entre les différents observateurs.

Discussion sur la diminution de la divergence au début des séquences

Nous notons pour les six séquences présentées une diminution de la valeur de la divergence de KL au début de leur affichage. Cette diminution a généralement duré pendant les 20 premières images de la séquence. Nous rappelons que la séquence débute après quatre secondes d'images grises durant lesquelles les observateurs promènent leur regard autour de la région centrale de l'image. À l'apparition du contenu de la séquence vidéo, les mécanismes *bottom-up* se déclenchent attirant ainsi l'attention des observateurs et par suite causant une diminution de la valeur de la divergence de KL. Environ 800 ms après l'activation des mécanismes *bottom-up*, la valeur de la divergence reprend une allure normale impliquant probablement l'activation des mécanismes *top-down* de l'attention visuelle.

6.4.2 Pertes, attention visuelle et qualité visuelle

Les conclusions dégagées des expérimentations réalisées ne sont valables que pour un même motif et un même nombre de pertes. Le but principal de ce chapitre étant l'obtention des séquences de saillance d'un ensemble de vidéos, l'interaction entre les facteurs identifiés ci-haut (contenu uniforme, absence d'une région d'intérêt intrinsèque au contenu et mouvement de la région d'intérêt) et la qualité d'usage n'est pas étudiée de près. Cependant, ces facteurs fournissent des pistes qui aident à comprendre cette problématique.

Tout d'abord, la visibilité de la perte ou de son effet dégradant contribue au jugement de qualité mais cette contribution est assez complexe. Nous ne pouvons pas considérer qu'une perte n'attirant pas l'attention de l'observateur n'influence pas la qualité visuelle. En effet, il se peut que cette perte soit visible pour l'observateur (en vision périphérique) et qu'elle occasionne une gêne exprimée lors de l'évaluation subjective de qualité.

Par ailleurs, une perte qui attire l'attention de l'observateur n'influence pas nécessairement son jugement de qualité. L'apparition d'une dégradation quelques secondes après le début de la visualisation d'une scène déclenche un mécanisme *bottom-up* qui vient perturber le mécanisme *top-down* déjà activé. Ce dernier concentrant l'attention sur la région d'intérêt, on peut penser que la qualité est plus influencée par le mécanisme *top-down* que par le mécanisme *bottom-up*.

Ensuite, la relation entre la durée des dégradations causées par les pertes et la qualité d'usage n'est pas simple. Il existe un délai entre l'apparition de l'effet de perte et la décision du SVH de déplacer l'attention visuelle ou non. Dans le cas positif, il faut en plus une durée minimale de fixation de la région dégradée pour bien évaluer l'effet de la dégradation. Entre temps, la dégradation peut devenir très atténuée ou même disparaître, par exemple à cause de la fin de la propagation temporelle. Comment la qualité est-elle perçue dans ce cas-là ? Le jugement de qualité est-il aussi sévère en présence de la perte qu'en son absence ?

Conclusion

Nous avons introduit dans ce chapitre la notion d'attention visuelle et ses mécanismes de sélection au niveau du SVH. Nous avons également présenté deux approches de modélisation de l'attention visuelle. La première approche concerne les modèles *top-down* où l'observateur visualise une scène suivant des critères cognitifs haut niveau. L'approche *bottom-up* considère l'attraction de l'attention visuelle exclusivement par la saillance des objets d'une scène.

Ensuite, nous avons décrit les méthodes subjectives de mesure de l'attention visuelle pour des vidéos. Bien que ces méthodes impliquent la mise en œuvre de moyens humains et matériels importants, elles sont fiables car elles permettent d'obtenir la vérité terrain concernant l'attractivité visuelle d'une vidéo.

Le premier objectif du test de suivi des mouvements oculométriques que nous avons réalisé était la création d'une base de vidéos couplées avec leurs séquences de saillance. Le second objectif était d'établir une relation entre le contenu de la vidéo et la visibilité des pertes de paquets. Nous avons montré dans ce cadre qu'une perte loin de la région d'intérêt n'attire généralement pas l'attention de l'observateur. Ceci peut toutefois arriver dans certaines conditions comme par exemple l'apparition d'une dégradation dans une région uniforme (exemple de la séquence *Harp*). Pour les séquences contenant une action rapide centrée autour d'un objet de la scène (*Rugby* par exemple), l'attention des observateurs est plus difficilement attirée que pour les séquences ne contenant pas une région d'intérêt intrinsèque (séquence *Above marathon*). Nous avons enfin proposé des éléments de réponse à la problématique de la qualité visuelle et des pertes de paquets, vue sous l'angle de l'attention visuelle.

L'importance perceptuelle de la position spatiale de la perte dans l'image a déjà été montrée dans le chapitre 5. La mesure de l'attention visuelle et par extension l'identification des régions spatiales les plus importantes de l'image ont été détaillées dans ce chapitre. Nous pouvons maintenant appliquer à ces régions des mécanismes de protection contre les pertes de paquets en vue de maintenir une qualité visuelle satisfaisante. Ces mécanismes font l'objet du chapitre suivant.

Quatrième partie

Protection de régions d'intérêt au sein de flux vidéos

Chapitre 7

Amélioration de la robustesse de flux vidéos H.264/AVC par protection des régions d'intérêt

Introduction

Différentes méthodes d'amélioration de la qualité d'usage ont été présentées dans le chapitre 2 de ce mémoire. Parmi ces méthodes, nous retrouvons les mécanismes qui opèrent au niveau du codeur pour améliorer la robustesse des flux vidéos codés en H.264/AVC contre les pertes de paquets. Le principe de tels mécanismes est de coder l'information en prévision des pertes qui auront lieu dans le réseau de paquets.

L'influence de la perte de paquets sur la qualité d'usage dépend de plusieurs facteurs dont la position spatiale de la perte dans l'image. Une perte cause généralement une gêne plus importante quand elle a lieu dans la région de l'image attirant l'attention de l'observateur qu'à l'extérieur de cette région. L'importance perceptuelle de la région d'intérêt de la vidéo a été déjà soulignée dans la section 5.4 du chapitre 5. Par ailleurs, la gêne peut être amplifiée par la propagation temporelle des dégradations liées à la perte dans plusieurs images. Dans la norme H.264/AVC, le phénomène de propagation est accentué du fait de l'utilisation intensive des prédictions intra et inter.

Nous proposons dans cette quatrième partie des méthodes de protection de flux vidéos qui prennent en compte les caractéristiques du contenu vidéo et la propagation de la perte. La méthode présentée dans ce chapitre restreint les types de prédiction possibles dans la région d'intérêt de l'image de manière à ce que cette dernière soit codée exclusivement en mode intra. L'idée est de tirer parti des possibilités du codeur pour augmenter la robustesse des régions

d'intérêt de la vidéo contre les pertes de paquets. Une extension de cette méthode est proposée dans le chapitre 8. Les travaux présentés dans cette partie peuvent être rapprochés d'un codage vidéo à base d'objets comme celui autorisé par la partie 2 de la norme MPEG-4 [Cai10].

Nous commençons le chapitre par la description des deux étapes de notre approche : la détermination des régions d'intérêt à protéger et l'application de l'algorithme de restriction de prédiction aux macroblocs de ces régions. Ensuite, nous évaluons les performances de notre approche en analysant la qualité objective d'une séquence et sa qualité subjective. Nous proposons également une extension de l'approche pour améliorer ses performances.

7.1 Description de l'approche proposée

La méthode que nous décrivons dans cette section s'applique aux régions d'intérêt de l'image. Ces régions sont déterminées à partir des points de fixation d'observateurs humains obtenus lors des tests oculométriques décrits au chapitre 6. L'utilisation de la vérité terrain nous permet de valider notre méthode d'une manière pertinente par rapport à la réalité de l'attention visuelle humaine.

7.1.1 Des séquences de saillance aux régions d'intérêt

Les séquences de saillance d'une vidéo représentent l'évolution dans le temps des régions de l'image qui attirent le plus l'attention de l'observateur moyen. Notre but est d'appliquer dans le codeur une protection aux macroblocs de la région d'intérêt. Le codage vidéo se fait sur la base d'une unité de codage qui est le macrobloc (16×16). Les données oculométriques étant disponibles pour des pixels (chaque pixel a une valeur de saillance), il faut donc passer à un traitement au niveau des macroblocs.

Ceci se fait premièrement par binarisation des séquences de saillance en comparant les valeurs de saillance de l'image à une valeur seuil. Seules les valeurs supérieures à ce seuil forment alors la région d'intérêt (RdI) de l'image. Une visualisation préliminaire de cartes de saillance expérimentales a été réalisée pour déterminer ce seuil empirique. Cette même visualisation a permis de s'assurer que la région d'intérêt n'occupe pas une partie importante de l'image pour éviter une dégradation significative de la qualité (plus de 20% de l'image dans notre cas).

Ayant classifié les pixels de l'image, nous considérons un macrobloc comme étant d'intérêt si au moins la moitié de ses pixels ont une valeur supérieure au seuil. Ce critère n'a été appliqué que pour les macroblocs situés relativement loin de la région la plus saillante de l'image. En effet, l'algorithme de création des séquences de saillance décrit à la sous-section 6.3.6.1 du chapitre précédent implique que la saillance des pixels diminue à mesure qu'on ne s'éloigne de la zone la plus saillante. Pour les macroblocs les plus saillants de l'image, leurs pixels sont donc généralement assez saillants pour qu'ils aient des valeurs supérieures au seuil.

(a) L'image 74. (b) L'image 82. (c) L'image 144.

FIG. 7.1 – Les cartes de saillance (haut) et les cartes de macroblocs correspondantes (bas) des images 74, 82 et 144 de la séquence *Harp*.

Quelques exemples du rendu du choix de seuil et des macroblocs d'intérêt sont donnés figure 7.1.

7.1.2 Algorithme restrictif de prédiction

Dans le but d'éviter la propagation d'erreurs d'image en image liée à la prédiction, nous proposons de coder tous les macroblocs de ces régions en mode intra. Ainsi, les régions d'intérêt sont indépendantes des images précédentes ou suivantes. Pour ce faire, nous implantons dans la version 14.0 du codeur de référence (JM) notre algorithme qui opère comme suit : pour chaque macrobloc d'une image B ou P, l'algorithme vérifie si le macrobloc appartient ou non à la RdI de l'image. Cette vérification se fait par comparaison des coordonnées du macrobloc à celles issues des données en sortie de l'oculomètre. Si le macrobloc appartient à la RdI, son type de prédiction est forcé en intra.

Une illustration du mode d'opération de l'algorithme restrictif est donnée figure 7.2. Dans l'image B ou P, les macroblocs de couleur grise appartiennent à la RdI de l'image. L'image de gauche est l'image de référence et ses macroblocs perdus sont marqués par une croix. Nous remarquons sur cette figure que les prédictions inter concernent uniquement les macroblocs

voisins des macroblocs gris qui eux sont codés en mode intra. Ainsi, la perte de macroblocs dans l'image de référence ne se répercute pas directement dans la RdI de l'image codée en mode inter.

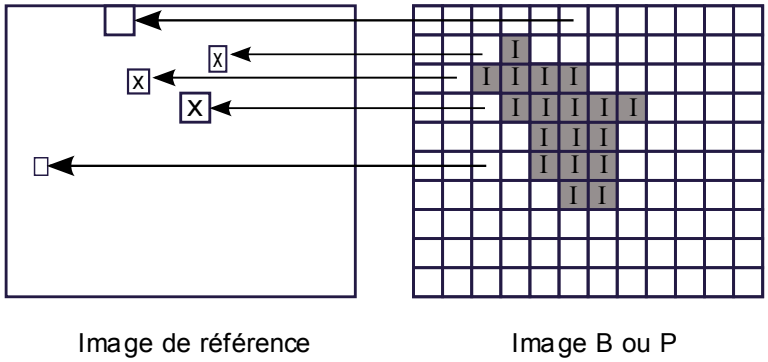


FIG. 7.2 – Illustration du codage restrictif intra. Les macroblocs grisés appartiennent à la RdI de l'image B ou P et par suite sont codés en mode intra. Les croix et les flèches indiquent respectivement des macroblocs perdus et quelques prédictions inter.

Surcoût de l'algorithme

En forçant le choix d'un mode de prédiction pour des macroblocs dans les images B et P qui n'est pas nécessairement le meilleur, l'efficacité de codage diminue. En effet, si le codage se fait à débit variable, la taille de la vidéo peut augmenter si un nombre de macroblocs est codé en intra alors qu'il était censé l'être en inter. Si le codage se fait à débit constant, la qualité de la vidéo peut diminuer à cause du débit supplémentaire occasionné par les macroblocs codés en intra. Ce surcoût de débit conduit à l'utilisation de paramètres de quantification plus élevés pour quelques macroblocs de l'image pour respecter la contrainte de débit fixée. D'où la possibilité de dégradation de qualité entre la séquence codée avec un codage classique et la séquence codée avec notre algorithme à débit identique. Les conséquences de ce surcoût sont étudiées dans la section 7.2.

Arrêt de la propagation spatiale des dégradations

La propagation temporelle de l'erreur dans les RdI est complètement stoppée quand l'algorithme restrictif est utilisé à cause du codage en mode intra dans les images B et P. Cependant, cet algorithme n'est pas très efficace contre la propagation spatiale de l'erreur. En effet, une slice peut contenir des macroblocs appartenant à la RdI et d'autres macroblocs qui appartiennent au reste de l'image. Prenons alors le cas où les macroblocs situés au-dessus et à gauche d'un macrobloc appartenant à la RdI d'une image B ou P n'appartiennent pas à la RdI. Ces macroblocs

peuvent être codés en mode inter et éventuellement utiliser comme références des macroblocs perdus (et donc compensés) ou affectés par la propagation de la perte. Comme la prédiction intra est spatiale et confinée à l'intérieur de la même slice, l'utilisation par des macroblocs de la RdI de tels macroblocs pour le codage intra conduit à l'introduction de la dégradation au sein même de la RdI.

Pour contourner ce problème, nous proposons d'établir un "voisinage de sécurité" autour de chaque macrobloc de la RdI. Le but de ce voisinage est de réduire l'impact de l'utilisation, lors de la prédiction intra, de macroblocs affectés par la propagation. Il est formé par les macroblocs situés au-dessus du macrobloc (à gauche, au centre et à droite) et à sa gauche. Ces macroblocs sont alors également codés en mode intra. Le voisinage peut être formé de moins de quatre macroblocs si l'un d'eux appartient à la RdI (et donc est nécessairement codé en intra). La figure 7.3 illustre un exemple de RdI avec son voisinage de sécurité. En comparant à la figure 7.2, nous notons que les voisins directs des macroblocs à la frontière de la RdI sont codés en mode intra et que les prédictions inter des macroblocs du voisinage n'existent plus. La propagation temporelle de l'erreur n'atteint plus alors la RdI de l'image.

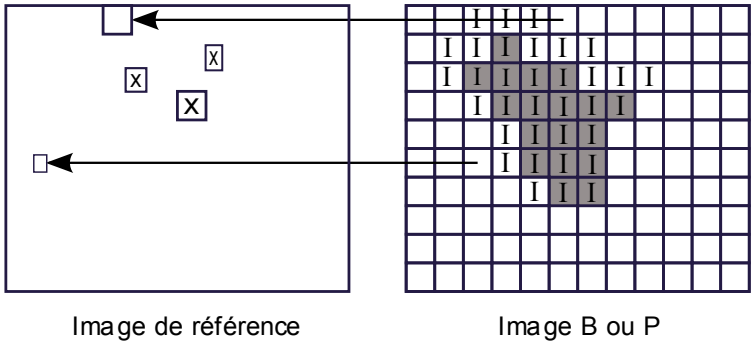


FIG. 7.3 – Codage restrictif intra avec voisinage de sécurité. Les macroblocs voisins directs des macroblocs de la RdI sont eux aussi codés en mode intra.

7.2 Première évaluation de performances

Pour évaluer les performances de la méthode de robustesse proposée, nous devons examiner si la propagation d'erreurs est interrompue dans la RdI de l'image. Pour ce faire, nous introduisons des pertes de paquets dans la RdI et nous évaluons ensuite la qualité de la vidéo à l'aide d'une métrique objective de qualité. La séquence *Harp* à la résolution SD (720×576) est utilisée. Le cadre général de notre travail est donné figure 7.4.

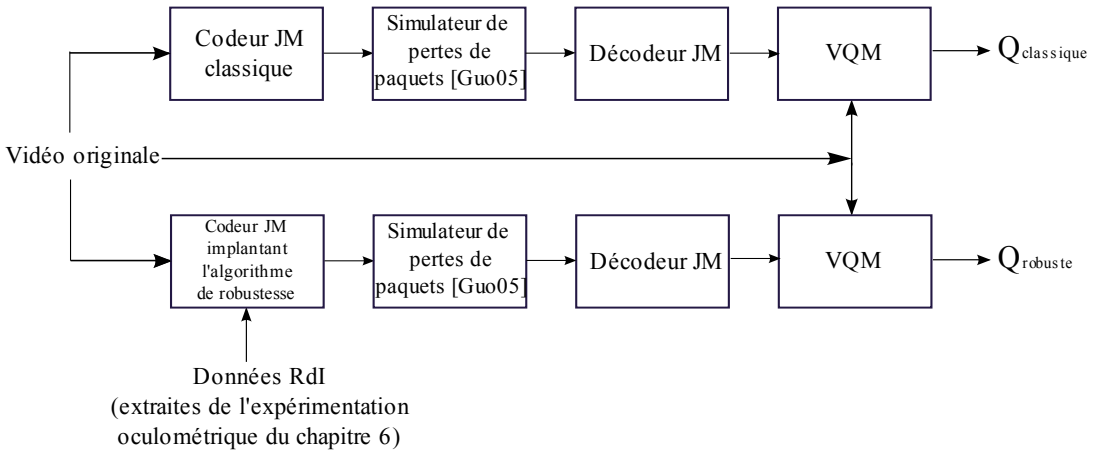


FIG. 7.4 – Le cadre général du travail présenté dans ce chapitre.

7.2.1 Motifs de pertes

Nous utilisons la version améliorée du simulateur de pertes du JVT décrite à la section 5.2 du chapitre 5. Ce simulateur nous permet de contrôler finement les slices à perdre. Comme les macroblocs de la RdI ne sont pas confinés dans une seule slice, la RdI s'étend sur plusieurs slices comme illustré figure 7.5. Dans cet exemple, les macroblocs en gris qui forment la RdI de l'image sont au nombre de 8. Deux de ces huit macroblocs appartiennent à la slice 1, cinq à la slice 2 et un seul à la slice 3. La perte d'une de ces slices affecte des macroblocs de la RdI mais aussi d'autres macroblocs qui ne sont pas classés comme ayant un intérêt perceptuel.

Nous testons deux motifs de pertes : l'un concerne trois slices consécutives et l'autre cinq slices non consécutives de la même image I. Ces slices ont la particularité de contenir des macroblocs appartenant à la RdI. Le choix d'une image I émane du fait que nous voulons contrer la propagation de l'erreur donc nous nous situons dans un cas de propagation maximale. Nous réalisons les pertes de slices dans la cinquième image I de la séquence *Harp* (97ème image sur 220), qui se situe autour de la moitié de la durée totale de la vidéo (dix secondes).

7.2.2 La métrique objective de qualité VQM

VQM (*Video Quality Metric*) [Pin04] est une métrique objective de qualité développée par l'ITS (*Institute for Telecommunication Sciences*), la branche d'ingénierie du NTIA (*National Telecommunications and Information Administration*) au ministère américain du commerce.

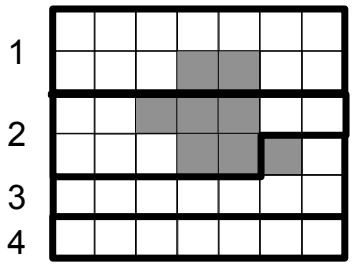


FIG. 7.5 – Un exemple d'une RdI (macroblocs en gris) qui s'étend sur trois slices consécutives.

Cette métrique, standardisée par l'ITU [itu04a, itu04b], est à référence réduite. Elle extrait des traits caractéristiques d'une séquence vidéo dégradée et de sa référence pour évaluer l'amplitude des dégradations. VQM est inspiré de phénomènes haut niveau de la perception visuelle. En particulier, il intègre la non-linéarité de la réponse du SVH à différents stimuli et de la réaction d'un observateur face aux dégradations. Les différentes phases d'opération de VQM sont données figure 7.6.

Tout d'abord, un alignement est effectué pour que les vidéos à comparer soient en correspondance spatiale et en correspondance temporelle. Ensuite, VQM divise la vidéo de référence et la vidéo dégradée en régions spatio-temporelles et en extrait des traits caractéristiques comme le gradient spatial, la chrominance, le contraste et l'information temporelle. Les valeurs de ces traits caractéristiques sont comparées et les paramètres obtenus sont combinés pour donner une note représentant la distorsion perçue. La note est comprise entre 0 et 1 et correspond au MOS de la séquence donné par un groupe d'observateurs lors de tests subjectifs. Par exemple, des notes de distorsion $0,1$ et $0,7$ correspondent respectivement à des valeurs de MOS de $4,6$ (bonne à excellente qualité) et $2,2$ (qualité médiocre) sur une échelle à cinq niveaux de qualité.

VQM rend possible la mesure de qualité sur une région spatio-temporelle définie en amont du calcul de la note. Ceci est dû au principe de fonctionnement de la métrique qui extrait les caractéristiques par fenêtres spatio-temporelles élémentaires. Une fenêtre de pixels est définie par ses deux dimensions spatiales et sa dimension temporelle.

La possibilité de mesurer la qualité sur une partie de l'image est évidemment d'un grand intérêt pour notre travail car elle nous permet d'évaluer la qualité de la RdI. Par ailleurs, VQM intègre plusieurs algorithmes optimisés pour différentes applications comme les visioconférences ou la télévision numérique. L'optimisation concerne la vitesse de calcul et le spectre de qualités rencontrées sur un service particulier. Nous choisissons d'utiliser le modèle télévision de VQM qui mesure les effets perceptuels des dégradations typiques de systèmes numériques (flou, effets de blocs, *etc.*).

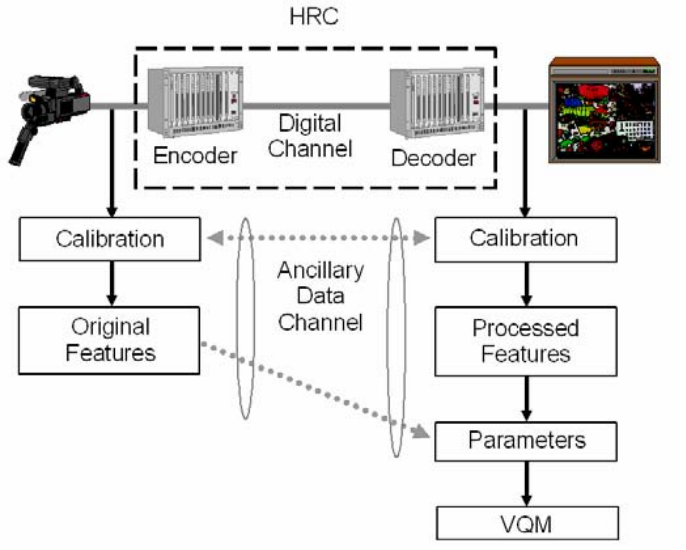


FIG. 7.6 – Le diagramme de blocs du critère objectif de qualité VQM [Pin04].

7.2.3 Résultats expérimentaux et discussion

Nous utilisons dans ce test la séquence *Harp* codée à l'aide du JM à un débit de 4 Mbit/s. À ce débit, la qualité de la vidéo est bonne et donc l'attention de l'observateur n'est pas perturbée par les défauts de codage. Nous choisissons cette séquence car son contenu implique le visage d'une personne qui attire naturellement l'attention d'un observateur. De plus, la dégradation de la RdI dans ce cas est facilement reconnaissable et engendre une gêne significative. La séquence *Harp* constitue donc un support idéal pour évaluer les performances de notre approche.

Le décodeur JM et ses algorithmes de compensation de pertes sont utilisés pour décoder la vidéo. La séquence *Harp* est codée deux fois en plus du codage classique : une première fois selon l'algorithme restrictif et une seconde fois utilisant la variante de l'algorithme comprenant le voisinage de sécurité. Ces deux modes de codage sont respectivement notés "codage RdI 1" et "codage RdI 2" dans la suite. L'évaluation de qualité des séquences codées se fait par rapport à la séquence de référence non compressée.

Performances de l'algorithme proposé

Nous représentons sur la figure 7.7 les notes de distorsion données par VQM pour chaque schéma de codage. Ces notes sont calculées en utilisant toute l'image (spatialement) et sur toute la durée de la vidéo. Nous remarquons sur cette figure que les notes des séquences non affectées par les pertes sont très proches. Elles sont respectivement $0,1$; $0,12$ et $0,13$ pour le codage classique, le codage RdI 1 et le codage RdI 2. Ces distorsions correspondent à des

notes de qualité (MOS) de $4,61$; $4,52$ et $4,47$ sur une échelle de 1 à 5. Ces valeurs, situées entre les catégories de qualité "bonne" ($MOS = 4$) et "excellente" ($MOS = 5$), montrent que l'algorithme n'introduit pas une dégradation de qualité significative au débit utilisé et pour la taille de RdI considérée.

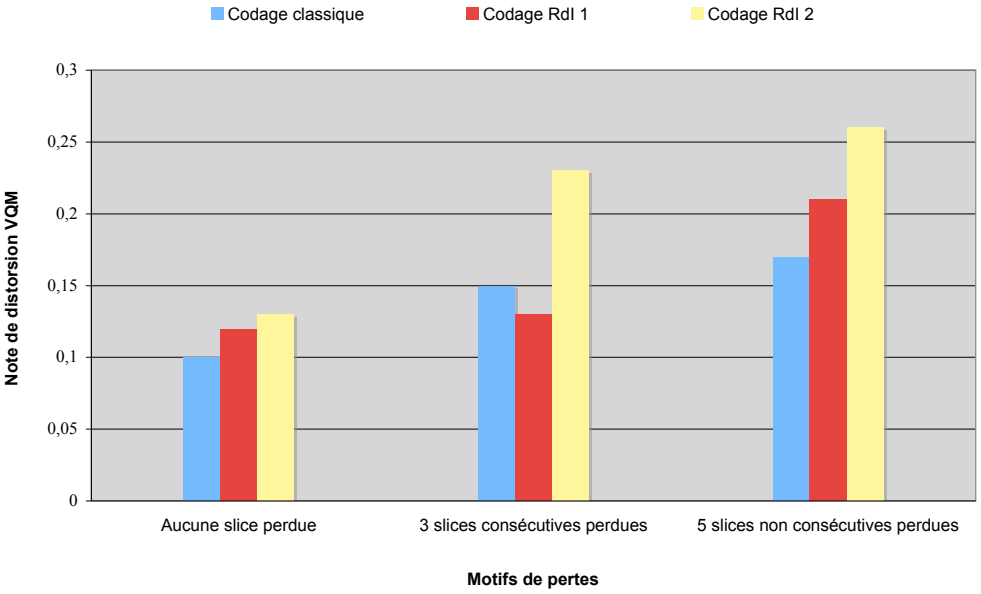


FIG. 7.7 – Les notes de distorsion des trois schémas de codage pour les deux motifs de pertes et le cas sans pertes.

Nous notons aussi que pour les deux motifs de pertes, la distorsion est généralement plus grande quand l'algorithme restrictif et sa variante sont utilisés. Ce résultat surprenant pourrait impliquer que l'algorithme proposé n'arrive pas à arrêter la propagation de l'erreur dans la séquence vidéo. Cependant, la comparaison visuelle des performances du codage classique et du codage RdI 1 pour la séquence *Harp* illustrée sur la figure 7.8 prouve le contraire. Le cadre vert représente la région à laquelle nous nous intéressons et qui appartient à la RdI de l'image. Quand une perte a lieu dans une image I, la propagation est atténuée au fur et à mesure que les images sont éloignées de cette image I. Dans le cas du codage classique, cette atténuation est progressive alors que dans le cas du codage RdI 1, la qualité augmente brusquement à partir de l'image 99 qui est juste à une distance de 2 images de l'image I ayant perdu trois slices.

L'évaluation de la qualité de la vidéo sur la totalité de ses résolutions spatiales et temporelles ne semble pas pertinente par rapport à notre approche qui a un effet local dans la RdI de l'image.

L'amélioration de qualité dans la RdI est ainsi "écrasée" par la mesure globale de la qualité. Nous effectuons donc une évaluation de qualité sur des résolutions spatiales et temporelles réduites.

Réduction spatiale de la fenêtre d'évaluation

Sur la figure 7.9, nous présentons les notes de distorsion calculées cette fois sur la région d'intérêt de l'image. Le but de cette évaluation est de dégager une note de distorsion qui reflète le plus fidèlement la qualité de la RdI. Le comportement noté sur cette figure est maintenant inversé : la distorsion est plus faible quand l'algorithme et sa variante sont utilisés. Même si les distorsions mesurées par VQM sont inférieures à la distorsion mesurée dans le cas d'un codage classique, il existe une incohérence entre les deux motifs de pertes. Ceci est dû à la dépendance entre la distorsion et le contenu des slices perdues qui ne sont pas les mêmes pour les deux motifs. Pour la perte de trois slices consécutives, les pertes sont condensées au niveau de la zone de mesure ce qui conduit parfois à une note de distorsion élevée.

Réduction spatio-temporelle de la fenêtre d'évaluation

Nous réduisons la région temporelle dont la qualité est évaluée par VQM pour concentrer l'évaluation autour de la perte de paquets. En effet, le but de l'expérimentation est d'évaluer les performances de notre algorithme en présence de propagation d'erreurs. Nous diminuons donc la durée sur laquelle est appliqué VQM au minimum autorisé (4 secondes). À la fréquence d'affichage de 25 ips, cette durée minimale représente 100 images. Nous y incluons évidemment le GOP dans lequel les pertes ont lieu. Nous gardons comme région spatiale à évaluer les RdI des cent images.

Les diagrammes en bâtons présentés sur la figure 7.10 montrent l'efficacité de l'algorithme dans la RdI. Ainsi, la note de distorsion pour la perte de trois slices consécutives est de $0,34$ pour un codage classique alors qu'elle diminue à $0,14$ pour le codage RdI 1.

Comparaison visuelle des trois schémas de codage

Sur la figure 7.11, les trois versions de l'image 103 de la séquence *Harp* sont illustrées. Nous remarquons dans l'image de la figure 7.11 que la RdI (le visage de la harpiste) présente des dégradations dues à la compensation des pertes ce qui n'est pas le cas pour les images des figures 7.11.b et 7.11.c. Par ailleurs, l'utilisation du voisinage de sécurité ne semble pas apporter un grand intérêt. L'effet de blocs qui apparaît sur les figures 7.11.b et 7.11.c est dû aux

FIG. 7.8 – La séquence *Harp* ayant perdu cinq slices consécutives de l'image 97. De haut en bas, les images 97, 99, 105 et 116 de la version codée par un codage classique (gauche) et par le codage robuste (droite).

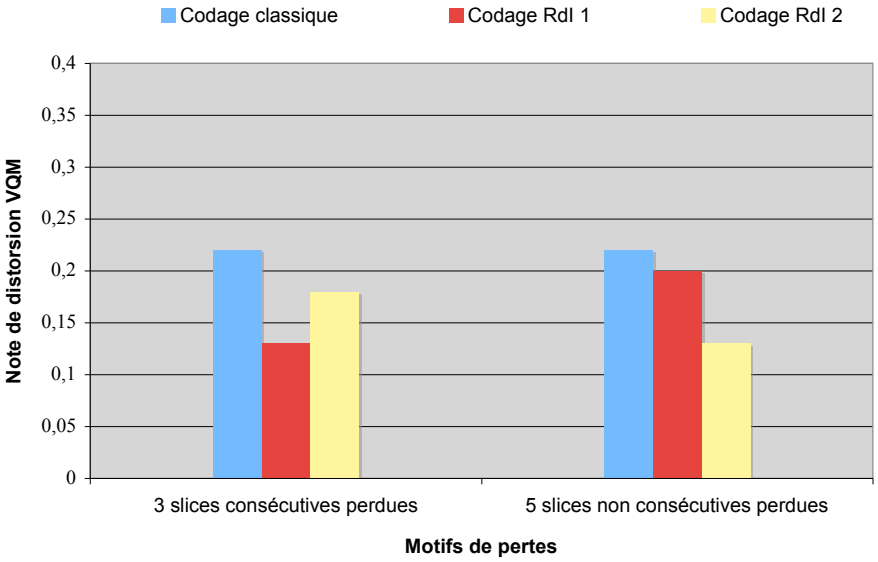


FIG. 7.9 – Les notes de distorsion des trois schémas de codage pour les deux motifs de pertes. VQM est appliqué au niveau de la RdI des images uniquement (moyenne spatiale).

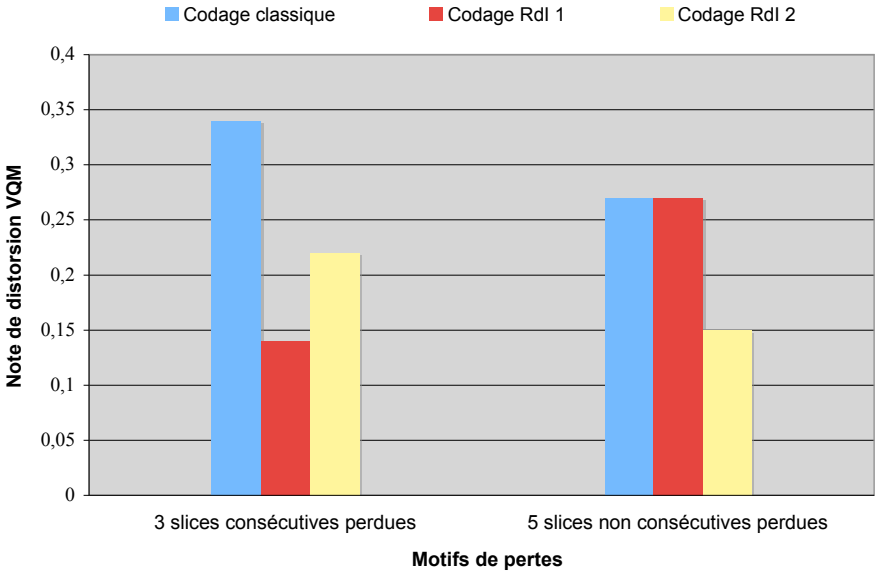


FIG. 7.10 – Les notes de distorsion des trois schémas de codage pour une application de VQM sur la RdI de cent images de la vidéo (moyenne spatio-temporelle).

(a) Codage classique. (b) Codage RdI 1. (c) Codage RdI 2.

FIG. 7.11 – L'image 103 de la séquence *Harp*, ayant perdu cinq slices consécutives de l'image 97, codée avec (a) un codage classique ; (b) le codage robuste et (c) le codage robuste avec prise en compte du voisinage de la RdI.

dépendances spatiales entre les macroblocs de la RdI et les macroblocs adjacents à la RdI. Ces derniers peuvent en effet être prédits en mode inter à partir d'une référence elle-même affectée par la perte. Dans ce cas, le codage intra crée un effet de blocs qui apparaît dans la RdI de l'image. Les notes de distorsion élevées illustrées sur la figure 7.7 pourraient être dues au fait que VQM pénalise les effets de blocs plus que les autres distorsions.

7.3 Amélioration des performances

Notre algorithme d'amélioration de la robustesse n'est pas efficace contre toutes les formes de propagation de l'erreur. En particulier, la propagation intra-image (spatiale) de l'erreur n'est pas arrêtée. Pour améliorer les performances de cet algorithme, nous proposons l'utilisation du codage intra contraint qui est une technique de robustesse de la norme H.264/AVC pour réduire les dépendances spatiales entre la RdI et son voisinage.

7.3.1 Limitations du codage restrictif intra

Le défaut majeur de notre algorithme restrictif réside dans une diminution limitée des effets perceptuels de la propagation spatiale de l'erreur. Pour évaluer l'importance de la propagation spatio-temporelle des dégradations et en particulier celle de la propagation spatiale, nous avons développé un outil qui reflète les dépendances intra et inter-images des macroblocs de la vidéo. Cet outil, implanté dans le décodeur JM, permet de suivre l'effet de propagation de la perte d'une ou de plusieurs slices.

7.3.1.1 Identification des dépendances intra et inter-images

Le principe de fonctionnement de cet outil est le suivant : en entrée, nous spécifions le numéro de la slice dont nous voulons identifier les dépendances spatio-temporelles. Nous indiquons aussi jusqu'à quel ordre de dépendance temporelle nous voulons suivre les dégradations.

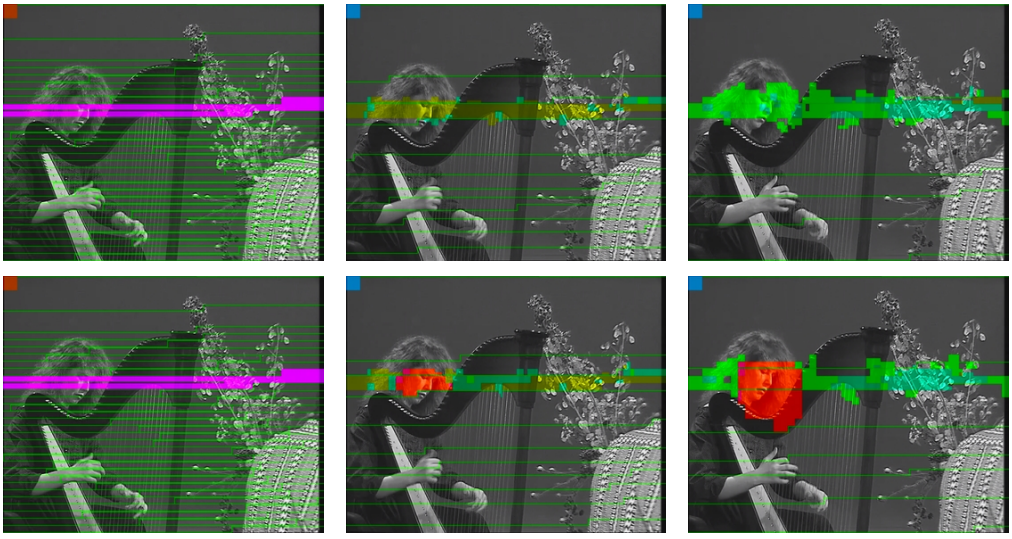
Le premier ordre de prédiction inter concerne les macroblocs prédits à partir de la slice indiquée. Un macrobloc est dit d'ordre n avec $n \geq 2$ s'il est prédit d'un macrobloc d'ordre $(n-1)$. Par exemple, le second et le troisième ordre signifient respectivement une prédiction à partir d'un macrobloc de premier et de second ordre. Une analyse visuelle préliminaire des dégradations dues à des prédictions inter d'ordre supérieur à 3 a montré que ces dernières ne contribuent pas d'une manière significative à la perte de qualité. Au niveau de la propagation spatiale, nous identifions deux cas de figures : la prédiction intra à partir d'un macrobloc codé en intra et la prédiction intra à partir d'un macrobloc codé en inter (quelque soit son ordre).

Quand le décodage du flux binaire commence, chaque macrobloc de la slice spécifiée est marqué pour suivre son éventuelle utilisation comme référence de prédiction. Dès qu'un macrobloc appartenant à une autre image utilise un des macroblocs marqués dans le processus de prédiction inter, il est à son tour marqué. Le même principe est appliqué aux prédictions intra. En sortie du décodeur modifié, nous recueillons l'information sur les dépendances spatio-temporelles sous deux formes. La première forme, textuelle, nous permet d'entreprendre une analyse des relations entre les macroblocs de la vidéo. La seconde forme est visuelle et se traduit par une séquence vidéo où les macroblocs sont colorés selon leur type et ordre de prédiction.

7.3.1.2 Suivi de la propagation spatio-temporelle des dégradations

Nous donnons figure 7.12 deux exemples de propagation spatio-temporelle des dégradations dues à la perte de trois slices dans la sixième image I de la séquence *Harp*. Le premier exemple illustre les dépendances des macroblocs pour un codage classique. La couleur rose indique les macroblocs perdus (7.12.a). Les couleurs jaune, bleue et verte représentent respectivement le premier, le second et le troisième ordre de prédiction inter (7.12.b et 7.12.c). Nous remarquons dans les images B de cet exemple une propagation des dégradations dans et autour de la région occupée par les slices perdues dans l'image I. La répartition des types de prédiction dépend de la distance entre l'image prédite en inter et l'image de référence : nous retrouvons une majorité de macroblocs jaunes dans l'image 123 et une majorité de macroblocs bleus et verts dans l'image 128.

Les macroblocs prédits en mode intra à partir de macroblocs codés en inter et intra sont respectivement désignés par les couleurs rouge et orange. Ils sont présents dans les images 123 et 128 de l'exemple du codage robuste. La présence de tels macroblocs dans des images prédites en inter est due à deux facteurs : l'application de notre algorithme de restriction des types de prédiction et le choix du type de prédiction intrinsèque au codeur. Le premier facteur influence exclusivement les macroblocs de la RdI alors que le second agit à l'extérieur de cette région.



(a) L'image I d'indice 121. (b) L'image B d'indice 123. (c) L'image B d'indice 128.

FIG. 7.12 – Simulation de la propagation spatio-temporelle des dégradations dans le cas d'un codage classique (haut) et d'un codage robuste (bas). La couleur du rectangle situé au coin supérieur gauche indique le type de l'image (I, B ou P).

Les images présentées sur la figure 7.13 illustrent l'impact de la perte des trois slices de l'image I de la figure 7.12.a.

Le tableau 7.1 présente les dépendances liées aux macroblocs des trois slices perdues dans l'exemple de la figure 7.12. Ces dépendances sont données pour le codage robuste proposé. Une dépendance de la forme I(T) signifie que le macrobloc est codé en mode intra à partir d'un macrobloc prédit en inter qui utilise un macrobloc de référence perdu ou compensé. La notation I(I) désigne une prédiction intra basée sur un macrobloc codé aussi en mode intra. Le premier, second et troisième ordre de prédiction sont respectivement notés T1, T2 et T3. Les pourcentages sont calculés par rapport au nombre total de macroblocs dépendant des slices perdues.

Nous notons que la majorité des dépendances I(I) ont lieu dans la RdI du fait de la disponibilité de macroblocs adjacents codés en mode intra. Les dépendances temporelles représentent approximativement le même pourcentage. La part des dépendances I(T), bien qu'inférieure aux autres, semble avoir un impact considérable sur la qualité visuelle. En effet, ces dépendances

(a) L'image I d'indice 121. (b) L'image B d'indice 123. (c) L'image B d'indice 128.

FIG. 7.13 – Comparaison des dégradations dans le cas d'un codage classique (haut) et d'un codage robuste (bas) pour les images de la figure 7.12.

TAB. 7.1 – La distribution des dépendances des macroblocs appartenant au GOP ayant trois slices de son image I perdue.

Type de dépendance	I(I)	I(T)	T1	T2	T3
Nombre de macroblocs affectés	1473	126	1432	1861	1776
Pourcentage de chaque type de dépendance	$22,1\%$	$1,9\%$	$21,5\%$	$27,9\%$	$26,6\%$

entraînent une propagation spatiale de l'erreur touchant une grande partie des macroblocs codés en intra qui représentent $22,1\%$ des macroblocs affectés par la propagation des dégradations.

7.3.2 Modification de la méthode de robustesse

L'analyse effectuée ci-avant nous conduit à l'utilisation d'une caractéristique de codage de la norme H.264/AVC : le codage intra contraint (*constrained intra prediction*). Ce paramètre de codage peut être spécifié en amont du codage pour qu'il puisse être pris en compte par le codeur de référence. Le principe de la prédiction intra contrainte est d'interdire au codage intra l'utilisation de données résiduelles et d'échantillons appartenant à des macroblocs voisins codés en mode inter. En d'autres termes, un macrobloc codé en intra n'utilise que des macroblocs codés eux-mêmes en intra dans la même slice.

Le surcoût de cette technique est évalué par Dhondt *et al.* [Dho07] quand l'actualisation par insertion de macroblocs codés en intra dans les images B et P est utilisée. Les résultats de

l'analyse montrent que la prédiction contrainte est coûteuse surtout pour les séquences codées à bas débit (entre $1,5\%$ et 23%). Ceci est expliqué par le fait que cette insertion se fait parmi des macroblocs quasiment tous codés en inter donc la prédiction intra utilise nécessairement le mode 2 (DC). Ce mode implique, en l'absence d'échantillons de référence pouvant être utilisés, que la valeur de prédiction de tous les pixels doit être fixée à une valeur constante (128 dans l'implantation du JM). L'erreur de prédiction (résidu d'information) est alors conséquente ce qui se répercute en un débit supplémentaire de codage.

Dans notre cas, même si le nombre de macroblocs dans la RdI est conséquent (et par suite le nombre de codages intra), les macroblocs sont confinés dans une région spatiale précise. Cette disposition offre souvent au macrobloc à coder des macroblocs voisins codés en intra. Par conséquent, la prédiction intra contrainte n'affectera que les macroblocs qui sont à la frontière entre la RdI et le reste de l'image. Pour les macroblocs situés à l'extérieur de la RdI et qui étaient initialement codés en intra, ils le seront toujours si ce choix est considéré par le codeur comme bénéfique (en termes de débit-distorsion). En pratique, le nombre de tels macroblocs diminue significativement car l'absence de macroblocs voisins codés en mode intra oblige le codeur à utiliser le mode 2 (DC). Ce mode étant généralement plus coûteux que les modes inter, il est décliné par le codeur au profit d'une prédiction en inter.

7.3.2.1 Surcoût du codage intra contraint

Comme nous l'avons évoqué précédemment, le surcoût de l'utilisation de cet outil à débit constant se répercute sur la qualité de la séquence. Nous effectuons donc une évaluation objective de qualité de trois séquences codées par les trois schémas de codage suivants : le codage classique, le codage RdI 1 et le codage RdI 1 utilisant un codage intra contraint. Les résultats de l'évaluation sont présentés sur la figure 7.14. Comme première constatation, nous pouvons dire que la qualité initiale des séquences codées est très bonne puisque les notes de distorsion sont toutes inférieures à $0,12$ (qui correspond à un MOS supérieur à 4). Les variations de qualité entre les différentes versions d'une séquence sont minimes pour les séquences *Harp* et *Formula 1*. Cette variation, même si remarquable dans le cas de la séquence *Canoe*, n'est pas significative du fait qu'elle se situe en haut de l'échelle de qualité où les variations n'ont pas une grande influence sur le jugement d'un observateur.

7.3.2.2 Performances en présence de pertes

La figure 7.15 montre les notes de distorsion VQM calculées sur la région spatiale de l'image dégradée par la perte de paquets (contenant la RdI). Nous pouvons remarquer que le codage *RdI 1 contraint* ne contribue pas significativement à la diminution des distorsions dues à la propagation spatiale des dégradations. Ainsi, pour les séquences *Harp* et *Canoe*, une très légère

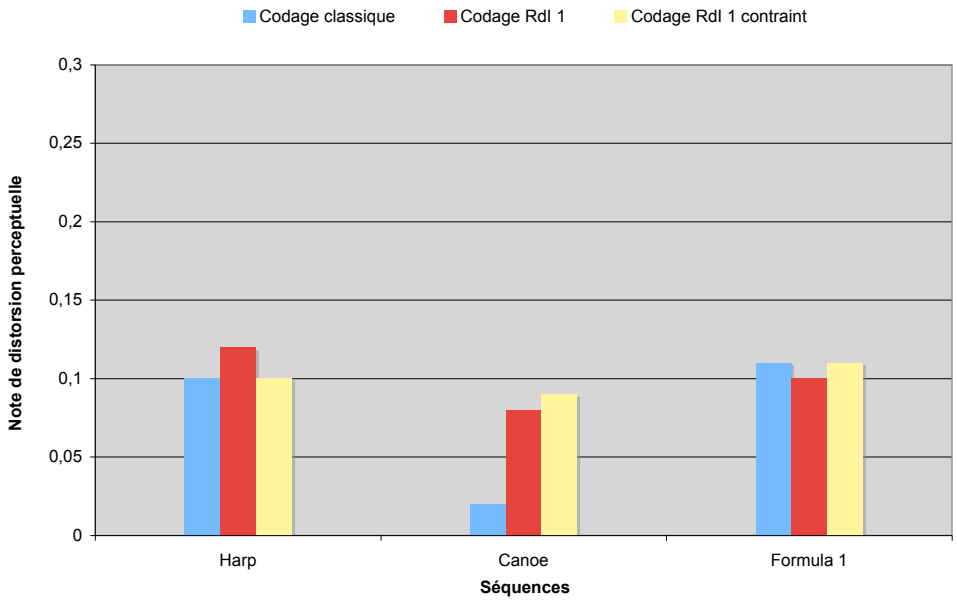


FIG. 7.14 – Les notes des séquences *Harp*, *Canoe* et *Formula 1* fournies par VQM pour les trois schémas de codage.

diminution est notée $(0,02)$ alors qu'une très faible augmentation des distorsions est notée pour la séquence *Formula 1* $(0,01)$. Ce constat nous amène à comparer la qualité subjective obtenue pour les deux schémas de codage en présence de pertes.

Pour évaluer visuellement l'amélioration de qualité due à l'utilisation du codage intra contraint, nous donnons figure 7.16 trois exemples d'images extraites de trois séquences différentes. Chaque séquence est codée une fois avec le codage RdI 1 et une autre fois avec le codage intra contraint. Le même motif de pertes, qui consiste à perdre trois slices consécutives d'une image I, est appliqué aux trois séquences. Nous remarquons que la qualité de la RdI, centrée autour des visages des personnes, est meilleure dans les images de (b) que celles de (a). L'absence de prédictions intra basées sur des macroblocs prédits en mode inter est vérifiée à l'aide de l'outil de suivi de propagation.

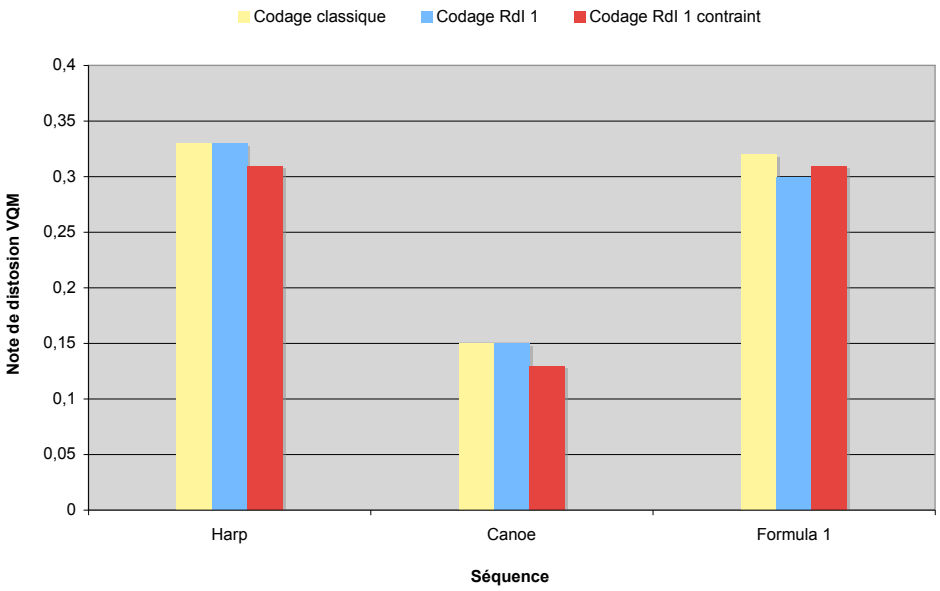


FIG. 7.15 – Les notes de distorsion des séquences *Harp*, *Canoe* et *Formula 1* pour les trois schémas de codage en présence de trois slices consécutives perdues.

7.4 Discussion générale

Nous discutons dans cette section de la complexité de notre approche en termes de temps nécessaire au codage de la vidéo et de la pertinence de la métrique de qualité utilisée (VQM). Nous présentons également, dans une perspective de comparaison, des travaux concomitants au nôtre.

7.4.1 Complexité de notre approche

L'approche que nous adoptons pour améliorer la robustesse de flux vidéos contre les pertes de paquets se décompose en deux étapes. Les régions d'intérêt de la vidéo sont tout d'abord identifiées puis la méthode de robustesse est appliquée aux RdI des images.

Dans ce travail, la réalisation de la première phase est complètement séparée de la phase de codage du fait qu'elle consiste en des tests de suivi des mouvements des yeux. Idéalement, la création des séquences de saillance doit être effectuée durant une phase de pré-traitement de la vidéo dans le codeur. Un modèle d'attention visuelle implanté dans le codeur augmentera évidemment la complexité du système et par suite le temps de codage de la vidéo.

Par ailleurs, l'algorithme restrictif n'altère pas radicalement le comportement du codeur. Il intervient juste au moment du choix du mode de prédiction pour imposer une prédiction intra au lieu d'une prédiction inter. La conséquence de cette modification du mode de codage

(a) Codage RdI 1.

(b) Codage robuste contraint.

FIG. 7.16 – Les images 203, 79 et 130 des séquences *Harp*, *Canoe* et *Formula 1* respectivement (de haut en bas) affectées par la perte de trois slices et codées avec (a) le codage robuste initalement proposé et (b) le codage robuste contraint.

est une diminution du temps de codage des macroblocs de la RdI. En effet, l'estimation de mouvement étant une des étapes de codage les plus longues, son remplacement par un codage intra réduit fortement le temps de codage. Cette réduction dépendra aussi de la taille de la RdI en macroblocs.

7.4.2 Choix de la métrique de qualité

Les notes de distorsion pour tous les motifs de pertes sont généralement petites ce qui implique que la qualité est au-dessus du seuil d'acceptabilité. Cependant, la qualité visuelle n'est pas aussi bonne que ne le prédit VQM. Les images de la séquence *Harp* données figure 7.8 illustrent cette différence entre la qualité objective et la qualité subjective. Il est aussi important de rappeler qu'un observateur est plus sensible aux dégradations quand elles touchent des régions contenant des formes humaines. Cet aspect n'est pas considéré par VQM.

De plus, VQM n'assigne pas un poids plus important à la RdI de l'image comme le ferait un observateur. Il traite toutes les régions de l'image d'une façon égale ce qui ne correspond pas vraiment à la réalité du jugement humain de qualité qui adopte une approche très hiérarchique. Une métrique de qualité qui prend en compte la RdI de l'image doit, en combinant les traits caractéristiques, attribuer un poids important aux dégradations qui ont lieu dans la région d'intérêt. L'exemple de la figure 7.11 montre par ailleurs que la préservation de la forme de la RdI est visuellement plus acceptable que sa déformation, même si un effet de blocs apparaît dans le premier cas.

7.4.3 Comparaison de notre méthode avec d'autres techniques d'amélioration de robustesse

En parallèle à nos travaux de recherche, d'autres approches ont été proposées (dans la même période d'étude) pour améliorer la robustesse des régions d'intérêt de la vidéo contre les pertes de paquets. Ces approches utilisent essentiellement l'organisation flexible de macroblocs (FMO) de la norme H.264/AVC pour assembler les macroblocs d'une image dans les slices selon des critères variés. Ces critères sont généralement l'impact de la perte de chaque macrobloc sur la qualité globale de la vidéo ou l'appartenance du macrobloc à la région fixée par l'observateur. Après l'identification de la correspondance entre chaque macrobloc et sa slice, l'outil FMO est utilisé pour effectuer l'encapsulation des macroblocs dans leurs slices correspondantes. Cette séparation au niveau du codage source permet par exemple l'application d'une protection inégale au niveau du codage canal [Ara06].

Approches non fondées sur l'attention visuelle

Dhondt *et al.* [Dho06] proposent un modèle de robustesse pour vidéos basé sur le calcul de la distorsion engendrée par la perte des macroblocs de l'image. Cette distorsion dépend de la

distorsion totale des pixels d'un macrobloc et du nombre de fois que chaque pixel est utilisé en tant que référence pour d'autres pixels dans la même image ou dans d'autres images. Les macroblocs sont alors groupés dans cinq slices par ordre d'importance, ceux engendrant la plus grande distorsion étant les plus importants. Les résultats de la simulation de taux de pertes de 5%, 10% et 20% montrent que l'application de l'approche proposée améliore la qualité de la vidéo codée, ici mesurée en PSNR, de 4 dB par rapport à l'utilisation du type 1 de FMO. Dans le même esprit, les techniques de slices redondantes et de FMO sont utilisées dans [Bac06] pour offrir une protection plus grande aux régions d'intérêt de l'image.

Dans [Che07b], une technique de robustesse pour un codage graduable basé sur les régions d'intérêt est proposée. Cette technique consiste à diviser le flux vidéo en deux couches distinctes ayant des priorités différentes. La première couche contient la région d'intérêt et la seconde couche, moins importante, est composée de l'arrière-plan de l'image. Cette séparation est effectuée à l'aide de l'outil FMO mais nécessite des changements au niveau des structures du codeur et du décodeur qui rendent ce dernier non conforme à la norme. Les dépendances entre les couches sont supprimées pour éviter que l'effet d'une perte ne se propage de la couche d'arrière-plan à la couche d'intérêt. Même si ce processus diminue l'efficacité de codage, il permet d'obtenir une qualité supérieure ($\Delta PSNR = 2$ dB) en présence d'erreurs de transmission par rapport à un codage classique.

L'intérêt majeur des approches proposées dans [Dho06, Bac06, Che07b] est la simplicité relative de leur mise en œuvre. Cependant, aucun aspect perceptuel n'est considéré lors de la détermination des régions d'intérêt de l'image ce qui éloigne ces approches de la réalité de l'attention visuelle humaine et par suite met en question leur efficacité. Les approches décrites dans la suite se basent sur l'attention visuelle pour identifier les régions de l'image à protéger.

Approches fondées sur l'attention visuelle

L'atténuation de l'impact de la propagation des dégradations peut être effectuée à l'aide de l'insertion périodique de macroblocs codés en intra, autorisée par la norme H.264/AVC. Dans [Ma09], un système complexe est proposé pour améliorer la robustesse du flux dans un environnement de transmission sans fil. Ce modèle couple une carte perceptuelle d'attention visuelle à l'état des pertes sur le canal pour guider l'insertion des macroblocs codés en intra. L'attention est modélisée sur la base de trois critères considérés comme les plus attractifs : la peau humaine, le mouvement et la région centrale de l'image. Un algorithme de détection des visages et des mains est utilisé pour vérifier si le macrobloc représente un tel contenu ou non. Un poids est assigné à chacun des critères et une carte perceptuelle est ainsi établie. Les macroblocs considérés comme appartenant à la région attirant l'attention de l'observateur sont codés en intra. Les performances de cette approche, toujours mesurées en PSNR, montrent une supériorité significative atteignant les 4 dB par rapport à l'insertion aléatoire de macroblocs

intra.

Chen *et al.* [Che07a] proposent aussi d'améliorer la robustesse des flux vidéos par insertion de macroblocs intra en se basant sur la modélisation de l'attention visuelle d'Itti [Itt98]. Des modes de prédictions avec restriction sont utilisés pour atténuer la propagation spatio-temporelle. Ainsi, une région d'intérêt dans une image codée en inter ne peut être prédite à partir d'une image non actualisée par des macroblocs intra. De même, le codage intra image est forcé d'utiliser des échantillons codés en mode intra uniquement ce qui évite l'utilisation d'échantillons potentiellement affectés par la propagation des pertes de paquets. La période d'actualisation est constante (4 images) pour un GOP de longueur 30. La région d'intérêt de l'image, dérivée à partir du modèle d'Itti, est fixée à 25% de la taille de l'image. À un taux de pertes de paquets de 20%, cette approche montre de meilleures performances ($\Delta PSNR = 7$ dB) que l'actualisation périodique de $\frac{1}{9}$ des macroblocs de chaque image.

Les deux travaux décrits dans cette sous-section traitent le problème de la propagation temporelle des dégradations dans les régions d'intérêt d'une façon similaire à notre approche. Le modèle d'attention visuelle utilisé dans [Che07a] est toutefois plus fiable que celui de [Ma09] car ce dernier n'est pas générique (limité à trois critères). Par ailleurs, l'évaluation de l'efficacité de ces approches est effectuée au travers d'une évaluation objective de qualité par le PSNR. D'après notre propre expérience tirée du travail présenté dans ce chapitre, l'évaluation de qualité n'est optimale que si elle considère la région d'intérêt de l'image. Ce constat va motiver l'utilisation des tests subjectifs dans le chapitre suivant.

Conclusion

Nous avons proposé dans ce chapitre une approche perceptuelle à l'amélioration de la robustesse des flux vidéos contre les pertes de paquets. Cette approche considère deux aspects des pertes de paquets : la position spatiale des dégradations dans l'image et la propagation spatio-temporelle engendrée. L'amélioration de la robustesse de la vidéo se fait ainsi par l'atténuation des dégradations dans les régions perceptuellement importantes de l'image.

Nous avons déterminé les régions d'intérêt de la vidéo en se basant sur les données oculométriques du test réalisé au chapitre précédent. Notre algorithme d'atténuation de la propagation spatio-temporelle des dégradations consistait à coder la RdI des images B et P en mode intra. Une première évaluation des performances a montré que l'algorithme proposé arrête la propagation temporelle dans la RdI mais n'est pas efficace contre la propagation spatiale des dégradations.

Nous avons alors modifié la méthode d'amélioration de la robustesse de façon à ce que le codage intra n'utilise pas des échantillons codés en mode inter et donc potentiellement affectés par les pertes de paquets. Bien que les deux méthodes ne génèrent pas un surcoût significatif aux

débits utilisés, la qualité subjective des vidéos est meilleure quand le codage intra contraint est utilisé. Nous avons aussi proposé un outil mettant en évidence les dépendances spatio-temporelles dans une vidéo.

L'étude menée dans ce chapitre était limitée à un nombre restreint de séquences vidéos et de motifs de pertes. De plus, l'évaluation de la qualité des vidéos par VQM ne prend pas en compte la présence d'une RdI dans la vidéo. Dans le chapitre suivant, nous évaluons l'efficacité d'une approche fondée sur la protection des RdI d'une manière plus extensive, au travers de tests subjectifs de qualité.

Par ailleurs, la séparation de la RdI du reste de l'image semble être primordiale pour complètement stopper la propagation spatiale de l'erreur. Cet aspect sera également considéré dans le chapitre suivant.

Chapitre 8

Extension de la méthode d'amélioration de robustesse des flux vidéos H.264/AVC

Introduction

Les limitations de la méthode d'amélioration de robustesse proposée dans le chapitre précédent sont essentiellement dues à la dépendance spatiale entre la région d'intérêt et son voisinage. Pour pallier ces limitations, nous adoptons une approche basée sur la séparation de la région d'intérêt du reste de l'image au niveau du codage. Cette séparation est effectuée par utilisation de l'organisation flexible de macroblocs (FMO) qui permet d'encapsuler la RdI dans une ou plusieurs slices indépendantes des autres slices de l'image.

Nous commençons ce chapitre par un rappel sur la technique de robustesse FMO déjà présentée au chapitre 2. Nous proposons ensuite deux méthodes pour améliorer la robustesse de la vidéo en protégeant sa RdI de la propagation des dégradations. Enfin, nous décrivons le test subjectif mis en place pour évaluer l'efficacité des méthodes proposées sur six séquences vidéos à contenus variés (quantité de mouvement, niveau de texture, *etc.*).

8.1 L'outil FMO au service des régions d'intérêt

Pour améliorer la robustesse des flux vidéos contre les pertes de paquets, nous avons proposé précédemment une méthode visant l'arrêt de la propagation temporelle des dégradations. Dans cette section, nous décrivons une extension de cette méthode utilisant l'outil FMO. L'apport de cette méthode est une séparation spatiale de la région d'intérêt du reste de l'image. Cette

séparation au niveau des slices de l'image confère à la RdI un cloisonnement spatio-temporel. De plus, nous proposons une variante de la méthode qui repose sur la création d'une dépendance entre les RdI pour éviter toute propagation des dégradations du reste de l'image à la RdI.

8.1.1 Rappel sur l'organisation flexible de macroblocs FMO

Les techniques de codage robuste offertes par la norme H.264/AVC servent à assurer la protection du flux vidéo contre d'éventuelles erreurs de transmission. L'outil FMO est généralement utilisé pour modifier l'encapsulation des macroblocs d'une image dans les slices de manière à ce que la perte d'une slice ne dégrade pas un groupe de macroblocs consécutifs. L'intérêt de cette approche est d'augmenter l'efficacité de la compensation des macroblocs perdus par l'utilisation de macroblocs voisins correctement reçus.

L'utilisation de l'outil FMO n'est autorisée que dans le profil de codage *Extended Profile* de la norme H.264/AVC. Ce profil diffère du *High Profile* utilisé jusqu'à présent sur deux points.

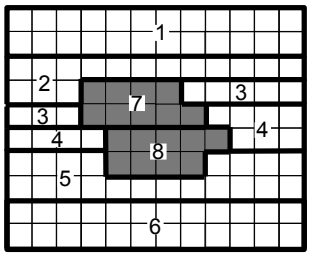
Premièrement, le profil *Extended* ne permet pas l'utilisation de la transformée entière 8×8. Cette transformation a été introduite dans la norme H.264/AVC, initialement basée sur une transformée 4×4, pour augmenter l'efficacité de la compression des régions uniformes dans les vidéos haute résolution (HD et plus). Pour les séquences étudiées dans ce chapitre, l'impact de l'utilisation exclusive de la transformée entière 4×4 sur la qualité visuelle est négligeable.

D'autre part, le profil *Extended* n'autorise pas l'utilisation du codage entropique CABAC. Comme nous l'avons déjà vu à la section 1.3 du chapitre 1, CABAC est plus performant que CAVLC surtout pour les hautes résolutions [Mar03, Maz09]. Cependant, les débits de codage élevés que nous utilisons laissent penser que la différence de performances ne sera pas significative au niveau de la qualité finale des vidéos. Le surcoût de la méthode proposée est étudié en détail dans la section 8.3.2.

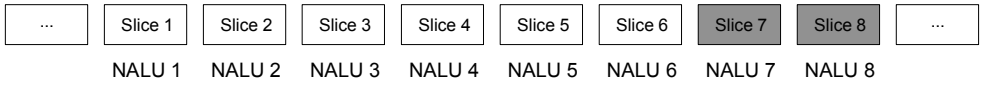
L'efficacité du codage est aussi réduite quand FMO est utilisé du fait du partitionnement de l'image en slices non consécutives. Ce partitionnement limite les possibilités de prédiction aux frontières des slices. De plus, cette remise à zéro du contexte de codage interrompt le codage entropique adapté au contexte. Le surcoût dû à l'utilisation de FMO dépend du type choisi (*cf.* sous-section 2.2.2.1 du chapitre 2).

8.1.2 Codage intra dans les régions d'intérêt

La figure 8.1.a illustre un exemple d'utilisation de FMO pour encapsuler les macroblocs de la RdI dans des slices séparées. Les macroblocs de couleur grise forment la RdI de l'image composée de deux slices dans cet exemple. La disposition des slices de l'image dans des NALU pour former le flux binaire est illustrée figure 8.1.b. Nous pouvons voir dans cet exemple que l'ordre des slices dans le flux binaire dépend du groupe de slices auquel elles appartiennent. Ainsi, les slices du premier groupe (numérotées de 1 à 6) sont ordonnées par ordre de codage



(a) Une image constituée de huit slices.



(b) Les slices de l'image disposées dans des NALU dans le flux binaire. Les slices 7 et 8 correspondent à la RdI.

FIG. 8.1 – La répartition spatiale des slices dans l'image (a) et leur disposition dans le flux vidéo (b). Les NALU en gris contiennent les slices de la RdI de l'image colorées en gris dans (a).

puis les slices du second groupe (correspondant à la RdI et numérotées de 7 à 8) sont ajoutées pour compléter la représentation binaire de l'image.

Comme nous l'avons montré dans le chapitre précédent, la propagation spatiale de l'erreur ne peut être arrêtée complètement à l'intérieur d'une slice. La présence, à l'intérieur d'une même slice, de macroblocs appartenant à la RdI et de macroblocs appartenant au reste de l'image pose en effet problème. La séparation de la RdI par utilisation de l'outil FMO permet, en plus de la protection contre la propagation spatiale des dégradations, l'application ultérieure d'un traitement particulier. Les slices contenant les macroblocs de la RdI peuvent par exemple être protégées d'une manière différente que les autres slices de l'image.

Parmi les sept types de FMO présentés dans le chapitre 2, nous choisissons d'utiliser le type 6. Ce type permet de spécifier explicitement à quelle slice appartient chaque macrobloc de l'image. Dans le cadre de ce travail, l'appartenance d'un macrobloc à une slice dépend de l'appartenance ou non de ce macrobloc à la RdI de l'image. Les RdI sont déterminées à partir des données recueillies lors des tests de suivi des mouvements oculaires. Au moment du codage, les coordonnées des macroblocs appartenant à la RdI sont fournies au codeur.

Deux groupes de slices sont ainsi créés : le premier contient les slices du reste de l'image et le second celles de la RdI. Pour chaque macrobloc de la RdI, l'algorithme décrit dans le chapitre 7 est utilisé pour forcer son mode de codage en intra.

Les cartes de correspondance entre les macroblocs et les slices d'une image sont formées au codage. Elles sont transmises dans des NALU de type *Picture Parameter Set* (PPS) pour

indiquer au décodeur la répartition des slices dans l'image. Généralement, une seule NALU PPS correspondant à un grand nombre d'images est incluse dans le flux binaire pour éviter les surcoûts. Cependant, dans notre travail, l'association d'une NALU PPS à chaque image est nécessaire car la RdI de l'image n'est pas fixe pour toute la durée de la séquence vidéo (elle change à chaque image). Pour les débits et la résolution utilisés (respectivement $4-6$ Mbit/s et 720×576), ces NALU engendrent généralement un surcoût de débit compris entre $0,5\%$ et 3% du débit requis par l'image associée.

8.1.3 Dépendance entre les régions d'intérêt

La propagation des dégradations du reste de l'image à la RdI peut nuire à la qualité visuelle d'une vidéo. Une solution à ce problème peut être la création d'une dépendance entre les macroblocs des RdI des images codées en mode inter et des RdI des images de référence. Nous proposons donc un algorithme qui opère de la façon suivante : pour les macroblocs des RdI des images B et P, le choix des macroblocs à utiliser comme référence pour le processus de prédiction inter est limité à ceux appartenant à la RdI de l'image de référence. L'outil FMO est toujours utilisé pour séparer les slices de la RdI des autres slices de l'image.

Pour les macroblocs utilisables comme référence, le codeur choisit le mode de codage qui minimise la fonction de coût comme expliqué au chapitre précédent. Ce mode peut être intra ou inter, auquel cas le macrobloc de référence appartient obligatoirement à la RdI. Si la RdI de l'image de référence n'est pas dans l'espace de recherche des prédictions inter, le mode intra est selectionné. Un exemple illustrant la restriction de prédiction inter est donné figure 8.2.

Sur cette figure, les macroblocs colorés en gris appartiennent aux RdI des images. Dans 8.2.a, les deux macroblocs gris utilisent comme référence des macroblocs n'appartenant pas à la RdI de l'image de référence. Ces prédictions sont modifiées dans 8.2.b de façon à ce que les deux macroblocs de référence utilisés appartiennent désormais à la RdI. La troisième prédiction inter illustrée reste inchangée car elle ne concerne pas un macrobloc de la RdI de l'image codée.

Cette approche permet d'éviter que la propagation des dégradations ne touche les RdI des images quand les pertes sont situées à l'extérieur de la RdI dans l'image de référence. Cependant, une perte de slices appartenant à la RdI engendre une propagation des dégradations essentiellement concentrée dans les RdI des images codées en mode inter. Ceci est dû à la forte dépendance créée par l'algorithme de restriction de prédiction. L'impact d'une telle perte pourrait être préjudiciable sur la qualité visuelle. Pour pallier cette faiblesse, il est nécessaire d'appliquer une protection inégale de l'information au niveau du codage canal pour mieux protéger les paquets contenant les slices relatives aux différentes RdI de l'image.

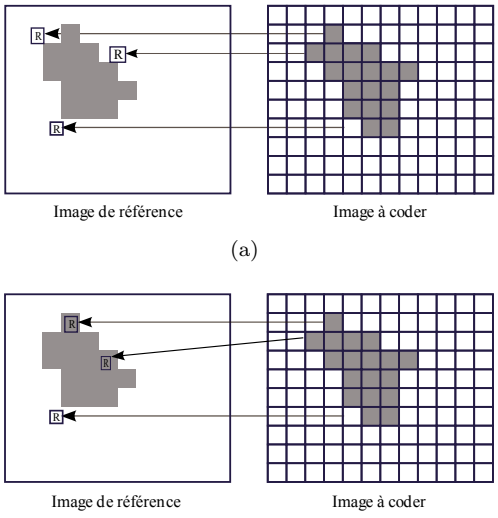


FIG. 8.2 – Illustration de la restriction de prédiction inter. Les macroblocs en gris correspondent aux RdI. Les prédictions à partir de macroblocs se situant à l'extérieur de la RdI de l'image de référence dans (a) sont modifiées après l'application de l'algorithme dans (b).

8.2 Tests subjectifs de qualité

Pour évaluer l'efficacité de l'algorithme proposé, il faut vérifier si la protection de la RdI de l'image aide à la préservation de la qualité globale de l'image. L'impact de la RdI étant le plus fidèlement évalué par des tests subjectifs, nous optons pour ce choix comme moyen d'évaluation de qualité.

La méthode d'évaluation par catégories absolues ACR est utilisée pour ce test. L'échelle catégorielle à cinq niveaux qui sert à noter la qualité des séquences vidéos est illustrée figure 8.3. Vingt-cinq observateurs ayant une vision normale ou parfaitement corrigée ont participé au test. Aucune de ces personnes ne travaille dans un domaine en relation avec le traitement d'image ou de vidéo. Le test est conduit selon la recommandation BT.500-11 de l'ITU [itu02].

8.2.1 Objectifs visés

Les objectifs du test subjectif de qualité sont les suivants :

– Quantifier l'importance perceptuelle de la RdI d'une image par rapport à la qualité globale de la vidéo. Dans la sous-section 5.4.3.6 du chapitre 5, nous avions déjà montré que la position spatiale de la perte dans l'image influence la dégradation de qualité.

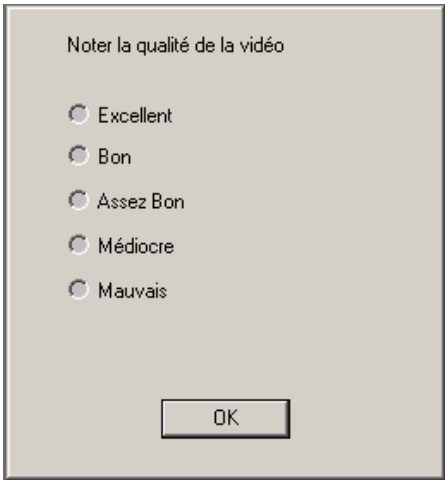


FIG. 8.3 – L'échelle de qualité à cinq catégories utilisée dans le test subjectif.

– Étudier le surcoût induit par les deux méthodes d'amélioration de robustesse proposées dans le cas où il n'y a aucune perte de paquets. Comme les séquences sont codées à débit constant, le surcoût se manifeste par une dégradation de qualité due au codage.
– Évaluer les performances des deux méthodes en présence de pertes de paquets et pour des motifs de pertes variés.

8.2.2 Codage des séquences de test

Les six séquences *Captain*, *Group disorder*, *Harp*, *Canoe*, *Formula 1* et *Rugby* à la résolution SD (720×576) sont incluses dans le test subjectif de qualité. Elles sont codées à débit fixe quel que soit le schéma de codage utilisé. Les débits sont choisis de façon à garantir une bonne qualité pour les séquences codées. Le débit de codage est de 4 Mbit/s pour les séquences *Captain* et *Harp* et de 6 Mbit/s pour *Group disorder*, *Canoe*, *Formula 1* et *Rugby*. Le codage se fait à l'aide du codeur JM avec une longueur de GOP de 24 images, une structure IDRBBP... et le profil de codage *Extended Profile*. Le codage entropique CAVLC est utilisé. Le nombre maximal d'images de référence est fixé à 5 et la taille de la fenêtre de recherche à 32. La taille d'une slice est fixée à 1450 octets. Le décodeur JM est utilisé avec ses algorithmes de compensation de pertes.

Les schémas de codage testés sont donnés dans le tableau 8.1. Le *codage classique* consiste à coder la séquence en *High Profile* avec CABAC sans application d'aucun outil de robustesse.

Le *codage RdI* sépare la RdI du reste de l'image à l'aide de l'outil FMO sans appliquer un traitement particulier pour améliorer la robustesse du flux. Ce schéma de codage est inclus dans le test essentiellement pour évaluer le surcoût de l'utilisation de FMO.

Le *codage RdI intra* est similaire au *codage RdI* sauf qu'il force le codage de la RdI en mode

intra arrêtant toute propagation spatio-temporelle des dégradations dans la RdI.

Le *codage RdI inter* crée des dépendances entre les RdI comme décrit à la sous-section 8.1.3 pour éviter la propagation des dégradations du reste de l'image à la RdI.

TAB. 8.1 – Les schémas de codage évalués.

Schéma de codage	FMO	Codage robuste	Type de codage robuste
Codage classique	Non	Non	-
Codage RdI	Oui	Non	-
Codage RdI intra	Oui	Oui	Codage intra de la RdI
Codage RdI inter	Oui	Oui	Inter-dépendance des RdI

8.2.3 Motifs de pertes utilisés

Les motifs de pertes sont choisis de manière à obtenir une propagation temporelle de l'effet de perte dans la RdI et dans les autres régions de l'image. Le tableau 8.2 résume les motifs de pertes utilisés. Les pertes sont toutes localisées dans les images I pour obtenir une propagation maximale des dégradations.

TAB. 8.2 – Les motifs de pertes utilisés.

Motif	Nombre d'images concernées par les pertes	Position temporelle	Position spatiale	Amplitude (slices)
RdI	1	GOP 5 ou 6	RdI	$4-11$ (dépend du contenu)
autourRdI	1	GOP 5 ou 6	Autour de la RdI	8
multipleRdI	3	GOP 4, 5 et 6	RdI	$11-32$ (dépend du contenu)
multipleAutourRdI	3	GOP 4, 5 et 6	Autour de la RdI	24

Le motif noté *RdI* représente la perte de toutes les slices appartenant à la RdI de la cinquième ou la sixième image I de la séquence vidéo. Les durées des vidéos étant de huit ou dix secondes, nous évitons ainsi que la perte ait lieu au début de la séquence ou à sa fin.

Le tableau 8.3 résume les tailles des RdI exprimées en nombre de slices ainsi que le pourcentage qu'elles représentent du nombre total de slices dans l'image I. Comme le nombre de slices dans l'image dépend du contenu et du débit utilisé, la séquence *Group disorder* contient plus de slices que la séquence *Captain*. Ceci se traduit par une différence minime de pourcentage pour un nombre triple de slices dans la RdI.

Le motif *autourRdI* consiste à perdre huit slices situées autour de la RdI. Ce nombre est choisi tel qu'il soit compris dans l'intervalle des tailles de RdI en slices données dans le tableau 8.3.

Pour le schéma *codage RdI intra*, nous utilisons deux motifs de pertes supplémentaires pour tester l'efficacité de l'approche en présence de pertes régulières. Ces motifs consistent à perdre la RdI ou la région autour de la RdI dans trois images I successives (la quatrième, la cinquième et la sixième). Ils sont respectivement notés *multipleRdI* et *multipleAutourRdI*. Les deux HRC *multiple* permettent d'avoir des scénarios de propagation temporelle des pertes pour une durée pouvant aller jusqu'à trois secondes qui est la durée approximative de trois GOP.

TAB. 8.3 – La taille en slices des RdI des images dans lesquelles ont lieu les pertes pour les six séquences étudiées.

Séquence	Captain	Group disorder	Harp	Canoe	Formula 1	Rugby
Nombre de slices RdI	4	11	7	5	9	8
Pourcentage du nombre total	$12,1\%$	$15,1\%$	$9,7\%$	$8,3\%$	$9,6\%$	12%

Le choix des motifs de pertes est induit par la nature des pertes de paquets observées sur des réseaux de paquets. Ces pertes sont généralement groupées ou en d'autres termes ont lieu "en rafales". Ceci est essentiellement dû aux congestions survenant au niveau des routeurs qui impliquent la perte d'un nombre de paquets consécutifs dans la file d'attente. Les motifs choisis pour ce travail se rapprochent des motifs 9 et 10 du chapitre 5 (*cf.* tableau 5.2) qui correspondent respectivement à la perte de quatre et dix paquets pour un codage à 4 Mbit/s. L'impact des pertes en rafales est l'élimination de NALU consécutives au niveau du flux binaire qui ne correspondent pas nécessairement à des slices adjacentes au niveau de l'image. Par exemple, la perte des NALU 2, 3 et 4 sur la figure 8.1 équivaut à la perte de la région spatiale située autour de la RdI de l'image. De même, la perte des NALU 7 et 8 implique la perte de la totalité de la RdI. La figure 8.4 illustre des exemples de dégradations ayant lieu dans la RdI de l'image (8.4.b) et autour de cette région (8.4.c).

Nombre total de PVS

Le tableau 8.4 résume les HRC du test. Les HRC 1 à 10 sont appliqués à chacun des six contenus obtenant ainsi $10 \times 6 = 60$ séquences dégradées. En intégrant les séquences codées avec les quatre schémas de codage et n'ayant pas subi de pertes de paquets (HRC 0) à l'ensemble des vidéos, le nombre total de PVS devient $60 + 24 = 84$. Ces séquences sont présentées aux observateurs dans un ordre aléatoire.

(a) L'image 73 sans aucune slice perdue.

(b) L'image 73 avec sa RdI perdue.

(c) L'image 73 avec 8 slices autour de la RdI perdues.

FIG. 8.4 – L'image 73 de la séquence *Rugby* (a) codée et n'ayant pas subie des pertes de slices ; (b) ayant perdue la totalité des slices de sa RdI et (c) ayant perdue 8 slices adjacentes à sa RdI.

TAB. 8.4 – Les HRC du test.

HRC	Codage	Motif de pertes
0	Tous les codages	-
1	*Codage classique*	*RdI*
2		*autourRdI*
3	*Codage RdI*	*RdI*
4		*autourRdI*
5	*Codage RdI inter*	*RdI*
6		*autourRdI*
7	*Codage RdI intra*	*RdI*
8		*autourRdI*
9		*multipleRdI*
10		*multipleAutourRdI*

8.3 Résultats expérimentaux

8.3.1 Importance perceptuelle de la région d'intérêt

Pour quantifier l'impact perceptuel de la position spatiale de la perte dans l'image, nous comparons la qualité d'une séquence affectée par le motif de pertes *RdI* à la qualité d'une séquence (représentant le même contenu) ayant subie des pertes de type *autourRdI*. Nous proposons de définir le rapport IE (Intérieur-Extérieur) entre deux séquences ayant le même contenu comme suit :

$$IE = \frac{nombre\ de\ slices\ perdues\ dans\ la\ RdI\ de\ la\ premi\grave{e}re\ s\acute{e}quence}{nombre\ de\ slices\ perdues\ autour\ de\ la\ RdI\ de\ la\ seconde\ s\acute{e}quence} \tag{8.1}$$

Une valeur de IE inférieure à 1 indique qu'il y a plus de pertes autour de la RdI que dans la RdI. Au contraire, un rapport IE supérieur à 1 indique que le nombre de slices perdues autour de la RdI est inférieur au nombre de pertes dans la RdI. Par exemple, un IE égal à $0,5$ signifie que les pertes autour de la RdI sont deux fois plus importantes que les pertes dans la RdI.

8.3.1.1 Étude du cas $IE < 1$

La figure 8.5 donne les MOS de trois séquences, pour lesquelles la valeur de IE est inférieure à 1, affectées par les deux motifs de pertes *autourRdI* et *RdI*. Les séquences sont codées avec un codage *RdI* permettant de contrôler finement la position spatiale de la perte par rapport à la RdI de l'image. Les valeurs de IE sont respectivement $0,5$; $0,62$ et $0,88$ pour les séquences *Captain*, *Canoe* et *Harp*. Par exemple, dans le cas de la séquence *Captain*, le nombre de slices perdues à l'extérieur de la RdI représente le double des pertes dans la RdI.

Nous remarquons que la qualité des trois séquences est meilleure (à différents degrés) pour le motif de pertes *autourRdI* que pour *RdI*. Dans le cas de la séquence *Captain*, la perte de quatre slices dans la RdI dégrade la qualité de $0,3$ unité MOS par rapport à la perte de huit slices autour de la RdI. Cette différence de qualité augmente pour les séquences *Canoe* et *Harp* (respectivement $0,45$ et $0,5$) qui ont des IE plus grands et donc plus de pertes dans la RdI.

Contrairement aux observations de la sous-section 5.4.3.6 du chapitre 5, les différences de qualité dues au changement de la position spatiale des pertes ne sont généralement pas très grandes dans les exemples présentés ici. Ceci est dû au fait que les positions spatiales étudiées (la RdI et la région autour de la RdI) sont adjacentes et ne sont donc pas assez distinctes. Selon l'amplitude de la perte, les motifs de pertes *RdI* et *autourRdI* peuvent être jugées de la même façon par les observateurs. La taille de la RdI est aussi un facteur qui contribue à son importance perceptuelle. Cet aspect est discuté à la section 8.4.

8.3.1.2 Étude du cas $IE \geq 1$

Pour les trois séquences ayant une valeur de IE supérieure ou égale à 1, il est normal que leur qualité soit meilleure quand la perte a lieu autour de la RdI plutôt qu'à l'intérieur de la RdI. Sur la figure 8.6, les deux séquences *Group disorder* et *Formula 1* confirment cette tendance avec des ΔMOS respectifs de $0,5$ et $0,85$. Cependant, la séquence *Rugby* a une qualité presque égale pour les deux motifs *RdI* et *autourRdI* alors que le nombre de slices perdues est le même. Ceci peut être dû à la rapidité du mouvement du ballon dans *Rugby* qui implique une variation rapide de la RdI entre deux images consécutives. Par suite, la RdI va se déplacer vers la région

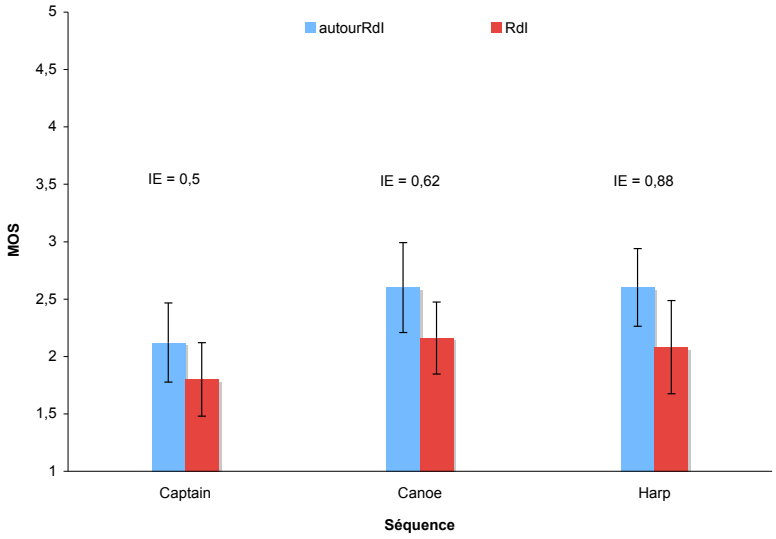


FIG. 8.5 – Les valeurs des MOS de trois séquences affectées par les motifs de pertes *RdI* et *autourRdI* et ayant des rapports IE inférieurs à 1. IE représente le rapport du nombre de slices perdues dans la RdI sur le nombre de slices perdues à l'extérieur de la RdI.

contenant les dégradations correspondant au motif *autourRdI* ce qui contribue à la diminution de qualité pour ce motif.

8.3.2 Surcoût du codage *RdI intra* en l'absence de pertes

Avant d'évaluer l'efficacité des différentes méthode d'amélioration de robustesse, nous souhaitons quantifier le surcoût introduit par d'une part l'utilisation de l'outil FMO et d'autre part le *codage RdI intra*. Ce surcoût se manifeste par une dégradation de qualité à débit de codage constant par rapport à un codage classique.

8.3.2.1 Dégradation de qualité due à l'utilisation de FMO

Nous donnons figure 8.7 les notes de qualité des six séquences codées à l'aide des trois schémas de codage : *codage classique*, *codage RdI* et *codage RdI intra*. Pour chacune des séquences, le diagramme de gauche correspond au cas sans pertes de paquets (HRC 0).

Nous remarquons sur les six diagrammes que la qualité en absence de pertes est sensiblement la même pour les schémas de codage *codage RdI* et *codage classique*. L'utilisation de FMO n'est donc pas coûteuse en termes de qualité. L'explication est la suivante : le type 6 est utilisé dans

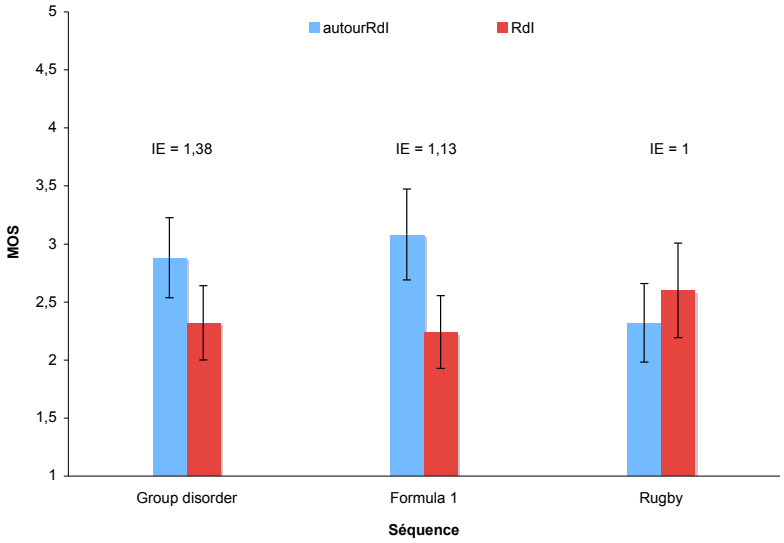


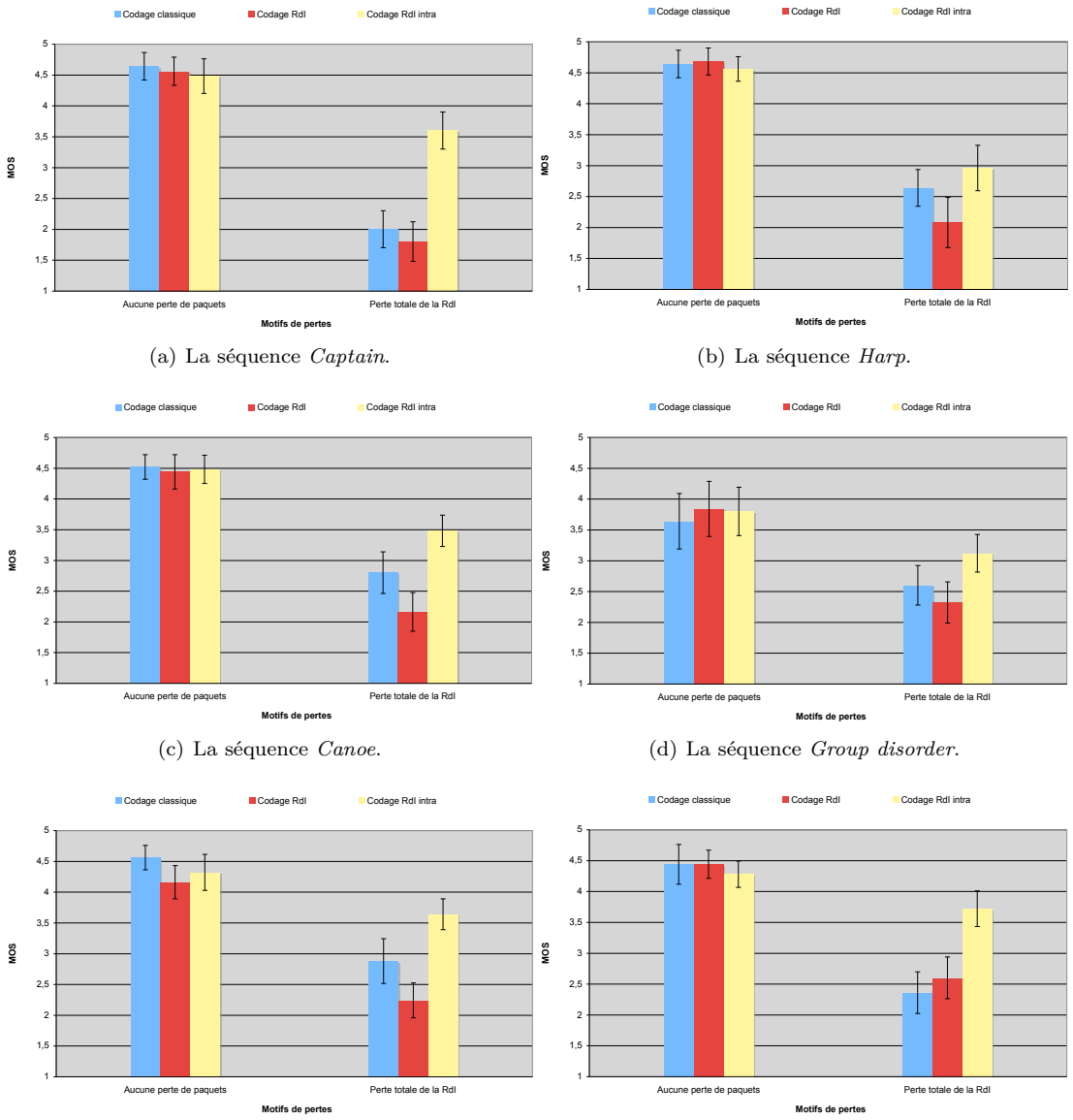
FIG. 8.6 – Les valeurs des MOS de trois séquences affectées par les motifs de pertes *RdI* et *autourRdI* et ayant des rapports IE supérieurs ou égaux à 1.

ce travail pour assembler les macroblocs de la région d'intérêt dans un groupe de slices séparé du groupe contenant les slices du reste de l'image. Ceci ne créé pas un effet de dispersion des macroblocs qui peut diminuer l'efficacité de codage. Comme la majorité des macroblocs adjacents appartiennent à la même slice, le codage intra peut tester tous les types de prédiction possibles pour choisir le meilleur candidat.

8.3.2.2 Dégradation de qualité due au codage *RdI intra*

La diminution de qualité est due à l'utilisation intensive des modes de codage intra dans les RdI. Nous notons sur la figure 8.7 que les MOS des séquences sont presque égales pour les schémas de codage *codage RdI intra* et *codage classique*. De plus, la qualité s'approche de $4,5$ pour toutes les séquences sauf *Group disorder* qui affiche une qualité initiale légèrement supérieure à $3,5$. Cette dernière valeur est expliquée par la grande quantité de mouvement contenue dans la séquence *Group disorder* qui rend sa qualité moyenne au débit choisi.

La faible diminution de qualité observée pour le HRC 0 est due aux débits et aux résolutions utilisées. En effet, comme nous voulions évaluer l'efficacité de notre méthode en présence de pertes, nous avons limité les dégradations dues au codage en utilisant des débits élevés adaptés à la résolution SD. Les dégradations observées sont essentiellement dues aux motifs de pertes et donc aux erreurs de transmission.



(a) La séquence *Captain*.

(b) La séquence *Harp*.

(c) La séquence *Canoe*.

(d) La séquence *Group disorder*.

(e) La séquence *Formula 1*.

(f) La séquence *Rugby*.

FIG. 8.7 – Les valeurs des MOS des six séquences codées par les trois schémas de codage. Dans chaque graphe, les diagrammes de gauche et de droite correspondent respectivement au cas sans pertes et au motif de pertes *RdI*.

8.3.3 Évaluation des performances des méthodes proposées

8.3.3.1 Performances de *RdI intra* par contenu

Les six diagrammes de droite sur la figure 8.7 donnent les notes de qualité des vidéos codées selon les trois schémas de codage en présence d'une perte dans la RdI de l'image. Nous remarquons d'abord que le schéma de codage robuste *codage RdI intra* arrive à maintenir un niveau de qualité moyen ($MOS \simeq 3$) pour les séquences *Harp* et *Group disorder*. Pour les quatre autres séquences (*Group disorder*, *Canoe*, *Formula 1* et *Rugby*), la qualité est au moins égale à $3,5$ ce qui montre que notre méthode a de bonnes performances indépendamment du contenu de la séquence vidéo. La diminution de qualité notée pour la séquence *Group disorder* est due au grand nombre de slices perdues (11 selon le tableau 8.3). Dans le cas de la séquence *Harp*, les dégradations impliquent le visage d'une personne ce qui créé une gêne importante pour l'observateur.

Nous comparons ces résultats, pour le même motif de pertes, aux notes de qualité des séquences codées en *codage classique* et *codage RdI*. Pour ces deux schémas de codage, la qualité est pour toutes les séquences en-deça du seuil d'acceptabilité qui correspond à un MOS égal à 3.

8.3.3.2 Performances de *RdI intra* sur tous les contenus

Dans le but d'obtenir une valeur moyenne de différence entre ces schémas de codage et le *codage RdI intra*, nous présentons sur la figure 8.8 les notes moyennes de qualité associées à chaque schéma de codage. La note moyenne de qualité d'un schéma de codage est calculée sur tous les observateurs et sur tous les contenus. Dans notre cas, nous avons 25 observateurs ayant noté chacun 6 séquences (contenus) codées par le même schéma de codage ce qui nous permet d'avoir 150 notes de qualité pour chaque algorithme.

Les schémas *codage classique*, *codage RdI* et *codage RdI intra* ont respectivement des MOS de $2,6$; $2,2$ et $3,4$. Cette différence entre d'une part les deux premiers schémas et d'autre part le troisième schéma est largement due à l'arrêt de la propagation temporelle des dégradations dans la RdI offerte par ce dernier schéma. La propagation est stoppée suite au codage en mode intra des macroblocs de la RdI et à la dépendance de ces macroblocs vis-à-vis des autres macroblocs de l'image. Cependant, la propagation temporelle des dégradations à l'extérieur de la RdI n'est pas exclue. En effet, des macroblocs appartenant au voisinage direct de la RdI peuvent être prédits en mode inter à partir de macroblocs dans la RdI. Il apparaît à travers ce test subjectif que ce facteur ne joue pas un rôle prédominant dans l'influence du jugement de qualité de l'observateur.

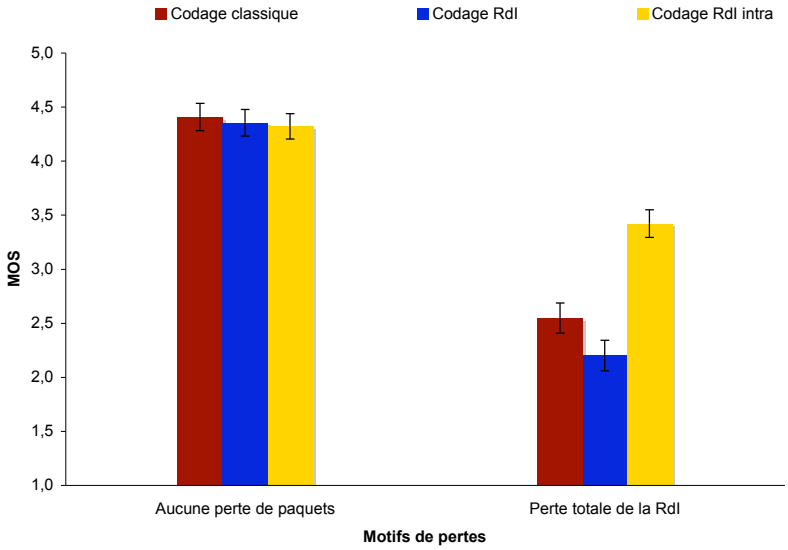


FIG. 8.8 – La qualité moyenne des séquences pour chaque schéma de codage et pour HRC 0 et le motif *RdI*.

TAB. 8.5 – Les notes de qualité des schémas *codage classique* et *codage RdI intra* pour la perte de la RdI.

		Motif de pertes	
		RdI	*multipleRdI*
Codage	*Codage classique*	2,5	N/A
	Codage RdI intra	3,4	2,5

Performances de *RdI intra* pour le motif *multipleRdI*

Par ailleurs, nous évaluons la robustesse de la méthode proposée en présence de plusieurs pertes affectant les RdI de plusieurs images. Pour le motif de pertes *multipleRdI*, la note moyenne de qualité des six séquences est égale à 2,5 pour le schéma *codage RdI intra* (*cf.* tableau 8.5). Cette valeur indique d'une part que la perte de la RdI dans trois images I consécutives réduit fortement la qualité même si un codage robuste est utilisé. D'autre part, cette note est égale à la note moyenne de qualité du schéma *codage classique* pour la perte de la RdI d'une seule image (motif *RdI*). Une séquence codée avec le schéma *codage RdI intra* tolère donc trois fois plus de pertes dans la RdI qu'une séquence codée sans application d'aucun outil de robustesse.

(a) Image 97 sans aucune perte de paquets.

(b) Image 97 avec perte totale de sa RdI.

(c) Image 98 selon le schéma *codage RdI*.

(d) Image 98 selon le schéma *codage RdI intra*.

FIG. 8.9 – Les images 97 (haut) et 98 (bas) de la séquence *Captain* affectée par la perte de sa RdI.

Exemple visuel montrant l'efficacité du *codage RdI intra*

Sur la figure 8.9, nous donnons un exemple visuel de la propagation temporelle des dégradations et de son cloisonnement. Les figures 8.9.a et 8.9.b illustrent respectivement l'image 97 de la séquence *Captain* sequence sans et avec perte de sa RdI. Cette image étant une image I, les dégradations doivent se propager aux images suivantes. Ceci est en effet le cas sur la figure 8.9.c qui illustre l'image 98 (de type B) de la séquence codée avec le *codage RdI*. Nous pouvons voir sur cette image une sévère dégradation de la RdI. Pour l'image 98 de la séquence codée avec le *codage RdI intra*, les dégradations sont fortement réduites (figure 8.9.d).

8.3.3.3 Performances du schéma *codage RdI*

Il est intéressant aussi de commenter la différence de qualité observée sur la figure 8.8 entre les schémas *codage classique* et *codage RdI*. L'application de l'outil FMO sans aucune intervention

au niveau des prédictions semble dégrader la robustesse du flux contre les pertes de paquets. L'explication est la suivante : l'organisation flexible de macroblocs a été conçue pour disperser les dégradations dues à la perte d'une ou de plusieurs slices. En assemblant les macroblocs de la RdI dans des slices qui sont éventuellement perdues, les dégradations sont concentrées dans la région fixée par l'observateur ce qui conduit à une gêne importante. Dans le cas du codage classique, toute la RdI ne peut être perdue de manière précise du fait que les macroblocs appartenant à la RdI ne sont pas séparés des autres macroblocs de l'image.

8.3.3.4 Performances de *RdI intra* en présence de pertes autour de la RdI

Le schéma *codage RdI intra* agit exclusivement dans la RdI de l'image. La préservation de la bonne qualité de la RdI doit toutefois se refléter sur la qualité visuelle globale.

Pertes autour de la RdI dans une seule image

La figure 8.10 donne les notes moyennes de qualité des six séquences obtenues pour les motifs de pertes *autourRdI* et *multipleAutourRdI*. Comme nous pouvons le constater sur la partie gauche de la figure, la méthode de robustesse n'apporte pas d'amélioration à la qualité visuelle par rapport aux deux autres schémas de codage. Ceci est probablement dû à la taille de la RdI qui ne représente pas une partie assez importante de l'image. Par suite, les pertes observées autour et dans la RdI s'assimilent et l'observateur est également gêné par les deux motifs de pertes.

Pertes autour de la RdI dans trois images I consécutives

Le motif *multipleAutourRdI* n'a été appliqué que sur les séquences codées avec le schéma *codage RdI intra*. Le diagramme illustré sur la partie droite de la figure 8.10 confirme l'inefficacité de ce schéma de codage face à trois pertes régulières qui touchent les régions entourant la RdI dans trois images I ($MOS = 1,7$). D'autre part, ce résultat met en évidence encore une fois l'importance de l'effet de gêne créé par une propagation prolongée des dégradations. Dans ce cas, la perte de huit slices dans trois images I consécutives dégrade fortement la qualité et gêne l'observateur même si la RdI reste intacte.

8.3.3.5 Performances du *codage RdI inter*

Les performances du schéma de codage *codage RdI inter* ne sont pas satisfaisantes en présence de pertes comme le montre la figure 8.11 ($MOS \leq 2,5$). Sur cette figure, les notes de qualité moyennes de ce schéma sont comparées à celles des schémas *codage RdI* et *codage RdI intra.* Nous remarquons que les performances des schémas *codage RdI* et *codage RdI inter* sont très

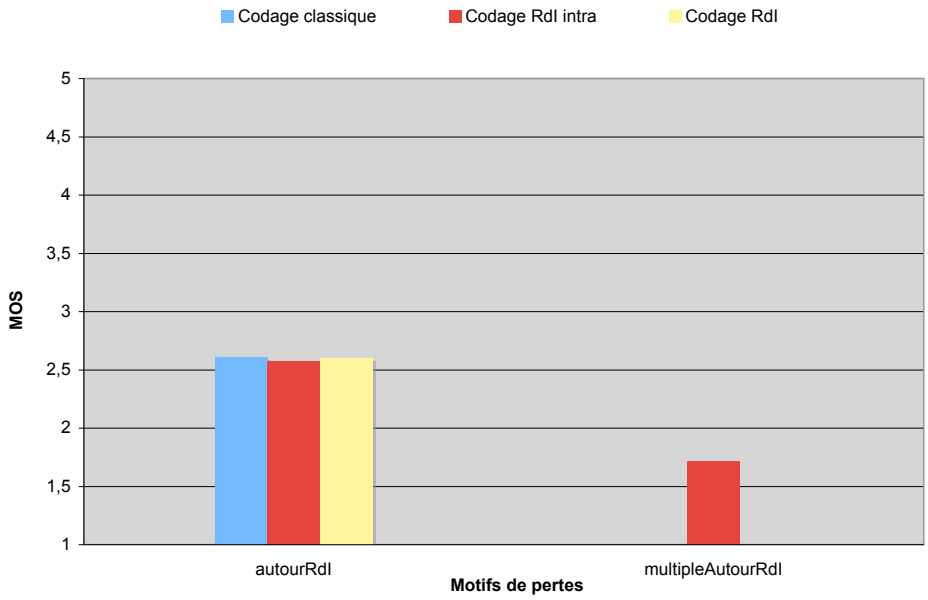


FIG. 8.10 – La qualité moyenne des séquences pour chaque schéma de codage et pour les motifs *autourRdI* et *multipleAutourRdI*.

proches en présence ou non de pertes de paquets. Ce résultat est particulièrement surprenant dans le cas du motif *autourRdI* pour lequel le *codage RdI inter* était censé protéger la RdI des propagations de dégradations.

Une explication de la similarité entre les performances des deux schémas de codage est la suivante. La position spatiale de la RdI ne change pas catégoriquement entre deux images consécutives. Par suite, les macroblocs de la RdI codés en mode inter utilisent des macroblocs de la RdI comme référence de prédiction pour un schéma *codage RdI*. Le fait de restreindre les prédictions inter n'influence alors plus que quelques macroblocs d'où le comportement similaire des deux schémas de codage.

8.4 Discussion et perspectives

Les résultats présentés dans ce chapitre montrent que notre approche a de bonnes performances quand les pertes affectent la RdI de l'image. Nous discutons dans ce qui suit des limites de ce travail et de ses possibles extensions.

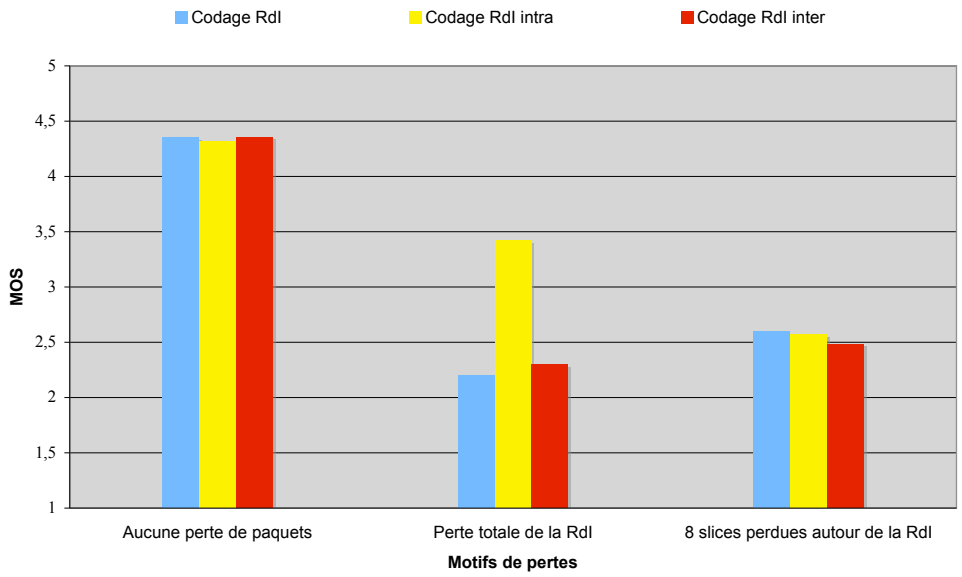


FIG. 8.11 – Comparaison des résultats du schéma *codage RdI inter* avec les schémas *codage RdI* et *codage RdI intra* pour HRC 0 et les motifs de pertes *RdI* et *autourRdI*.

Taille de la RdI

D'après les résultats présentés dans ce chapitre, il est clair que la taille de la RdI est un paramètre important à considérer dans l'application de notre méthode de robustesse. Nous donnons dans le tableau 8.6 les tailles des RdI des images touchées par les pertes pour les six séquences étudiées. Ce pourcentage est calculé à partir du rapport du nombre de macroblocs dans la RdI sur le nombre de macroblocs dans l'image. Pour la résolution 720×576, nous avons dans une image 1620 macroblocs de taille 16×16 pixels chacun. Les pourcentages donnés dans ce tableau sont cohérents avec ceux du tableau 8.3. Ce constat était attendu du fait que les slices ont des tailles approximativement égales.

Les valeurs rapportées dans le tableau 8.6 montrent qu'en général la RdI n'occupe pas plus de 15% du plan de l'image avec des régions représentant 8% seulement de l'image. Les pourcentages sont exprimés par rapport au nombre total de macroblocs dans l'image. Ces petites tailles de RdI jouent évidemment un rôle dans la diminution du surcoût de codage car elles impliquent le codage en intra d'un nombre inférieur de macroblocs. Cependant, si la RdI ne couvre pas la totalité d'un objet, elle peut perdre son sens et ne plus représenter un support adéquat pour l'application de notre méthode d'amélioration de robustesse.

Au niveau du codage source, la qualité initiale de la vidéo est d'autant plus mauvaise que la RdI est grande. Une adaptation conjointe de la taille de la RdI à la protection canal à ajouter est une solution envisageable.

Codage conjoint source-canal

L'application du schéma *codage RdI intra* contribue à l'augmentation de la robustesse de la séquence vidéo. Ce codage robuste impose aussi une égalité entre les différents paquets de l'image adaptant ainsi le flux vidéo aux réseaux *best effort*. Plus précisément, les dégradations résultant de pertes dans la RdI ne créent plus une gêne supérieure à celle due à la perte de n'importe quel paquet de l'image. La perte de paquets contenant des slices de la RdI est ainsi tolérable car elle est compensée par la restriction de prédiction.

Au niveau du codage canal, il n'est plus nécessaire d'appliquer plus de protection aux paquets de la RdI. Au contraire, nous pouvons nous permettre de protéger moins bien ces paquets au profit d'une augmentation de la protection allouée aux autres paquets de l'image. Ainsi, dans une image I, les paquets qui n'appartiennent pas à la RdI sont protégés plus que les paquets de la RdI car leur perte engendre une propagation temporelle des dégradations. En effet, comme les RdI des images B et P sont codées en mode intra, l'impact de la perte de la RdI d'une image I ne se répercute qu'à l'extérieur des RdI (et faiblement) dans les images codées en mode inter. Un critère de répartition du débit alloué au codage canal pourrait être la position de la slice par rapport à la RdI où les slices reçoivent plus de protection à mesure qu'ils sont plus près de la RdI.

Un exemple d'un tel schéma de codage conjoint source-canal est illustré sur la figure 8.12. Les slices adjacentes à la RdI grise sont colorées en gris clair. Les slices 1 et 6 sont les slices les plus loin de la RdI et par suite sont moins protégées que les slices 2, 3, 4 et 5. Ceci est traduit par des taux de redondance différents des codes correcteurs appliqués. Les slices 7 et 8 de la RdI ne sont pas protégées du tout car leur perte n'influence pas les RdI des images B et P.

Réalisme des motifs de pertes

Les motifs de pertes choisis s'approchent de la réalité des pertes observées sur un réseau de paquets. La caractéristique des pertes en rafale est particulièrement mise en avant dans ce travail. Cependant, il reste à voir si les pertes arrivent assez souvent dans la RdI pour justifier la mise en œuvre d'une telle solution. La réponse à cette question dépend du canal de transmission,

TAB. 8.6 – La taille des RdI des images dans lesquelles ont lieu les pertes pour les six séquences étudiées.

Séquence	Captain	Group disorder	Harp	Canoe	Formula 1	Rugby
Nombre de macroblocs RdI	181	249	159	137	158	197
Pourcentage du nombre total	11,2%	15,3%	9,8%	8,4%	9,8%	12,2%

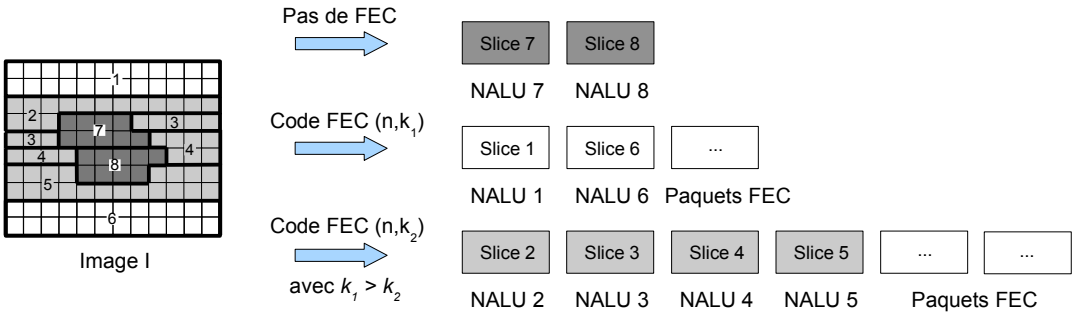


FIG. 8.12 – Codage conjoint source-canal prenant en compte la position spatiale des slices dans l'image. Les slices 7 et 8 ne sont pas protégées au niveau du canal (codage source robuste) tandis que les slices numérotées de 2 à 5 sont fortement protégées (pas de codage source robuste).

de la taille de la RdI et du codage canal. En supposant que le canal assurant le transport de la vidéo est imposé au préalable, les paramètres que nous pouvons modifier sont les deux derniers.

Si une approche de protection inégale comme celle proposée sur la figure 8.12 est adoptée au niveau du canal, un motif de pertes n'a pas la même influence sur la qualité visuelle s'il touche les paquets de la RdI ou ceux des autres régions de l'image. Ainsi, il est plus probable que la RdI soit touchée quand des paquets sources (non redondants) sont perdus. L'augmentation de la taille de la RdI pourrait être aussi un moyen augmentant sa vulnérabilité aux pertes du fait qu'elle occupera un nombre plus grand de slices et par suite de paquets.

Conclusion

Nous avons présenté dans ce chapitre une extension de notre méthode d'amélioration de la robustesse des vidéos contre les pertes de paquets. Cette extension consiste principalement à séparer la RdI des autres régions de l'image en utilisant l'outil FMO de H.264/AVC. L'étude de performances, réalisée au travers d'un test subjectif de qualité, montre que cette approche est efficace contre la propagation spatio-temporelle des dégradations quand elle est couplée à un codage intra dans la RdI. En effet, une amélioration d'une unité MOS est observée quand cette méthode est utilisée par rapport aux schémas de codage classiques. De plus, la dégradation de qualité engendrée a un impact négligeable sur la qualité subjective des vidéos. Si une diminution de qualité est toutefois amenée à apparaître, elle peut être compensée par une diminution de la redondance allouée aux paquets de la RdI lors du codage canal.

Conclusion générale et perspectives

Les travaux de recherche menés dans le cadre de cette thèse s'articulent autour de troix axes principaux : l'amélioration de la qualité de service et d'usage par transformation Mojette, l'identification d'une hiérarchie perceptuelle au niveau des sources vidéos et la proposition de méthodes de protection adaptées.

Dans les deux premières parties de ce mémoire, nous avons traité la problématique de la transmission de contenus multimédias sur IP. Nous nous sommes intéressés en particulier à l'amélioration de la qualité de service et de la qualité d'usage. Nous avons proposé dans un premier temps l'utilisation de la transformation Mojette comme opérateur de codage réseau dans un contexte d'optimisation de l'usage de la bande passante. Cette approche générique au problème de la qualité de service ne prend pas en compte les propriétés de la source ni les mécanismes de la perception de la qualité.

Ce constat nous a amené à utiliser cette même transformation Mojette dans le cadre d'une protection inégale perceptuelle d'un flux hiérarchique JPEG 2000. Cette protection est allouée en fonction de la contribution des sous-flux à la qualité visuelle finale de l'image, mesurée à l'aide de la métrique de qualité C4. La transformation Mojette a été utilisée comme code correcteur d'erreurs $(1+\epsilon)$MDS. L'intérêt d'une telle approche est qu'elle est conjointe source-canal et tient compte de la hiérarchie de la source.

Dans la troisième partie, nous avons étudié, au travers de tests subjectifs d'évaluation de qualité, l'impact des dégradations dues aux pertes de paquets sur la qualité d'une vidéo à la résolution SD. Nous avons montré que des pertes pouvant atteindre 40 paquets consécutifs juste avant le changement de scène ne causaient aucune dégradation de qualité significative. Cependant, la perte de 20 paquets dans la première image de la nouvelle scène peut réduire la qualité de la vidéo de plus de deux unités MOS.

Nous avons également établi deux relations entre la qualité visuelle d'une part et la propagation temporelle des dégradations et leur position spatiale dans l'image d'autre part. La première relation implique que la durée de la propagation des dégradations influence la qualité plus que l'amplitude des pertes. Ainsi, les observateurs sont plus gênés par quatre pertes régulières de 10 paquets chacune que par la perte de 40 paquets consécutifs.

La seconde relation a été établie en variant la position spatiale des pertes dans l'image entre l'arrière-plan de l'image et la région attirant l'attention de l'observateur. Les résultats des tests subjectifs ont montré que la qualité d'une séquence affectée par 20 paquets dans l'arrière-plan de l'image est supérieure de plus d'une unité MOS à la qualité de la même séquence ayant 10 paquets perdus dans sa région d'intérêt.

Ces deux relations que nous avons identifiées ont guidé notre travail qui s'est concentré sur la hiérarchisation d'une source vidéo en vue du développement de méthodes de protection perceptuelles pertinentes. Nous avons donc réalisé des tests de suivi des mouvements oculaires

pour déterminer les régions saillantes de la vidéo. Ces tests nous ont permis, en outre de la création d'une base de vidéos (SD et HD) avec leurs séquences de saillance, d'identifier l'impact du contenu d'une séquence sur l'attractivité des dégradations dues aux pertes de paquets. Par exemple, le mouvement de la région d'intérêt rend plus difficile l'attraction de l'attention de l'observateur par des dégradations ayant lieu loin de la zone d'action.

Ayant hiérarchisé le contenu des vidéos, nous avons proposé dans la quatrième partie de protéger les régions d'intérêt contre la propagation spatio-temporelle des dégradations. Cette protection est fondée sur le codage en mode intra des macroblocs de la région d'intérêt dans les images B et P. Une première évaluation de performances sur un nombre limité de séquences a montré que l'utilisation de cette méthode ne causait pas une diminution de la qualité visuelle à débit constant. De plus, les évaluations objectives et subjectives de qualité ont indiqué une amélioration de la qualité dans la région d'intérêt par rapport à un codage classique.

Nous avons alors conduit des expérimentations subjectives de qualité sur six séquences à la résolution SD avec deux motifs de pertes différents. La première conclusion dégagée met en évidence l'importance perceptuelle de la région d'intérêt. Ainsi, pour la séquence *Captain*, perdre quatre slices dans la région d'intérêt dégrade la qualité au moins autant que la perte de huit slices à l'extérieur de la région d'intérêt.

Ensuite, nous avons montré que l'utilisation de notre méthode de protection aux débits choisis ne dégrade pas la qualité de la séquence dans le cas sans perte de paquets par rapport à un codage classique ($MOS > 4$ pour les deux schémas de codage).

Enfin, nous avons montré l'efficacité de notre approche quand les pertes ont lieu dans la région d'intérêt avec une différence de qualité moyenne d'une unité MOS.

Perspectives

Le travail présenté dans ce mémoire peut être étendu dans plusieurs directions.

- Effets perceptuels des pertes de paquets : il serait intéressant, maintenant que les principaux facteurs entrant en jeu sont identifiés, d'étudier l'interaction entre ces facteurs d'une part et la qualité visuelle d'autre part. Par exemple, l'introduction de pertes dans la région d'intérêt de l'image pour quelques images avant le changement de scène permettrait de mieux comprendre l'influence conjointe de ces deux facteurs sur la qualité perceptuelle. Une meilleure compréhension de cette interaction facilite l'élaboration de métriques objectives de qualité pour des vidéos affectées par les pertes de paquets.
- Attention visuelle, pertes et qualité visuelle : le test conduit dans le chapitre 6 s'est limité à l'étude de la visibilité et de l'attractivité des pertes de paquets quand elles sont à l'extérieur de la région d'intérêt de l'image. Le test subjectif de qualité présenté dans le chapitre 8 comprenait un motif unique de pertes à l'extérieur de la région d'intérêt. Pour évaluer

l'impact de ce type de motif sur la qualité visuelle, il est nécessaire de réaliser des tests subjectifs pendant lesquels la durée de propagation des dégradations et leurs distances de la région d'intérêt sont variées. De plus, il est important d'établir une relation entre la visibilité d'une perte et son impact perceptuel. En effet, la perception d'une dégradation n'implique pas indubitablement une gêne chez l'observateur.

- Intégration d'un modèle d'attention visuelle dans la méthode de robustesse : dans ce travail, la détermination des régions saillantes de l'image a été effectuée à partir des données recueillies lors des tests oculométriques. La conception d'un système de protection semblable à celui que nous avons proposé implique l'automatisation de cette étape. Ceci peut être fait à l'aide d'un modèle objectif de saillance (par exemple [Meu06]) implanté dans le codeur. Le modèle aura comme tâche l'identification des régions perceptuellement importantes en amont de l'application de l'algorithme de robustesse.
- Intégration d'une graduabilité dans le codage source robuste : comme nous l'avons évoqué dans la discussion de la section 8.4 du chapitre 8, la taille de la RdI contribue simultanément à la diminution de la qualité initiale et à la robustesse en présence de pertes. Une variante des méthodes de robustesse proposées dans ce travail pourrait être la création de régions d'intérêt de taille plus grande mais qui ne soient pas codées intégralement en mode intra. Ainsi, le centre de cette région sera codé en mode intra et sa périphérie sera actualisée par l'insertion de macroblocs codés en intra qui amélioreront l'efficacité de la compensation de pertes au décodage.

Bibliographie

[Agr06] D. Agrafiotis, D. R. Bull and C. N. Canagarajah. Enhanced Error Concealment With Mode Selection. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(8) :pp. 960–973, août 2006.

[Ahl00] R. Ahlswede, N. Cai, S.-Y. R. Li and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46(4) :pp. 1204–1216, juillet 2000.

[Alb94] A. Albanese, J. Blomer, J. Edmonds and M. Luby. Priority Encoding Transmission. Tech. rep., International Computer Science Institute, TR-94-039, Berkeley, août 1994.

[Alb96] A. Albanese, J. Blomer, J. Edmonds, M. Luby and M. Sudan. Priority Encoding Transmission. *IEEE Transactions on Information Theory*, 42(6) :pp. 1737–1744, novembre 1996.

[Alo96] N. Alon and M. Luby. A Linear Time Erasure-Resilient Code with Nearly Optimal Recovery. *IEEE Transactions on Information Theory*, 42(6) :pp. 1732–1736, juillet 1996.

[Apo02] J. Apostolopoulos, T. Wong, W. Tan and S. Wee. Network Coding for Large Scale Content Distribution. In *Proceedings of Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM*, vol. 3, pp. 1736–1745, juin 2002.

[Ara06] H. K. Arachchi, W. A. C. Fernando, S. Panchadcharam and W. A. R. J. Weerakkody. Unequal Error Protection Technique for ROI Based H.264 Video Coding. In *Proceedings of Canadian Conference on Electrical and Computer Engineering*, pp. 2033–2036, mai 2006.

[Bab03] M. Babel, O. Déforges and J. Ronsin. Décomposition pyramidale à redondance minimale pour compression d'image sans perte. In *Colloque du Groupe de Recherche et d'Etudes du Traitement du Signal et des Images, GRETSI*, 2003.

[Bac06] P. Baccichet, S. Rane and B. Girod. Systematic Lossy Error Protection based on H.264/AVC Redundant Slices and Flexible Macroblock Ordering. *Journal of Zhejiang University - Science A*, 7(5), mai 2006.

[Ban05] S. Bandyopadhyay, Z. Wu, P. Pandit and J. Boyce. Frame Loss Error Concealment for H.264/AVC. JVT-P072, juillet 2005.

[Ban06] S. K. Bandyopadhyay, Z. Wu, P. Pandit and J. M. Boyce. An Error Concealment Scheme for Entire Frame Losses for H.264/AVC. In *IEEE Sarnoff Symposium*, mars 2006.

[Bel03] S. Belfiore, M. Grangetto, E. Magli and G. Olmo. Spatiotemporal error concealment with optimized mode selection and application to H.264. *Elsevier Journal of Signal Processing : Image Communication*, 18(10) :pp. 907–923, novembre 2003.

[Bla98] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang and W. Weiss. An Architecture for Differentiated Services. RFC 2475, Informational, IETF, décembre 1998.

[Bor98] M. S. Borella, D. Swider, S. Uludag and G. B. Brewster. Internet Packet Loss : Measurement and Implications for End-to-End QoS. In *Proceedings of International Conference on Parallel Processing*, pp. 3–12, août 1998.

[Bor01] C. Bormann. RObust Header Compression (ROHC) : Framework and four profiles : RTP, UDP, ESP, and uncompressed. RFC 3095, IETF Standards Track, juillet 2001.

[Bou04] I. Bouazizi and M. Günes. Distortion-Optimized FEC for Unequal Error Protection in MPEG-4 Video Delivery. In *Proceedings of Ninth International Symposium on Computers and Communications*, vol. 2, pp. 615–620, juillet 2004.

[Bra94] R. Braden, D. Clark and S. Shenker. Integrated Services in the Internet Architecture : an Overview. RFC 1633, Informational, IETF, juin 1994.

[Bra97] R. Braden, L. Zhang, S. Berson, S. Herzog and S. Jamin. Resource ReSerVation Protocol (RSVP). RFC 2205, IETF Standards Track, septembre 1997.

[Cai10] X. Cai, F. H. Ali and E. Stipidis. Object-based video coding with dynamic quality control. *Elsevier Image and Vision Computing*, 28(3) :pp. 285–297, mars 2010.

[Car04] M. Carnec. Critères de qualité d'images couleur avec référence réduite perceptuelle générique. Thèse de doctorat, Université de Nantes, juillet 2004.

[Car05] M. Carnec, P. L. Callet and D. Barba. Visual Features for Image Quality Assessment with Reduced Reference. In *Proceedings of International Conference on Image Processing*, vol. 1, pp. 421–424, septembre 2005.

[Car06] T. A. Carlson, H. Hogendoorn and F. A. J. Verstraten. The speed of visual attention : What time is it ? *Journal of Vision*, 6(12) :pp. 1406–1411, décembre 2006.

[Che98] T. Chen and R. R. Rao. Audio-visual integration in multimodal communication. *Proceedings of the IEEE*, 86(5) :pp. 837–852, mai 1998.

[Che07a] Q. Chen, Z. Chen, X. Gu and C. Wang. Attention-Based Adaptive Intra Refresh for Error-Prone Video Transmission. *IEEE Communications Magazine*, 45(1) :pp. 52–60, janvier 2007.

[Che07b] Q. Chen, L. Song, X. Yang and W. Zhang. Robust Region-of-Interest Scalable Coding with Leaky Prediction in H.264/AVC. In *Proceedings of IEEE Workshop on Signal Processing Systems*, pp. 357–362, octobre 2007.

[Cho03] P. A. Chou, Y. Wu and K. Jain. Practical Network Coding. In *51st Allerton Conference on Communication, Control and Computing*, octobre 2003.

[Chr00] C. Christopoulos, A. Skodras and T. Ebrahimi. The JPEG2000 Still Image Coding System : An Overview. *IEEE Transactions on Consumer Electronics*, 46(4) :pp. 1103–1127, novembre 2000.

[Cla99] M. Claypool and J. Tanner. The Effects of Jitter on the Perceptual Quality of Video. In *Proceedings of the seventh ACM international conference on Multimedia*, vol. 2, pp. 115–118, novembre 1999.

[Con04] C. E. Connor, H. E. Egeth and S. Yantis. Visual Attention : Bottom-Up Versus Top-Down. *Current Biology*, 14 :pp. 850–852, octobre 2004.

[Deb05] S. Deb, M. Effros, T. Ho, D. R. Karger, R. Koetter, D. S. Lun, M. Medard and N. Ratnakar. Network coding for wireless applications : A brief tutorial. In *Proceedings of International Workshop on Wireless Ad-hoc Networks*, mai 2005.

[Dho05] Y. Dhondt, P. Lambert, S. Notebaert and R. V. de Walle. Flexible macroblock ordering as a content adaptation tool in H.264/AVC. In *Proceedings of SPIE Multimedia Systems and Applications VIII*, vol. 6015, pp. 44–52, octobre 2005.

[Dho06] Y. Dhondt, P. Lambert and R. V. de Walle. A Flexible Macroblock Scheme for Unequal Error Protection. In *Proceedings of International Conference on Image Processing*, pp. 829–832, octobre 2006.

[Dho07] Y. Dhondt, S. Mys, K. Vermeirsch and R. V. de Walle. *Constrained Inter Prediction : Removing Dependencies Between Different Data Partitions*, vol. 4678 of *Lecture Notes in Computer Science*, pp. 720–731. Springer Berlin, 2007.

[Duc79] C. E. Duchon. Lanczos Filtering in One and Two Dimensions. *Journal of Applied Meteorology*, 18(8) :pp. 1016–1022, août 1979.

[ets09] Digital Video Broadcasting (DVB) ; Framing structure, channel coding and modulation for digital terrestrial television. European Telecommunications Standards Institute (ETSI) standard, ETSI EN 300 744 V1.6.1, janvier 2009.

[Fan05] T. Fang and L.-P. Chau. A novel unequal error protection approach for error resilient video transmission. In *Proceedings of IEEE International Symposium on Circuits and Systems*, vol. 4, pp. 4022–4025, mai 2005.

[Far04] M. Farias. No-reference and reduced reference video quality metrics : New contributions. Thèse de doctorat, University of California, 2004.

[Feg05] R. Feghali, D. Wang, F. Speranza and A. Vincent. Quality Metric for Video Sequences with Temporal Scalability. In *Proceedings of International Conference on Image Processing*, vol. 3, pp. 137–140, septembre 2005.

[Fen98] C. Fenimore, J. Libert and S. Wolf. Perceptual Effects of Noise in Digital Video Compression. In *140th Society of Motion Picture and Television Engineers (SMPTE) Technical Conference*, octobre 1998.

[Fra06] C. Fragouli, J.-Y. L. Boudec and J. Widmer. Network Coding : An Instant Primer. *ACM SIGCOMM Computer Communication Review*, 36(1) :pp. 63–68, 2006.

[Gha06] M. M. Ghandi and M. Ghanbari. Layered H.264 video transmission with hierarchical QAM. *Journal of Visual Communication and Image Representation : Special Issue on emerging H.264/AVC video coding standard*, 17(2) :pp. 451–466, avril 2006.

[Gir99] B. Girod, K. Stuhlmüller, M. Link and U. Horn. Packet Loss Resilient Internet Video Streaming. In *Proceedings of SPIE Visual Communications and Image Processing*, vol. 3653, pp. 833–844, janvier 1999.

[Gka05] C. Gkantsidis and P. R. Rodriguez. Network Coding for Large Scale Content Distribution. In *Proceedings of Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM*, vol. 4, pp. 2235–2245, mars 2005.

[Goh08] K. H. Goh, D. J. Wu, J. Y. Tham, T. K. Chiew and W. S. Lee. Real-Time Software MPEG-2 to H.264 Video Transcoding. In *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 165–168, juin 2008.

[Goy01] V. Goyal. Multiple Description Coding : Compression Meets the Network. *IEEE Signal Processing Magazine*, 18(5) :pp. 74–93, septembre 2001.

[Gué01] J. Guédon, B. Parrein and N. Normand. Internet distributed image information system. *Integrated Computer-Aided Engineering*, 8(3) :pp. 205–214, août 2001.

[Gué09] J. Guédon. *The Mojette Transform : Theory and Applications*. Wiley-ISTE, février 2009.

[Gul07] S. R. Gulliver and G. Ghinea. The Perceptual and Attentive Impact of Delay and Jitter in Multimedia Delivery. *IEEE Transactions on Broadcasting*, 53(2) :pp. 449–458, juin 2007.

[Guo05] Y. Guo, H. Li and Y.-K. Wang. SVC/AVC Loss Simulator. Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, JVT-Q069, octobre 2005.

[Hal04] T. Halbach and S. Olsen. Error Robustness Evaluation of H.264/MPEG-4 AVC. In *Proceedings of SPIE Visual Communications and Image Processing*, vol. 5308, pp. 617–627, janvier 2004.

[Han01] D. S. Hands and S. E. Avons. Recency and Duration Neglect in Subjective Assessment of Television Picture Quality. *Applied Cognitive Psychology*, 15(6) :pp. 639–657, novembre 2001.

[Han07] C. Han and J. Liu. New Temporal Error Concealment Method for H.264 Based on Motion Strength. In *Proceedings of International Conference on Computational Intelligence and Security Workshops*, pp. 858–861, décembre 2007.

[Har04] O. Harmanci and A. Tekalp. Optimization of H264 for Low Delay Video Communications over Lossy Channels. In *Proceedings of International Conference on Image Processing*, vol. 5, pp. 3209–3212, octobre 2004.

[Hil06] O. I. Hillestad, O. Jetlund and A. Perkis. RTP-Based Broadcast Streaming of High Definition H.264/AVC Video : An Error Robustness Evaluation. *Journal of Zhejiang University - Science A*, 7(1) :pp. 19–26, janvier 2006.

[Ho05] T. Ho, B. Leong, Y.-H. Chang, Y. Wen and R. Koetter. Network monitoring in multicast networks using network coding. In *Proceedings of IEEE International Symposium on Information Theory*, pp. 1977–1981, septembre 2005.

[HT08] Q. Huynh-Thu and M. Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, 44(13) :pp. 800–801, juin 2008.

[Hua07] Y. Huang and J. G. Apostolopoulos. A Joint Packet Selection/Omission and FEC System for Streaming Video. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 845–848, avril 2007.

[Hua08] S.-C. Huang and S.-Y. Kuo. *Optimization of Spatial Error Concealment for H.264 Featuring Low Complexity*, vol. 4903 of *Lecture Notes in Computer Science*, pp. 391–401. Springer Berlin, 2008.

[Itt98] L. Itti, C. Koch and E. Niebur. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11) :pp. 1254–1259, novembre 1998.

[itu96] ITU-T Recommendation P.800, Methods for subjective determination of transmission quality. International Telecommunication Union-Standardization Sector, août 1996.

[itu98] ITU-R Recommendation BT.1359-1, Relative timing of sound and vision for broadcasting. International Telecommunication Union-Radiocommunication Sector, novembre 1998.

[itu99] ITU-T Recommendation P.910, Subjective video quality assessment methods for multimedia applications. International Telecommunication Union-Standardization Sector, septembre 1999.

[itu02] ITU-R Recommendation BT.500-11, Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union-Radiocommunication Sector, juin 2002.

[itu04a] ITU-R Recommendation BT.1683, Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference. International Telecommunication Union-Radiocommunication Sector, janvier 2004.

[itu04b] ITU-T Recommendation J.144, Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference. International Telecommunication Union-Standardization Sector, mars 2004.

[itu04c] ITU-T Recommendation Y.2001, General overview of ngn. International Telecommunication Union-Standardization Sector, décembre 2004.

[itu07] ITU-T Recommendation H.264, Advanced video coding for generic audiovisual services. International Telecommunication Union-Standardization Sector, novembre 2007.

[Iva06] M. Ivanovici and R. Beuran. User-Perceived Quality Assessment for Multimedia Applications. In *Proceedings of the 10th International Conference on Optimization of Electrical and Electronic Equipments*, vol. 4, pp. 55–60, mai 2006.

[Jia99] W. Jiang and H. Schulzrinne. QoS Measurement of Internet Real-Time Multimedia Services. Rapport technique CUCS-015-99. Disponible sur `http://www.cs.columbia.edu/techreports/cucs-015-99.pdf`, 1999.

[Jor09] T. R. Jordan, K. B. Paterson, S. Kurtev and M. Xu. Do fixation cues ensure fixation accuracy in split-fovea studies of word recognition ? *Elsevier Neuropsychologia*, 47(8-9) :pp. 2004–2007, juillet 2009.

[jpe04] JPEG 2000 image coding system : Core coding system. ISO/IEC Standard 15444-1, Joint Technical Committe 1 (JTC) & Subcommittee 29 (SC), 2004.

[Kan04] S. Kanumuri, P. Cosman and A. Reibman. A Generalized Linear Model for MPEG-2 Packet-Loss Visibility. In *Proceedings of the 14th International Packet Video Workshop*, décembre 2004.

[Kan06] S. Kanumuri, S. Subramanian, P. Cosman and A. Reibman. Predicting H.264 Packet Loss Visibility using a Generalized Linear Model. In *Proceedings of International Conference on Image Processing*, pp. 2245–2248, octobre 2006.

[Kat78] M. Katz. *Questions of Uniqueness and Resolution in Reconstruction from Projections*. Lecture Notes in Biomathematics. Springer-Verlag New York, 1978.

[Kat05] S. Katti, D. Katabi, W. Hu, H. Rahul and M. Médard. The Importance of Being Opportunistic : Practical Network Coding for Wireless Environments. In *Allerton Conference on Communication, Control and Computing*, septembre 2005.

[Kat07] B. Katz, S. Greenberg, N. Yarkoni, N. Blaunstien and R. Giladi. New Error-Resilient Scheme Based on FMO and Dynamic Redundant Slices Allocation for Wireless Video Transmission. *IEEE Transactions on Broadcasting*, 53(1) :pp. 308–319, mars 2007.

[Kat08] S. Katti, H. Rahul, H. Wenjun, D. Katabi, M. Medard and J. Crowcroft. XORs in the Air : Practical Wireless Network Coding. *IEEE/ACM Transactions on Networking*, 16(3) :pp. 497–510, juin 2008.

[Kim05] D. Kim, S. Yang and J. Jeong. A New Temporal Error Concealment Method for H.264 using Adaptive Block Sizes. In *Proceedings of International Conference on Image Processing*, vol. 3, pp. 928–931, septembre 2005.

[Kim08] 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 *Proceedings of the fourth International Conference on Networked Computing and Advanced Information Management*, vol. 2, pp. 719–725, septembre 2008.

[Kin09] A. Kingston, F. Autrusseau, E. Grall, T. Hamon and B. Parrein. *Mojette-based Security*, chap. 10, pp. 239–267. In "The Mojette Transform : Theory and Applications". Wiley-ISTE, 2009.

[Kul51] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1) :pp. 79–86, mars 1951.

[Lac09] J. Lacan, V. Roca, J. Peltotalo and S. Peltotalo. Reed-Solomon Forward Error Correction (FEC) Schemes. RFC 5510, IETF Standards Track, avril 2009.

[Lam93] W. M. Lam, A. R. Reibman and B. Liu. Recovery of lost or erroneously received motion vectors. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 417–420, avril 1993.

[Lam06] P. Lambert, W. D. Neve, Y. Dhondt and R. V. de Walle. Flexible macroblock ordering in H.264/AVC. *Elsevier Journal of Visual Communication and Image Representation*, 17(2) :pp. 358–375, avril 2006.

[Lan05] K. A. Lane. *Developing Ocular Motor and Visual Perceptual Skills : An Activity Workbook*. SLACK Incorporated, 2005.

[Lar06] P. Larsson, N. Johansson and K.-E. Sunell. Coded Bi-directional Relaying. In *IEEE 63rd Vehicular Technology Conference*, vol. 2, pp. 851–855, mai 2006.

[Lei94] C. Leicher. Hierarchical Encoding of MPEG Sequences Using Priority Encoding Transmission (PET). Tech. rep., International Computer Science Institute, TR-94-058, Berkeley, novembre 1994.

[Lia03] Y. Liang, J. G. Apostolopoulos and B. Girod. Analysis of Packet Loss for Compressed Video : Does Burst-Length Matter ? In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. 684–687, avril 2003.

[Lia06] Y. J. Liang, K. El-Maleh and S. Manjunath. Upfront Intra-Refresh Decision for Low-Complexity Wireless Video Telephony. In *IEEE International Symposium on Circuits and Systems*, mai 2006.

[Lub97] M. G. Luby, M. Mitzenmacher, A. Shokrollahi, D. A. Spielman and V. Stemann. Practical Loss-Resilient Codes. In *Proceedings of 29th Annual ACM Symposium on Theory of Computing*, pp. 150–159, mai 1997.

[Lub02] M. G. Luby. LT Codes. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, pp. 271–280, novembre 2002.

[Ma09] H. Ma, X. Meng and Y. Chen. An Attention-Based Network-Aware Adaptive Intra Refresh Method for Wireless Video Communications. In *Proceedings of International Conference on Communications and Mobile Computing*, vol. 2, pp. 225–229, janvier 2009.

[Man03] B. R. Manor and E. Gordon. Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Elsevier Journal of Neuroscience Methods*, 128(1-2) :pp. 85–93, septembre 2003.

[Mar03] D. Marpe, H. Schwarz and T. Wiegand. Context-Based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7) :pp. 620–636, juillet 2003.

[Mar04] M. G. Martini and M. Chiani. Rate-Distortion Models for Unequal Error Protection for Wireless Video Transmission. In *IEEE 59th Vehicular Technology Conference*, vol. 2, pp. 1049–1053, mai 2004.

[Maz09] C. Mazataud and B. Bing. A Practical Survey of H.264 Capabilities. In *Proceedings of Seventh Annual Communication Networks and Services Research Conference*, pp. 25–32, mai 2009.

[Meu05] O. L. Meur. Attention sélective en visualisation d'images fixes et animées sur écran : modèles et évaluation de performances - applications. Thèse de doctorat, Université de Nantes, octobre 2005.

[Meu06] O. L. Meur, P. L. Callet, D. Barba and D. Thoreau. A Coherent Computational Approach to Model Bottom-Up Visual Attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5) :pp. 802–817, mai 2006.

[Moh00] A. E. Mohr, E. A. Riskin and R. E. Ladner. Unequal Loss Protection : Graceful Degradation of Image Quality over Packet Erasure Channels Through Forward Error Correction. *IEEE Journal on Selected Areas in Communications*, 18(6) :pp. 819–828, juin 2000.

[Moh02] S. Mohamed and G. Rubino. A Study of Real-Time Packet Video Quality Using Random Neural Networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(12) :pp. 1071–1083, décembre 2002.

[Nem05] O. Nemethova, A. Al-Moghrabi and M. Rupp. Flexible Error Concealment for H.264 Based on Directional Interpolation. In *Proceedings of International Conference on Wireless Networks, Communications and Mobile Computing*, vol. 2, pp. 1255–1260, juin 2005.

[Ngu07] D. Nguyen, T. Nguyen and X. Yang. Multimedia Wireless Transmission with Network Coding. In *Proceedings of the 16th International Packet Video Workshop*, pp. 326–335, novembre 2007.

[Ngu08] D. T. Nguyen, M. Hayashi and J. Ostermann. Adaptive Error Protection for Scalable Video Coding Extension of H.264/AVC. In *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 417–420, juin 2008.

[Nin07] A. Ninassi, O. L. Meur, P. L. Callet and D. Barba. Does where you Gaze on an Image Affect your Perception of Quality ? Applying Visual Attention to Image Quality Metric. In *Proceedings of International Conference on Image Processing*, vol. 2, pp. 169–172, septembre 2007.

[Nin09] A. Ninassi. De la perception locale des distorsions de codage à l'appréciation globale de la qualité visuelle des images et vidéos. apport de l'attention visuelle dans le jugement de qualité. Thèse de doctorat, Université de Nantes, mars 2009.

[Nor97] N. Normand. Représentation d'images et distances discrètes basées sur les éléments structurants à deux pixels. Thèse de doctorat, Université de Nantes, 1997.

[Nor06] N. Normand, A. Kingston and P. Évenou. A Geometry Driven Reconstruction Algorithm for the Mojette Transform. In *Discrete Geometry for Computer Imagery*, vol. 4245, pp. 122–133, octobre 2006.

[Oro04] J. Orozco and D. Ros. DiffServ-Aware Streaming of H.264 Video. Publication interne 1627, Systèmes communicants - Projet Armor, Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), juin 2004.

[Osb98] W. Osberger and A. J. Maeder. Automatic Identification of Perceptually Important Regions in an Image. In *Proceedings of 14th International Conference on Pattern Recognition*, vol. 1, pp. 701–704, août 1998.

[Par01] B. Parrein. Description multiple de l'information par transformaion mojette. Thèse de doctorat, Université de Nantes, novembre 2001.

[Par02] D. Parkhurst. Selective attention in natural vision : Using computational models to quantify stimulus-driven attentional allocation. Thèse de doctorat, The Johns Hopkins University, 2002.

[Pax99] V. Paxson. End-to-end Internet packet dynamics. *IEEE/ACM Transactions on Networking*, 7(3) :pp. 277–292, juin 1999.

[Péc08] S. Péchard. Qualité d'usage en télévision haute définition : évaluations subjectives et métriques objectives. Thèse de doctorat, Université de Nantes, octobre 2008.

[Per05] F. Pereira. Sensations, perceptions and emotions : towards quality of experience evaluation for consumer electronics video adaptations. In *First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, janvier 2005.

[Pin04] M. Pinson and S. Wolf. A New Standardized Method for Objectively Measuring Video Quality. *IEEE Transactions on Broadcasting*, 50(3) :pp. 312–322, septembre 2004.

[Pin08] S. Pinneli and D. M. Chandler. A Bayesian Approach to Predicting the Perceived Interest of Objects. In *Proceedings of International Conference on Image Processing*, pp. 2584–2587, octobre 2008.

[Pos80a] M. I. Posner. Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32(1) :pp. 3–25, 1980.

[Pos80b] J. Postel. User Datagram Protocol. RFC 768, IETF Standards Track, août 1980.

[PV04a] R. R. Pastrana-Vidal, J. C. Gicquel, C. Colomes and H. Cherifi. Frame Dropping Effects on User Quality Perception. In *5th International Workshop on Image Analysis for Multimedia Interactice Services*, avril 2004.

[PV04b] R. R. Pastrana-Vidal, J. C. Gicquel, C. Colomes and H. Cherifi. Sporadic Frame Dropping Impact on Quality Perception. In *Proceedings of SPIE Human Vision and Electronic Imaging*, vol. 5292, pp. 182–193, janvier 2004.

[PV04c] R. R. Pastrana-Vidal, J. C. Gicquel, C. Colomes and H. Cherifi. Temporal Masking Effect on Dropped Frames at Video Scene Cuts. In *Proceedings of SPIE Human Vision and Electronic Imaging*, vol. 5292, pp. 194–201, janvier 2004.

[Qu03] Q. Qu, Y. Pei and J. W. Modestino. Robust H.264 video coding and transmission over bursty packet-loss wireless networks. In *IEEE 58th Vehicular Technology Conference*, vol. 5, pp. 3395–3399, octobre 2003.

[Rah06] T. Rahrer, R. Fiandra and S. Wright. Triple-play Services Quality of Experience (QoE) Requirements. TR-126, Architecture & Transport Working Group, DSL Forum, décembre 2006.

[Ram06] M. Ramon, F.-X. Coudoux, M. Gazalet, M. Gharbi and P. Corlay. Systematic Lossy Error Protection of Video based on Reduced Spatial Resolution. In *Proceedings of 13th IEEE International Conference on Electronics, Circuits and Systems*, pp. 850–853, décembre 2006.

[Ray09] K. Rayner, T. J. Smith, G. L. Malcolm and J. M. Henderson. Eye movements and visual encoding during scene perception. *Psychological Science*, 20(1) :pp. 6–10, janvier 2009.

[Ree60] I. S. Reed and G. Solomon. Polynomial Codes Over Certain Finite Fields. *SIAM Journal of Applied Mathematics*, 8(2) :pp. 300–304, juin 1960.

[Rei07] A. Reibman and D. Poole. Predicting Packet-Loss Visibility Using Scene Characteristics. In *Proceedings of the 16th International Packet Video Workshop*, pp. 308–317, novembre 2007.

[Riz97] L. Rizzo. Effective Erasure Codes for Reliable Computer Communication Protocols. *ACM SIGCOMM Computer Communication Review*, 27(2) :pp. 24–36, avril 1997.

[Rot02] H. L. Roth, A. N. Lora and K. M. Heilman. Effects of monocular viewing and eye dominance on spatial attention. *Brain*, 125(9) :pp. 2023–2035, septembre 2002.

[Sal98] P. Salama, N. B. Shroff and E. J. Delp. *Signal Recovery Techniques for Image and Video Compression and Transmission*, chap. "Error Concealment in Encoded Video Streams", pp. 199–234. Kluwer Academic Publishers, 1998.

[Sal99] D. D. Salvucci. Mapping Eye Movements to Cognitive Processes. Thèse de doctorat, Carnegie Mellon University, mai 1999.

[Sal00] D. D. Salvucci and J. H. Goldberg. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Proceedings of the Eye Tracking Research and Applications Symposium*, pp. 71–78, novembre 2000.

[Say07] M. Sayit and G. Seckin. Scalable Video with Raptor for Wireless Multicast Networks. In *Proceedings of the 16th International Packet Video Workshop*, pp. 336–341, novembre 2007.

[Sch03] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson. RTP : A Transport Protocol for Real-Time Applications. RFC 3550, IETF Standards Track, juillet 2003.

[Sch07] H. Schwarz, D. Marpe and T. Wiegand. Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9) :pp. 1103–1120, septembre 2007.

[Sef07] H. Seferoglu and A. Markopoulou. Opportunistic Network Coding for Video Streaming over Wireless. In *Proceedings of the 16th International Packet Video Workshop*, pp. 191–200, novembre 2007.

[Ser05] M. Servières. Reconstruction tomographique Mojette. Thèse de doctorat, Université de Nantes, décembre 2005.

[She09] X. Shen, A. Wonfor, R. V. Penty and I. H. White. Receiver Playout Buffer Requirement for TCP Video Streaming in the presence of Burst Packet Drops. In *London Communications Symposium*, 2009.

[Shi01] J. Shin, J. Kim and C.-C. Kuo. Quality-of-Service Mapping Mechanism for Packet Video in Differentiated Services Network. *IEEE Transactions on Multimedia*, 3(2) :pp. 219–231, juin 2001.

[Sho06] A. Shokrollahi. Raptor Codes. *IEEE Transactions on Information Theory*, 52(6) :pp. 2551–2567, juin 2006.

[Ste96] R. Steinmetz. Human Perception of Jitter and Media Synchronization. *IEEE Journal on Selected Areas in Communications*, 14(1) :pp. 61–72, janvier 1996.

[Sto03] T. Stockhammer, M. M. Hannuksela and T. Wiegand. H.264/AVC in Wireless Environments. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7) :pp. 657–673, juillet 2003.

[Sto04] T. Stockhammer and M. Bystrom. H.264/AVC Data Partitioning for Mobile Video Communication. In *Proceedings of International Conference on Image Processing*, vol. 1, pp. 545–548, octobre 2004.

[Sul03] G. Sullivan, T. Wiegand and K.-P. Lim. Joint Model Reference Encoding Methods and Decoding Concealment Methods. JVT-I049, septembre 2003.

[Sul04] G. J. Sullivan, P. Topiwala and A. Luthra. The H.264/AVC Advanced Video Coding Standard : Overview and Introduction to the Fidelity Range Extensions. In *Proceedings of SPIE Applications of Digital Image Processing XXVII*, vol. 5558, pp. 454–474, août 2004.

[Sus00] J.-F. Susbielle. *Internet multimédia et temps réel*. Editions Eyrolles, 2000.

[Tan03] A. S. Tanenbaum. *Computer Networks, 4/E*. Prentice Hall, Vrije University, Amsterdam, The Netherlands., 2003.

[Tau00] D. Taubman. High Performance Scalable Image Compression with EBCOT. *IEEE Transactions on Image Processing*, 9(7) :pp. 1158–1170, juillet 2000.

[Tho97] K. Thompson, G. J. Miller and R. Wilder. Wide-Area Internet Traffic Patterns and Characteristics. *IEEE Networks*, 11(6) :pp. 10–23, novembre 1997.

[Tho06] N. Thomos, S. Argyropoulos, N. V. Boulgouris and M. G. Strintzis. Robust transmission of H.264/AVC streams using adaptive group slicing and unequal error protection. *EURASIP Journal on Applied Signal Processing*, Article ID 51502, vol. 2006, 13 pages, 2006.

[Tho09] N. Thomos, J. Chakareski and P. Frossard. Randomized network coding for UEP video delivery in overlay networks. In *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 730–733, juin 2009.

[Til08] T. Tillo, M. Grangetto and G. Olmo. Redundant Slice Optimal Allocation for H.264 Multiple Description Coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(1) :pp. 59–70, janvier 2008.

[vD99] P. M. J. van Diepen, L. Ruelens and G. d'Ydewalle. Brief foveal masking during scene perception. *Elsevier Acta Psychologica*, 101(1) :pp. 91–103, mars 1999.

[vdBL96] C. van den Branden Lambrecht and O. Verscheure. Perceptual Quality Measure Using a Spatio-Temporal Model of the Human Visual System. In *Proceedings of SPIE Human Vision and Electronic Imaging*, vol. 2668, pp. 450–461, janvier 1996.

[vqe00] Final Report on the Validation of Objective Models of Video Quality Assessment. Video Quality Experts Group, juin 2000.

[Wan03] Z. Wang, H. R. Sheikh and A. C. Bovik. *The Handbook of Video Databases : Design and Applications*, chap. 41 : Objective Video Quality Assessment, pp. 1041–1078. CRC Press, septembre 2003.

[Wan04] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli. Image Quality Assessment : From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4) :pp. 600–612, avril 2004.

[Wan05] Y. Wang, A. Reibman and S. Lin. Multiple Description Coding for Video Delivery. *Proceedings of the IEEE*, 93(1) :pp. 57–70, janvier 2005.

[Wan07] H. Wang, S. Xiao and C.-C. J. Kuo. Robust and Flexible Wireless Video Multicast with Network Coding. In *Proceedings of IEEE Global Telecommunications Conference, GLOBECOM*, pp. 2129–2133, novembre 2007.

[Wat01] A. B. Watson, J. Hu and J. F. Mcgowan. DVQ : A digital video quality metric based on human vision. *Journal of Electronic Imaging*, 10 :pp. 20–29, 2001.

[Wen99] S. Wenger. Error Patterns for Internet Experiments. ITU-T VCEG, Q15-I-16r1, octobre 1999.

[Wen03] S. Wenger. H.264/AVC over IP. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7) :pp. 645–656, juillet 2003.

[Wen05] S. Wenger, M. Hannuksela, T. Stockhammer, M. Westerlund and D. Singer. RTP Payload Format for H.264 Video. RFC 3984, IETF Standards Track, février 2005.

[Wie03] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra. Overview of the H.264/AVC Video Coding Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7) :pp. 560–576, juillet 2003.

[Xu08] Y. Xu and Y. Zhou. Adaptive Temporal Error Concealment Scheme for H.264/AVC Video Decoder. *IEEE Transactions on Consumer Electronics*, 54(4) :pp. 1846–1851, novembre 2008.

[Yam05] K. Yamagishi and T. Hayashi. Analysis of psychological factors for quality assessment of interactive multimodal service. In *Proceedings of SPIE Human Vision and Electronic Imaging*, vol. 5666, pp. 130–138, janvier 2005.

[Yam07] T. Yamada, Y. Miyamoto and M. Serizawa. No-Reference Video Quality Estimation Based on Error-Concealment Effectiveness. In *Proceedings of the 16th International Packet Video Workshop*, pp. 288–293, novembre 2007.

[Zha06] Y. Zhang, W. Gao and D. Zhao. Joint Data Partition and Rate-Distortion Optimized Mode Selection for H.264 Error-Resilient Coding. In *Proceedings of the IEEE International Workshop on Multimedia Signal Processing*, pp. 248–251, octobre 2006.

[Zha09] X. Zhang, X.-H. Peng, D. Wu, T. Porter and R. Haywood. A Hierarchical Unequal Packet Loss Protection Scheme for Robust H.264/AVC Transmission. In *IEEE Consumer Communications and Networking Conference*, janvier 2009.

Résumé

Le trafic multimédia sur IP connaît une forte croissance ces dernières années grâce à l'émergence de services comme la TV sur IP ou la vidéo à la demande (*Video on Demand*). Cependant, la Qualité d'Usage (QdU) associée à ce type de trafic n'est pas garantie, principalement à cause de la fluctuation de la Qualité de Service (QdS). Pour assurer un service de qualité acceptable, il est possible d'améliorer les paramètres de QdS ou même d'améliorer directement la QdU. Dans cette thèse, nous nous intéressons à l'étude de l'impact perceptuel de la variation de la QdU et à son amélioration. Nous proposons tout d'abord d'utiliser la transformation Mojette, une transformation de Radon discrète exacte, comme opérateur de *network coding*. Cette technique vise l'amélioration de la QdS en optimisant l'utilisation de la bande passante disponible. Nous proposons également une méthode de protection inégale perceptuelle de flux hiérarchiques par transformation Mojette. Ensuite, nous étudions les effets perceptuels des pertes de paquets sur des vidéos codées en H.264/AVC au travers de tests subjectifs d'évaluation de qualité. Ces tests mènent à l'identification de l'importance de la position spatiale de la perte dans l'image. Nous conduisons alors des expérimentations oculométriques pour identifier les régions d'intérêt de la vidéo. Partant d'une hiérarchie de la source guidée par ces régions d'intérêt, nous proposons des méthodes de protection perceptuelles inégales. Ces techniques de codage robuste, mettant en œuvre l'outil *Flexible Macroblock Ordering* (FMO) de H.264/AVC, sont fondées sur l'arrêt de la propagation spatio-temporelle des dégradations. L'évaluation de performances montre que les méthodes proposées sont efficaces contre les pertes de paquets ayant lieu dans les régions d'intérêt de la vidéo.

Mots-clés : qualité d'usage, pertes de paquets, attention visuelle, codage robuste H.264/AVC.

Abstract

Multimedia traffic over IP has grown significantly in recent years due to the proliferation of IPTV channels and Video on Demand (VOD) services. However, the Quality of Experience (QoE) associated with multimedia services can vary dramatically because of the fluctuating Quality of Service (QoS). Improving QoS parameters or acting directly on the QoE ensures an acceptable quality level. In this thesis, we study the perceptual impact of QoE fluctuation and the possible ways of improving it. We first propose to use an exact and discrete Radon transform, the Mojette transform, as a network coding tool. Network coding aims at enhancing QoS by optimizing the use of available bandwidth. We also propose a perceptual unequal error protection method for scalable bitstreams based on the Mojette transform. Second, we study the perceptual effects of packet loss on H.264/AVC encoded videos by means of subjective quality tests. The test results show that the spatial position of the loss in the image is important w.r.t. the overall visual quality. This leads us to conduct eye tracking tests in order to identify the regions of interest of a video sequence. We then propose perceptual unequal error protection methods on the basis of the hierarchy determined by the regions of interest. These error resilient techniques target halting the spatio-temporal error propagation using H.264/AVC Flexible Macroblock Ordering (FMO). Performance evaluation results show that the proposed methods are efficient against packet loss occurring in the regions of interest of the video.

Keywords : Quality of Experience, packet loss, visual attention, H.264/AVC resilient coding.