



HAL
open science

Schéma de protection systématique à scalabilité fréquentielle d'un flux vidéo compressé basé sur le codage de Wyner-Ziv

Marie Ramon

► **To cite this version:**

Marie Ramon. Schéma de protection systématique à scalabilité fréquentielle d'un flux vidéo compressé basé sur le codage de Wyner-Ziv. *Electronique*. Université de Valenciennes et du Hainaut-Cambrésis, 2007. Français. NNT : 2007VALE0021 . tel-03001730

HAL Id: tel-03001730

<https://hal.science/tel-03001730>

Submitted on 12 Nov 2020

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

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



Université de Valenciennes et du Hainaut Cambrésis

THESE DE DOCTORAT

Discipline : Electronique, optronique et systèmes

Par

Marie Ramon

SCHEMA DE PROTECTION SYSTEMATIQUE A SCALABILITE FREQUENTIELLE D'UN FLUX VIDEO COMPRESSE BASE SUR LE CODAGE DE WYNER-ZIV

Soutenue le 19 juin 2007 devant la commission d'examen :

Rapporteurs :

Dominique BARBA

Professeur des Universités, IRCCyN Nantes

Rodolphe VAUZELLE

Professeur des Universités, SIC Chasseneuil

Membre invité

M. Nicolas GAULLIER

Chef de projet ingénierie, Cognacq-Jay Image

Directeurs de thèse

M. Marc GAZALET

Professeur des Universités, IEMN DOAE Valenciennes

M. François-Xavier COUDOUX

Professeur des Universités, IEMN DOAE Valenciennes

Institut d'Electronique, de Microélectronique et de Nanotechnologies (UMR 8520)
Département O.A.E, Le Mont Houy, 59300 Valenciennes

Université de Valenciennes et du Hainaut Cambrésis

THESE DE DOCTORAT

Discipline : Electronique, optronique et systèmes

Par

Marie Ramon



**SCHEMA DE PROTECTION SYSTEMATIQUE A SCALABILITE FREQUENTIELLE
D'UN FLUX VIDEO COMPRESSE BASE SUR LE CODAGE DE WYNER-ZIV**

Soutenue le 19 juin 2007 devant la commission d'examen :

Rapporteurs :

Dominique BARBA

Professeur des Universités, IRCCyN Nantes

Rodolphe VAUZELLE

Professeur des Universités, SIC Chasseneuil

Membre invité

M. Nicolas GAULLIER

Chef de projet ingénierie, Cognacq-Jay Image

Directeurs de thèse

M. Marc GAZALET

Professeur des Universités, IEMN DOAE Valenciennes

M. François-Xavier COUDOUX

Professeur des Universités, IEMN DOAE Valenciennes

Institut d'Electronique, de Microélectronique et de Nanotechnologies (UMR 8520)
Département O.A.E, Le Mont Houy, 59300 Valenciennes

Remerciements

Mes plus sincères remerciements vont à François-Xavier Coudoux pour m'avoir proposé ce sujet et avoir su orienter mes recherches. Je voudrais aussi le remercier pour son dynamisme et sa bonne humeur communicative, sa patience et sa bienveillance qui m'ont permis de surmonter les divers obstacles rencontrés.

J'adresse également toute l'expression de ma reconnaissance respectueuse à Marc Georges Gazalet, qui par ses compétences scientifiques a fortement enrichi ma formation et apporté la rigueur nécessaire aux aspects théoriques de cette thèse.

Ils ont été des enseignants passionnants puis des directeurs de thèse attentifs, c'est donc à leur grande compétence que ce travail doit son commencement et son aboutissement.

M. Barba m'a fait l'honneur d'accepter d'être président de mon jury de thèse et rapporteur de celle-ci. Pour cela, ainsi que pour ses critiques constructives sur mon manuscrit, je lui exprime ma profonde gratitude. Je remercie M. Vauzelle pour avoir accepté d'être rapporteur de ma thèse, ainsi que pour ses commentaires très pertinents sur mon mémoire. Je remercie également M. Gaullier d'avoir accepté de faire partie de mon jury de thèse. L'intérêt qu'ils ont manifesté envers ce travail constitue en soi une véritable récompense.

Je remercie finalement :

- L'ensemble des personnels du laboratoire pour leur accueil chaleureux, en particulier l'équipe de Transmission Numérique des Signaux et Images (TNSI), Charlène, David et Kahena avec qui j'ai partagé le bureau pendant ces années avec beaucoup de plaisir et Patrick Corlay (aka Charlie) pour son sens de l'humour impayable,
- Edith, Anne et Véronique pour leur efficacité souriante, même face aux délais les plus courts,
- Le service de reprographie pour l'édition et l'assemblage des exemplaires,
- Ma famille pour m'avoir fourni les ressources humaines et matérielles nécessaires au bon déroulement d'un tel projet, mon père et mon grand père notamment qui ont été mes premiers guides dans l'apprentissage des sciences et à qui je dédie ce mémoire,
- Les amis, en particulier Steph (*The most beautiful discovery true friends make is that they can grow separately without growing apart. [Elisabeth Foley]*), Soft et Nad pour les moments de détente partagés, Gregwar et Tr8n pour le soutien musical qu'ils m'apportent continuellement.
- Et Khal, merci pour ta petite présence apaisante...

Table des matières

Introduction	1
1 Contrôle d'erreur dans les communications vidéo	5
1.1 Préambule.....	6
1.1.1 Réseaux de communication audiovisuelle.....	7
1.1.2 Principes de compression vidéo.....	9
1.2 Codage de canal.....	14
1.2.1 Retransmission sous contrainte de délai	15
1.2.2 Codes correcteurs « par anticipation ».....	16
1.3 Codages conjoints source canal	18
1.3.1 Modification du codage de source	19
1.3.2 Modification du codage de canal.....	23
1.4 Dissimulation d'erreur.....	28
1.4.1 Principe général de la dissimulation d'erreur.....	29
1.4.2 Méthodes de récupération de la texture.....	31
1.4.3 Récupération des vecteurs de mouvement et mode de codage.....	32
1.5 Conclusion	34
2 Théorème de Wyner-Ziv et applications.....	37
2.1 Bases théoriques du codage source avec information adjacente.....	39
2.1.1 Théorème de Slepian-Wolf.....	41
2.1.2 Théorème de Wyner-Ziv.....	43
2.2 Mises en œuvre des théorèmes et application à la compression.....	45
2.2.1 Principe de mise en œuvre des théorèmes.....	45
2.2.2 Codage inter images	51
2.3 Application à la protection contre les erreurs.....	55
2.3.1 Codage scalable.....	56
2.3.2 Codage de Wyner-Ziv résistant aux erreurs	59
2.3.3 Codage systématique résistant aux erreurs	61
2.4 Conclusion	65
3 Schéma de protection spatiale inégale systématique proposé et analyse	67
3.1 Présentation du système proposé	68
3.1.1 Décimation DCT.....	69
3.1.2 Comparaison Spatial / SNR.....	73
3.1.3 Récapitulatif du fonctionnement du système proposé.....	77
3.2 Modèle analytique	79

3.2.1	Définition du canal de transmission	79
3.2.2	Calcul de distorsion.....	81
3.2.3	Prise en compte de la propagation des erreurs	83
3.3	Optimisation du schéma.....	84
3.3.1	Détermination du nombre de slices de parité	85
3.3.2	Détermination de la description optimale	87
3.3.3	Effet de la résolution spatiale de l'image d'origine.....	90
3.4	Conclusion	92
4	Résultats analytiques et expérimentaux	93
4.1	Description des expériences proposées.....	93
4.1.1	Contexte général.....	94
4.1.2	Mise en paquets	96
4.1.3	Cas d'une transmission réelle.....	98
4.2	Résultats du modèle analytique	100
4.2.1	Analyse de la résistance aux erreurs du schéma.....	101
4.2.2	Effet de la longueur des salves d'erreur	104
4.2.3	Effet d'une augmentation du débit alloué à la protection.....	107
4.3	Résultats expérimentaux	109
4.3.1	Résultats de simulation	109
4.3.2	Résultats de transmission	117
4.4	Conclusion	123
	Conclusion et perspectives.....	125
	Annexes	129
	Annexe I : transmission par paquet.....	131
	Annexe II : Définition du canal de Gilbert-Elliott	135
	Annexe III : Détermination de la distorsion moyenne d'une image du GOP.....	139
	Annexe IV : Détail des séquences utilisées	143
	Annexe V : Résultats de simulations	147
	Annexe VI : Résultats d'expériences.....	153
	Liste des figures	161
	Liste des tableaux.....	165
	Références	167

Introduction

La convergence entre les secteurs des télécommunications, de l'audiovisuel et de l'information à travers les nombreuses applications de l'Internet ou le déploiement de la télévision numérique est en train de se concrétiser réellement [RIV02]. Dans ce contexte, l'intégration des flux audiovisuels au sein des réseaux de communication Internet fixe ou mobile constitue aujourd'hui un enjeu technologique majeur. On assiste depuis quelques années à une convergence des réseaux de diffusion sur la base du protocole Internet et les transmissions par commutation de paquets se présentent comme une voix d'avenir pour les communications audiovisuelles. Elles offrent une grande souplesse dans le choix du médium de transmission physique sous-jacent puisque les paquets IP peuvent être facilement fragmentés et transportés dans des trames DVB mais aussi dans des cellules ATM, transmis sur les réseaux sans fil 3G, etc.

En contrepartie, les applications multimédias communicantes sont aujourd'hui confrontées à des problèmes de fiabilité de transmission sur des réseaux aux performances et topologies variables, incluant des environnements multi-points hétérogènes, fixes ou mobiles. Le taux d'erreur résiduel des signaux transmis vu par le décodeur est souvent loin d'être négligeable. Or les exigences de l'utilisateur en terme de qualité de service (QoS) vont en s'accroissant au travers d'applications multimédia de plus en plus gourmandes telles que la télévision numérique sur ADSL ou sur IP, le streaming, la vidéo sur demande ou la diffusion en haute définition. Pour que la vidéo soit correctement décodée dans un environnement propice aux erreurs, l'encodeur, le décodeur et le réseau de communication doivent être conçus de manière optimisée et souvent conjointe.

Lorsque l'on cherche à optimiser la transmission d'un flux vidéo, il faut d'abord établir les besoins de l'application en termes de débit souhaité et de pertes tolérables. Pour des communications vidéo efficaces, la réduction du débit vidéo brut est une étape nécessaire mais pas suffisante. L'usage intensif du codage prédictif et des codes à longueur variable (« Variable Length Code », VLC) rend le signal vidéo compressé extrêmement sensible aux erreurs de transmission qui peuvent alors se propager aux informations décodées par la suite. Cela a conduit les standards de compression à structurer le train binaire en paquets délimités par des marqueurs de synchronisation au sein desquels l'erreur reste confinée.

Garantir l'intégrité de ces paquets nécessite l'usage de techniques de contrôle d'erreur dédiées. On citera notamment la retransmission et les codes correcteurs par anticipation (Forward Error Correction, FEC). Mais les limitations avérées de ces techniques (délai et anticipation du taux d'erreur) ne permettent pas d'obtenir des résultats satisfaisants dans le cadre d'une transmission vidéo sur un canal de caractéristiques variables. La majorité des recherches de ces dernières années a donc été consacrée au développement des techniques plus complexes mais plus souples telles que la protection hiérarchique d'un flux scalable. Récemment, une solution a été proposée qui permet d'obtenir des performances similaires sans impliquer la complexité et perte d'efficacité de codage associée. Le schéma SLEP (Systematic Lossy Error Protection) utilise les codes correcteurs et les principes du codage source avec information adjacente pour protéger hiérarchiquement le flux vidéo transmis sans que ce dernier ne soit modifié.

Le présent manuscrit est organisé comme suit :

Le premier chapitre présente un état de l'art des techniques de compression vidéo résistantes aux erreurs. Nous commençons par un descriptif rapide des réseaux de communication vidéo et leurs caractéristiques. Les contraintes relevées en termes de bande passante et fiabilité confirment la nécessité d'effectuer un codage source pour réduire le débit vidéo et un codage canal pour lutter contre les erreurs de transmission. Le codage de canal peut être effectué séparément du codage source ou conjointement à ce dernier. Dans ce cas, l'ajout d'informations redondantes au flux tient compte de l'importance de l'élément à protéger. C'est le cas du codage scalable qui a reçu une grande attention des chercheurs ces dernières années sans que son efficacité soit jamais satisfaisante.

Le chapitre II présente les principes et récentes applications du codage de Wyner-Ziv. En 1973, Slepian et Wolf [SLE73] ont montré que le codage séparé de deux sources corrélées pouvait être aussi performant qu'un codage conjoint de ces deux sources, si le décodage était conjoint. Des résultats équivalents pour le codage avec pertes ont été obtenus peu après par Wyner et Ziv [WYN76]. L'ensemble a conduit à un nouveau paradigme de codage dans lequel les statistiques de la source sont exploitées au décodeur. Les développements récents portent sur le codage de sources distribuées, la compression d'une source unique et le codage par couches. La dualité entre codage source avec information adjacente et codage de canal a permis le développement de techniques de protection hiérarchique d'un flux vidéo MPEG non modifié.

Les deux chapitres bibliographiques précédents nous ont amené à considérer un schéma de protection basé sur le codage de Wyner-Ziv qui permet une dégradation progressive de la qualité vidéo quand le taux d'erreur augmente, sans altérer l'efficacité du codage source. Parmi les différentes formes de scalabilité existantes, la scalabilité spatiale permet l'adaptation du format

de diffusion à différents types de récepteurs. Cette fonctionnalité est particulièrement intéressante dans le contexte actuel du passage à la HD et du déploiement de terminaux mobiles capables de diffuser de la vidéo. Le chapitre III détaille l'adaptation du schéma initial à une scalabilité spatiale et analyse son fonctionnement afin d'en optimiser les paramètres.

Finalement, le chapitre IV compare les performances obtenues par les différents flux de protection issus du schéma développé et par un schéma de protection classique, après transmission sur une liaison sans fil. La qualité de l'image affichée et la résistance aux erreurs du schéma sont évalués après application au flux généré de motifs d'erreurs modélisés, simulés et mesurés. Les effets du taux d'erreur moyen, de la longueur moyenne d'une salve d'erreur et de la taille de paquet utilisée sont commentés. Les résultats présentés démontrent les multiples intérêts de notre système.

1 Contrôle d'erreur dans les communications vidéo

La conception d'un système de transmission vidéo numérique efficace constitue actuellement un problème complexe pour diverses raisons. Tout d'abord, la bande passante du canal constitue une ressource coûteuse et limitée. Or la numérisation d'un signal vidéo analogique conduit à un volume de données binaires énorme, souvent incompatible avec le stockage et la transmission des fichiers générés. L'usage de techniques de compression qui réduisent la quantité d'information nécessaire à la description du contenu vidéo est donc indispensable. Elles seules permettent un acheminement rapide et efficace des informations sur le support de transmission choisi. Mais le flux compressé, plus compact, est aussi plus sensible aux perturbations survenant au sein de la chaîne de communication. La perte d'un élément aura ainsi un impact plus important sur la qualité de la vidéo reconstruite.

Par ailleurs, les systèmes de communication multimédia s'appuient sur l'utilisation d'une grande variété de supports physiques de transmission (lignes téléphoniques, liaisons hertziennes, ...) de caractéristiques variables en termes de taux d'erreur et de délais. Leur conception doit donc prendre en compte les erreurs de transmission potentielles en s'adaptant aux caractéristiques du support de transmission utilisé. La mise en forme du flux avant la transmission (multiplexage avec les données utilisateur, le son et préparation des unités de transmission) doit également intégrer un module de gestion de ces erreurs. En pratique, on applique un codage de canal aux données binaires résultant du codage source. On trouve essentiellement deux méthodes : la première consiste à retransmettre les informations perdues, si une voie de retour existe et si le délai induit est compatible avec l'application visée. Sinon on peut, par anticipation, adjoindre aux données compressées des données permettant la détection et la correction des erreurs. Mais cette méthode se révèle sous optimale en pratique à cause de la variabilité des conditions de transmission.

Une autre approche consiste à réaliser l'adaptation des paramètres de transmission et de compression conjointement. De cette façon, on peut modifier le codeur source afin de limiter la sensibilité du flux aux erreurs et / ou adapter le codage de canal au contenu du flux compressé.

Finalement, le succès de ces méthodes n'étant pas garanti, la solution consiste à dissimuler les erreurs résiduelles à la réception, grâce à la redondance du signal vidéo et parfois à des informations subsidiaires disponibles. Les techniques de dissimulation d'erreur ne consomment pas de bande passante, mais augmentent la complexité du décodeur.

Un système de communication implique donc typiquement cinq étapes (cf. Figure 1.1). En pratique, les données subissent une étape préalable de mise en forme (répartition en paquets) avant d'être envoyées sur le réseau de communication. La plupart des applications intègrent la mise en paquets et le codage canal dans le codeur source pour former une unique couche d'adaptation au réseau physique.

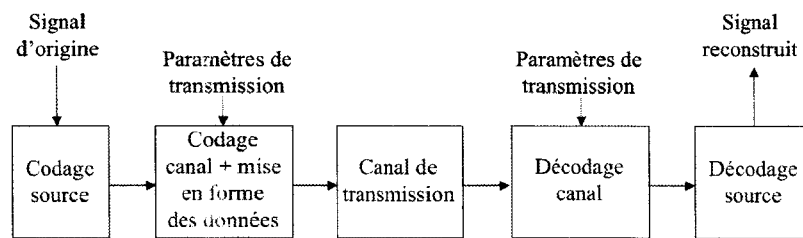


Figure 1.1 : Schéma bloc d'un système de transmission typique.

Les diverses étapes de cette chaîne de communication respectent des contraintes bien spécifiques que nous allons décrire dans la suite de ce chapitre. Après un aperçu général des principales applications de transmission vidéo nous rappelons les principes de la compression vidéo. Puis les techniques de contrôle d'erreur pour les communications audiovisuelles sont décrites. Le codage de canal et le codage conjoint source canal seront présentés et illustrés par les principales méthodes existantes dans la littérature. Finalement, en cas d'échec des méthodes précédentes, la solution consistant à dissimuler les erreurs résiduelles avant affichage est abordée.

1.1 Préambule

Les réseaux de diffusion audiovisuelle mettent en œuvre des technologies de plus en plus variées en termes d'architecture, de format de données, d'équipements terminaux, etc. Il convient de considérer cette hétérogénéité avant d'envisager la transmission de l'information vidéo à communiquer. En effet, les difficultés rencontrées dans la mise en place des services de transmission vidéo ont un impact sur la qualité d'expérience perçue par l'utilisateur final qui ne peut être négligé.

La qualité perçue est aussi fonction de l'algorithme de compression sous-jacent. D'une part la réduction de débit induit des distorsions caractéristiques et d'autre part, la structure du flux compressé le rend très vulnérable aux erreurs de transmission. Beaucoup de standards de compression actuels se basent sur la transformée en cosinus discrète (DCT). C'est le cas des

normes JPEG et MPEG qui nous serviront de modèles pour illustrer les principes la réduction du débit vidéo numérique dans la deuxième partie de ce paragraphe.

1.1.1 Réseaux de communication audiovisuelle

Dans ce paragraphe, nous décrivons les principaux types de réseaux de vidéo communication existant et les paramètres caractérisant leurs performances. Chaque type de réseaux de vidéo communication et de protocoles associés possède ses propres caractéristiques en termes de bande passante, délai, gigue et taux de pertes. Le Tableau 1.1 résume les caractéristiques des applications principales des communications vidéo concernant les standards, réseaux, tailles de paquet et type d'erreurs associé.

Application et standard	Protocole de multiplexage	Standard de codage vidéo	Débit vidéo typique	Taille de paquet	Caractéristiques des erreurs
Vidéoconférence sur ISDN (H.320)	H.221	H.261, H.263	64-384 kbps	N/A	taux d'erreur binaire = 10^{-10} - 10^{-8}
Vidéoconférence sur PSTN (H.324)	H.223	H.263	20 kbps	100 octets	Très peu d'erreurs et de paquets perdus
Vidéoconférence mobile (H.324 sans fil)	H.223 sans fil	H.263	10-300 kbps	100 octets	taux d'erreur binaire = 10^{-5} - 10^{-3} et quelques pertes de paquets
Vidéoconférence sur des réseaux à commutation de paquets (H.323)	H.225/RTP/UDP/IP	H.261, H.263, H.262	10-1000 kbps	1500 octets max	taux d'erreur binaire = 0 à 30% et pertes de paquets
TV terrestre / câble / satellite (DVB)	système MPEG2	Vidéo MPEG2	6-12 Mbps	188 octets	Presque sans erreurs, dépend des conditions météorologiques
Vidéo conférence sur ATM (H.310, H.321)	H.222.0	H.262	1-12 Mbps	53 octets (cellule)	Presque sans erreurs, taux d'erreur cellules = 10^{-6} - 10^{-4}

Tableau 1.1 : Caractéristiques des applications de communication vidéo principales [WAN02].

Le flux vidéo numérique est acheminé sur un réseau de diffusion dont les caractéristiques physiques sont imparfaites. Lors de la transmission, le signal subit des déformations liées à ces caractéristiques (fonction de transfert, bruit) et les données reconstruites au récepteur risquent de différer de celles émises : il y a alors erreur de transmission. Les erreurs de transmission peuvent être classées en 2 catégories:

- Les erreurs binaires : elles sont causées par les imperfections du canal physique de transmission. On notera que dans le cas d'une transmission par paquets, si le bit corrompu ne peut être corrigé, le paquet complet de données est le plus souvent supprimé. Le cas d'une erreur binaire se rapporte alors à un cas d'effacement.
- Les effacements sont le résultat des pertes de paquets dans les réseaux à commutation de paquets comme Internet, des salves d'erreurs dans les médiums de stockage dues à des défauts physiques ou des pannes-système de courte durée. Ils peuvent également être la conséquence d'un évanouissement ou de mauvaises conditions climatiques dans le cas de liaisons sans fil.

De plus, les applications vidéo, à l'exception du simple téléchargement, nécessitent une transmission temps réel. Dans le cas d'une application interactive en particulier, un délai de transmission excessif nuit à la qualité de service. Plusieurs facteurs contribuent au délai de transmission d'une vidéo depuis la source jusqu'à la destination [WAN02]:

- le délai de traitement du codeur, y compris l'acquisition des données dans le cas de vidéos générées en live (vidéo conférence, diffusion en direct),
- la durée de stockage en mémoire tampon à l'encodeur (pour atténuer les variations de débit du flux compressé),
- le délai de transmission (qui inclut la transmission elle-même mais surtout les mises en attente et retransmissions éventuelles),
- la durée de stockage en mémoire tampon au décodeur (pour atténuer la gigue),
- le délai de traitement du décodeur (y compris le stockage avant diffusion à taux d'image constant).

Dans certains cas, comme celui du streaming sur Internet, la source est pré-compressée, le délai d'encodage n'est donc pas une source de problème. L'usage d'une mémoire tampon au décodeur permet de relâcher encore la contrainte de délai, d'atténuer la gigue et de diffuser la vidéo à un débit d'image constant. Actuellement, les ressources (puissance de calcul, rapidité d'exécution, capacité de stockage, taille de la mémoire tampon) disponibles pour le grand public rendent raisonnable l'hypothèse que les délais liés à l'encodeur, au décodeur et aux mises en mémoire associées sont acceptables pour l'application concernée. Néanmoins, si le délai maximum autorisé par l'application est dépassé, les données seront considérées comme perdues. Le récepteur devra traiter le problème de délai comme un cas d'erreur.

1.1.2 Principes de compression vidéo

Prenons l'exemple d'une séquence vidéo brute, de structure d'échantillonnage 4:2:2, numérisée avec une précision de 8 bits par échantillon, son codage demande un débit instantané de 27 Moctets/s (ou 216 Mbits/s). Les problèmes de stockage qui en découlent sont évidents. Concernant la transmission, d'après le critère de Nyquist, la bande passante minimale nécessaire est alors de 108 Mhz. Cette valeur, à comparer avec le cas analogique qui ne requiert que 6 MHz, justifie l'usage de techniques de compression qui permettent de réduire la quantité de données binaires nécessaires à la représentation du signal vidéo.

Le codage source ou compression se fonde sur la théorie de l'information initiée par Shannon [SHA48]. On définit l'entropie de la source comme la quantité d'information moyenne qui caractérise une source d'information. Elle correspond à la limite théorique du débit minimal nécessaire et suffisant au codage sans pertes d'une source S . Le théorème du codage de source établit qu'il existe un nombre entier N tel que cette limite inférieure peut être atteinte en encodant S^N (source adjointe d'ordre N). Ce théorème sous-entend que le prix de l'optimalité est une complexité accrue. En pratique, les techniques de compression sans pertes éliminent les différents types de redondance: spatiale, temporelle, statistique et subjective inhérentes au flux vidéo et tendent vers ce résultat qui permet d'en évaluer les performances.

Dans le cas d'une compression vidéo avec pertes, dite également non conservative ou destructive, on s'attache aux informations qui possèdent un caractère exploitable et informatif pour l'œil. Les techniques de compression avec pertes suppriment les détails non perceptibles à peu visibles par le système visuel humain (SVH) et préservent les informations les plus pertinentes. Le débit minimal nécessaire est fonction de la distorsion ciblée (rate-distorsion function, notée $R(D)$).

Dans la suite de ce paragraphe, nous nous attacherons à illustrer les principes de la compression d'image et de séquences animées. Pour cela, nous prendrons comme modèles les algorithmes JPEG (Joint Photographic Expert Group) et MPEG (Moving Picture Expert Group) puisqu'ils sont les plus largement répandus dans les systèmes actuels de transmission d'images. Finalement, les effets de la compression sur la qualité perçue par l'utilisateur seront détaillés.

1.1.2.1 Compression d'image fixe

De nombreux algorithmes de compression ont été proposés dans la littérature depuis la fin des années 70 [NET95] et [BHA96]. La norme JPEG en particulier, créée il y a maintenant plus de vingt ans par l'ISO (International Standard Organisation Group), offre l'un des meilleurs rapports qualité / débit de l'image disponible [ISO92].

La première étape est la décorrélation spatiale des informations. On utilise pour cela une transformation qui permet d'exprimer les variations spatiales ou composantes fréquentielles du signal image. Cette transformation est appliquée à l'image par blocs élémentaires de 8 x 8 pixels, afin notamment de réduire la complexité d'implémentation. En pratique, la plupart des algorithmes utilisent la transformée en cosinus discrète (Discrete Cosine Transform, DCT). Les 64 coefficients de la matrice résultante traduisent la participation de motifs fréquentiels horizontaux et verticaux au contenu du bloc à coder (cf. Figure 1.2). Dans une image vidéo naturelle, les motifs de plus basse fréquence concentrent la majorité de l'énergie du bloc.

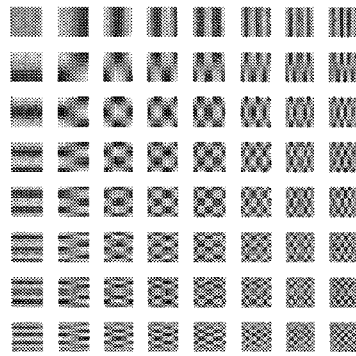


Figure 1.2 : Vecteurs de base de la DCT.

Dans le cas d'un codage destructif, une sélection des informations les plus pertinentes, basée sur des critères psychovisuels, est réalisée par le biais de la quantification. Les coefficients de plus haute fréquence (diagonales supérieures gauches), correspondant aux détails les plus fins de l'image, seront associés à un pas de quantification plus important donc codés avec une précision moindre. À l'issue de la quantification, la plupart des coefficients haute fréquence ont des valeurs très faibles ou nulles. La participation des motifs DCT au contenu visuel du bloc est approximée ce qui génère des défauts caractéristiques (effet de bloc, motifs DCT parasites) comme on peut l'observer sur la Figure 1.3 :

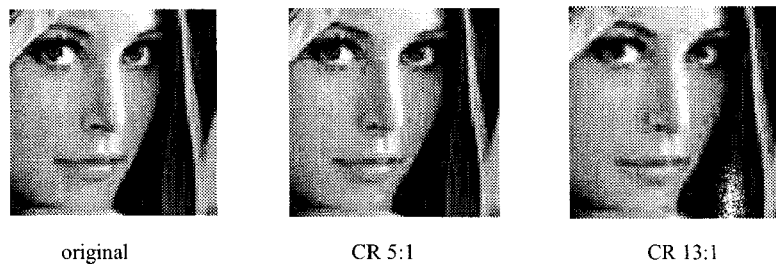


Figure 1.3 : Illustration de l'effet de la quantification sur la qualité de l'image décodée et le taux de compression (CR).

Après quantification, la composante continue (coefficient DC) du bloc est codée différemment et les 63 coefficients restants (coefficients AC) sont lus en zigzag pour sérialiser la matrice. Les coefficients non nuls sont ainsi regroupés au début du vecteur constitué. La fin du vecteur ne contient alors que des zéros, qui seront remplacés par le symbole EoB, placé à la suite du dernier coefficient non nul.

Afin de tirer profit des suites de coefficients nuls qui existent dans la série de coefficients AC du bloc après lecture en zigzag, on effectue un codage par plage (Run Length Coding, RLC). Le principe consiste à coder le nombre de coefficients nuls consécutifs et la valeur du coefficient non nul qui les suit. Pour finir, un codage à longueur variable (Variable Length Code, VLC) permet de s'assurer que les paires run-level les plus fréquentes seront associées aux mots code les plus courts. L'algorithme JPEG utilise principalement le codage de Huffman, pour lequel des tables de correspondance ont été déterminées à l'aide des statistiques générales des images à coder, bien que l'utilisateur puisse définir des tables propres à son application.

1.1.2.2 Codage du mouvement

Les travaux du groupe d'experts MPEG ont donné naissance à plusieurs normes parmi lesquelles la norme MPEG-1 [ISO93], dont le principe est repris ici afin d'expliquer le processus de compression vidéo. On peut décomposer le codage en deux parties: le codage intra-images (similaire à l'algorithme JPEG) et le codage inter-images (codage des variations entre images successives). La Figure 1.4 donne un aperçu des différents composants d'un codeur MPEG qui seront explicités dans la suite de ce paragraphe.

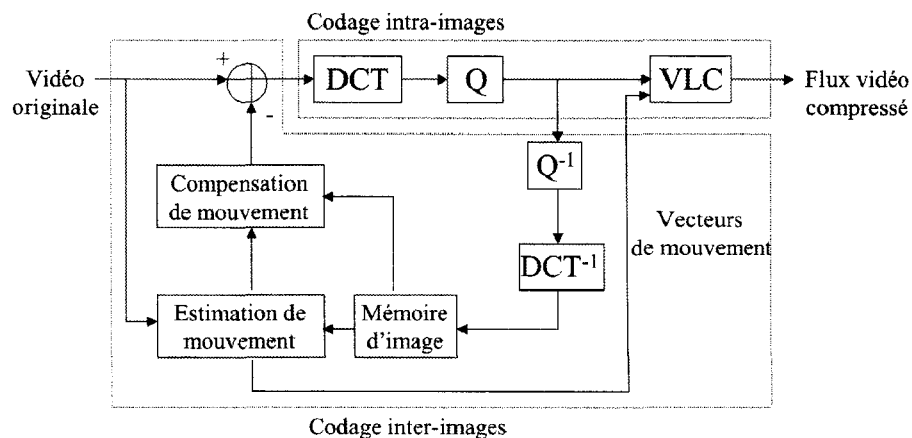


Figure 1.4 : Schéma bloc d'un codeur MPEG.

Afin d'accroître l'efficacité de la compression, le codage inter-images est utilisé pour supprimer les redondances temporelles. Le principe est de ne coder que les informations qui diffèrent entre deux images successives, correspondant aux objets en mouvement de la scène. La technique la plus couramment employée est la mise en correspondance de blocs ou « block matching » [RIB97]. La recherche des zones de variation se fait typiquement sur les blocs de 16 x 16 pixels appelés aussi macrobloc (MB): on va chercher, à l'intérieur d'une fenêtre de l'image de référence, le macrobloc de contenu le plus semblable à celui du macrobloc de l'image à coder. On en déduit le vecteur de mouvement correspondant (motion vector, MV), c'est l'estimation de mouvement. Ensuite, on calcule la différence entre le bloc à coder et son meilleur correspondant, cette erreur de prédiction est codée de façon similaire au codage intra-image et transmise avec le vecteur de mouvement. Si, pour un bloc donné, la compensation de mouvement n'a pas donné de résultat exploitable, le bloc sera codé sans référence.

Chaque séquence vidéo est découpée en groupes d'images (Group of Picture, GOP). Un groupe d'image MPEG est constitué de trois types d'image:

- les images I (intra), codées sans référence à d'autres images comme en JPEG,
- les images P (prédites), codées par rapport à une image de référence de type I ou P,
- les images B (bidirectionnelles), codées par interpolation entre les deux images de type I ou P qui l'encadrent.

On notera que le standard MPEG1 définit un quatrième type d'image : les images D, semblables aux images I mais qui ne retiennent que le coefficient DC de chaque bloc.

Deux paramètres définissent dans la norme l'enchaînement des différents types d'image:

- M est le nombre d'images entre deux images P (ou I et P) successives,
- N est le nombre d'images entre deux images I successives.

On notera que l'ordre de transmission des images doit prendre en compte cette structure afin que les données des images de référence soient disponibles au décodeur lors du décodage des images codées en inter (images P et B).

1.1.2.3 Inconvénients de la compression

Une fois les différentes étapes de compression accomplies, le débit est réduit de sorte à ne pas dépasser la bande passante du canal. En contrepartie, la qualité de l'image reconstruite est altérée. En effet, après quantification, le nombre de bits nécessaires à la description du contenu d'un bloc de 8 x 8 pixels est réduit. La quantification introduit une approximation puisque les

détails les plus fins du bloc ont été supprimés. Cette erreur également appelée “bruit de quantification” va avoir un impact considérable sur la qualité visuelle de la séquence vidéo reconstruite. Un taux de compression plus élevé peut être obtenu en appliquant une quantification plus sévère, mais les distorsions induites seront d’autant plus visibles. La diminution de la qualité de l’image reconstruite est irréversible.

L’autre effet pervers de la compression est qu’elle augmente la sensibilité du flux final aux erreurs binaires. Clairement, chaque bit du flux compressé a un poids informatif plus important. De plus, une erreur dans le train binaire peut engendrer une désynchronisation du décodeur. En effet, une simple erreur sur un bit d’un code à longueur variable peut altérer le décodage du flux jusqu’au point de synchronisation suivant, rendant les bits intermédiaires inutilisables. Enfin, un échantillon erroné peut affecter le décodage des échantillons voisins (codage différentiel de la composante continue) et des échantillons suivants (propagation temporelle due à la compensation de mouvement) comme le montrent les Figure 1.5 et 1.6 réciproquement :



Figure 1.5 : Effet de la propagation d’une erreur binaire liée à l’usage de techniques de compression sur une image fixe.

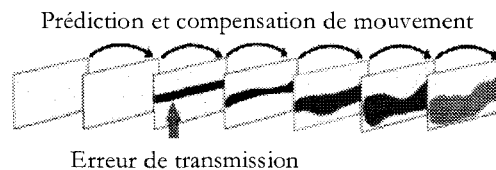


Figure 1.6 : Illustration de la propagation d’une erreur liée à l’usage de techniques de prédiction temporelle.

Le signal vidéo numérique doit subir une compression pour atteindre des débits permettant son transport sur les réseaux de communication. La compression avec pertes des données induit une dégradation de la qualité de la vidéo reconstruite. Mais la transmission du flux vidéo résultant sur un canal de transmission imparfait cause elle aussi des distorsions dans la vidéo reconstruite. De

nombreux travaux de recherche ont été entrepris depuis quelques années afin d'améliorer la qualité visuelle de l'image / vidéo reconstruite après transmission puis décompression, en limitant les distorsions dues au codage de source et au bruit du canal. La présentation de ces travaux fait l'objet des paragraphes suivants.

1.2 Codage de canal

On a vu dans le paragraphe précédent comment adapter le débit de la source à la capacité du canal, au prix d'une perte de qualité de la vidéo reconstruite et d'une aggravation de la sensibilité du flux aux erreurs de transmission. Le transport du flux vidéo sur un canal non fiable peut causer une diminution dramatique de la qualité vidéo à la réception car une erreur de transmission se propagera spatialement et temporellement. C'est pourquoi le système de transmission se doit de tenir compte de ces phénomènes en intégrant un ou plusieurs modules de gestion des erreurs.

Le signal issu du codeur source subit une mise en forme (constitution des paquets) avant d'être émis sur le médium physique de transmission. Quand la bonne délivrance des données ne peut être garantie, une étape supplémentaire est nécessaire : le codage de canal. Afin de rendre le flux compressé plus résistant au bruit de transmission, différentes techniques sont mises en oeuvre, basées sur la répétition des informations de la source. On notera que la bande passante disponible est limitée, en consacrer une partie à la protection contre les erreurs impose de réduire le débit des données utiles.

Le codage de canal se base sur le théorème du codage de canal de Shannon [SHA48] pour les canaux à perturbations. Il stipule que, pour une source de débit D inférieur à la capacité C du canal, on peut toujours trouver un code de longueur N qui permettra d'obtenir une probabilité d'erreur résiduelle au décodage aussi petite que l'on veut. En conséquence, comme pour le codage de source, si on veut un débit D aussi proche que possible de la capacité C , il faudra une longueur de code la plus grande possible. Le cas réel peut donc tendre vers le cas idéal à condition d'accepter une complexité croissante. Autrement dit, une approximation du cas idéal devrait donc avoir pour propriété que le mot code altéré reste identifiable au détecteur. Cette propriété peut être vérifiée au prix d'une certaine redondance introduite par le codage.

En pratique, cela signifie que, pour une certaine probabilité d'erreur du canal, l'ajout de la quantité adéquate de redondance, correctement exploitée au décodeur, permet la bonne réception du message malgré le bruit. Deux paramètres déterminent la quantité de redondance à ajouter : le taux d'erreur et la capacité du canal.

En se basant sur cette assertion, deux méthodes ont été dérivées afin de combattre les erreurs de transmission : l'application d'un code correcteur et la retransmission. Toutes deux consistent en une duplication des informations, donc en l'ajout de redondance. La première repose sur l'anticipation de la probabilité d'erreur et code le signal source de sorte à générer un signal redondant. La redondance est ensuite exploitée au décodeur pour détecter et corriger les erreurs d'où son nom de correction d'erreur par anticipation (FEC, Forward Error Correction). La seconde consiste à réitérer l'émission des informations corrompues. La retransmission (ARQ, Automated Repeat reQuest) a lieu en réponse à un message de non-réception du décodeur. Ces deux méthodes sont exposées dans la section suivante.

1.2.1 Retransmission sous contrainte de délai

Le schéma de retransmission conventionnel consiste en une ré-émission immédiate des données perdues à la réception d'une requête du récepteur (« Automated Repeat reQuest », ARQ, [LIN84]). Les données retransmises rejoindront le récepteur après un délai le plus souvent inacceptable pour une application temps-réel. Plus le délai d'aller-retour (Round Trip Time, RTT) est grand, moins la technique sera efficace.

Cependant, dans certains environnements (Local Area Network) les hôtes sont peu éloignés et le RTT est plus court. D'autre part, dans le cas de la diffusion ou du streaming de vidéo sur Internet, un délai de quelques secondes peut être toléré grâce à la présence d'une mémoire tampon importante au récepteur. Ces cas de figure autorisent quelques retransmissions sous contrainte de délai. Deux approches sont à distinguer : le contrôle depuis l'émetteur pour l'envoi de la requête et le contrôle depuis le récepteur pour la ré-émission des données. Le principe reste dans les deux cas d'estimer le temps de trajet et de ne tenter la retransmission que si les données peuvent arriver avant leur date d'utilisation [WUD00]. La condition généralement retenue est la suivante :

<p>À l'émetteur :</p> <p>Si $(T_C + T_0 + D_S) < T_d'(N)$, Envoyer la requête de retransmission.</p>	<p>Au récepteur :</p> <p>Si $(T_C + RTT + D_S) < T_d(N)$, Ré-émettre le paquet N.</p>
--	---

Où :

- T_C est l'instant présent,
- T_0 est la durée d'un aller simple estimée,
- RTT est la durée d'un aller-retour estimée,
- $T_d(N)$ est l'instant prévu pour la diffusion du paquet N,
- $T_d'(N)$ est une estimation de $T_d(N)$,
- D_S est la marge d'erreur.

L'usage de la retransmission requiert la mise à disposition d'une voie de retour (feedback) permettant au récepteur de communiquer le statut du paquet reçu à l'émetteur qui n'existe pas toujours. Une autre critique faite à la retransmission est qu'elle aggrave une situation de congestion. En effet, sans contrôle approprié, elle participe à la surcharge du réseau et en conséquence augmente le taux de paquets perdus.

1.2.2 Codes correcteurs « par anticipation »

Dans des environnements où le canal de retour est de capacité très limitée ou inexistant, la retransmission est inutilisable. Les codes correcteurs par anticipation ou FEC peuvent être utilisés pour améliorer ou remplacer d'autres méthodes de contrôle d'erreur. Ils sont bien connus pour la détection et la correction des erreurs et des effacements dans les communications de données [LIN83]. Il existe essentiellement deux grandes classes de codes correcteurs : les codes convolutifs ou convolutionnels [VIT71] ou [PRO95] et les codes en blocs auxquels nous nous intéressons dans la suite de ce paragraphe.

La classe des codes en blocs [PRO95] encode des messages de taille fixée k par des blocs de taille fixe n . Autrement dit, un code correcteur en bloc de paramètres (n, k) prend en entrée des segments de k mots du flux issu du codeur source, et crée typiquement en sortie un bloc de n mots. Dans la pratique, les mots correspondent généralement à des regroupements d'éléments binaires sous forme d'octets. Le codeur est dit « systématique » quand la sortie du codeur est composée des k symboles d'entrée non modifiés et de $(n-k)$ symboles additionnels redondants [RFC02]. Au décodeur, un décodage de ces n symboles permet de détecter / corriger un nombre d'erreurs limité par la capacité correctrice du code. Si celle-ci n'est pas dépassée, on pourra recréer une copie exacte des k symboles de la source.

Au niveau du récepteur, on estime que le mot de code émis, x_0 , est vraisemblablement celui se trouvant le plus près du mot reçu, $x_0 + e$. Ce procédé conduit à une erreur de décodage à chaque fois que la distance de Hamming d_0 entre le code émis et le code reçu est supérieure à la distance d_1 séparant le code reçu et un code erroné x_1 . Autrement dit, la détection échouera quand l'erreur e est telle que $d(x_0, x_0 + e) > d(x_1, x_0 + e)$ comme le schématise la Figure 1.7.

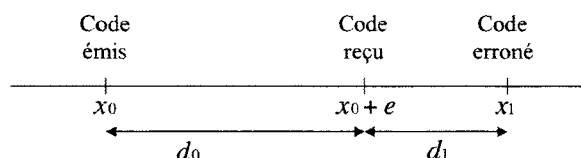


Figure 1.7 : Illustration du principe d'estimation de la valeur du symbole reçu.

Cela ne peut pas se produire si le poids de e est inférieur à la capacité de correction du code notée t , avec:

$$t = \left\lfloor \frac{\min_{x_1 \neq x_0 \in C} d(x_1, x_0)}{2} \right\rfloor = \left\lfloor \frac{d_{\min}}{2} \right\rfloor, \text{ où } d_{\min} \text{ s'appelle la distance minimale.}$$

La capacité de correction est donc le poids en dessous duquel toutes les erreurs peuvent être corrigées. Dans le cas d'un code en blocs, on vérifie que : $d_{\min} = n - k$ et $t = \left\lfloor \frac{n - k}{2} \right\rfloor$.

Les codes correcteurs peuvent être appliqués directement aux bits du flux compressé (cf. Figure 1.8a) et sont alors efficaces contre les erreurs binaires. Mais si les erreurs se produisent par salves ou si un paquet complet est perdu, cette protection s'avèrera inutile. En effet, il est beaucoup plus difficile de restaurer le bloc puisque l'erreur affecte des centaines de bits consécutifs, dépassant la capacité correctrice du code. Habituellement, pour lutter contre ce type d'erreur, on applique le FEC à travers les paquets ce qui équivaut à un brassage (ou entrelacement) de degré la taille du paquet. À la réception, le désentrelacement permet de répartir les erreurs de sorte à ce qu'un paquet perdu ne représente plus qu'un symbole du bloc protégé (cf. Figure 1.8b). On notera que la quantité d'information devant être stockée avant de procéder au codage du bloc augmente, ce qui induit un délai supplémentaire.

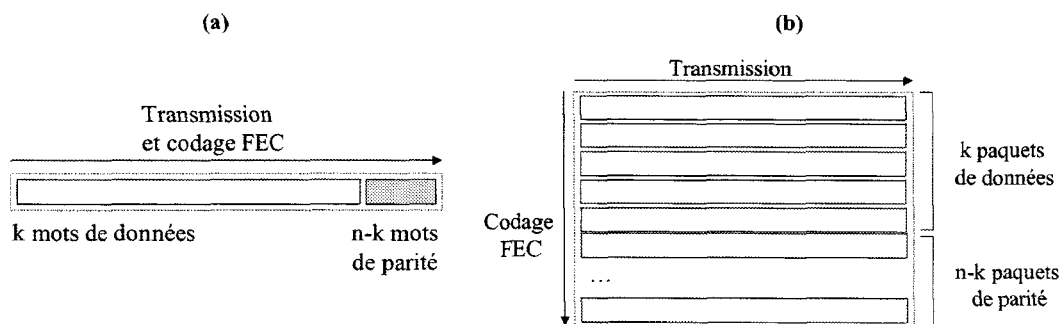


Figure 1.8 : Modes d'application du FEC : (a) aux données binaires et (b) à travers des paquets de données.

Il existe des moyens plus sophistiqués pour ajouter des symboles redondants sous forme de parité. Par exemple, si les k symboles sont organisés en matrice de taille $p \times p$, avec $p = \sqrt{k}$, alors on peut appliquer le code correcteur aux lignes et aux colonnes de la matrice ce qui constitue un code produit (product code, [STA03]). Chaque ligne (colonne) de la matrice sera alors reconstituée à partir d'un sous-ensemble de p symboles horizontalement (verticalement). Ce procédé est extensible à plus de deux dimensions, néanmoins l'usage de ces constructions doit être adapté aux motifs de pertes qui seront rencontrés. Les codes correcteurs peuvent aussi

être appliqués adaptivement en fonction de l'importance des éléments à protéger, on parle alors de protection aux erreurs inégale (cf. Chapitre I, paragraphe 3.2.2).

Certaines couches protocoles de bas niveau du réseau (couche de liaison) peuvent procéder à contrôle d'intégrité des paquets et à la suppression des paquets corrompus, avant le décodage FEC. Dans ce cas, la position des paquets perdus est connue et l'on parle de décodage à effacement (erasure decoding, [RIZ97]). La connaissance de la position des symboles en erreur permet de doubler la capacité correctrice du code.

La quantité et la répartition des erreurs de transmission dans le flux reçu n'est pas connue du codeur. En l'absence d'erreurs, cette protection sera superflue et les ressources mal utilisées. Si la quantité d'erreurs dépasse les prévisions, la protection se révélera insuffisante, voire inefficace. Dans la mesure où la qualité de service vidéo est fortement dépendante de la constance de la qualité affichée, cette solution n'est pas satisfaisante.

1.3 Codages conjoints source canal

Un codeur source doit compresser l'information autant que possible pour un niveau de distorsion donné (cf. Chapitre I, paragraphe 1.2) et, ensuite, le codeur de canal ajoute de la redondance au flux compressé de sorte à ce que le débit total (codage source + canal) ne dépasse pas la capacité du canal. Le théorème de séparation de Shannon établit que codage canal et codage source peuvent, sous certaines conditions, être conçus séparément sans perte de performances [SHA48]. Cependant, un tel système n'est optimal que sous les hypothèses suivantes :

- des tailles de bloc infinies (donc un délai et une complexité infinis),
- la stationnarité de la source et du canal,
- l'égale importance de tous les symboles émis.

Ces hypothèses ne sont pas vérifiées en pratique et ce modèle s'avère très vite sous optimal. C'est pourquoi le codage conjoint source canal est souvent un schéma plus viable. Dans ce cas, le codage canal est effectué conjointement au codage source. Une coopération entre codec et protocoles-réseau est également envisageable. Le codage conjoint permet d'optimiser globalement la chaîne de communication grâce à la prise en compte des caractéristiques du canal dans la conception des algorithmes de codage source ou des caractéristiques de la source dans la conception des algorithmes de codage canal. Ces méthodes permettent d'améliorer la résistance du flux vidéo en contrôlant sa sensibilité aux erreurs de transmission. Elles font en général

l'objet d'une option dans un mode ou un profil spécifique des standards de compression actuels, le Tableau 1.2 en liste quelques implémentations.

Standard	FEC	Mots de synchronisation	Codes à longueur variable réversibles	Data Partitioning	Sélection de l'image de référence
H.263+	oui	oui	oui	non	oui
H.26L	oui	oui	oui	oui	oui
MPEG-2	non	oui	non	oui	oui
MPEG-4	non	oui	oui	oui	oui

Tableau 1.2 : Quelques outils de codage robuste aux erreurs disponibles dans les standards actuels de compression vidéo.

Nous donnons, dans la suite de cette section, quelques exemples pratiques de mise en œuvre de ces techniques et leur éventuelle intégration dans les normes MPEG2 [ISO94], MPEG4 [ISO04] et H263 [ITU98].

1.3.1 Modification du codage de source

Dans les paragraphes suivants, nous décrivons différentes méthodes de codage où l'efficacité du codeur de source est volontairement diminuée de sorte à générer un flux dont la sensibilité aux erreurs sera atténuée. Une redondance résiduelle choisie est tolérée qui limite la propagation spatiale / temporelle des erreurs, ou insérée périodiquement pour permettre au décodeur de se resynchroniser plus rapidement.

1.3.1.1 Isolement de l'erreur

La sensibilité du flux vidéo codé aux erreurs de transmission est principalement due à l'usage des codes à longueur variable. Une erreur sur un bit d'un mot code peut affecter le décodage des mots codes suivants, même si ceux-ci sont reçus correctement. Les techniques d'isolement d'erreur ont pour but de réduire l'effet d'une erreur de transmission à une région limitée dans le temps ou l'espace. La philosophie est essentiellement de compartimenter les mots du code: on peut insérer des marqueurs de resynchronisation dans le flux compressé et / ou segmenter les données. Ces techniques ont été intégrées dans les normes MPEG-4 et H.263.

Une approche simple et efficace pour améliorer la résistance du flux aux erreurs est l'insertion périodique de marqueurs de resynchronisation qui ont la propriété d'être facilement distingués des mots code du flux, même si ces derniers sont corrompus (cf. Figure 1.9). Habituellement, les

en-têtes sont attachés immédiatement après ces informations de resynchronisation, ainsi le décodage peut reprendre dès qu'un marqueur est détecté. Des marqueurs plus longs et plus fréquents permettront au décodeur une resynchronisation plus rapide, limitant la taille de la région affectée dans l'image reconstruite. À l'évidence, ces insertions réduisent l'efficacité de codage en augmentant la quantité de bits à transmettre, un système de codage vidéo doit donc en faire usage avec les précautions nécessaires. La syntaxe du flux MPEG2 prévoit l'insertion de marqueurs de synchronisation de 32 bits de longueur conçus de sorte à rester identifiables en cas d'erreur sur le flux [CHE95].

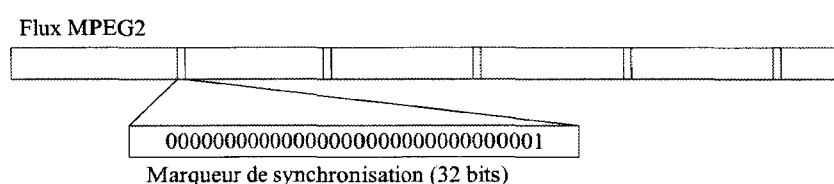


Figure 1.9 : Insertion de marqueurs de synchronisation dans le flux MPEG2.

Les données situées entre deux marqueur de resynchronisation peuvent être segmentées en unités logiques plus fines. Par exemple, les modes de codage résistant des standards MPEG4 et H.263 utilisent des marqueurs secondaires qui séparent les en-têtes de macrobloc et les vecteurs de mouvement des coefficients DCT des macroblocs d'une tranche ou d'un GOB. En cas d'erreur sur l'unité contenant les coefficients DCT, l'en-tête et les informations de mouvement pourront ainsi être décodées. Le standard MPEG2 autorise la découpe d'une image en tranches, chaque tranche étant codée indépendamment du reste de l'image et précédée d'un start code [CHE95].

1.3.1.2 Codage binaire robuste

En addition à l'isolation des erreurs, on peut aussi directement modifier les méthodes d'encodage binaire de sorte à générer un flux plus résistant aux erreurs de transmission. Nous présentons ci-après deux méthodes de ce type.

Les codes à longueur variable réversibles (Reversible Variable Length codes, RVLC, [WEN98]) permettent au décodeur de décoder non seulement les informations consécutives à un marqueur de resynchronisation (MR) mais aussi les informations précédant le marqueur de resynchronisation suivant comme le montre la Figure 1.10.

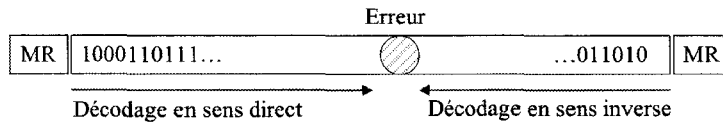


Figure 1.10 : Illustration du décodage des codes à longueur variable réversibles.

Si les codes à longueur variable réversibles sont utilisés, le décodeur pourra récupérer un plus grand nombre de bits correctement reçus et la zone de l'image affectée par l'erreur sera réduite. Ceci permet une meilleure localisation des erreurs au prix d'une légère augmentation de la complexité du codec. Les codes à longueur variable réversibles ont été adoptés par MPEG4 et H.263 conjointement à l'usage des marqueurs de resynchronisation et la segmentation des données.

La seconde méthode que nous citerons repose sur une réorganisation des codes issus du codeur entropique. Le codage entropique résistant aux erreurs (Error Resilient Entropy Coding, EREC, [RED96]) réordonne les bits codés de sorte à permettre au décodeur de se resynchroniser au début de chaque bloc. Plus spécifiquement, les blocs issus du codage à longueur variable de macroblocs ou GOB sont distribués dans des emplacements (slots) de taille fixe. Initialement, chacun des blocs de longueur variable est placé dans un compartiment puis les données des blocs de longueur supérieure à celle des slots sont réparties dans les slots incomplets selon une séquence de décalage prédéfinie jusqu'à ce que tous les bits soient inclus (cf. Figure 1.11). La taille constante des blocs de données obtenus permet au décodeur de se resynchroniser au début de chaque bloc. De plus, comme EREC n'insère aucun bit dédié à la synchronisation, la redondance générée est négligeable.

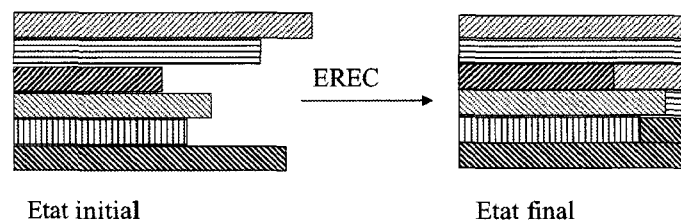


Figure 1.11: Réorganisation des bits du flux codé par EREC.

1.3.1.3 Prédiction résistante aux erreurs

Dans les paragraphes précédents, nous nous sommes intéressés aux techniques de codage robuste qui combattent la propagation des erreurs liée à l'usage des codes à longueur variable. Une autre cause de propagation des erreurs de transmission est l'usage de la prédiction

temporelle. En effet, si la reconstruction d'une image diffère de l'image dont l'encodeur dispose, les images de référence à partir de ce point seront différentes de celles présentes à l'encodeur, en conséquence, toutes les images reconstruites seront dégradées. L'utilisation de la prédiction spatiale pour le coefficient DC et les vecteurs de mouvement cause également une propagation de l'erreur mais restreinte au contenu d'une même image.

Les techniques de prédiction résistantes aux erreurs contraignent la boucle de prédiction de sorte à limiter la propagation de l'erreur à un court intervalle de temps. Une méthode consiste à insérer périodiquement dans le flux des images ou macroblochs codés en intra, donc sans référence. Le rafraichissement intraframe périodique a été intégré au standard MPEG2 [CHE95]. Le codage d'une image entière en mode intra augmente le débit, le codage intra de tranches ou de macroblochs est plus ajustable et suffisamment efficace. Deux paramètres sont à déterminer en fonction des conditions de transmission: le nombre de macroblochs à coder en intra et leur répartition spatiale [STU00].

Une autre approche pour limiter la portée de la propagation d'erreur est de diviser les données en segments et de n'autoriser la prédiction qu'au sein d'un même segment. Ainsi, une erreur sur un segment n'affectera pas les autres segments. Par exemple, une image peut être divisée en plusieurs régions (tranches ou GOB) et la région 1 ne peut être prédite qu'à partir des informations de la région 1 de l'image précédente. Similairement, il est possible de séparer les images d'indice pair des images d'indice impair. Ces méthodes ont été incorporées dans le standard H.263 sous les noms respectifs « independant segment decoding » et « redundant video coding ».

1.3.1.4 Codage à descriptions multiples

Une approche alternative consiste à utiliser le codage à descriptions multiples (Multiple Description Coding, MDC) : plusieurs flux sont alors générés à partir de la vidéo originale et transmis sur des canaux séparés (cf. Figure 1.12). À la destination, la qualité du signal reconstruit sera fonction du nombre de descriptions reçues : acceptable si une seule description est disponible et incrémentée pour chaque description supplémentaire.

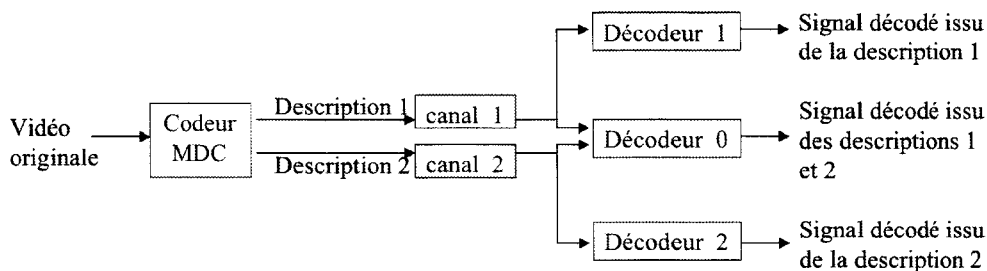


Figure 1.12 : Schéma bloc d'un codec à descriptions multiples.

Pour que chaque description fournisse à elle seule une qualité visuelle acceptable, toutes les descriptions doivent contenir les informations fondamentales de la source, elles sont donc fortement corrélées. Cette corrélation est la source de la redondance inhérente au codage MDC, donc de son inefficacité en l'absence d'erreurs de transmission. Mais elle permet surtout de ne pas dépendre de la capacité du réseau à fournir une communication fiable. En effet, si toutes les descriptions ont une importance équivalente, il n'est pas nécessaire de fiabiliser la transmission de l'une d'entre elles en particulier. Les approches de mise en œuvre du MDC sont nombreuses [WAN05] et incluent par exemple l'échantillonnage temporel entrelacé qui a été intégré à la norme H.263 sous l'appellation de « redundant video coding » déjà évoqué auparavant.

1.3.2 Modification du codage de canal

Dans les paragraphes suivants, nous listons des méthodes de codage dont l'objectif principal est d'appliquer une protection plus importante aux informations considérées comme fondamentales pour le décodage et l'affichage du contenu vidéo. L'application de ce codage de canal hiérarchique implique un partitionnement des données du flux préalable. Cette approche permet d'obtenir une qualité vidéo décodée variant en fonction des conditions de transmission.

1.3.2.1 Mise en paquet résistante aux erreurs et multiplexage

La perte d'un paquet d'information peut avoir différentes conséquences sur la vidéo reconstruite selon la méthode de mise en paquet utilisée. Il est important de s'arranger pour qu'une erreur puisse être isolée dans une région la plus limitée possible. Autrement dit, le flux compressé ne doit pas être réparti à l'aveugle dans des paquets de tailles identiques. Il est préférable de construire les paquets en fonction de l'algorithme d'encodage sous-jacent, pour qu'un paquet contienne un (ou plus) blocs de données codés indépendamment [CLA90]. Par exemple, un paquet devra coïncider avec une ou plusieurs tranches (slices), avec une possible répétition de

l'en-tête d'image au début de chaque paquet. Pour une vidéo codée en MPEG2, la taille des tranches en nombre de macroblochs inclus peut être fixe ou adaptative de sorte à permettre une mise en paquet efficace.

Dans des réseaux qui autorisent des tailles de paquets variables, utiliser une taille maximale permet de minimiser le surcoût relatif aux en-têtes et au processus de mise en paquet. Dans le cas d'Internet, la taille de paquet maximale (« Maximum Transmission Unit », MTU) la plus fréquente est 1500 octets. En général, la suite de protocoles RTP/UDP/IP est utilisée pour le transport de vidéo sur Internet. Les spécifications RTP incluent la définition des formats de données à utiliser (cf. Annexe I et [RFC98]). Afin de limiter le surcoût lié à l'ajout des en-têtes, en particulier dans le cas d'applications bas débit, on peut utiliser des techniques de compression d'en-tête [CIV00].

Pour prévenir la perte de blocs de données contigus des suites de la perte d'un unique paquet, on peut procéder à un brassage des données avant leur mise en paquet. Les données de blocs adjacents seront alors placées dans des paquets différents, de cette façon un bloc perdu sera entouré de blocs reçus et la dissimulation d'erreur sera facilitée. Un exemple est la mise en paquets séparés des trames paires et impaires d'une séquence d'image [VIL99]. Notons que le brassage à la couche transport impose au codeur source de limiter la prédiction aux blocs qui seront mis en paquet séquentiellement.

Un autre facteur qui peut influencer les caractéristiques d'une communication vidéo est le multiplexage des données vidéo et autres (son, contrôle) et la protection des en-têtes. En effet, une erreur sur l'en-tête de multiplexage peut mener à une mauvaise distribution des données vidéo, donc à la perte d'un large morceau de données au récepteur. Pour éviter ce genre d'incidents, chaque paquet est précédé d'un long mot de synchronisation et l'en-tête est fortement protégé par un code correcteur. Le standard H.223 de multiplexage pour les communications multimédias bas débit en est un bon exemple [ITU96].

1.3.2.2 Transmission hiérarchique et protection inégale aux erreurs

Les bits d'un flux vidéo compressé n'ont pas tous la même importance. Par exemple, l'altération de l'en-tête d'image est bien plus néfaste que celle du contenu d'un bloc car elle empêcherait le décodage de l'image complète, de même le contenu d'une image I aura plus d'impact sur la qualité de la vidéo décodée que celui d'une image P ou B. Quand un codage scalable est utilisé, la couche de base contient les informations fondamentales sur le contenu vidéo alors que les couches d'amélioration ne font qu'y apporter des précisions.

Quand le contrôleur de transport ou l'émetteur peut assigner une priorité appropriée à différents éléments d'un flux de données, on parle de protection inégale aux erreurs (Unequal Error Protection, UEP). Les informations considérées comme prioritaires sont alors délivrées avec un taux d'erreur bien moindre, ce qui est une forme de contrôle d'erreur qui tient compte de la structure de la source. Différents systèmes de transmission peuvent implémenter cette hiérarchisation de différentes façons.

La hiérarchisation de la transmission peut intervenir lors de la modulation. Une modulation multiporteuse peut être birésolution, elle transporte alors deux flux en utilisant une constellation unique. En augmentant la distance entre les nuages de points de la constellation, la détection de la valeur des bits codant le cadrant est améliorée au détriment de celle des bits codant la position dans le cadrant. La Figure 1.13 présente l'exemple d'une 4/16 QAM qui permet la transmission de 2 bits de haute priorité et 2 bits de basse priorité par symbole de 4 bits [GOU05].

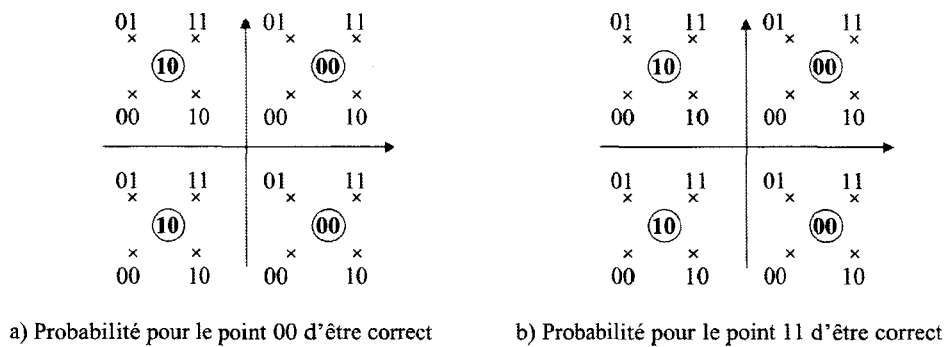


Figure 1.13 : Illustration des zones de décision pour une 4/16 QAM pour a) les bits de haute priorité et b) les bits de basse priorité.

La priorisation du transport peut aussi être réalisée par l'utilisation de plusieurs niveaux de puissance d'émission dans un réseau sans fil. Dans les réseaux ATM, il existe un bit dans l'en-tête de chaque cellule qui code la priorité et permet, quand une congestion réseau se produit, au noeud de décider quelles cellules seront supprimées en premier. Sur Internet, les spécifications RTP définissent des mécanismes génériques pour le transport d'un flux vidéo scalable [RFC96]. Un autre moyen de fournir différents niveaux de protection est d'utiliser différents protocoles de transport, par exemple TCP pour une petite quantité d'information haute priorité (comme le contrôle de session) avant le début de la session et UDP pour l'information basse priorité restante pendant la session [CIV01].

Quel que soit le type de réseau, la priorisation peut être réalisée par l'adaptation des paramètres de la technique contrôle d'erreur utilisée. Par exemple, on peut mettre en place une protection par FEC forte et des retransmissions pour les informations fondamentales et une protection par

FEC plus faible, sans retransmission, pour le reste du flux. Le nombre de tentatives de retransmission et de copies du paquet par tentative peut être déterminé en fonction de l'importance des données. Dans certains cas extrêmes, on pourra même dupliquer les informations les plus importantes. Une double transmission de l'en-tête d'image et de la matrice de quantification a été proposée pour les vidéos MPEG [SUN94].

1.3.2.3 Codage scalable et UEP

Le bénéfice de la protection hiérarchique est basé sur la séparation du contenu vidéo en plusieurs flux d'importances différentes et sur la capacité du système de transmission à garantir la bonne réception de celui contenant les informations indispensables au décodage de la vidéo. Dans un codeur scalable, un de ces flux permet de reconstruire une image cohérente, d'une qualité minimale garantie. Un codeur scalable peut être à granularité grossière (cas de deux ou trois flux distincts) ou fine (accès à une qualité qui est fonction du nombre de bits reçus).

L'approche la plus souvent retenue pour la réalisation du codage scalable est le codage par couche (« Layered Coding », LC). Dans le cas birésolution, la couche de base (« base layer ») est identifiée comme contenant les informations les plus importantes, elle fournit une qualité faible mais acceptable ; le flux d'amélioration (« enhancement layer ») contient les informations restantes et permet d'incrémenter le niveau de qualité décodé. La couche de base est identique au flux généré par un codeur monocouche de bas débit / basse résolution, elle peut donc être décodée indépendamment des couches d'amélioration, la réciproque étant fautive.

Pour fournir une résistance accrue aux erreurs par rapport à un codeur classique, le codage scalable doit donc être couplé à une protection inégale contre les erreurs (Unequal Error Protection, UEP). Ainsi, les deux types de flux issus d'un codage par couche sont transmis de sorte à ce que les erreurs soient concentrées sur les couches d'amélioration et que la couche de base ne soit pas corrompue (cf. Figure 1.14). Par exemple, on peut combiner codage scalable et retransmission à priorité en plaçant les paquets de la couche de base en tête de la file de transmission pour augmenter le nombre de tentatives de retransmission et retransmettre à chaque tentative plusieurs copies du paquet de la couche de base alors qu'un paquet de la couche d'amélioration ne sera pas retransmis [SMI94].

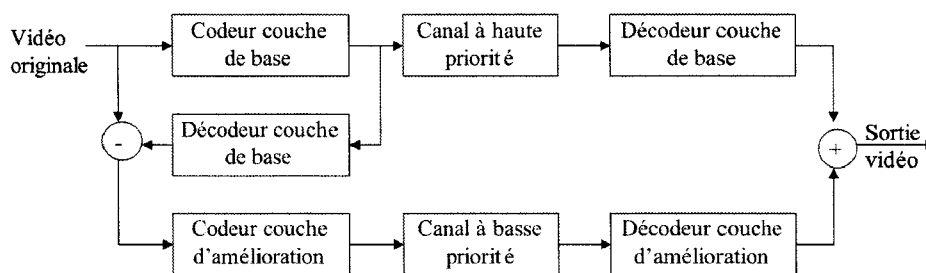


Figure 1.14 : Schéma bloc d'un codeur par couches avec transport hiérarchique.

La scalabilité permet l'adaptation à différentes bandes passantes, caractéristiques d'erreur ou puissance de calculs du récepteur, en faisant varier le nombre de couches transmises et / ou la protection qui leur est attribuée.

Les différentes implémentations de la scalabilité incluent typiquement:

- Scalabilité de qualité ou SNR : elle fournit différentes résolutions d'amplitude grâce à une quantification de pas variable (dans le domaine original ou transformé).
- Scalabilité spatiale : elle fournit différentes résolutions spatiales ou tailles d'image. La couche de base est obtenue par sous échantillonnage de l'image originale et les couches d'amélioration contiennent la différence entre l'original et la version interpolée de la couche de base.
- Scalabilité temporelle : elle fournit différentes résolutions temporelles ou fréquences d'affichage des images. Le principe est le même que pour la scalabilité spatiale à la différence que l'on utilise un sous-échantillonnage temporel et non spatial des données.

Pour éviter la propagation des erreurs, la couche de base sert de référence pour la prédiction et la compensation de mouvement, ce qui a pour effet de diminuer l'efficacité du codeur, déjà réduite par la transmission multiple des en-têtes (un en-tête par couche). Dans la couche supérieure, la prédiction de mouvement peut être déduite de celle de la couche inférieure. Les performances sont alors légèrement dégradées, due à la restriction imposée à la prédiction.

L'intérêt des codeurs scalables dans un environnement relativement sans erreurs est limité par les performances du codec scalable lui-même. Si le taux d'erreur est grand, alors c'est la qualité de la couche la plus basse seule, sans aucune couche d'amélioration, qui limite la performance. Les modes de scalabilité SNR, temporel et spatial font partie des options du standard MPEG2.

1.3.2.4 Le data partitionning

Les informations au sein d'un flux vidéo n'ont pas la même importance. Les vecteurs de mouvement, par exemple, ont plus d'impact sur la qualité de la vidéo décodée que les coefficients DCT, et les coefficients DCT basse fréquence ont plus d'importance que les coefficients haute fréquence. Le partitionnement des données permet de diviser un flux vidéo non scalable en deux parties d'inégale importance et peut, de ce point de vue, être considéré comme apportant une forme de scalabilité. Le flux codé est simplement séparé au niveau du point de coupure (Priority Break Point, PBP) en deux trains (cf. Figure 1.15):

- le train de base (TB) comprend les premiers coefficients du flot quantifié et codé et autres informations (MV, type de MB, ...),
- le train complémentaire (TC) contient les autres coefficients du bloc et le EoB.

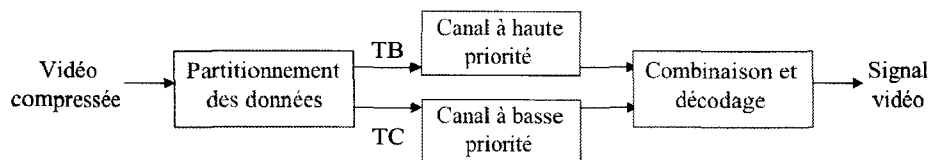


Figure 1.15 : Schéma bloc d'une transmission en mode data partitionning.

Le data partitionning n'est pas vraiment un codage hiérarchique mais, comme pour le codage scalable, le train de base fournit une qualité minimale pouvant être améliorée par le train complémentaire. Si la transmission peut être effectuée de sorte à ce que la couche basse ne subisse aucune erreur mais que ces erreurs frappent la couche d'amélioration, on obtient une meilleure qualité. On peut par exemple appliquer une modulation hiérarchique aux bits du TB et du TC [GOU05]. L'avantage de ce procédé est qu'il n'introduit qu'un léger surcoût de débit, donc sa performance dans le cas d'une transmission sans erreur est proche de celles d'un codage non-scalable. Le data partitionning a été intégré comme option de codage à la norme MPEG2.

1.4 Dissimulation d'erreur

Comme mentionné dans les sections précédentes, le flux vidéo compressé peut être endommagé pendant la transmission à cause du manque de fiabilité des réseaux de communication. Les techniques de contrôle d'erreur précédentes visent à minimiser l'effet d'erreurs de transmission

sur la qualité de la vidéo décodée. Mais il est impossible de garantir le succès de ces techniques, erreurs et pertes résiduelles existeront toujours au décodeur.

Quand un échantillon (ou bloc d'échantillons) de l'image est manquant au décodeur, celui-ci tente d'en estimer le contenu. Ces méthodes relèvent de la dissimulation d'erreur, elles ne consomment pas de bande passante, mais augmentent la complexité du décodeur. Elles se basent sur l'utilisation de la corrélation temporelle et / ou spatiale entre échantillons voisins et peuvent opérer aussi bien dans le domaine spatial que dans le domaine fréquentiel.

1.4.1 Principe général de la dissimulation d'erreur

Dans cette section, nous nous intéressons aux techniques permettant de dissimuler l'impact d'une erreur de transmission sur la qualité visuelle perçue. La dissimulation d'erreur repose sur :

- l'exploitation des redondances statistiques résiduelles dans la séquence vidéo compressée qui persistent à l'issue du codage de source,
- la tolérance du système visuel humain à un certain degré de distorsion.

La dissimulation des erreurs implique tout d'abord leur détection qui peut intervenir au niveau transport ou au décodeur [ZHU00]. Au niveau transport, deux types d'outils peuvent aider à la détection des pertes de données. Dans le cas d'une communication basée sur la transmission de paquets, l'attribution d'un numéro de séquence (protocole de transport RTP) permet l'identification des paquets perdus. Alternativement, lors du décodage FEC [LIN83], l'intégrité du flux peut être confirmée et dans le cas contraire, l'erreur sera localisée. Au niveau du décodeur, la connaissance de la syntaxe du flux permet de repérer les aberrations (valeurs d'échantillons hors domaine, nombre d'éléments décodés incorrect) qui n'ont pas mené à un échec du décodage (VLC inexistant). D'autre part, dans la mesure où les erreurs de transmission touchent le flux compressé, les artefacts résultants dans le domaine spatial sont particuliers, des métriques de qualité spécifiques peuvent être employées pour les quantifier [HHU06]. Dans ce cadre de détection, les méthodes de dissimulation s'appliquent autant aux artefacts de codage qu'aux artefacts de transmission.

Nous nous limitons dans la suite de cette section aux méthodes de dissimulation d'erreur qui ne nécessitent pas d'interaction entre encodeur et décodeur. Si tel est le cas, leur coopération permet d'améliorer considérablement les performances grâce à l'usage de la retransmission des informations perdues (cf. paragraphe 1.3.1.1) ou la modification du comportement de l'encodeur ou du codeur de transport [WAN98].

Une fois les zones à traiter répertoriées, il faut choisir la méthode de dissimulation adéquate. L'efficacité d'une méthode de dissimulation dépend des caractéristiques de la scène à restaurer :

- la taille de la région endommagée (du simple pixel à plusieurs images consécutives),
- le contenu de cette région (quantité de détails, corrélation avec le voisinage, mouvement).

Comme précisé dans les sections précédentes, un unique bit en erreur peut corrompre une région entière de la séquence à cause de l'usage des codes à longueur variable. La taille de la région endommagée des suites de la propagation d'une erreur de transmission est fonction des techniques de robustesse utilisées et de leur paramétrage. Si l'erreur affecte une partie du contenu d'un macrobloc, la dissimulation pourra estimer les informations manquantes à partir des macroblocs voisins dans les domaines spatial, temporel ou fréquentiel [ZHU00]. Dans le cas où une image entière (et les images du GOP prédites à partir de celle-ci) est affectée, les images précédentes et / ou suivantes peuvent être utilisées. Le standard MPEG2 permet la dissimulation d'erreur par prédiction spatio-temporelle et par substitution depuis l'image de référence précédente. Bien sûr, les informations des images adjacentes peuvent être utilisées en plus des blocs voisins disponibles ce qui permettra une estimation plus fine. La Figure 1.16 montre l'amélioration de la qualité visuelle d'une image diffusée grâce à une dissimulation d'erreur, même simpliste :

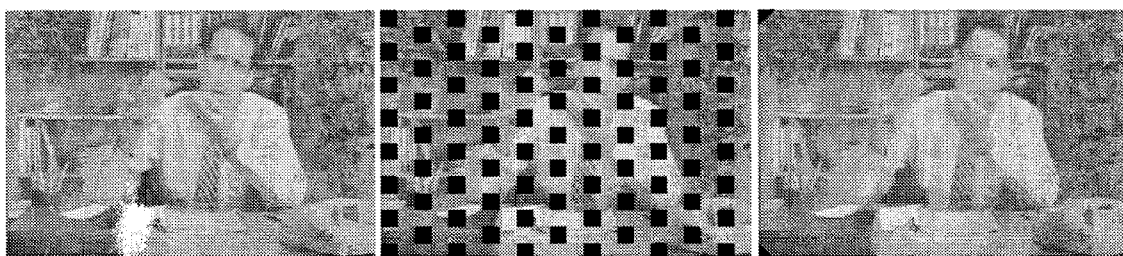


Figure 1.16 : Illustration de la dissimulation par interpolation spatiale à l'aide d'un flitre médian : image originale, corrompue et après dissimulation.

La présence / l'absence de mouvement dans la scène à restaurer est un autre paramètre à prendre en compte dans le choix de la technique de dissimulation à employer. La dissimulation par prédiction temporelle ou remplacement de bloc est efficace dans les zones statiques de l'image. Pour les zones peu animées de l'image, le bloc de substitution peut être désigné par le vecteur de mouvement, s'il est disponible. Mais dans les zones à fort mouvement, une interpolation spatiale à partir des blocs avoisinants donne de meilleurs résultats.

Les techniques de dissimulation peuvent aussi être plus raffinées et chercher à restaurer le flux en prenant en compte sa structure. Par exemple concernant les codeurs basés sur l'application d'une DCT à des en blocs de l'image et sur la compensation de mouvement, un macrobloc est caractérisé par trois types d'information :

- la texture du contenu, incluant les valeurs des coefficients DCT du bloc codé ou de l'erreur de prédiction,
- les informations de mouvement essentiellement par le biais des vecteurs de mouvement des images prédites
- le mode de codage du bloc (I, P, B).

Des techniques de dissimulation d'erreur spécifiques ont été définies qui permettent d'estimer ces différents types d'information, nous en présentons quelques exemples dans le paragraphe suivant. Ces méthodes peuvent également être combinées pour une meilleure efficacité.

1.4.2 Méthodes de récupération de la texture

Toutes les techniques développées se basent sur la propriété de variation régulière des signaux image et vidéo naturels. Nous nous sommes limités ici aux techniques les plus simples comme l'interpolation spatiale et l'interpolation temporelle des valeurs des pixels. Le problème de ces techniques réside dans la sélection du filtre d'interpolation à utiliser.

1.4.2.1 Algorithme de variation maximale régulière (Maximally Smooth Recovery, MSR) [WAN93] :

Les données voisines d'un flux vidéo varient le plus souvent avec régularité dans les domaines spatial et temporel. Les valeurs à restaurer sont obtenues en minimisant les mesures des différences d'intensité entre la valeur estimée et les valeurs du voisinage bloc par bloc dans les deux domaines. La mesure utilisée peut être du premier ordre pour limiter la complexité de l'algorithme, mais on peut utiliser une mesure d'ordre supérieur pour limiter l'effet de flou qui touche alors les contours éventuellement présents dans le bloc perdu.

1.4.2.2 Algorithme de projection sur des ensembles convexes (Projection onto convex sets, POCS) [SUN95] :

Cette méthode utilise les huit blocs autour du bloc perdu pour déterminer s'il s'agit d'un bloc monotone ou d'un bloc contenant un bord. Dans le deuxième cas, l'orientation du bord est quantifiée et ramenée à une des huit directions principales et un filtre passe-bande correspondant à cette direction est appliqué dans le domaine transformé. Si le bloc est monotone, on lui applique un filtre passe-bas. Ensuite, les valeurs du bloc dans le domaine spatial sont tronquées à $[0, 255]$. Ces deux étapes sont répétées jusqu'à ce que le résultat soit stable.

Ces méthodes ont quelques limitations non négligeables. Tout d'abord, elles supposent que les informations de mouvement, de mode de codage et blocs du voisinage ont été bien reçues. Cette hypothèse n'est pas garantie : en particulier, les lignes ou blocs nécessaires aux calculs sont rarement tous disponibles. Par exemple, si un paquet transport est perdu, plusieurs macroblocs de données le seront. Donc un macrobloc perdu sera souvent précédé d'un macrobloc perdu, la ligne de gauche sera absente et les critères de correspondance ne pourront pas tous être spécifiés.

1.4.3 Récupération des vecteurs de mouvement et mode de codage

Si le bloc perdu appartient à une image où la quantité de mouvement est faible, la valeur du vecteur de mouvement peut être négligée sans conséquence dramatique. De la même façon, on peut considérer que le mode du bloc perdu était un codage en intra et se contenter d'en évaluer la texture grâce à une interpolation spatiale.

Mais les méthodes de recouvrement de la texture donnent de bien meilleurs résultats si les modes de codage et vecteurs de mouvement peuvent être exploités pour affiner la prédiction. Le vecteur de mouvement peut être utilisé pour trouver le bloc, dans une image précédente, qui fournira une bonne estimation de l'information perdue. Si l'estimation du vecteur de mouvement perdu est précise, elle peut considérablement améliorer la dissimulation dans les zones à fort mouvement de l'image. Comme précédemment, la propriété de variation graduelle du contenu des images naturelles peut servir de base à la récupération de ces informations. Nous donnons dans la suite de cette section quelques exemples de telles techniques.

1.4.3.1 Récupération du mode de codage d'un bloc [SUN92] :

Le mode de codage d'un bloc dans le cas d'un codage MPEG2 peut être estimé à l'aide des modes des blocs situés au-dessus et en dessous du bloc considéré grâce aux tables de la Figure 1.17:

	Forw.	Intra
Forw.	Forw.	Forw.
Intra	Forw.	Intra

Image P

	Forw.	Backw.	Inter	Intra
Forw.	Forw.	Inter	Inter	Forw.
Backw.	Inter	Backw.	Inter	Backw.
Inter	Inter	Inter	Inter	Inter
Intra	Forw.	Backw.	Inter	Intra

Image B

Figure 1.17 : Illustration de la récupération du mode de codage d'un bloc en MPEG2.

1.4.3.2 Dissimulation par compensation de mouvement (Motion Compensated Concealment, MCC) [HUH06] :

Cette méthode utilise les vecteurs de mouvement des macroblocs situés au-dessus ou au-dessous d'un macrobloc perdu pour prédire le vecteur de mouvement manquant car les macroblocs latéraux sont le plus souvent perdus. Cependant, elle n'est pas applicable dans le cas où le bloc perdu est entouré de blocs codés en intra puisque aucun vecteur de mouvement n'est alors transmis. On notera que le standard MPEG2 permet de générer des vecteurs de mouvement pour les macroblocs codés en intra. Bien sûr, ces vecteurs de mouvement doivent être transmis séparément des autres informations du bloc. Leur seul intérêt étant la dissimulation d'erreur, il faut qu'ils restent disponibles en cas de pertes.

1.4.3.3 Algorithme de correspondance des bords (Boundary Matching Algorithm, BMA) [LAM93]:

Cet algorithme a pour objectif d'exploiter le fait que le meilleur bloc remplaçant un bloc perdu sera fortement corrélé avec les pixels qui forment son contour dans l'image courante. L'algorithme prend les lignes de pixels situées au-dessus, en dessous et à gauche du macrobloc perdu et calcule l'erreur quadratique entre ces lignes et les lignes correspondantes au bord du macrobloc issu d'une image précédemment décodée. Le meilleur correspondant est le macrobloc qui minimise cette erreur quadratique : le macrobloc de l'image précédente testé et désigné par un vecteur de mouvement estimé (cf. Figure 1.18). La recherche du meilleur vecteur peut se limiter à un nombre restreint de candidats : celui de bloc de même emplacement dans l'image précédente, ceux des macroblocs du voisinage, etc.

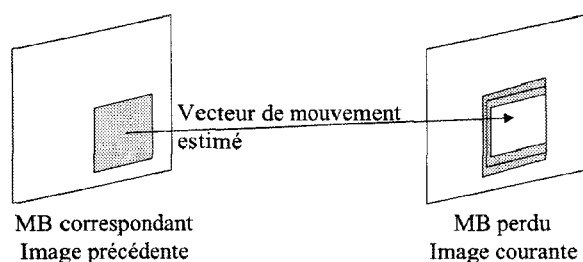


Figure 1.18 : Illustration du "Block Matching Algorithm".

1.4.3.4 Estimation des vecteurs de mouvement au décodeur (Decoder Motion Vector Estimation, DMVE) [HUH06]:

Quand une erreur s'est produite, plusieurs lignes (de 2 à 8) autour du macrobloc perdu (dessus, dessous et à gauche) sont considérées. L'algorithme procède ensuite à une recherche complète dans l'image précédente de la meilleure correspondance avec les lignes disponibles de l'image actuelle. Le macrobloc correspondant le mieux au contour est considéré comme le meilleur correspondant du macrobloc perdu. La complexité de mise en œuvre associée au grand nombre de tests que nécessite cette recherche est néanmoins prohibitive. Une méthode de recherche rapide a été développée pour laquelle la recherche ne porte que sur un nombre restreint de correspondants potentiels.

DMVE simplifié :

On part de l'hypothèse que le vecteur de mouvement du macrobloc situé juste au-dessus du macrobloc perdu peut être utilisé pour la dissimulation d'erreur à compensation de mouvement. Parce que la corrélation spatiale est généralement importante entre les champs de vecteurs de mouvement typiques, cette hypothèse est le plus souvent raisonnable. Mais quand elle n'est pas vérifiée, on voit apparaître des artefacts gênants.

On notera que, en pratique, les informations de mouvement, mode et texture sont stockées séquentiellement et intégralement perdues en cas de perte du bloc. Si de plus le réseau de communication est à commutation de paquets, alors on aura aussi probablement perdu les blocs avoisinants. Si on a appliqué un partitionnement des données, alors les vecteurs de mouvement, modes de codage voire un petit nombre de coefficients DCT basse fréquence seront disponibles et pourront aider à la récupération de la texture.

Les techniques abordées ci-dessus sont très élaborées et augmentent substantiellement la complexité du décodeur. C'est pourquoi nous les considérons, dans la suite de ce travail, comme des fonctionnalités optionnelles. Si les ressources du récepteur permettent leur utilisation, la qualité de l'image diffusée n'en sera que meilleure.

1.5 **Conclusion**

Nous avons vu dans ce chapitre les différents outils visant à assurer une transmission efficace d'une source vidéo numérique sur un réseau à perturbations. La bande passante étant limitée, la première étape est l'adaptation du débit de la source à la capacité du réseau. Le codage de source ou compression élimine un maximum de redondance, voire supprime les informations les moins

pertinentes. Mais, au-delà de la perte de qualité contrôlée qui en découle, la sensibilité du flux aux erreurs de transmission augmente alors dramatiquement. Le codage de canal, visant à protéger le flux compressé, devient donc indispensable.

On cherche à déterminer la paire codeur source – codeur canal optimale, qui permet de minimiser les distorsions résultant de la quantification et des erreurs de transmission. De nombreuses techniques ont été développées pour contrôler l'occurrence ou l'impact des erreurs de transmission sur la qualité de la vidéo décodée. Parmi celles-ci, le codage scalable a monopolisé l'attention ces dernières années. Son objectif est de garantir une qualité de base qui sera incrémentée graduellement si les conditions de transmission le permettent. Il constitue une solution prometteuse, bien que la complexité et le surcoût de débit induits en limitent encore aujourd'hui les applications.

Dernièrement, l'intérêt de la communauté s'est tourné vers un nouveau pôle de recherche: le codage de Wyner-Ziv. Bien que les résultats théoriques aient été établis dans les années soixante-dix, leurs mises en oeuvre sont récentes et en plein essor. Les techniques de compression développées permettent notamment de déplacer la complexité de l'encodeur vers le décodeur et sont donc mieux adaptées aux nouveaux scénarios de transmission uplink. Nous proposons donc d'en décrire les principes et premières implémentations dans le chapitre suivant.

2 Théorème de Wyner-Ziv et applications

Les ressources actuelles des terminaux et réseaux de communication mobiles en termes de bande passante permettent la transmission de contenus audiovisuels. Mais la puissance de calcul liée à la consommation et la mémoire intégrée de ces terminaux mobiles restent pourtant limités pour des raisons de portabilité. D'autre part, les réseaux de capteurs (environnement multi-vues) se développent pour des applications variées (détection d'agents biologiques, vidéosurveillance, mesures environnementales) [XIO04]. L'énergie dont dispose ces capteurs est limitée et leur réapprovisionnement problématique, ce qui influe sur leur capacité à traiter / transmettre les données. L'usage de techniques de compression permettant de réduire le volume de données générées constitue une partie de la solution. Cependant, les standards de compression actuels utilisent un codage interframe prédictif, la complexité de l'encodeur est alors fort élevée. Dans les cas où l'encodage doit être réalisé sur des terminaux à faible capacité, un encodeur plus simple, même au prix d'un décodeur plus complexe mis en œuvre au niveau de la station de base, serait souhaitable (cf. Figure 2.1). Cette considération a conduit à envisager un système basé sur un encodage intra image et un décodage inter image.

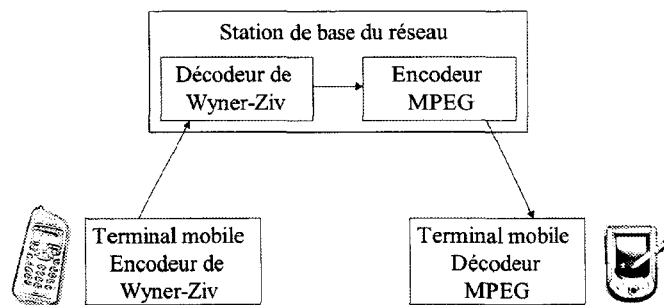


Figure 2.1 : Scénario de transcoding pour les architectures de réseau sans fil.

Le codage de sources distribuées (Distributed Source Coding, DSC) est un domaine de recherche récent qui s'intéresse à la communication de sources multiples corrélées. Le DSC est basé sur le théorème de Slepian-Wolf [SLE73] et le théorème de Wyner-Ziv [WYN76] respectivement dans le cas du codage sans perte et avec pertes. Ces théorèmes datant des années 1970 établissent qu'un codec où les signaux seraient encodés séparément mais décodés conjointement peut être aussi efficace qu'un codage / décodage conjoint. L'exploitation de la corrélation des sources doit alors se faire uniquement à l'encodeur.

Il a fallu de nombreuses années avant qu'une méthode soit développée qui permette d'exploiter les statistiques d'une source au décodeur. La corrélation entre les sources permet d'assimiler le problème du codage source à un problème de codage canal. En effet, une source X à encoder peut être vue comme une version bruitée de la source adjacente Y . Cette dualité a conduit à l'utilisation de codes correcteurs d'erreurs pour la réduction de la quantité d'information à transmettre. Le décodage de canal appliqué à l'information adjacente Y et au syndrome généré à partir de X permet de corriger la version de X "corrompue" qu'est Y . Cela permet de réduire la complexité du codeur, au prix d'une augmentation de celle du décodeur. Des propositions de mise en œuvre appliquées à la compression d'une ou plusieurs sources commencent à approcher d'assez près les possibilités théoriques.

Plus particulièrement, une nouvelle méthode de codage robuste a été dérivée du paradigme de Wyner-Ziv. En effet, l'application d'une protection hiérarchique à un flux vidéo impliquait jusqu'ici un partitionnement (data partitionning, codage par couches) préalable des données. Mais ces modifications du codage source causent une perte d'efficacité (répétition des en-têtes) et une augmentation de la complexité du codec. Afin de garantir la bonne réception des informations grossières du flux vidéo, on peut appliquer un code correcteur à une version bas débit judicieusement choisie. Si le code correcteur appliqué est systématique (cf. Chapitre I, paragraphe 3.1.2), les données vidéo protégées sont transmises sans modification sur le canal avec ajout d'informations de parité. Or on peut considérer que les données vidéo bas débit sont incluses dans la version haut débit correspondante, il n'est donc pas utile de transmettre les deux flux puisque l'un est déductible de l'autre. On transmet donc le flux haut débit conjointement à la parité portant sur la version bas débit. Au décodeur, en cas d'erreur de transmission, la version grossière sera déduite de la version fine de la vidéo transmise et corrigée grâce à la parité. Dans la mesure où la protection ne porte pas directement sur l'ensemble du flux binaire issu du codeur source mais autorise un degré de distorsion supplémentaire en cas d'erreurs de transmission, le débit nécessaire peut être réduit.

L'organisation du chapitre est la suivante : les bases théoriques du codage source avec information adjacente, les théorèmes de Wyner-Ziv et Slepian-Wolf, sont présentées. Puis, les principales techniques de mise en œuvre pour la compression de séquences vidéo avec encodeur de faible complexité sont décrites. Finalement, l'application de ces principes au codage résistant aux erreurs est explicitée puisqu'il représente le thème central de ce manuscrit.

2.1 Bases théoriques du codage source avec information adjacente

Le codage de sources distribuées a pour objectif la réduction du débit nécessaire à la description de plusieurs signaux issus de sources corrélées comme, par exemple, les séquences issues d'un réseau de capteurs. Le codage avec information adjacente est un cas particulier du codage de sources distribuées où l'une des sources, ici nommée X est codée en considérant la disponibilité de la source Y uniquement au décodeur.

Dans ce paragraphe, les bases théoriques sur lesquelles le codage de sources distribuées (Distributed Source Coding, DSC) et, par conséquent, le codage source avec information adjacente (Source Coding with Side Information, SCS) sont basés, sont décrites. Considérons un système de communication ayant à transmettre deux séquences corrélées X et Y , on cherche à déterminer quelle est la limite inférieure sur le débit à achever pour permettre leur reconstruction. Deux cas sont successivement abordés, le cas d'un codage sans pertes avec le théorème de Slepian-Wolf [SLE73] et le cas d'un codage avec pertes basé sur le théorème de Wyner-Ziv [WYN76].

Scénario et problématique :

On définit deux variables aléatoires X et Y , chacune étant indépendante et identiquement distribuée (i.i.d.), à valeurs discrètes, provenant de deux sources corrélées. On note :

- $p_{XY}(x, y) = \Pr(X = x, Y = y)$, leur distribution conjointe,
- $p_X(x) = \sum_y p_{XY}(x, y)$ et $p_Y(y) = \sum_x p_{XY}(x, y)$, leurs distributions respectives,
- $p_{X|Y}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)}$ et $p_{Y|X}(y|x) = \frac{p_{XY}(x, y)}{p_X(x)}$, leurs distributions conditionnelles.

On considère tout d'abord le cas où les encodeurs et décodeurs ne peuvent pas communiquer (cf. Figure 2.2). Cela revient à encoder séparément chaque source sans tenir compte de l'existence de l'autre ni de leur corrélation. Trivialement, la totalité de l'information de chaque source devra être communiquée au décodeur pour que la reconstruction du signal encodé soit possible. On sait que le débit minimum permettant la caractérisation sans perte d'une source à coder S correspond à l'entropie de cette source $H(S) = -\sum_s p_S(s) \log p_S(s)$.

On note $H(X)$ et $H(Y)$ les entropies respectives des sources X et Y , $H(X,Y)$ l'entropie conjointe, $H(X|Y)$ et $H(Y|X)$ les entropies conditionnelles associées.

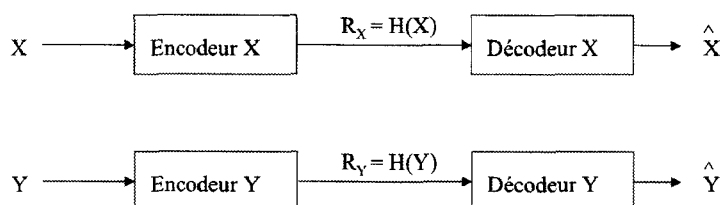


Figure 2.2 : Schéma bloc du codage / décodage séparé de deux sources corrélées X et Y .

Supposons maintenant que ces signaux corrélés proviennent de deux sources qui communiquent à l'encodeur et au décodeur (cf. Figure 2.3). Si l'on transmet un débit total $R = H(X) + H(Y)$ l'information commune aux deux séquences est transmise deux fois : une première fois avec X , une seconde avec Y . En exploitant les probabilités conjointes des deux signaux, le débit minimal total peut être réduit à l'entropie conjointe $H(X,Y)$. Par exemple, si la source Y est transmise avec un débit $H(Y)$, alors l'entropie conditionnelle $H(X|Y)$ suffira à combler l'incertitude restante sur la source X , et inversement.

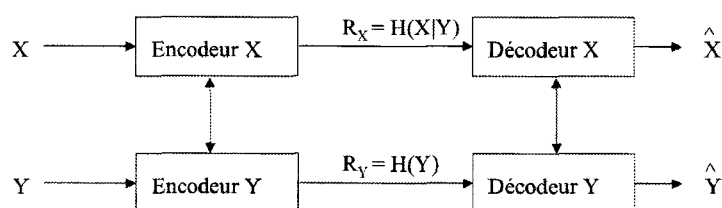


Figure 2.3 : Schéma bloc du codage / décodage conjoint de deux sources corrélées X et Y .

L'encodeur utilise les probabilités conjointes et diminue le débit nécessaire à la représentation des deux sources en supprimant la redondance qui existe entre les deux signaux. Mais si on considère le cas où les deux sources sont encodées séparément mais décodées conjointement (cf. Figure 2.4), la question de la réduction du débit de transmission total se pose.

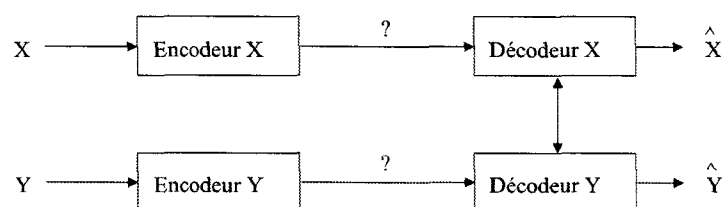


Figure 2.4 : Schéma bloc du codage séparé / décodage conjoint de deux sources corrélées X et Y .

On a vu que si les décodeurs ne communiquent pas, le débit minimal de transmission de la source X est $H(X)$. Mais dans le cas contraire, la présence de Y au niveau du décodeur permet de diminuer le débit nécessaire à la reconstruction de la source. Si on ne transmet qu'une information partielle sur X , les deux sources étant corrélées, il est possible d'utiliser Y pour lever l'indétermination résiduelle.

2.1.1 Théorème de Slepian-Wolf

Slepian et Wolf se sont intéressés à la généralisation des résultats du codage sans perte d'une source d'information au cas de deux sources d'information X et Y corrélées [SLE73]. Leur article établit les ensembles de débits R_X et R_Y possibles pour divers cas de couple encodeur / décodeur qui sont résumés par la Figure 2.5.

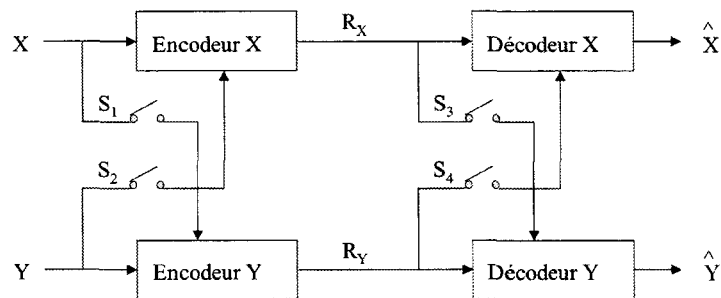


Figure 2.5 : Différents cas de codage de sources corrélées [SLE73].

Les interrupteurs S_i ($i=1, \dots, 4$) modélisent la communication ($S_i = 1$, interrupteur fermé) ou séparation ($S_i = 0$, interrupteur ouvert) des encodeurs et décodeurs. Parmi les configurations étudiées, on retrouve par exemple les cas bien connus du codage séparé 0000 ($S_1 = S_2 = S_3 = S_4 = 0$) et du codage conjoint 1111 ($S_1 = S_2 = S_3 = S_4 = 1$) de deux sources corrélées. Le cas 0011 ($S_1 = S_2 = 0, S_3 = S_4 = 1$) est de loin le plus intéressant de l'article, le plus innovant : il correspond au codage de sources distribuées (cf. Figure 2.4).

Comme première étape, on considère le cas 0001, où la source Y est transmise et le décodeur de la source X a connaissance de la version décodée de Y (cf. Figure 2.6). On sait que la source Y doit être transmise avec un débit minimal de $H(Y)$ au décodeur et on s'intéresse au débit minimal de transmission de la source X . Ce cas particulier du codage de deux sources corrélées est appelé codage de source avec information adjacente.

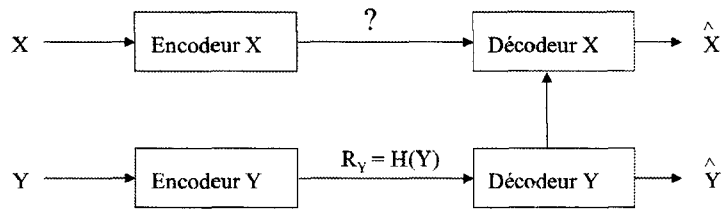


Figure 2.6 : Cas du codage de la source X avec information adjacente Y.

Dans cette configuration, en considérant Y comme une version bruitée de X, les résultats du codage de canal peuvent être utilisés. Soit un canal discret sans mémoire de probabilité de transition $p_{Y|X}(y|x)$ et soit R_Y suffisant pour garantir la réception sans erreur de la source Y au décodeur. Slepian et Wolf montrent que :

Pour ϵ_1 et ϵ_2 arbitrairement petits, la séquence $\{X_k\}$ peut être encodée par blocs de longueur N (N suffisamment grand) et que le décodeur (qui a accès à $\{Y_k\}$) peut alors reconstruire une séquence $\{\hat{X}_k\}$ telle que : $\frac{1}{n} \sum_{k=1}^n \Pr(\hat{X}_k \neq X_k) \leq \epsilon_1$ et telle que le code de taille de bloc n a au plus $2^{N(H(X|Y)+\epsilon_2)}$ mots codes.

Autrement dit, il existe un code permettant de transmettre X à un débit $H(X|Y)$ et tel que X soit reconstruite avec un taux d'erreur arbitrairement petit [SLE73] en utilisant la corrélation entre cette source et la source Y uniquement au décodeur.

Par un raisonnement analogue, on établit le débit minimal pour la communication de la source Y, avec disponibilité de l'information sur la source X (cas 0010). Les cas 0010 et 0001 correspondent au codage asymétrique des deux sources et sont représentés respectivement par les points A et B sur la Figure 2.7. À partir des deux cas extrêmes précédents, on établit un ensemble de débits réalisables pour la compression symétrique de deux sources i.i.d. statistiquement dépendantes :

$$\begin{aligned} R_X + R_Y &\geq H(X,Y) \\ R_X &\geq H(X|Y) \\ R_Y &\geq H(Y|X) \end{aligned}$$

Equation 2.1: Théorème de Slepian-Wolf.

Ce qui signifie que tant que les débits individuels R_X et R_Y dépassent l'entropie conditionnelle, le débit total $R = R_X + R_Y$ peut être réduit à l'entropie conjointe $H(X,Y)$. La démonstration du théorème de Slepian-Wolf est basée sur le regroupement aléatoire des symboles de la source en

sous-ensembles distincts. La limite inférieure est accessible asymptotiquement, pour des codes de longueur infinie. On peut représenter l'ensemble des débits réalisables comme le montre la Figure 2.7 :

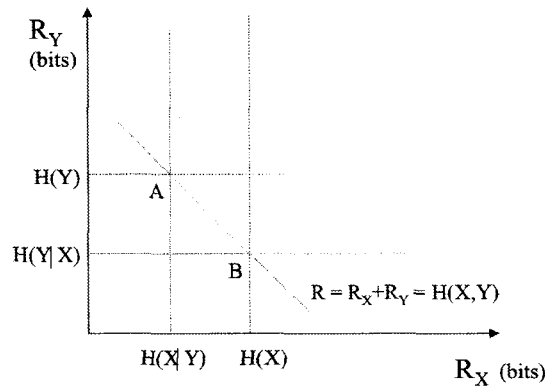


Figure 2.7 : Représentation de l'ensemble des débits réalisables pour la représentation de deux sources corrélées X et Y d'après le théorème de Slepian-Wolf [SLE73].

Le débit minimal nécessaire à la transmission de sources corrélées est donc l'entropie conjointe $H(X, Y)$ dans le cas d'un décodage conjoint, que les encodeurs communiquent ou non.

2.1.2 Théorème de Wyner-Ziv

Peu après que Slepian et Wolf aient établi le théorème sur le codage sans perte de sources distribuées, Wyner et Ziv se sont intéressés au cas de la compression avec pertes d'une source avec information adjacente. Dans ce cas, on cherche à déterminer le débit de transmission minimal $R(D)$ permettant de satisfaire un critère de distorsion D du signal reconstruit donné.

Soient X et Y deux séquences i.i.d. représentant respectivement la source à coder et l'information adjacente (cf. Figure 2.8). Les valeurs de la source X sont encodées sans accès aux valeurs prises par la source Y . Cependant, le décodeur détermine la reconstruction \tilde{X} de la source X en tenant compte de l'information apportée par la source Y .

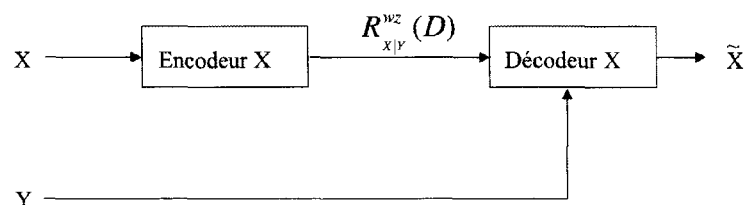


Figure 2.8 : Codage de la source X avec information adjacente Y disponible uniquement au décodeur.

On note $R_{X|Y}(D)$ le débit nécessaire à la reconstruction de la source X , avec un niveau de distorsion D dans le cas où Y est accessible à l'encodeur et au décodeur. On note $R_{X|Y}^{wz}(D)$ le débit nécessaire à la reconstruction de la source X , avec un niveau de distorsion D dans le cas où Y n'est accessible qu'au décodeur. On notera que d'après le théorème de Slepian-Wolf (cf. Equation 2.1), on a :

$$R_{X|Y}^{wz}(0) = R_{X|Y}(0) = H(X|Y)$$

Dans le cas général, le théorème de Wyner-Ziv [WYN76] établit l'inégalité suivante :

$$R_{X|Y}^{wz}(D) \geq R_{X|Y}(D), \text{ pour } D > 0$$

Équation 2.2 : Théorème de Wyner-Ziv.

Cependant, ils ont aussi démontré que cette perte d'efficacité, comparée au cas d'un encodage conjoint, peut être nulle. Par exemple, si la mesure de distorsion utilisée est l'Erreur Quadratique Moyenne (EQM) et que les sources sont gaussiennes et sans mémoire, les fonctions débit-distorsion des systèmes avec ou sans communication à l'encodeur sont les mêmes [WYN76], soit :

$$R_{X|Y}^{wz}(D) = R_{X|Y}(D) = H(X|Y)$$

Équation 2.3 : Théorème de Wyner-Ziv pour des sources gaussiennes et mesure de distorsion par EQM.

D'autres cas particuliers ont été étudiés pour lesquels la perte de débit due à l'absence de Y à l'encodeur a pu être majorée :

- $R_{X|Y}(D) - R_{X|Y}^{wz}(D) = 0$ dans le cas où Y est de distribution arbitraire et X est la somme de Y et d'un bruit gaussien indépendant [PRA03],
- $R_{X|Y}(D) - R_{X|Y}^{wz}(D) \leq 0.22 \text{ bits}$ si X et Y sont des sources binaires et la distorsion la distance de Hamming [ZAM96],
- $R_{X|Y}(D) - R_{X|Y}^{wz}(D) \leq 0.5 \text{ bits}$ si X et Y sont des sources continues et la mesure de distorsion de type EQM [ZAM96].

Nous avons cité quelques travaux visant à préciser les bornes débit-distorsion du codage avec information adjacente. Supposer l'information adjacente connue parfaitement du décodeur est une limitation vis-à-vis des scénarios pratiques de compression vidéo. Une extension de ce principe au cas où l'information adjacente est elle aussi à compresser avec un critère de fidélité est connue sous le nom de codage multi terminal ou codage de sources distribuées. De nombreuses autres extensions du problème ont été étudiées [GIR05]. Les résultats en termes de

région de débits réalisables sont encore évasifs, mais le potentiel de ce paradigme de codage ne fait aucun doute.

De plus, il s'agit de résultats uniquement théoriques et asymptotiques. En effet, l'existence d'un code permettant d'atteindre les performances voulues est prouvée, mais sa construction n'est pas explicitement donnée. C'est pourquoi ces théorèmes sont restés sans application pratique pendant plus de vingt ans. Dans la section suivante, nous abordons l'aspect mise en oeuvre de ces théorèmes et donnons quelques exemples des principales implémentations existantes.

2.2 Mises en œuvre des théorèmes et application à la compression

Nous avons vu dans le paragraphe précédant les bases théoriques qui ont motivé les recherches sur le codage de sources distribuées. Ces théorèmes datent des années 70 mais les systèmes pratiques permettant de mettre en œuvre les théories établies sont récents. En effet, la stratégie de codage envisagée est radicalement différente de celle utilisée dans les codecs standards. Les statistiques de la source ne sont pas exploitées à l'encodeur mais au décodeur. Le moyen concret de diminuer la quantité d'information à transmettre en se basant sur cette assertion n'est pas évident.

En reformulant le problème, il s'agit de transmettre une information partielle sur la source à compresser de sorte que, grâce à l'information adjacente dont il dispose, le décodeur puisse lever l'indétermination. Les techniques développées utilisent la dualité du codage de Wyner-Ziv avec le codage de canal, et identifient la source principale à une version bruitée de l'information adjacente. Les paragraphes suivants illustrent ces techniques d'abord à l'aide d'exemples simples. Par la suite, les principales mises en œuvre du codage de Wyner-Ziv pour la compression d'une source vidéo sont évoquées. Deux applications principales sont à distinguer : le codage vidéo avec un encodeur de faible complexité et le codage de sources multiples distribuées [XIO04], [GIR05]. Nous nous focaliserons sur la première approche.

2.2.1 Principe de mise en œuvre des théorèmes

Pradhan et Ramchandran furent les premiers à proposer une approche pratique au problème du codage de source avec information adjacente. La dualité entre le codage de Slepian-Wolf et les principes du codage de canal avait été utilisée pour dériver les bornes théoriques que peut atteindre le débit, mais jamais pour construire un code pratique. À partir d'exemples simples, ils élaborèrent l'algorithme DISCUSS (DIstributed Source Coding Using Syndromes) [PRA99]. Ces

exemples sont repris dans la suite de ce paragraphe afin d'illustrer le principe de codage utilisé dans les mises en oeuvre pratique.

2.2.1.1 Conception d'un codeur de Slepian-Wolf

Le théorème de Slepian-Wolf établit l'existence d'un code permettant de transmettre la source X à un débit $H(X|Y)$, que l'information adjacente Y soit disponible à l'encodeur ou non. L'exemple simple suivant permet d'illustrer les étapes menant à la construction d'un tel code [PRA99].

Supposons X et Y deux mots de 3 bits, corrélés dans le sens où X et Y diffèrent de 1 bit au plus, soit la distance de Hamming $d(X,Y) \leq 1$:

$$H(X) = H(Y) = 3 \text{ bits,}$$

$$X = Y \oplus \{000, 001, 010, 100\}.$$

Si Y est connu à l'encodeur et au décodeur, il est inutile d'utiliser 3 bits pour décrire X dans la mesure où il n'y a que 2 bits d'incertitude entre X et Y . S'il existe quatre possibilités pour la somme binaire modulo 2 de X et Y , alors $H(X|Y) = 2$ bits, il suffit de transmettre 2 bits au décodeur pour lever l'incertitude sur X .

D'après le théorème de Slepian-Wolf, si Y n'est connu qu'au décodeur, on doit pouvoir arriver au même résultat. Il faut donc s'arranger pour que l'incertitude sur X soit la plus facile à lever possible. À l'encodeur, on regroupe les huit mots codes possibles en quatre sous-ensembles, de sorte à ce que les valeurs associées deux à deux soient les plus distantes possible donc de distance de Hamming valant 3 (cf. Figure 2.9).

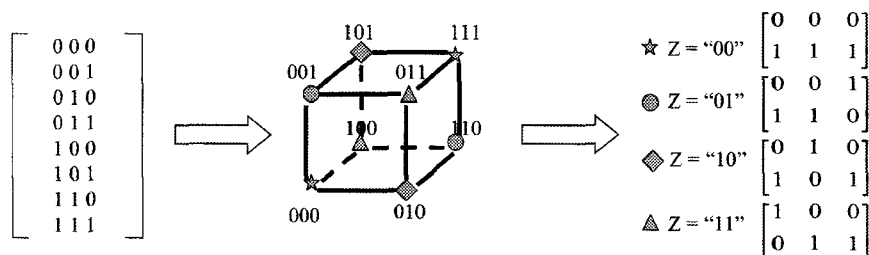


Figure 2.9 : Partitionnement des symboles en sous-ensembles.

On transmet alors au décodeur l'index du sous-ensemble auquel appartient X , donc on transmet 2 bits. Au décodeur, l'index du sous-ensemble transmis permet au décodeur de limiter

l'incertitude sur X à deux valeurs possibles, maximalement séparées. La comparaison de ces valeurs à celle prise par Y permet de déduire la valeur de X (cf. Figure 2.10).

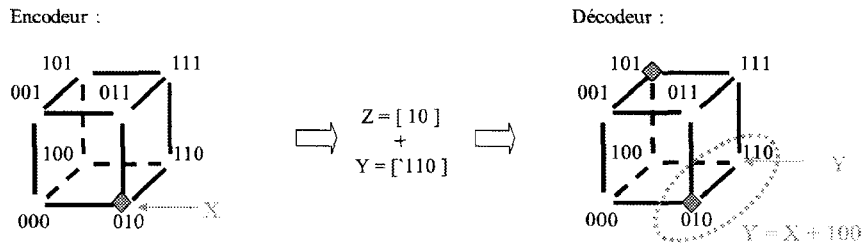


Figure 2.10 : Décodage de la valeur de X grâce à l'information adjacente Y.

On a bien réussi à réduire le débit de transmission de la source X dans le cas où l'information Y est disponible à l'encodeur et au décodeur, mais aussi dans le cas où elle n'est disponible qu'au décodeur. La méthode utilisée repose sur le regroupement des différents symboles possibles de l'alphabet de la source X en sous-ensembles distincts de mots codes maximalement séparés.

On peut généraliser cet exemple très simple au cas de symboles de n bits [XIO04]. Soient X et Y deux sources binaires de n bits, on a donc 2^n mots code possibles :

$$H(X) = H(Y) = n \text{ bits}$$

De nouveau, X et Y diffèrent d'une position au plus. On pose $n = 2^i - 1$ bits et connaissant Y, il y a 2^i possibilités pour la somme binaire modulo 2 de X et Y :

$$H(X|Y) = i \text{ bits}$$

$$H(X,Y) = H(Y) + H(X|Y) = n+i \text{ bits}$$

On crée 2^i sous-ensembles, chacun contenant 2^{n-i} mots code (cf. Figure 2.11) :

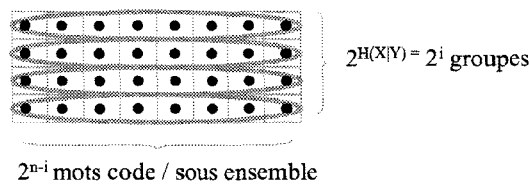


Figure 2.11 : Regroupement des symboles de la source en sous-ensembles.

On utilise i bits pour spécifier l'index d'un des 2^i sous-ensembles. Au décodeur, l'index et l'information adjacente Y permettent conjointement de décoder X. On obtient un facteur de compression $n : i$ sur le débit nécessaire à la transmission de la source X.

Intuitivement, on peut considérer la source à transmettre X comme une version bruitée de l'information adjacente Y . On peut alors assimiler le bruit de corrélation entre X et Y au bruit résultant de la transmission sur un canal virtuel qui à l'entrée X ferait correspondre Y (cf. Figure 2.12). Le codage de source avec information adjacente s'apparente donc à un codage de canal [SUJ00]. Autrement dit, si un code de canal est performant vis-à-vis du canal virtuel de corrélation, il sera un code performant pour le codage de source avec information adjacente. Deux approches existent : celle basée sur les syndromes et celle basée sur la parité. Toutes deux permettent d'indexer des sous-ensembles disjoints de mots codes, mais les bits de parité sont plus complexes à calculer [XUQ06].

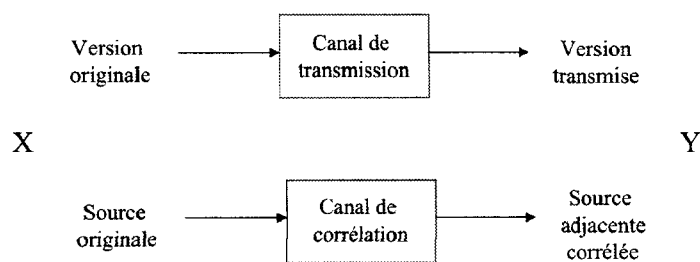


Figure 2.12 : Analogie entre codage de canal (en haut) et codage de source avec information adjacente (en bas).

Une mesure de la qualité d'un code de canal est sa capacité correctrice en cas d'effacement. Un code correcteur linéaire (n, k) a une capacité correctrice $(n-k)$ en cas d'effacements. Les 2^n mots du code sont partitionnés en 2^{n-k} sous ensembles disjoints de 2^k mots binaires, indexés par le syndrome, codé sur $(n-k)$ bits. On peut par exemple utiliser les $(n-k)$ mots de parité des codes systématiques pour spécifier le sous-ensemble auquel le mot code de n mots appartient. Le taux de compression obtenu est $n : n-k$.

NB : Plus la valeur $(n-k)$ sera grande, plus la distance entre les mots codes d'un même sous-ensemble sera grande, et plus le risque d'erreur d'estimation sur le mot reçu sera faible. On notera que plus l'information adjacente Y et la source à transmettre X sont corrélées, plus on pourra réduire le débit. En effet, dans le cas d'un codage de canal classique, si l'erreur sur X (valeur de Y) est plus faible, la quantité d'information à fournir pour la correction sera plus faible.

2.2.1.2 Conception d'un codeur de Wyner-Ziv

La méthode de répartition uniforme des mots codes en sous-ensembles exposée ci-dessus permet la compression dans le cadre du codage source avec information adjacente sans perte. C'est le codage de Slepian-Wolf. On cherche dans le cas du codage de Wyner-Ziv à étendre le

procédé à une compression avec pertes en autorisant une certaine distorsion entre la source et sa reconstruction. L'étape permettant la suppression des informations les moins pertinentes d'un signal vidéo est la quantification. C'est donc cette opération qu'il faut inclure au cas d'un codage avec information adjacente.

Traditionnellement, la compression est l'association d'un quantificateur et d'un codeur entropique. La quantification consiste à associer à un ensemble de valeurs proches une valeur unique moyenne. Dans le cas du codage source avec information adjacente, le codage entropique est réalisé par un codeur de Slepian-Wolf. Le codeur de Slepian-Wolf peut-être assimilé à un quantificateur [XIO04] puisqu'il associe à plusieurs valeurs distantes possibles de la source X une valeur unique. Le codage source destructif avec information adjacente se présente comme la combinaison d'une quantification fine (codage source) et d'une quantification grossière (codage canal). On parle de quantification emboîtée (Nested quantization [ZAM02], [XIO03]). La Figure 2.13 illustre le cas le plus simple de quantification emboîtée : cas d'un quantificateur scalaire à 4 sous-ensembles et un cas de quantification emboîtée à deux dimensions. La quantification grossière est imbriquée dans la quantification fine au sens où chaque point du quantificateur grossier est aussi un point du quantificateur fin, la réciproque étant fautive.

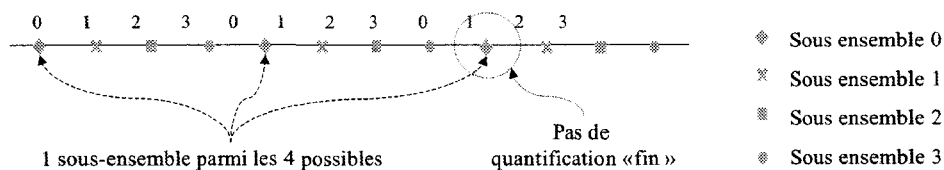


Figure 2.13 : Quantification scalaire emboîtée.

La conception d'un codeur de Wyner-Ziv implique donc deux étapes :

- La conception du quantificateur fin, qui doit minimiser la distorsion liée au codage de source tout en satisfaisant une contrainte de débit R comme pour un codage de source classique (fonction débit-distorsion $R(D_S)$).
- La conception du codeur de Slepian-Wolf adapté à l'alphabet issu du quantificateur, qui répartit les symboles quantifiés dans N sous-ensembles.

À la valeur originale X du symbole, la quantification uniforme de pas q associe la valeur X_q . Le codeur de Slepian-Wolf détermine les sous-ensembles de valeurs grâce à un codage de canal de distance minimale $d_{\min} = Nq$. L'index du sous-ensemble auquel X_q appartient est transmis. Au décodeur, l'index transmis désigne un sous-ensemble de candidats pour l'estimation de la valeur X_q . Finalement, l'information adjacente Y permet, grâce à sa corrélation avec X , de sélectionner

dans cet ensemble la valeur qui correspond le mieux à X . Le processus d'encodage / décodage dans le cas avec pertes est illustré par la Figure 2.14.

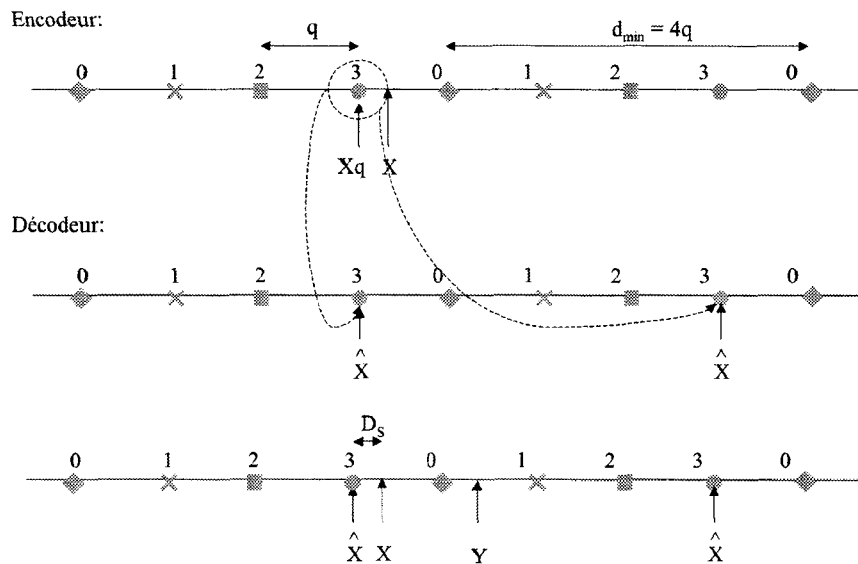


Figure 2.14 : Illustration du principe de codage / décodage de Wyner-Ziv, exemple simple d'un quantificateur "emboîté" scalaire à 4 sous-ensembles.

La distorsion globale associée à cette configuration admet deux composantes :

- La première, notée D_s , est liée à la quantification ; elle peut être réduite en réduisant le pas de quantification q .
- La seconde, notée D_C , est liée à une éventuelle erreur lors du décodage de canal (si la valeur de Y mène à une mauvaise identification de X). Elle peut être réduite en augmentant la distance entre deux symboles consécutifs d'un même sous-ensemble, donc en augmentant d_{\min} (cf. paragraphe 1.2.1.2).

Donc pour un N fixé, il existe un q optimal qui minimise la distorsion globale.

L'implémentation de l'algorithme DISCUS utilise une quantification scalaire et un codage de canal par modulation codée en treillis ("Trellis Coded Modulation", TCM). Ils se basent sur l'hypothèse que l'information adjacente Y est la somme de la source X et d'un bruit N où X est gaussienne, i.d.d., de moyenne nulle et variance unitaire et N est un processus gaussien i.d.d., de moyenne nulle et indépendant de X . Leur construction est évaluée à travers différentes valeurs des paramètres : nombre de niveaux de quantification (finesse du pas de quantification) et nombre de sous-ensembles (taux de protection du codage de canal). Ils comparent également leur configuration avec un partitionnement des valeurs quantifiées sans codage. Les résultats sont donnés en termes de distorsion et de probabilité d'échec du décodage et mettent en évidence le compromis à faire entre ces deux objectifs.

Les résultats de ce travail préliminaire ont permis de vérifier la faisabilité du schéma présenté à partir de composants du schéma très simples, même si performances théoriques par rapport à un codage / décodage conjoint classique restaient encore modestes. Leur schéma fut ensuite étendu au cas de la compression symétrique de sources distribuées (réseaux de capteurs) [PRA00], puis sous des contraintes de bande passante limitée [PRA02].

2.2.2 Codage inter images

De nombreuses recherches furent menées à partir du schéma DISCUS avec implémentation du codage de canal par des codes plus avancés comme les turbo-codes [AAR02]. Mais le principal défaut de ces implémentations résidait dans la non-exploitation des redondances temporelles des signaux à compresser. Les algorithmes demeuraient largement moins performants que ceux des codeurs prédictifs classiques. Les recherches s'orientèrent donc vers l'intégration du codage inter-images.

2.2.2.1 Travaux d'Aaron et Girod

Aaron et Girod ont proposé une architecture de codage intra images / décodage inter image en considérant les images impaires et paires d'un flux vidéo comme deux sources corrélées [AAR02b]. Les images impaires, appelées ici images clef, sont codées en intra et jouent le rôle d'information adjacente au décodeur. Les images paires sont encodées indépendamment des images impaires par le codeur de Slepian-Wolf (cf. Figure 2.15). Le décodeur procède à l'interpolation / extrapolation des images impaires pour former une estimation de l'image paire à reconstruire. Les bits de parités issus du codeur de canal sont délivrés au décodeur sur demande et lui permettent, grâce à l'information adjacente, de former les symboles décodés. Le processus de demande de parité / décodage est répété jusqu'à ce que la probabilité d'erreur sur la valeur de ces symboles soit acceptable. Des turbo-codes poinçonnés (Rate Punctured Turbo Codes, RPTC) sont utilisés pour leur souplesse en termes de débit : un sous-ensemble des bits de parité peut-être utilisé qui correspond à une protection plus faible qui est incrémentée à chaque bit de parité supplémentaire reçu.

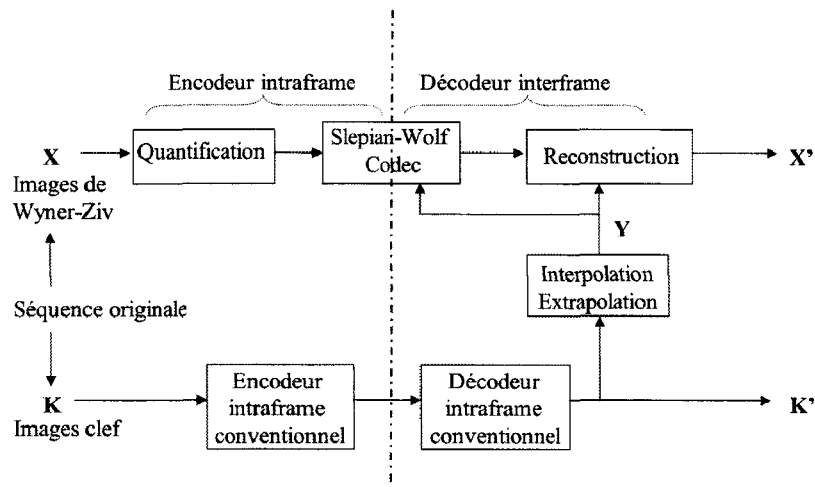


Figure 2.15 : Codeur intra images / décodeur inter images basé sur la séparation des images paires et impaires.

Les résultats de ce codage de Wyner-Ziv sont comparés à ceux du codec h.263. Les performances se situent entre celles d'une structure de GOP I-B-I-B (-6 dB) et celles d'une structure de GOP I-I-I-I (+ 4dB). En effet, le codec n'exploite pas les redondances spatiales des images, ce qui représente un indéniable handicap.

Ils généralisèrent la méthode énoncée au cas d'un sous-ensemble d'images clef quelconque (GOP quelconque) et une estimation des images Wyner-Ziv à partir de toutes les images reconstruites [AAR03]. Ils modifièrent ensuite la structure de leur schéma en l'appliquant au domaine transformé [AAR04], ce qui a permis d'exploiter les redondances spatiales des images et d'améliorer ainsi les performances du schéma de codage. Leur implémentation permet d'obtenir de meilleurs résultats qu'un codeur intra images (+12 dB) pour la même complexité. Les performances restent dépendantes de la technique d'interpolation (MC-I) / extrapolation (MC-E) utilisée au décodeur (cf. Figure 2.16). Dans cette figure, la méthode est appliquée soit au domaine pixel soit au domaine DCT.

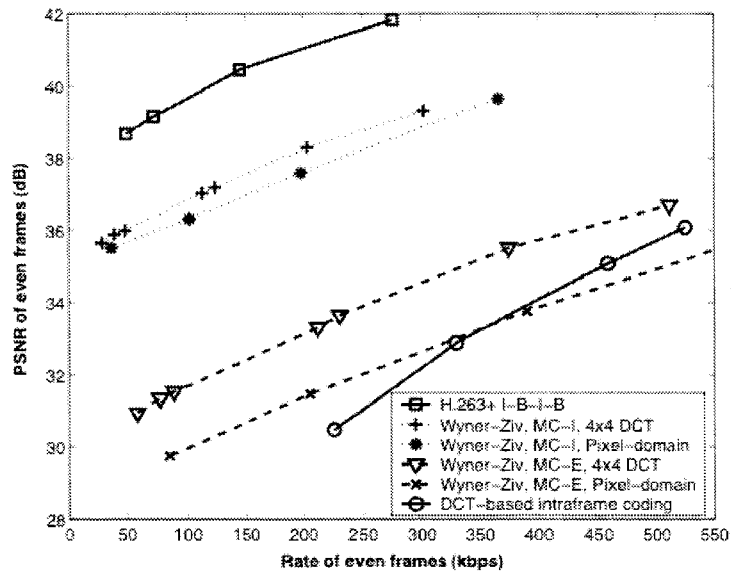


Figure 2.16 : Evaluation des performances du système proposé pour la séquence Foreman [AAR04].

2.2.2.2 PRISM (Power-efficient, Robust, hIgh-compression, Syndrome-based Multimedia coding) [PUR02]

PRISM est le deuxième schéma issu des travaux de recherche de Pradhan et Ramchandran. Les objectifs de cet algorithme sont les suivants:

- répartition flexible de la complexité entre encodeur et décodeur,
- taux de compression élevé,
- robustesse aux pertes de paquets et erreurs de transmission.

Afin de limiter la complexité de l'encodeur, l'estimation de mouvement est réalisée au décodeur. En déplaçant les plus lourdes tâches comme l'estimation de mouvement au décodeur, l'architecture PRISM allège les contraintes syntaxiques de l'encodeur. Le problème de la robustesse aux erreurs est associé au problème de la protection contre le bruit de corrélation entre les différentes sources. Le codage de canal protège le signal contre la somme de ces deux bruits sans distinction.

La principale innovation de ce schéma réside dans le codage inter-image (cf. Figure 2.17). En effet, jusqu'alors, l'inconvénient principal de tous les codecs basé sur la compression de sources distribuées était qu'ils se contentaient d'un codage intra image. Pour réaliser le codage inter images, l'idée directrice fut de considérer la corrélation entre un macrobloc à encoder X et son meilleur correspondant dans l'image précédente Y . X est encodé en intra et le codage de canal est utilisé pour lutter contre ce bruit de corrélation et partitionner les valeurs quantifiées de X à

transmettre. Au décodeur, l'usage d'un code à redondance cyclique (Cyclic Redundancy Check code, CRC) permet de sélectionner parmi les différentes versions décodées celle correspondant au vecteur de mouvement qui doit être utilisé. Ainsi, l'estimation et la compensation de mouvement sont reléguées au décodeur, allégeant la complexité de l'encodeur.

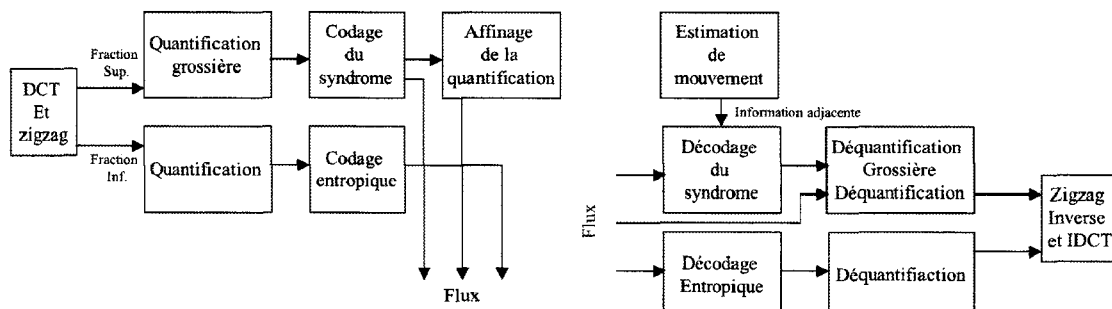


Figure 2.17 : Diagramme en bloc de l'encodeur et du décodeur PRISM [PUR02].

Les résultats de simulation montrent des performances qui se rapprochent de celles d'un codage inter images classique tout en restant inférieures de 2 à 3 dB. Les résultats obtenus pour la séquence Football sont reportés Figure 2.18 à titre d'exemple.

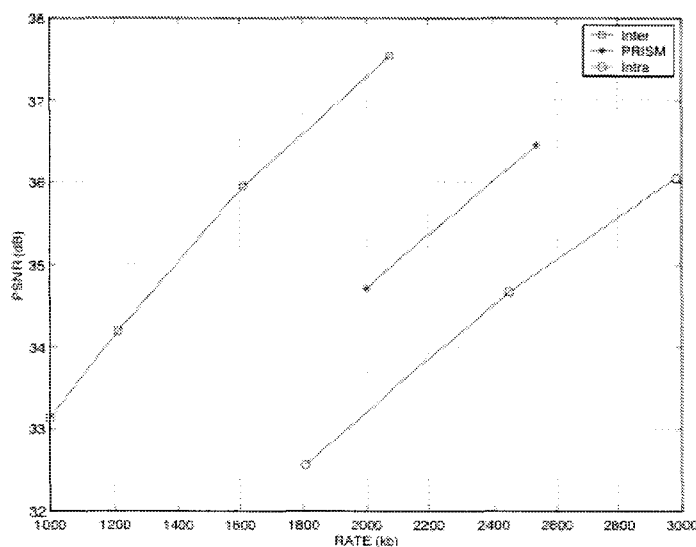


Figure 2.18 : Comparaison de l'algorithme PRISM et du codec H. 263+ pour la séquence Football [PUR02].

Le principe a été repris par Aaron, Rane et Girod [AAR04b] et les performances se situent entre 1 et 4 dB en dessous de celles du codec h.263. Cependant, l'objectif d'une distribution flexible de la complexité entre encodeur et décodeur est atteint. La robustesse du schéma aux pertes est elle aussi mise en évidence par rapport au codec h.263. La perte d'une image complète ne

menant qu'à une baisse de la qualité de l'ordre de 0,2 dB. L'usage du codage de canal permettant de corriger le bruit de corrélation est également efficace contre les erreurs de transmission.

Nous avons jusqu'ici principalement évoqué les avantages du codage de Wyner-Ziv concernant la simplicité de l'encodeur vidéo réalisable. Il est intéressant d'interpréter la construction d'un codeur de Wyner-Ziv sous l'angle de la protection inégale aux erreurs (cf. Chapitre I, paragraphe 3.2.2). Dans le cas du codage d'une source binaire par plans de bits, le MSB est plus important que le LSB. Il convient classiquement de lui attribuer une plus grande protection. Mais dans le cas du paradigme de Wyner-Ziv, les LSB de X et de Y sont moins corrélés que les MSB. Il convient donc d'accorder un débit supérieur au codage des LSB pour éviter une erreur de décodage, puisque leur estimation à partir de Y ne sera pas fiable. L'approche basée sur l'utilisation du codage de Wyner-Ziv est donc à priori en contradiction avec celle des systèmes de transmission classiques [PRA02]. Le paragraphe suivant décrit cette approche au travers des schémas de mise en œuvre du codage résistant aux erreurs dérivés du paradigme de Wyner-Ziv.

2.3 Application à la protection contre les erreurs

Pour faire face à des dégradations qui peuvent être brutales, l'approche actuelle en codage vidéo consiste généralement à générer des représentations scalables qui seront ensuite protégées de manière inégale en fonction de leur importance par l'application d'un codage de canal. Cette stratégie permet de limiter de manière efficace l'impact des erreurs de transmission, mais engendre le plus souvent une complexité accrue.

Dans le cas du codage source avec information adjacente, l'information adjacente est souvent assimilée à une version bruitée de la source. Le codage de Wyner-Ziv revient à appliquer un codage de canal pour protéger la source à encoder contre le bruit de corrélation. En augmentant le débit alloué à ce codage de canal, les informations de parité générées peuvent également lutter contre le bruit d'un canal de transmission. C'est en se basant sur cette observation que des schémas de codage résistants aux erreurs ont été développés. Ils font l'objet des paragraphes suivants.

2.3.1 Codage scalable

Une autre perspective d'intérêt liée au codage de Wyner-Ziv est l'intégration de la scalabilité. Il a été prouvé [PRA01] que les informations de parité d'un code à effacement pouvaient, dans le cadre d'un codage conjoint source-canal, servir à affiner la qualité de la description d'une source d'information. En général, l'usage d'un code correcteur permet de corriger les erreurs de transmission et de garantir une qualité maximale fixée à l'avance. Dans le cas où la capacité de correction est insuffisante, on assiste à une brusque chute de la qualité reconstruite (« cliff effect »). Lorsque le taux d'erreur est inférieur aux prévisions, cette protection est superflue. Mais pour une configuration Wyner-Ziv, la parité n'est pas que redondance, elle constitue une information (conditionnelle) sur la source à encoder. Ce résultat est valable pour des sources corrélées X et Y quelconques telles que $X = Y + Z$, où Z est gaussienne et indépendante de l'information adjacente [CHE04]. Dans ce cas, le codage par couche n'impose aucune perte de performance comparé à un codage non scalable.

Afin de rendre le codeur Wyner-Ziv scalable, on peut générer un flux de WZ par couches successives [CHE04]. L'encodage et le décodage sont effectués sur plusieurs niveaux et chaque décodeur utilise les informations générées par les niveaux de codage inférieurs et l'information adjacente. Pour leur implémentation, la quantification scalaire emboîtée est appliquée aux coefficients transformés de la source X . Le codage de Slepian-Wolf (cf. Figure 2.19), qui dans le cas présent exploite la corrélation entre les plans de bits quantifiés de X et l'information adjacente Y , est réalisé grâce aux codes LDPC (Low-Density Parity-Check codes). La scalabilité obtenue est à grain fin (Fine Grain Scalability, FGS).

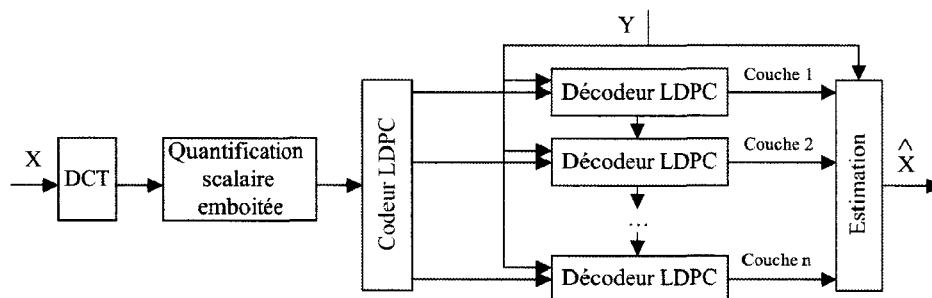


Figure 2.19 : Codeur de Slepian-Wolf par couche.

La corrélation entre couche de base et flux de Wyner-Ziv diminue au fur et à mesure que l'on passe du MSB au LSB. Les plans de bits les plus significatifs peuvent être supprimés lors de la quantification car ils sont essentiellement identiques à l'information adjacente. Les plans de bits les moins significatifs sont quantifiés à 0 et seuls les plans de bits médians sont transmis après compression par le codeur de Slepian-Wolf [XUQ04]. En considérant le flux issu d'un codeur

MPEG standard comme couche de base ou information adjacente, le flux Wyner-Ziv fournit une amélioration de la qualité qui est fonction du nombre de couches reçues (cf. Figure 2.20).

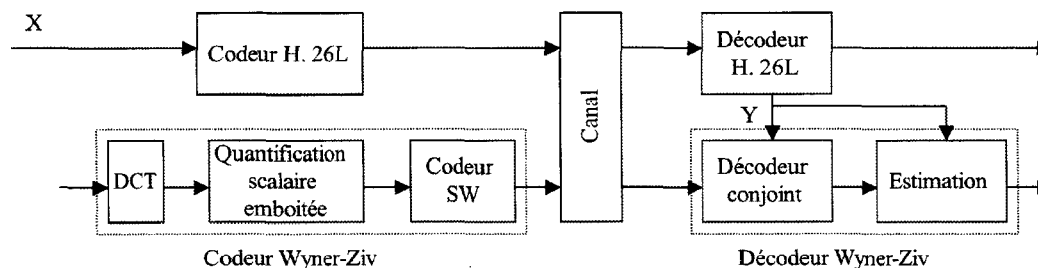


Figure 2.20 : Codeur de Wyner-Ziv par couches [XUQ04].

Les résultats montrent que le codage par couches ne subit quasi aucune perte par rapport au codage de Wyner-Ziv monocouche. D'autre part, la qualité de la vidéo reconstruite est proportionnelle au nombre de couches d'amélioration décodées. Les performances du système mis en place égalent celles du codage h.263 FGS, mais restent inférieures de 1 à 2 dB à celles du codec h.26L classique (cf. Figure 2.21). La quantité de mouvement de la séquence encodée joue sur la corrélation entre la source et l'information adjacente (couche de base), les performances sont donc plus élevées pour une séquence moins animée.

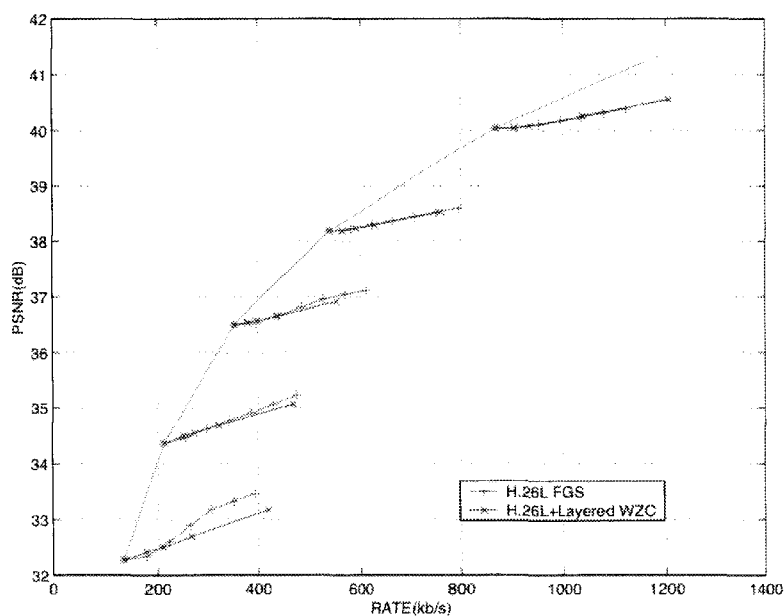


Figure 2.21 : Comparaison des performances du codeur h. 26L FGS et du codeur WZ par couches, utilisant des flux H26L de différents débits comme couche de base / information adjacente.

Dans cette version du système, on suppose une réception sans erreur des données. Néanmoins, la robustesse du système aux erreurs a été étudiée. La simulation est effectuée sur un GOP de 15 images de type IPPP... L'évolution du PSNR pour un taux de macroblocs en erreur de 1% à chaque image du GOP est reportée Figure 2.22. Le codec est naturellement plus robuste aux erreurs pour des performances tout à fait comparables en termes d'efficacité de codage. La propagation temporelle des erreurs est fortement atténuée.

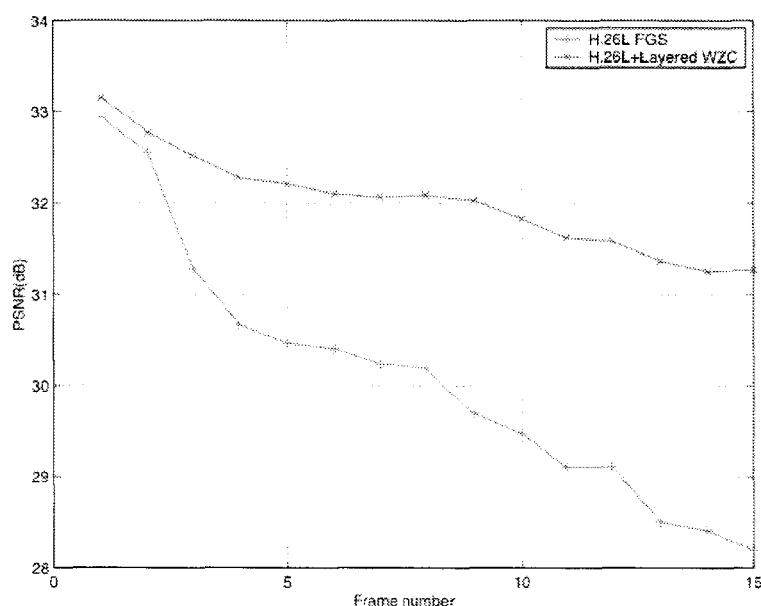


Figure 2.22 : Observation de l'évolution du PSNR dans le cas d'1% de macroblocs en erreur dans le cas d'un codeur WZ par couches et d'un codeur h. 26L FGS.

Considérant la robustesse aux erreurs inhérente à leur schéma, la transmission sur des canaux non fiables a été considérée [XUQ04b], [XUQ06]. Plutôt que de protéger le flux issu du codeur de Slepian-Wolf, l'idée est de n'utiliser qu'un seul codage de canal pour réaliser le codage entropique (codeur de Slepian-Wolf) et le code correcteur. Le canal virtuel de corrélation et le canal de transmission sont associés pour former un unique canal. La quantité de parité émise est supérieure à celle requise par le théorème de Slepian-Wolf et permet la protection contre les erreurs de transmission.

Différents cas de transmission ont été étudiés incluant les canaux bruités et les canaux à effacement [XUQ04]. Les résultats indiquent que l'implémentation de leur algorithme permet d'obtenir un débit supérieur au débit théorique de 0.07 bits/échantillon dans le cas d'un canal BSC de probabilité d'erreur 0,05. Dans le cas d'un canal à bruit blanc gaussien additif (AWGN), la perte n'est que de 0,05 bits/échantillon.

Dans le cas général, si la couche de base est corrompue, le décodage des couches d'amélioration les plus hautes échoue. Les couches d'amélioration basses permettent toujours d'obtenir une reconstruction de meilleure qualité et la propagation des erreurs est enrayée, même si les performances du système chutent globalement de 2 à 4 dB.

2.3.2 Codage de Wyner-Ziv résistant aux erreurs

Le codage de Wyner-Ziv possède une robustesse aux erreurs intrinsèque déjà mise en relief dans l'algorithme PRISM (paragraphe 2.2.2.2). Le bruit de corrélation entre deux images adjacentes est corrigé grâce au codage de canal. Dans le cas où un bruit de canal s'ajoute à ce bruit de corrélation, le schéma ne voit que la somme de ces deux composants et la distinction n'est pas nécessaire. Si le codage de canal est suffisamment fort, il n'y aura pas d'erreur de décodage. L'analogie entre codage de canal et codage source avec information adjacente a permis la conception de méthodes dédiées à la protection contre les erreurs. Un flux de Wyner-Ziv est alors généré en plus du flux vidéo principal. L'information portée par les bits issus du codage de Wyner-Ziv est utilisée pour corriger les erreurs de transmission.

La principale faiblesse du codage vidéo prédictif lors d'une transmission sur un canal bruité est la propagation temporelle des erreurs. Un échantillon erroné servira de prédicteur à un échantillon de l'image suivante dont la valeur décodée sera faussée. Pour lutter contre ce phénomène appelé erreur de glissement (ou drift error), on peut insérer périodiquement une information codée en intra. La boucle de prédiction temporelle est alors remise à zéro et la propagation de l'erreur interrompue (paragraphe 1.3.1.1). Mais le surcoût de débit engendré limite le recours à cette technique.

Le codage source prédictif a été reformulé comme un problème de codage source avec information adjacente ou codage de canal [SEH03]. En effet, l'information générée par un codeur de Wyner-Ziv permet de corriger l'erreur entre l'information adjacente et la source corrélée. Par extension, elle peut être utilisée pour corriger l'erreur entre un symbole prédit et son prédicteur. Si cette erreur est augmentée par une erreur de transmission, le problème sera résolu par la transmission d'un surplus de parité.

Cette remarque a permis de concevoir un codec pour lequel les prédicteurs utilisés à l'encodeur et au décodeur peuvent différer [SEH03], [SEH04]. Le flux est encodé par un codeur h.26L classique. Un nouveau type d'image est défini par les auteurs (images « peg ») et inséré périodiquement dans le flux. Pour ces images, le système transmet la valeur du symbole après codage inter et l'index du sous-ensemble auquel il appartient dans des paquets distincts. L'index du sous-ensemble est obtenu grâce aux turbo-codes [SEH03] ou codes LPDC [SEH04]. Au

décodeur, ces deux informations garantissent la bonne reconstruction du symbole en cas d'erreur (cf. Figure 2.23).

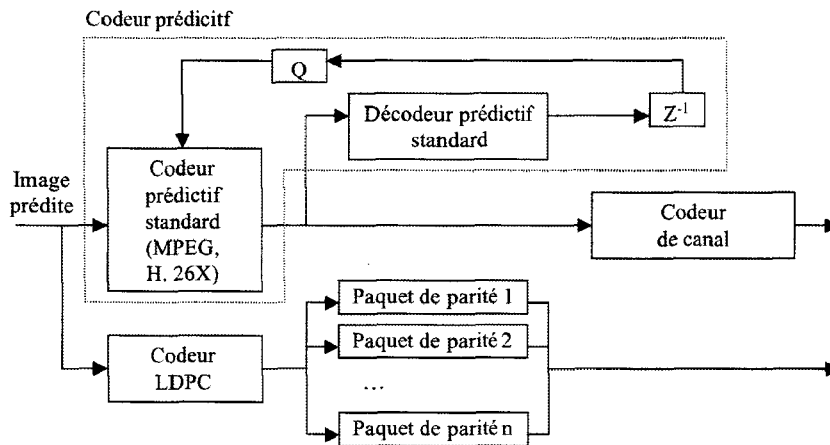


Figure 2.23 : Encodeur vidéo résistant à la propagation temporelle des erreurs [SEH04].

Le taux de protection fourni et la périodicité des images « peg » est déterminé en fonction du canal. L'index du sous-ensemble est supposé reçu sans erreur et nécessite donc une lourde protection par FEC. Les performances du système dépassent celle d'un rafraîchissement intra des images de 3 à 4 dB. La limite théorique d'un système idéal opérant avec un délai infini est situé 1 à 2,5 dB au-dessus des résultats obtenus (cf. Figure 2.24).

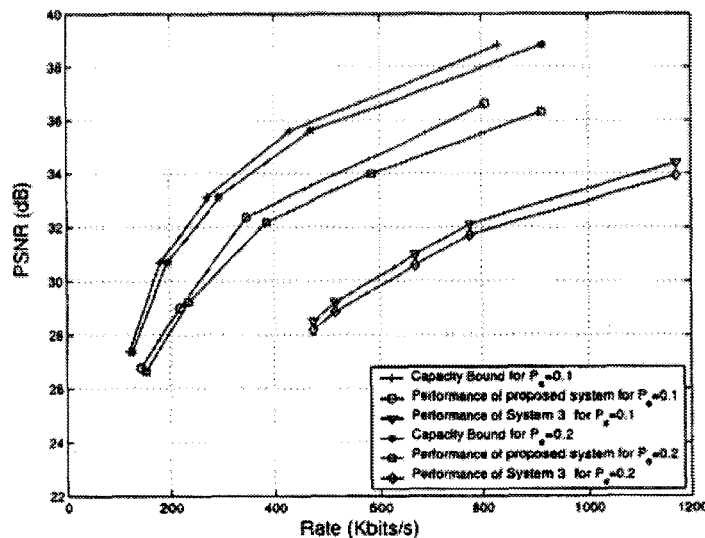


Figure 2.24 : Comparaison du PSNR de l'approche proposée par [SEH04] avec la limite théorique et l'insertion d'images codées en intra.

L'interprétation du codage prédictif comme la transmission d'un prédicteur sur un canal hypothétique dont la sortie est le symbole prédit a également été exploitée dans le cas d'un codage scalable [SEH04b]. En effet, quand le flux est accessible à différents degrés de fidélité, la sélection du prédicteur à utiliser est problématique. La réponse classique est la sélection systématique de la couche de base comme prédicteur, même si les couches d'amélioration sont disponibles. Cette solution engendre une perte qui peut être significativement réduite en utilisant un codage de Wyner-Ziv (gain de 4 dB).

2.3.3 Codage systématique résistant aux erreurs

Le codage systématique consiste à appliquer aux données à transmettre un code correcteur dont la parité est simplement multiplexée temporellement au flux d'entrée (cf. Chapitre I, paragraphe 2.2). L'avantage de ce procédé est sa parfaite compatibilité avec le système existant puisque la structure du flux transmis n'est pas modifiée. Classiquement, on transmet les données et la parité, le décodeur utilise cette parité pour corriger les erreurs sur les données transmises. Une alternative à ce schéma peut être apportée là encore par le codage de sources distribuées qui se substitue alors au codage canal.

2.3.3.1 Amélioration d'image analogique grâce à l'information adjacente numérique

L'idée fut introduite par Pradhan et Ramchandran pour protéger un flux analogique à l'aide d'informations disponibles via un canal numérique [PRA01b]. Plutôt que d'utiliser le paradigme de Wyner-Ziv pour compresser l'information, il est utilisé pour améliorer la qualité de l'information délivrée. La corrélation entre la version analogique et la version numérique d'une séquence vidéo permet de compresser la version numérique avec un encodeur de Wyner-Ziv (cf. Figure 2.25). À la réception, la version analogique, ou information adjacente, permet de déterminer quelle information afficher parmi celles désignées par le flux de Wyner-Ziv.

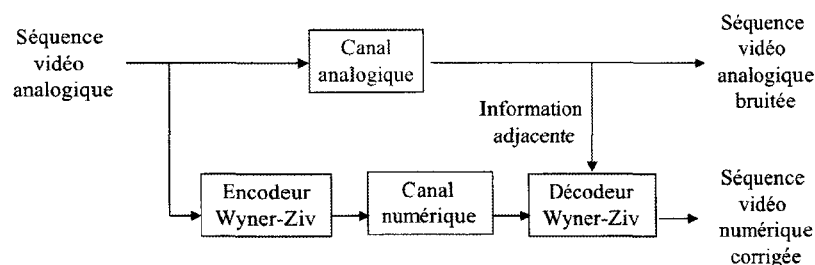


Figure 2.25 : Illustration du codage source canal systématique.

Comme nous allons le voir dans le paragraphe suivant, cette idée fut reprise et appliquée au cas de la transmission d'un flux vidéo numérique compressé. On adjoint à ce flux une protection générée grâce au codage de Wyner-Ziv. À la réception, la version corrompue du flux sert d'information adjacente pour le décodage du flux de Wyner-Ziv et une reconstruction de qualité supérieure est obtenue.

2.3.3.2 Systematic Lossy Error Protection (SLEP)

La portion systématique est constituée par un flux vidéo MPEG conventionnel transmis sur un canal bruité sans protection. Le flux adjacent est une représentation bas débit de la vidéo transmise générée indépendamment par l'encodeur de Wyner-Ziv. Le décodeur combine les deux flux reçus pour produire un signal vidéo de qualité supérieure X^* (cf. Figure 2.26).

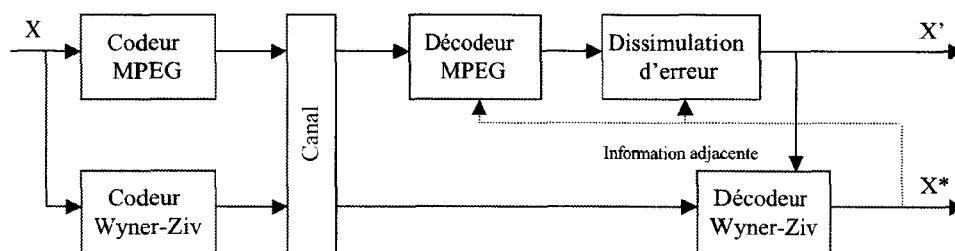


Figure 2.26 : Schéma de protection systématique d'un flux vidéo grâce au paradigme de Wyner-Ziv [AAR03b].

Dans leur première implémentation [AAR03b], le codeur de Wyner-Ziv est constitué d'un quantificateur grossier et les turbo-codes sont utilisés pour le codage de Slepian-Wolf. À la réception, le flux MPEG principal est décodé et les erreurs de transmission dissimulées. Au décodeur, la version requantifiée corrompue du flux MPEG est utilisée comme information adjacente pour décoder les bits du flux de Wyner-Ziv (cf. Figure 2.27). Si le débit WZ est suffisant pour le bruit du canal, ces bits permettent de corriger la version WZ corrompue qui est substituée aux portions en erreur du flux MPEG.

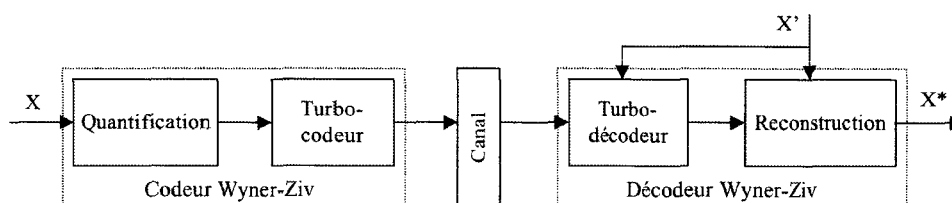


Figure 2.27 : Codeur de Wyner-Ziv basé sur une quantification scalaire et des turbo-codes [AAR03b].

La qualité de la vidéo décodée sera donc fonction de la quantité d'erreurs. Les zones non-décodables du flux MPEG sont remplacées par les zones correspondantes de la version Wyner-Ziv. Cette version étant grossièrement quantifiée, on observera une diminution du PSNR supplémentaire pour chaque zone remplacée. Ce système permet néanmoins de garantir une qualité reconstruite minimale tant que la capacité du code n'est pas dépassée.

On remarque que pour un même débit du flux Wyner-Ziv, plus la quantification est sévère, plus le codage de canal sera fort. En cas d'erreur sur le flux MPEG, si la protection est suffisante, la qualité décodée sera celle de la version quantifiée grossière. Si la protection est insuffisante, le décodage échouera et la qualité décodée sera déplorable. La qualité de la vidéo reconstruite est fonction de ses deux paramètres, il existe donc un compromis entre robustesse et distorsion.

Cette configuration peut-être étendue au cas de multiples flux Wyner-Ziv de taux de quantification différents comme le montre la Figure 2.28 pour deux niveaux de qualité. Le compromis protection / distorsion peut alors être adouci. Si le canal est bon (nombre d'erreurs peu élevé) on pourra décoder les deux flux de Wyner-Ziv et la qualité sera améliorée par rapport au cas d'un seul flux WZ. Si le canal se dégrade, on ne décodera plus que le flux le plus grossier et la qualité diminuera. On notera que la protection étant alors répartie entre flux de WZ fin et flux de WZ grossier, la résistance aux erreurs se dégrade par rapport au cas d'un seul flux WZ.

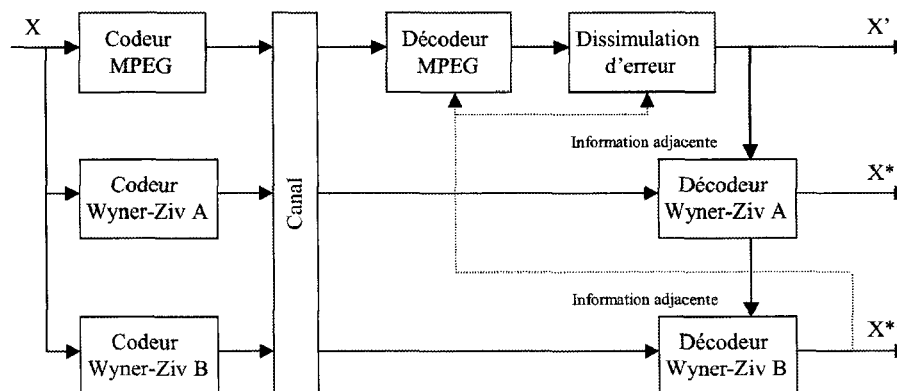


Figure 2.28 : Protection d'un flux vidéo par de multiples flux de Wyner-Ziv.

Le schéma décrit protège les bits du flux MPEG dans le domaine spatial. Une amélioration a été obtenue en appliquant le système au domaine transformé [RAN04]. Le codage de Slepian-Wolf est réalisé par un code de Reed-Solomon dont seule la parité est transmise. L'application des codes se fait à travers les slices d'une image codée (cf. Figure 2.29), ce qui correspond à un brassage de degré la taille d'une slice (ou tranche) et permet d'améliorer l'efficacité de la correction en cas de salves d'erreurs ou de canaux à effacement.

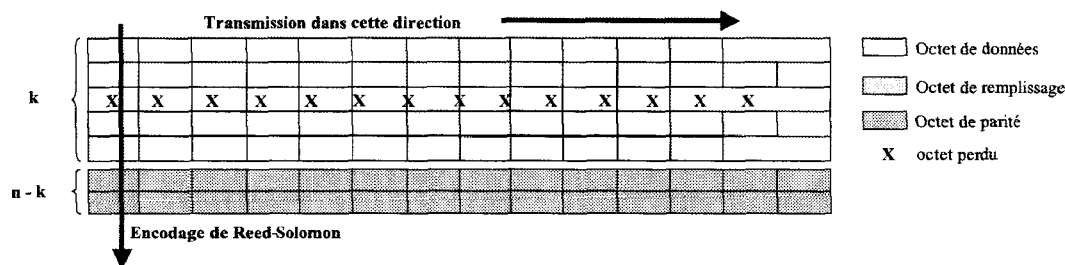


Figure 2.29 : Application des codes de Reed-Solomon à travers les slices d'une image.

Dans un premier temps, le flux de Wyner-Ziv était généré à partir d'une requantification grossière des coefficients dans le domaine transformé. Afin d'augmenter la corrélation entre flux principal et flux de Wyner-Ziv, le flux de Wyner-Ziv est généré en transcodant le flux MPEG principal (cf. Figure 2.30).

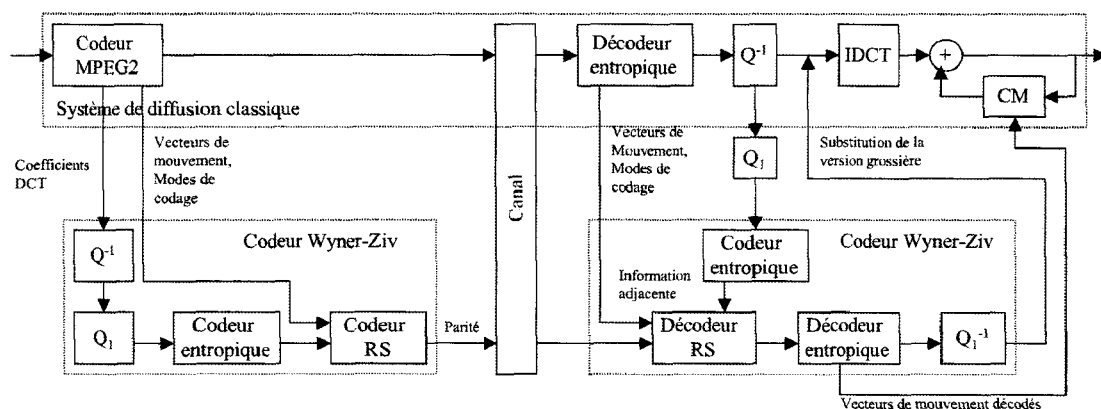


Figure 2.30 : Schéma de protection systématique avec pertes d'un flux vidéo MPEG utilisant les codes de Reed-Solomon.

Les résultats permettent de comparer le système avec le codage FEC classique. On notera que ce dernier est un cas particulier du schéma SLEP où la version grossière correspond à la version fine (pas de requantification). La requantification permet de diminuer le débit de l'information à protéger et d'augmenter la capacité correctrice à débit de la protection égal. La protection ne portant plus que sur les informations les plus importantes (au sens SNR) du flux original, le schéma apporte une protection inégale sans nécessité d'un codage scalable. Le « cliff effect » due à une protection par FEC simple peut ainsi être évité. Les dégradations sont douces pour un intervalle de taux d'erreur plus large que dans le cas FEC (cf. Figure 2.31). Ce système est donc particulièrement adapté aux canaux de caractéristiques d'erreurs variables.

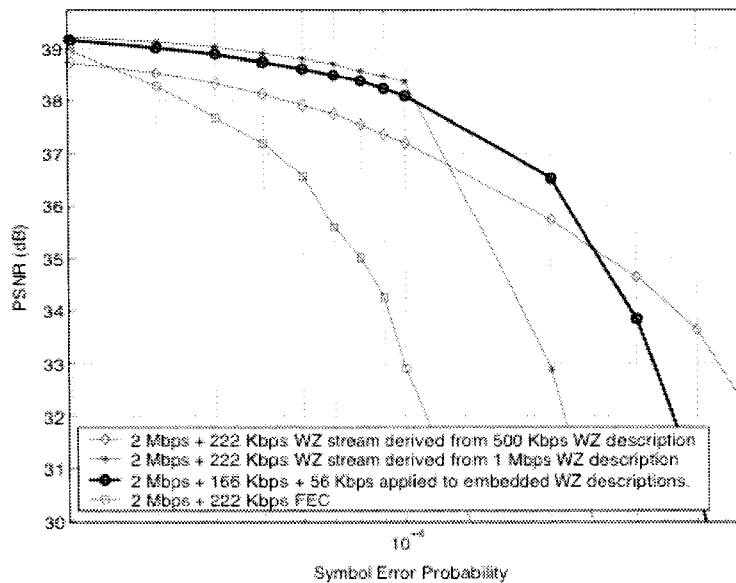


Figure 2.31 : Comparaison des performances du schéma pour une protection de qualité grossière, plus fine ou mixte [RAN05].

Une étude centrée sur les flux multiples a été réalisée [RAN05]. Les performances de ce schéma furent également évaluées en effectuant une comparaison avec un codage par couches [RAN05b]. Le cas d'une application à un scénario de transmission sur un réseau Ad Hoc a aussi été étudié [ZHU05]. Plus récemment, le système a été adapté au format MPEG4-AVC [RAN06]. L'option du standard permettant de générer des slices redondantes est utilisée et la protection porte alors sur ces dernières. Une analyse du fonctionnement de système a été développée [RAN04b], [RAN06b].

2.4 Conclusion

Nous avons évoqué au long de ce chapitre les bases théoriques, méthodes de construction et principales implémentations du codage de Wyner-Ziv. Parmi les différentes applications possibles, le schéma SLEP (cf. paragraphe 2.3.3.2) présente de nombreux avantages :

- la compatibilité avec les systèmes existants puisque le flux de Wyner-Ziv peut être ignoré si le décodeur ne sait l'exploiter,
- la simplicité de mise en œuvre puisque les blocs constitutifs du systèmes (quantificateur et codeur de Reed-Solomon sont bien connus),

- souplesse du schéma qui peut être adapté à différents formats du flux principal (MPEG 1, 2, 4),
- efficacité face aux erreurs d'un codage scalable sans en requérir la complexité et sans engendrer la perte d'efficacité usuelle.

C'est sur cette implémentation que le système présenté dans le chapitre suivant est basé. Or, on constate actuellement une prolifération des standards de diffusion et capacités d'affichage (par exemple TVHD, expérimentation de vidéo sur les terminaux mobiles avec le DVB-H). Un schéma de type SLEP qui offrirait des facilités de conversions de format constituerait à l'évidence un atout pour le développement futur de systèmes de vidéo communications. Nous avons donc adapté l'architecture ci-dessus à une scalabilité spatiale plutôt que SNR.

3 Schéma de protection spatiale inégale systématique proposé et analyse

Nous avons présenté, dans le chapitre précédent, les principes et différentes applications du codage de Wyner-Ziv. En particulier, les propriétés de résistance aux erreurs inhérentes à la construction de tels encodeurs ont été mises en évidence. Elles ont permis de développer un schéma de protection systématique d'un flux vidéo simple à implémenter et qui, comme un flux scalable, améliore la constance de la qualité reconstruite quand le taux d'erreur est variable (cf. Chapitre II, paragraphe 3.3.2).

L'implémentation proposée dans [RAN04] permet de garantir une qualité minimale basée sur un critère de type SNR. La version de qualité minimale (équivalent de la couche de base d'un codage scalable) est obtenue par requantification du flux vidéo principal. Étant donné l'existence de multiples formats de diffusion, en particulier dans le contexte du passage à la Haute Définition (HD), il existe de nombreux scénarios d'application qui nécessitent une conversion de format d'affichage. Nous nous sommes interrogés sur l'adaptation de la méthode à une protection inégale des données selon leur importance en termes de résolution spatiale. En cas d'erreurs de transmission, le flux Wyner-Ziv permet au récepteur de restaurer la description de basse résolution spatiale qui peut ensuite être affichée au format d'origine ou avec une taille d'image réduite. L'inconvénient de la complexité d'implémentation liée au filtrage spatial a pu être écarté grâce aux techniques de réduction de résolution dans le domaine transformé. Une étude comparative des flux après réduction de débit SNR et spatiale à débit égal nous a permis de valider notre approche.

La protection fournie par ce schéma varie en fonction de multiples paramètres comme le taux de réduction du débit, le taux de protection, le type d'erreur rencontré ou les propriétés de la séquence encodée. L'optimisation de la transmission doit donc passer par une étude approfondie de l'impact de ces paramètres sur la distorsion finale. Afin d'étudier les interactions des grandeurs concernées, nous avons développé un modèle analytique du système proposé grâce auquel les paramètres optimaux peuvent être déterminés.

Le chapitre est organisé comme suit : tout d'abord, nous présentons et justifions le nouveau schéma dérivé du schéma SLEP [RAN04]. Dans un deuxième temps, nous analysons les propriétés du canal de transmission et les étapes du transcodage. Finalement, ces informations permettent de modéliser l'ensemble de la chaîne de transmission et d'estimer ainsi la distorsion globale.

3.1 Présentation du système proposé

Nous nous proposons de développer un schéma SLEP qui utilise une scalabilité spatiale (cf. Figure 3.1). Dans ce schéma, un flux de protection additionnel est généré grâce au codage de Wyner-Ziv d'une version basse résolution spatiale du flux MPEG principal. Le flux principal non protégé et le flux de Wyner-Ziv sont transmis au décodeur. En cas d'erreur, le flux de Wyner-Ziv est décodé grâce au flux MPEG corrompu, ce qui permet de reconstruire les informations basse résolution perdues.

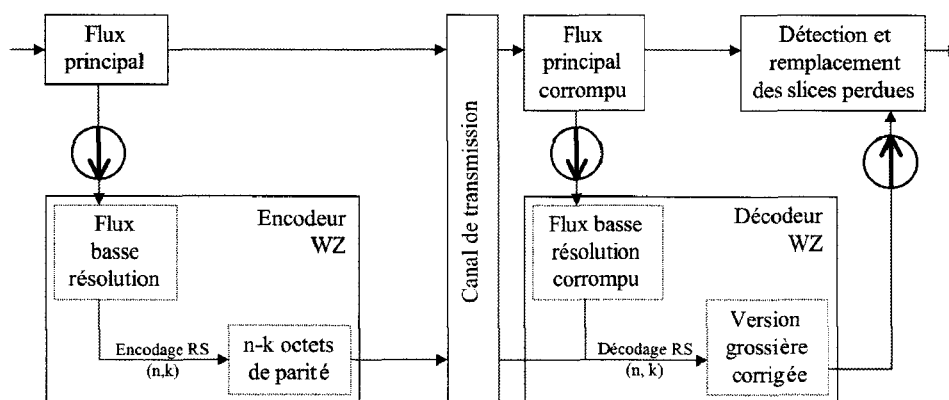


Figure 3.1 : Système proposé, protection contre les erreurs systématique basée sur une scalabilité spatiale.

Le premier problème rencontré est lié à la complexité d'implémentation. Le filtrage spatial est habituellement réalisé sur les pixels décodés du flux. Nous avons donc étudié une solution alternative consistant à décimer les coefficients dans le domaine DCT, permettant ainsi de diminuer la résolution spatiale sans passer par un décodage complet du flux.

D'autre part, la qualité visuelle d'un flux de basse résolution spatiale par rapport à celle d'un flux quantifié grossièrement, à débit égal, doit être évaluée. En effet, en cas d'erreur sur le flux principal, la version Wyner-Ziv est affichée. Il nous faut vérifier que la version basse résolution affichée satisfait des critères de qualité visuelle au moins comparables à la version requantifiée.

Ces deux aspects de l'implémentation de notre schéma sont présentés dans les paragraphes suivants. Finalement, le schéma original de protection inégale spatiale d'un flux vidéo compressé est détaillé.

3.1.1 Décimation DCT

Notre objectif est d'adapter le schéma SLEP à une scalabilité spatiale. Pour cela, le flux de Wyner-Ziv doit être construit non plus à partir d'une version requantifiée (cas SNR) du flux principal mais d'une version basse résolution (cas spatial). La quantification est réalisée dans le domaine transformé, elle ne nécessite que le décodage des codes à longueur variable. Le filtrage spatial classiquement utilisé impose un décodage complet du flux puisqu'il est réalisé au niveau pixel. La complexité de mise en oeuvre de la scalabilité spatiale doit être réduite.

La scalabilité spatiale est définie comme la représentation d'une même vidéo à différentes résolutions ou tailles d'image (cf. paragraphe 1.3.2.3). Concernant notre application, l'image bas débit obtenue doit être de résolution inférieure mais de taille identique à l'image originale. Dans le cas d'un codage par couches, la couche de base fournit une version de prévisualisation de l'image encodée à une résolution inférieure [WAN02b]. Le décodage de chaque couche d'amélioration permet d'augmenter la résolution spatiale de l'image reconstruite, jusqu'à atteindre la résolution de l'image d'origine.

La méthode classique d'obtention d'un flux à scalabilité spatiale est illustrée Figure 3.2 dans le cas de deux couches. La couche de base est générée par le codage mono résolution de l'image sous-échantillonnée spatialement. La version décodée de la couche de base est interpolée pour former une estimation de la version de résolution supérieure. Cette estimation est soustraite à l'image d'origine puis codée pour former la couche d'amélioration.

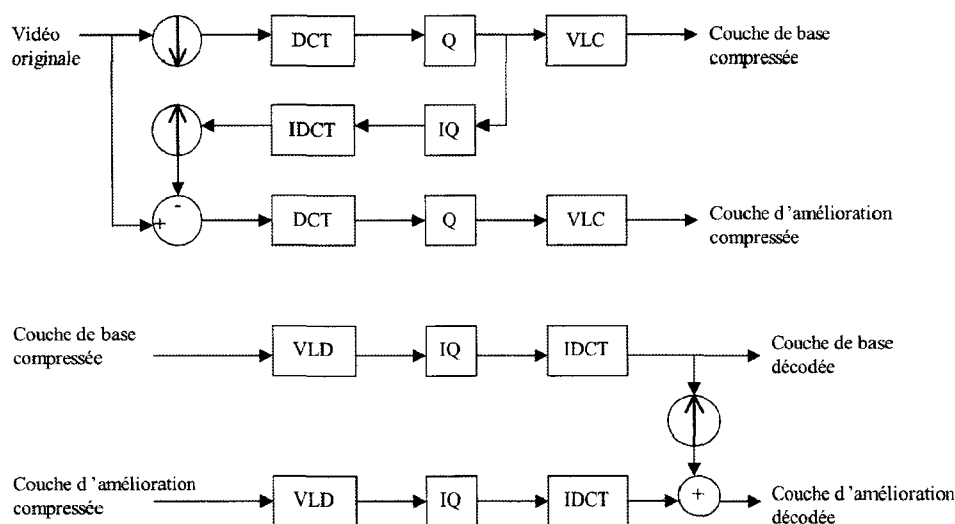


Figure 3.2 : Codeur (a) et décodeur (b) à scalabilité spatiale [WAN02b].

Cette architecture cascadée est très souple (en termes de tailles d'image accessibles) mais complexe à implémenter pour des applications temps réel. La génération des diverses résolutions impose de repasser dans le domaine spatial, donc de décoder puis ré-encoder complètement le flux. De plus, concernant notre application, la couche de base est de taille réduite et nécessitera une interpolation avant affichage.

Or, nous avons vu dans le premier chapitre que les coefficients d'un bloc après DCT traduisent la participation de motifs fréquentiels au contenu du bloc à coder. En utilisant le cas 1D comme illustration, on note N le nombre de pixels. Le signal dans le domaine spatial peut être vu comme une combinaison linéaire de vecteurs de base de la forme :

$$\Phi_{k,n} = \cos \frac{2\pi(n + \frac{1}{2})k}{2N} \tag{3.1}$$

On peut représenter les vecteurs de base de la DCT 1D appliquée à une ligne de huit pixels comme le montre la Figure suivante :

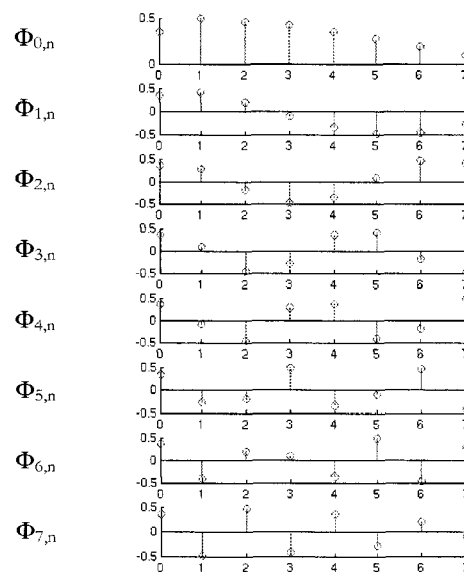


Figure 3.3 : Représentation des vecteurs de base de la DCT 1D.

Les k premiers coefficients codent les variations spatiales inférieures à $\frac{k}{2}$ alternances noir-blanc (pour la luminance par exemple) par bloc (cf. Chapitre I, Figure 1.5). D'après le théorème de

Shannon, k échantillons suffisent à représenter les variations codées par X_k . En supprimant les coefficients DCT d'indices supérieurs, on élimine les variations spatiales plus rapides, celles qui nécessiteraient plus de k pixels pour être représentées dans le domaine spatial. Il est possible de dériver de ces propriétés des architectures efficaces de conversion de résolution spatiale dans le domaine DCT [RAO91]. Par exemple, en forçant à zéro les coefficients représentant les hautes fréquences, on réduit la résolution spatiale de l'image (cf. Figure 3.4).

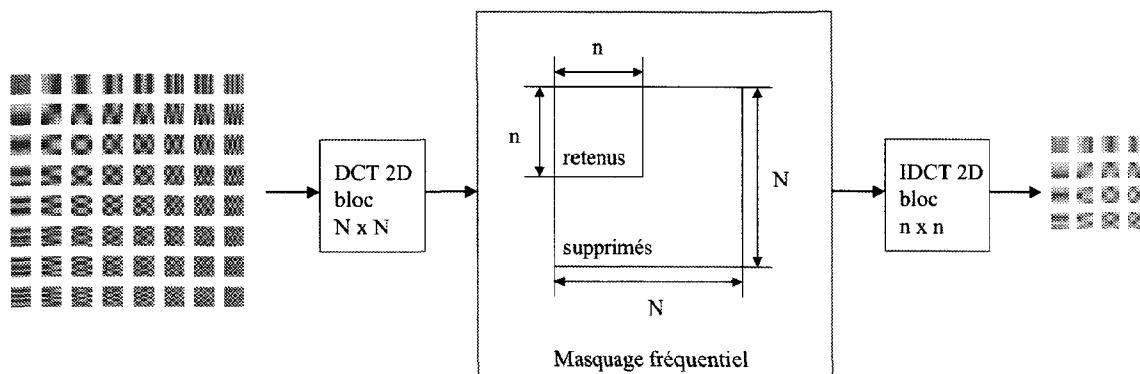


Figure 3.4 : Schéma de principe de la diminution de résolution spatiale par décimation des coefficients DCT [RAO91].

La partition des coefficients fréquentiels en différentes couches de résolutions spatiales variables est qualifiée de scalabilité fréquentielle [WAN02b]. La couche de base fournit alors une version de l'image dont la précision est améliorée avec le décodage des couches supérieures. Si l'on ne garde par exemple que les 4×4 premiers coefficients du bloc de 8×8 coefficients DCT (les coefficients restants sont mis à zéro par masquage fréquentiel), deux solutions existent pour reconstruire l'image finale :

- La reconstruction d'une image de taille réduite [RAO91], [BAO96], [SKO98] : ceci ne correspond pas à notre objectif.
- La reconstruction d'une image de résolution réduite mais de taille identique : on applique une IDCT de taille 8×8 au bloc de 8×8 coefficients issus du masquage fréquentiel (cf. Figure 3.5). On reconstruit alors une image de taille identique à l'originale, mais de résolution réduite de rapport $8 : 4$ horizontalement et verticalement.

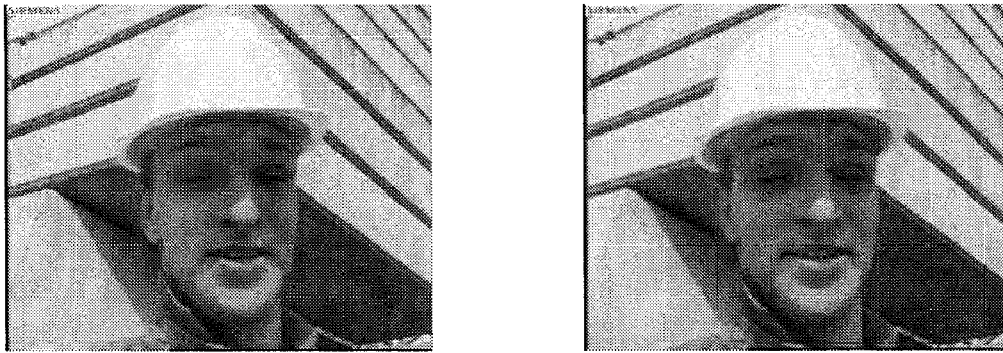
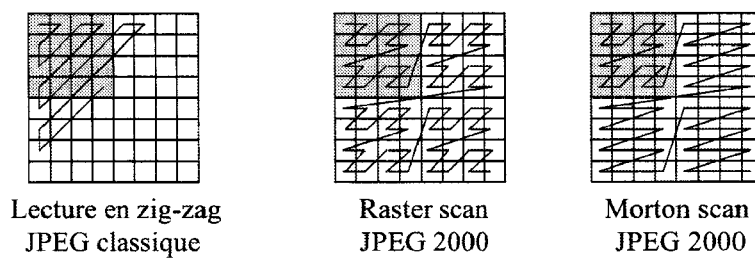


Image codée

Image après décimation DCT 8 :4

Figure 3.5 : Réduction de la résolution spatiale d'une image par masquage fréquentiel.

Le procédé ne requiert pas d'implémentation particulière de la DCT / IDCT. Le codage entropique classique peut être utilisé, même si une modification de l'ordre de sérialisation s'avère souhaitable. En effet, si la lecture en zigzag classique est utilisée, certains coefficients DCT systématiquement mis à zéro lors du masquage fréquentiel devront être codés. Afin de ne pas transmettre ces zéros, on peut utiliser par exemple les ordres de sérialisation du standard JPEG2000 [TAU02] (cf. Figure 3.6). D'après les expériences préliminaires réalisées, cette modification de l'ordre de lecture des coefficients DCT permettrait un gain de 10% sur le débit final pour les décimations de rapport 8 :2 à 8 :4. Nous avons cependant préféré garder le zigzag classique pour des raisons de compatibilité avec les systèmes existants. Si de plus, le codage de Huffman peut être modifié afin de mieux tirer parti des statistiques particulières du bloc à coder, un gain de débit supplémentaire est possible.



Lecture en zig-zag
JPEG classique

Raster scan
JPEG 2000

Morton scan
JPEG 2000

Figure 3.6 : Sérialisation des coefficients DCT après décimation.

Nous avons choisi la diminution de résolution avec conservation de la taille d'image d'origine pour sa simplicité d'implémentation. L'avantage pour notre application est que l'on n'aura pas à interpoler (sur-échantillonner) la version bas débit corrigée avant de la réinsérer dans le flux principal corrompu, la taille d'image n'ayant pas été modifiée. Plus généralement, si un des relais

de la transmission ou le décodeur lui-même est apte / amené à diminuer la taille de l'image afin de s'adapter au format du récepteur, la décimation DCT permettant d'améliorer la protection constitue également la première étape de cette opération. Les méthodes citées ci-dessus peuvent alors être utilisées avec une complexité réduite et la version du flux de taille réduite obtenue sera intégralement protégée. Le schéma résultant est donné Figure 3.7 [RAM05].

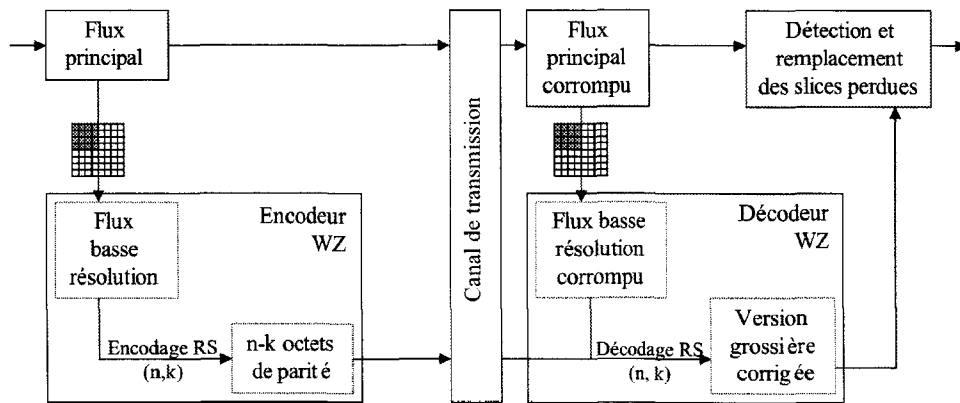


Figure 3.7 : Schéma de protection avec décimation DCT.

On notera que le masquage fréquentiel permet également d'alléger la complexité de la diminution de taille d'image dans le domaine fréquentiel [SKO98]. Si le format d'affichage doit être réduit à la réception ou en cours de transmission (transcodeur relais situé par exemple dans une station de base d'un réseau sans fil), la manipulation sera grandement simplifiée dans le cas où une décimation dans le domaine DCT a été appliquée à l'image auparavant.

3.1.2 Comparaison Spatial / SNR

Le schéma de codage résistant aux erreurs présenté ci-dessus est basé sur la protection d'une version bas débit de la vidéo encodée. Cette version peut être obtenue en augmentant le facteur de quantification (scalabilité SNR) ou en décimant les coefficients AC du bloc (scalabilité spatiale). Les méthodes d'obtention étant différentes, les versions bas débit résultantes ont des caractéristiques en termes de débit et de qualité visuelle différentes.

3.1.2.1 Critères de comparaison

On cherche à établir une comparaison entre les qualités de l'image décodée et corrigée dans le cas d'une scalabilité SNR et spatiale pour un débit de la protection fixé. Dans le schéma de protection SLEP, l'image basse résolution n'est pas transmise, le volume de données à considérer pour la protection est constitué des $(n - k)$ slices de parité seules. Les images bas

débit n'interviennent dans la reconstruction de l'image finale qu'en cas d'erreur et proportionnellement au nombre de slices corrigées.

→ Il faut donc évaluer la qualité de l'image basse résolution de chacune des méthodes abordées pour un débit de la protection et une résistance aux erreurs identiques. Pour obtenir un débit et une résistance aux erreurs identiques, le nombre et la taille des slices de parité doivent être les mêmes.

Or la longueur d'une slice de parité du flux WZ est la longueur maximale d'une slice de la version bas débit utilisée (cf. Chapitre II, Figure 2.28). Pour notre application, nous nous intéressons donc au débit de slice maximal plus qu'au débit de l'image codée. Les slices plus courtes seront complétées par des zéros jusqu'à obtention de la taille maximale avant que le codage de Reed-Solomon ne soit appliqué. Plus la taille de slice est variable, plus la protection sera appliquée aux zéros de remplissage et moins elle sera utile.

Prenons l'exemple d'une décimation de rapport 8 :2 en ligne et en colonne, appliquée à une image de la séquence foreman. L'image bas débit obtenue par requantification correspondante a été déterminée en se basant sur la longueur des slices de parité (longueur maximale des slices image après codage entropique). On constate que pour une longueur maximale de slice identique, le débit de l'image de basse résolution spatiale est nettement plus élevé que celui de l'image de basse résolution SNR (cf. Tableau 3.1).

	Cas SNR	Cas spatial
Longueur de slice maximale	1256 bits	1264 bits
Débit total de l'image basse résolution	22,371 Kbits	35,572 Kbits
Rapport lg. Max/ lg. Moy	2.0128	1.2742

Tableau 3.1 : Caractéristiques des images basses résolution obtenues par requantification (SNR) et masquage fréquentiel (spatial).

La longueur des slices après requantification est plus variable que celle des slices après décimation dans le domaine DCT. En effet, en limitant le nombre de coefficients DCT non nuls du bloc, le masquage fréquentiel limite la longueur maximale des slices après codage entropique. La quantité d'information portée par chaque slice de l'image est plus constante, moins dépendante du contenu. En conséquence, pour un même débit de parité, le débit image protégé est plus important dans le cas d'une image de résolution spatiale réduite. Une solution consisterait à appliquer un codage de type EREC (Error Resilient Entropy Coding, Chapitre I paragraphe 3.1.2, [RED96]) afin d'obtenir des slices de longueur fixe. La complexité d'implémentation induite n'est pas justifiée dans le cas de la scalabilité spatiale, mais pourrait l'être dans celui de la scalabilité SNR.

On ajoutera que pour un débit du flux WZ fixé, la protection est d'autant plus efficace que le nombre de slices de parité envoyé est grand. La longueur de slice maximale doit donc être la plus faible possible. Si on se base sur des images de basse résolution de même débit, la taille maximale de slice sera plus courte dans le cas d'une scalabilité spatiale et la protection sera donc plus efficace.

3.1.2.2 Qualité visuelle

Pour les deux types de construction du flux WZ présentées, les images obtenues font apparaître des artefacts caractéristiques distincts. Avec la décimation, les coefficients BF sont intacts (au lieu de modifiés légèrement) et les coefficients HF supprimés (au lieu de modifiés plus fortement).

Quand on augmente le facteur de quantification (diminue la qualité), on constate l'apparition de défauts caractéristiques liés au codage (cf. Figure 3.8). Deux types de distorsions caractéristiques se créent dans l'image reconstruite : une déformation du contenu de ces blocs et une discontinuité aux frontières entre deux blocs adjacents : l'effet de bloc. En effet, les blocs sont traités indépendamment les uns des autres et modifiés sans tenir compte de leur voisinage immédiat. Visuellement, l'effet de bloc a un impact visuel important car l'œil est sensible aux motifs de directions verticales et horizontales mais aussi à la périodicité du phénomène de fréquence spatiale proche de la sensibilité maximale de l'œil. Il est d'autant plus gênant que l'activité spatiale des pixels environnants dans l'image d'origine est faible (zones uniformes).



Image codée



Image après requantification

Figure 3.8 : Illustration de l'effet de la requantification (image extraite de la séquence Foreman).

Dans le cas d'une décimation DCT, on supprime les coefficients haute et moyenne fréquence de l'image. Les coefficients basse fréquence (non modifiés) sont donc seuls à représenter le contenu du bloc (cf. Figure 3.9). Hors les coefficients de plus hautes fréquences codent les motifs les plus

fins du bloc, les détails de l'image. Leur perte a pour effet un flou sur l'image, en particulier dans les régions d'activité spatiale modérée à haute, comme les zones fortement texturées et les contours d'objets de la scène.

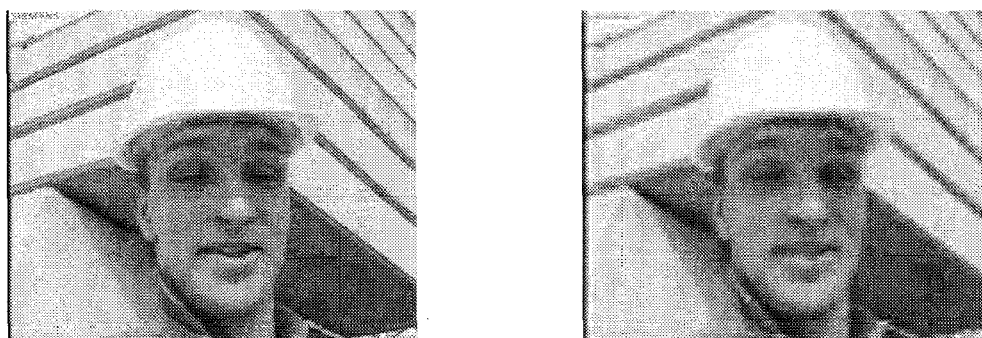


Image codée

Image après décimation DCT 8 :2

Figure 3.9 : Illustration de l'effet d'une décimation DCT.

L'évaluation de la qualité d'une image se fait en général par le calcul du PSNR, mesure objective la plus utilisée, bien que mal corrélée avec les opinions des observateurs [VQE00]. Les propriétés du système visuel humain (SVH) ne sont absolument pas prises en compte, le même poids est accordé aux différences visibles et à celles qui sont masquées par d'autres phénomènes sans distinction. Les artefacts présentés par les images issues des deux méthodes de réduction de débit étant de nature différente, la comparaison de qualité visuelle ne peut être effectuée par cet outil qui ne prend pas en compte le défaut le plus gênant visuellement : l'effet de bloc (très présent dans le cas SNR et peu présent dans le cas spatial).

Dans les zones de haute activité spatiale, l'effet de flou coïncide avec l'effet de bloc. La requantification modifie les coefficients basse et haute fréquence, proportionnellement à leur rang. La décimation, elle, conserve les informations BF et supprime toute information HF. Dans les zones de faible activité spatiale, la décimation DCT n'aura pas d'effet puisque les informations hautes fréquences sont inexistantes. Par contre, la requantification modifiera les informations BF et des effets de bloc apparaîtront.

On observera que l'image basse résolution présente des défauts plus diffus que celle issue d'une quantification sévère. Une image basse résolution apparaît plus « naturelle ». Si elle n'est pas directement comparée à l'originale, son aspect visuel est satisfaisant, alors que l'image requantifiée présente des artefacts de codages très saillants. Visuellement, la version de basse résolution spatiale est le plus souvent meilleure que la version requantifiée ou au moins de qualité acceptable. Nous concluons en ajoutant que, dans le cas d'une image haute définition, la

diminution de résolution spatiale est moins repérable. En effet, le bloc flou ne représente pas la même proportion d'information que dans le cas d'un format standard puisque la surface de l'image a augmenté. Rapporté à une même taille d'écran, les images HD contiennent un plus grand nombre de blocs, chaque bloc occupe donc une surface d'affichage inférieure.

3.1.3 Récapitulatif du fonctionnement du système proposé

Après avoir clarifié les blocs principaux constituant le schéma de protection, nous en détaillons les paramètres et leur influence sur les performances du système.

Paramètres de l'encodeur :

Nous avons choisi de nous baser sur un codeur MPEG2 car il représente aujourd'hui le standard de diffusion broadcast. Mais les résultats peuvent être transposés à d'autres standards de compression vidéo plus récents tels que H. 264 / MPEG4 AVC. L'encodage principal se fait par slices, nous nous limitons ici au cas général où une slice contient une ligne de macroblocs. Le codeur de Wyner-Ziv récupère les slices du flux MPEG2 principal et décode les codes à longueur variable. Les blocs de coefficients DCT quantifiés passent par le masquage fréquentiel. Les vecteurs de mouvement, modes de codage et paramètres de quantifications sont laissés inchangés. Un codage entropique classique de type Huffman est appliqué, on obtient des slices de longueur variable L_s . Le flux basse résolution est affecté par la distorsion, notée D_1 , due à la diminution de résolution. Le codage de Reed-Solomon de paramètres (n,k) est appliqué à travers les slices basse résolution de l'image. Les slices de parité constituant le flux de Wyner-Ziv sont ainsi obtenues.

→ Paramètres de l'encodeur : Longueurs des slices bas débit : L_s ,
Distorsion due à la diminution de résolution spatiale : D_1 ,
Paramètres du codeur de Reed-Solomon : (n, k) .

Mise en paquet et transmission :

On admet que chaque slice issue de l'encodeur (slices image fines et slices de parité issues de la version grossière) est encapsulée dans un paquet distinct et de taille adaptée. Cette mise en paquets de taille variable correspond au cas d'une transmission sur un réseau de transmission par paquet. Chaque paquet passe par les couches protocole RTP, UDP, IP et est affecté des en-têtes correspondants (soit 40 octets au total).

NB : Il est fortement souhaitable que la taille maximale d'une slice soit inférieure à la taille maximale autorisée sur l'ensemble du réseau (Maximum Transmission Unit, MTU). Cette taille correspond celle du tronçon le plus contraignant du chemin emprunté par les données. Dans le cas d'Internet, la limite correspond le plus souvent à la taille des paquets sur une connexion Ethernet soit 1500 octets. D'autre part, la taille du paquet doit être suffisante pour que le surcoût des en-têtes par paquet ne nuise pas à l'efficacité du système. En pratique, la taille de slice après codage varie en fonction du type d'image (I, P ou B) et de données (image ou parité), une taille de paquet fixe ne sera donc pas optimale (cf. Chapitre IV, paragraphe 1.2).

Les paquets sont transmis sur le canal physique. Selon les cas, ils seront éventuellement fragmentés en cellules de 48 octets (pour un réseau ATM), en paquets TS (cas du DVB) ou émis symbole par symbole (modulation multiporteuse sur ligne ADSL) et réassemblés à destination. Les informations de l'en-tête de chaque paquet (cf. Annexe I) permettent à la réception de contrôler l'intégrité du paquet reçu (CRC) et de détecter les paquets perdus (numéro de séquence). La position des paquets perdus est donc connue du décodeur et les paquets reçus sont considérés comme sans erreur.

Paramètres du décodeur :

Les paquets constituant le flux principal sont décodés. Si toutes les slices d'une image ne sont pas disponibles (paquet perdu), la version basse résolution est générée. Chaque slice reçue du flux principal subit alors une décimation DCT et un codage entropique. Le flux basse résolution recréé est identique à celui généré au décodeur à l'exception des slices en erreur. Soit T_s le taux d'erreur sur les slices du flux fin au décodeur, T_s est aussi le taux d'erreur sur les slices bas débit. Les slices image bas débit sont envoyées vers le décodeur de Reed-Solomon qui pourra restaurer autant de slices que le lui permet sa capacité correctrice. La position des paquets perdus étant connue, la capacité du code est $(n - k)$ (décodage à effacement). Les slices basse résolution corrigées sont substituées aux slices perdues du flux principal. Dans le cas contraire, le décodage de Reed-Solomon échoue, l'image est affectée par la distorsion due à la perte de la slice notée par la suite D_{Err} .

→ Paramètres du décodeur : Taux d'erreur slice ou paquet : $T_s = TEP$,
Distorsion due à la perte de l'image: D_{Err} .

Formulation du problème:

Tant que la capacité correctrice est supérieure au nombre de slices perdues, le décodage de Reed-Solomon permet de corriger les slices perdues. Au décodeur, une slice de l'image reconstruite

peut correspondre à une slice du flux principal, une slice de basse résolution spatiale ou une slice perdue. La distorsion d'une image après décodage de Wyner-Ziv est donc fonction du taux d'erreur, du code correcteur et des distorsions citées. Afin d'anticiper les différents effets de ces paramètres sur la qualité de la vidéo reconstruite, nous les analysons dans le paragraphe suivant.

3.2 Modèle analytique

Le décodeur vidéo présenté ci-dessus, lorsqu'une slice image du flux fin n'est pas disponible (paquet perdu), génère la version grossière de l'image. À l'aide des slices / paquets de parité reçues, un décodage de Reed-Solomon est appliqué à la version grossière corrompue de l'image. Si la capacité du code n'est pas dépassée, la version grossière sera restaurée et substituée aux slices perdues. Plusieurs facteurs participent à la distorsion de la vidéo reconstruite :

- La distorsion liée à la diminution de résolution,
- La distorsion liée à la perte des paquets,
- La distorsion liée à la propagation temporelle des erreurs.

3.2.1 Définition du canal de transmission

Le paquet est transmis sur le médium de transmission physique utilisé. Au récepteur, les paquets corrompus sont supprimés (on ne considère pas ici l'usage des techniques de retransmission, mal adaptées aux transmissions temps-réel). Certains paquets n'arrivent pas (ou trop tard) à destination à cause de congestions (filaire) ou d'évanouissements (sans fil) qui sont des phénomènes potentiellement durables. La modélisation du canal doit rendre compte de l'alternance de ces états (bon et mauvais).

Comme ces pertes sont corrélées, elles peuvent être modélisées par un canal de Gilbert ([GIL60], cf. Annexe II), constitué d'une chaîne de Markov à deux états. La Figure 3.10 montre une représentation du canal de Gilbert. Le canal est caractérisé par ses probabilités de transition de l'état « Good » (G) à l'état « Bad » (B) et inversement : P_{GB} et P_{BG} .

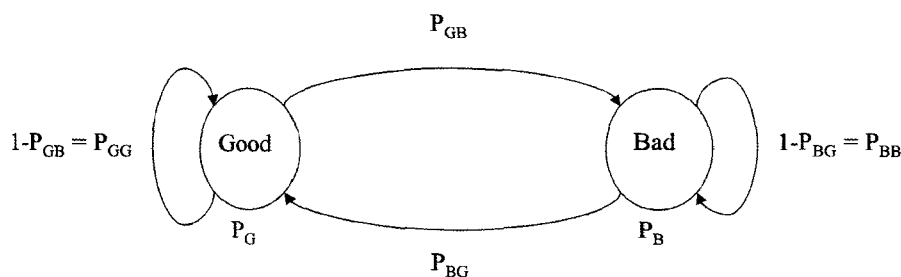


Figure 3.10 : Canal de Gilbert.

On note respectivement P_B et P_G la probabilité d'erreur symbole dans l'état « Bad » et la probabilité d'erreur symbole dans l'état « Good ». Si le canal est dans l'état « G », la transmission est fiable et le symbole reçu sera identique au symbole émis. Si le canal est dans l'état « B » lors de la transmission du symbole, le bruit est prédominant. Les bits du symbole reçu seront donc probablement corrompus et le symbole reçu différent du symbole émis. On considère en pratique que dans l'état G les symboles seront reçus sans erreur ($P_G = 0$) alors que dans l'état B ils seront systématiquement perdus ($P_B = 1$) (cf. Annexe II et [YEE95]). En conséquence, on peut exprimer les caractéristiques du canal sous forme du taux d'erreur symbole moyen P_e et de la longueur moyenne d'une salve d'erreur L_b valant respectivement :

$$P_e = P(B) \text{ et } L_b = \frac{1}{P_{BG}} \tag{3.2}$$

Avec : $P(B) = \frac{P_{GB}}{(P_{BG} + P_{GB})}$, la fraction du temps passée dans l'état « Bad »,

$P(G) = \frac{P_{BG}}{(P_{BG} + P_{GB})}$, la fraction du temps passée dans l'état « Good ».

On peut alors dériver la probabilité $P(m,n)$ de perdre $m - 1$ paquets parmi les $n - 1$ paquets suivant un paquet perdu (cf. Annexe II).

Le codage de Reed-Solomon est appliqué à travers les slices / paquets, le décodeur voit donc le taux d'erreur slice comme son taux d'erreur symbole d'entrée. On note T la capacité correctrice du code, avec $T = n - k$ puisque le décodage est à effacement (cf. Chapitre I, paragraphe 2.2).

On note F l'événement correspondant à l'échec du décodage de Reed-Solomon. La probabilité $P(F)$ que le décodage RS du bloc de n paquets échoue est:

$$P(F) = \sum_{i=T+1}^n P(i, n) \quad (3.3)$$

Le décodage RS n'aura lieu que si un paquet est perdu, événement noté S . On s'intéresse à l'évènement F/S : "le décodage RS échoue sachant que la slice est perdue", ce qui revient à évaluer la probabilité que sur les $n - 1$ slices restantes, au moins T soient perdues :

$$P(F/S) = \sum_{i=T}^{n-1} P(i, n) \quad (3.4)$$

On peut finalement évaluer les probabilités d'affichage des slices [RAM06]¹ comme le décrit la Figure 3.11 :

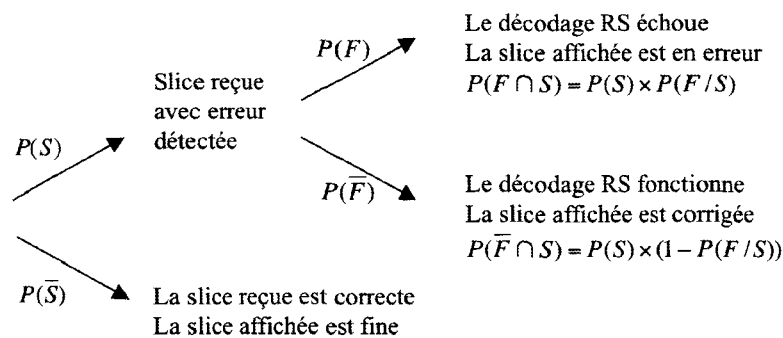


Figure 3.11 : Probabilités d'affichage des slices.

3.2.2 Calcul de distorsion

Pour mesurer l'impact visuel des modifications de l'image, la mesure la plus communément utilisée reste l'Erreur Quadratique Moyenne (EQM), ou le rapport signal sur bruit (Peak Signal to Noise Ratio, PSNR) qui en dépend directement. Dans le schéma étudié, l'effet de bloc n'est pas le défaut prédominant, le PSNR constitue une mesure simple, bien qu'imparfaite, de la qualité obtenue. La valeur du PSNR pour une image de taille $I \times J$ est définie à partir de l'Erreur Quadratique Moyenne (EQM) par :

¹ On notera qu'une approche similaire avait été proposée dans [RAN04b] pour le cas particulier où les erreurs sont indépendantes.

$$PSNR = 10 \times \log_{10}\left(\frac{255 \times 255}{EQM}\right) \quad \text{avec} \quad EQM = \frac{\sum_{i=1}^I \sum_{j=1}^J (x_{ij} - x'_{ij})^2}{I \times J} \quad (3.5)$$

La différence inter pixel ($x_{ij} - x'_{ij}$) est fonction de la valeur du pixel x_{ij} de l'image originale et de la valeur du pixel x'_{ij} de l'image décodée.

Si l'on considère la chaîne de transmission bout à bout, la dégradation du signal vidéo reconstruit est due à deux facteurs :

- Le codage de source destructif,
- Les erreurs / pertes lors de la transmission sur le canal.

Ces deux facteurs étant indépendants, ces deux contributions à la distorsion globale peuvent être calculées séparément. Dans la suite de ce paragraphe, on considère comme image de référence l'image issue du codage source. On s'intéresse à la distorsion (en termes d'EQM) due à la transmission, notée D_{tr} . La qualité d'une image décodée dépend donc de la probabilité que la valeur du pixel soit perdue et de la probabilité que le décodage de canal ne puisse la restaurer.

Dans un premier temps, on néglige la propagation temporelle liée au codage inter images. On note :

- D_1^T , l'EQM de transcoding entre l'image encodée finement et la version ré-encodée grossièrement, pour une image de type $T = \{I, P, B\}$.
- D_{err}^T , l'EQM due à la perte de l'information transmise de l'image de type T. Il s'agit de l'image pour une image I et de l'image différentielle pour une image P ou B,

NB : La valeur de D_{err}^T dépend de la méthode de dissimulation d'erreur utilisée (cf. Chapitre I, paragraphe 4). Exemples :

- Le pixel codé perdu x_{ijn} est remplacé par un pixel de valeur la teinte moyenne globale : une teinte gris moyen (pixels de valeur 128).
- Le pixel codé perdu x_{ijn} est remplacé par le pixel décodé correspondant de l'image précédente déjà reconstruite (previous frame error concealment).

Ces distorsions peuvent s'écrire respectivement :

$$D_{err}^T = EQM(x_{ijn}, 128) \quad \text{et} \quad D_{err}^T = EQM(x_{ijn}, x_{ijconc}) \quad (3.6)$$

Ces valeurs dépendent du contenu des slices, elle seront calculées et moyennées pour chaque structure de GOP testée, pour chaque séquence encodée.

On en déduit la distorsion sur l'image même qui résulte directement des erreurs de transmission pour une image de type $T = \{I, P, B\}$.

Dans le cas WZ, la distorsion moyenne de l'image de type T vaut :

$$D_{tr}^T = P_{wz}^T(\bar{F} \cap S) \times D_1^T + P_{wz}^T(F \cap S) \times D_{err}^T \quad (3.7)$$

Dans le cas particulier du FEC, la version grossière est la version fine ($D_1 = 0$). Si le décodage de canal fonctionne, la distorsion liée à la transmission sera nulle. On peut donc simplifier le calcul de la distorsion de l'image :

$$D_{tr}^T = P_{fec}^T(F \cap S) \times D_{err}^T \quad (3.8)$$

NB : on remarque que le terme $P^T(\bar{F} \cap S)$ n'apparaît pas dans l'expression de la distorsion du cas FEC. En effet, le cas où le décodage RS fonctionne ne génère alors aucune distorsion. Cependant, la probabilité d'échec du décodage RS est beaucoup plus élevée dans le cas FEC que dans le cas WZ car la protection est plus faible :

$$n_{wz} \gg n_{fec} \text{ donc } P_{fec}^T(F \cap S) \gg P_{wz}^T(F \cap S).$$

3.2.3 Prise en compte de la propagation des erreurs

Dans le cas d'un codage intra, toute l'information nécessaire à la reconstruction est transmise dans la même image. Considérons maintenant le cas où l'image est codée en inter : à l'erreur sur l'information transmise (image différentielle) s'ajoute l'erreur éventuelle sur l'image I ou P de référence. Dans ce deuxième cas, il faut prendre en compte la distorsion de l'image de référence pour obtenir la distorsion moyenne de l'image reconstruite. Nous prendrons l'exemple du GOP le plus couramment utilisé (cf. Figure 3.12)

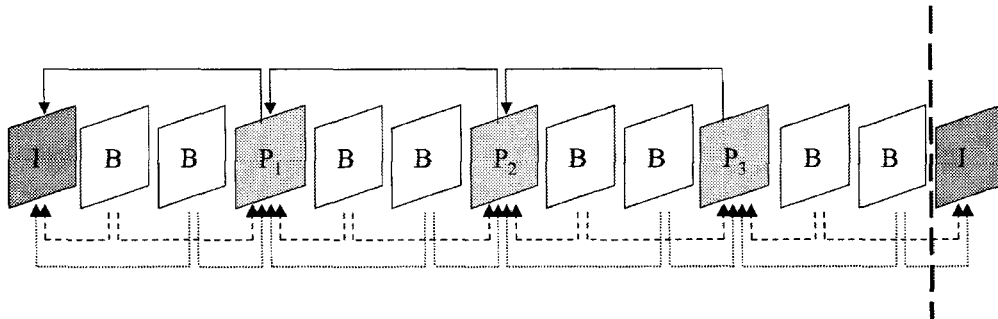


Figure 3.12 : Illustration de la structure du GOP $N = 12, M = 3$.

Une image I est codée sans appel à une référence, son EQM de transmission ne dépend donc que des erreurs affectant ses propres slices. Une image P est codée par prédiction par rapport à une référence qui est soit une image I, soit une image P. Son EQM résulte des erreurs de transmission de l'image différentielle et de celles de son image de référence. Les images B sont codées par interpolation de deux images voisines qui sont de type I ou P. Leur EQM résulte des erreurs de transmission de l'image différentielle et de ses images de référence. De plus, en pratique, les macroblocs des images codées en inter peuvent être codés en intra. Le codeur effectue ainsi un rafraîchissement périodique des slices par insertion de macroblocs codés en intra dans les images codées en inter.

On peut déduire de l'analyse de ces phénomènes un facteur de pondération de l'EQM moyenne calculée pour chaque type d'image qui permet de tenir compte de la propagation des erreurs au sein du GOP (cf. Annexe III). Globalement, la distorsion moyenne due à la transmission d'une image quelconque du GOP est :

$$D_r = \frac{1}{N} \times (nbI \times D_r^I \times w_i + nbP \times D_r^P \times w_p + nbB \times D_r^B \times w_b) \quad (3.9)$$

où :

D_r^T est l'EQM de transmission d'une image de type T (Equation 3.9).

N est le nombre d'image du GOP,

nbI , nbP et nbB sont respectivement le nombre d'images I, P et B du GOP,

w_i , w_p et w_b sont respectivement les poids des erreurs de transmission sur le GOP des images I, P et B. Ces poids dépendent de la structure du GOP et de β_0 et β_1 les taux de blocs non intra dans une image P ou B.

Nous avons dérivé un modèle pour l'estimation de la distorsion de l'image reconstruite au décodeur après transmission et correction grâce au schéma de protection basé sur un codage de Wyner-Ziv. Nous en étudions les applications dans le paragraphe suivant.

3.3 Optimisation du schéma

Le modèle analytique exposé au paragraphe précédent permet d'estimer la distorsion moyenne d'une image reconstruite après correction des erreurs grâce à un flux généré par codage de Wyner-Ziv. Ce modèle évalue la perte de qualité subie en fonction des paramètres du canal, du codage de Wyner-Ziv et des caractéristiques intrinsèques de la séquence. Il utilise donc des informations moyennes calculées à partir du flux MPEG2 principal. Les séquences utilisées appartiennent à des ensembles tests édités par le groupe d'expert VQEG (Video Quality Expert Group). Les valeurs moyennes données dans ce paragraphe sont calculées sur des séquences de

caractéristiques variées qui constituent une référence représentative d'un programme télévisuel multi genre (cf. Annexe IV).

Les valeurs calculées à partir des séquences encodées nécessaires à l'évaluation de la distorsion finale par le modèle analytique sont :

- Les longueurs de slices après codage entropique pour les diverses versions grossières. Elles permettent de déterminer le nombre de slices de parité à adjoindre au flux principal pour atteindre le débit visé.
- Les EQM dans le cas de la réduction de débit, de la perte de la slice, d'une dissimulation d'erreur. Elles permettent de calculer la distorsion moyenne de l'image reconstruite au décodeur en fonction des caractéristiques du canal.
- Le taux de bloc codé en intra dans les images inter (P et B). Il permet de préciser l'influence de la propagation d'erreur sur la distorsion moyenne d'une image reconstruite du GOP.

Grâce à ces valeurs intermédiaires, le modèle analytique évalue les contributions dues aux pertes de slices, substitutions par la version bas débit et dissimulation d'erreur, à la distorsion finale de l'image. Nous allons, dans les paragraphes suivants, détailler les effets des différents paramètres du système sur la distorsion finale.

3.3.1 Détermination du nombre de slices de parité

Nous cherchons à obtenir la qualité vidéo reconstruite la plus haute possible au décodeur. En considérant le flux principal (en termes de débit et distorsion) et la bande passante totale comme fixés, cela revient à minimiser la distorsion de transmission moyenne. Or la distorsion de transmission dépend de la qualité du canal de transmission et de la protection appliquée au flux MPEG2. Le canal est défini par le taux d'erreur paquet (ici considéré comme égal au taux d'erreur slice) et la longueur moyenne d'une salve d'erreur. La protection est fonction de la capacité correctrice du code, donc de la résolution de la description à laquelle il est appliqué. La démarche exposée dans le paragraphe suivant permet de déterminer la description la mieux adaptée à la construction du flux WZ en prenant en compte les conditions de transmission fixées.

Le flux de Wyner-Ziv est caractérisé par le débit de la version basse résolution spatiale et la capacité correctrice du codage de Reed-Solomon. Le débit du flux de Wyner-Ziv est défini à partir du nombre de slices de parité, $(n - k)$, et de leur longueur par:

$$R_{wz} = (n - k) \times \max(Ls) \quad (3.10)$$

Pour chaque description basse résolution, la longueur maximale des slices image est connue. Pour un débit de protection fixé, le nombre maximal de slices de parité que l'on peut adjoindre au flux principal est donc :

$$n = \frac{R_{wz}}{\max(Ls)} + k \quad (3.11)$$

Nous nous limitons aux descriptions obtenues par décimation de rapport 8 :8, 8 :4, 8 :2 et 8 :1 pour des questions de complexité et de compatibilité. En effet, si le nombre de coefficients DCT du bloc est une puissance de deux, le calcul de la DCT inverse permettant la reconstruction d'une image de taille réduite est simplifié [SKO98]. Nous avons mesuré la taille maximale des slices image moyenne de chaque résolution du flux Wyner-Ziv. Le Tableau ci-dessous illustre la réduction de taille des slices de parité pour un flux de protection Wyner-Ziv par rapport à celles d'un flux de protection FEC. La séquence vidéo est au format CIF et encodée à 2Mbps, ce qui correspond à un niveau de qualité satisfaisant.

	réduction de débit
16 coef. DCT	49,55%
4 coef. DCT	73,10%
1 coef. DCT	82,49%

Tableau 3.2 : Réduction de débit moyenne associée à la décimation DCT pour une séquence au format CIF@2Mbps.

La valeur de n sera d'autant plus grande que la longueur de slice maximale sera petite. Comme la protection peut être augmentée proportionnellement à la réduction de débit, la résistance aux erreurs peut être grandement améliorée.

Pour un débit du flux de Wyner-Ziv fixé, à chaque valeur de n correspond une longueur des slices de parité déterminée grâce à l'Equation 3.13. La Figure 3.13 représente les différentes valeurs de n et des longueurs de slices de parité que l'on peut obtenir dans le cas d'un flux principal au format CIF encodé à 2Mbps. Le taux de protection appliqué est de 10% afin d'adapter la protection la plus fine à un taux d'erreur d'au plus 10% tout en permettant à la protection la plus forte de lutter contre des taux d'erreur pouvant régulièrement atteindre 30% dans le cas d'une transmission Internet. On rappelle que pour le format CIF, chaque image contient 18 slices soit $k = 18$.

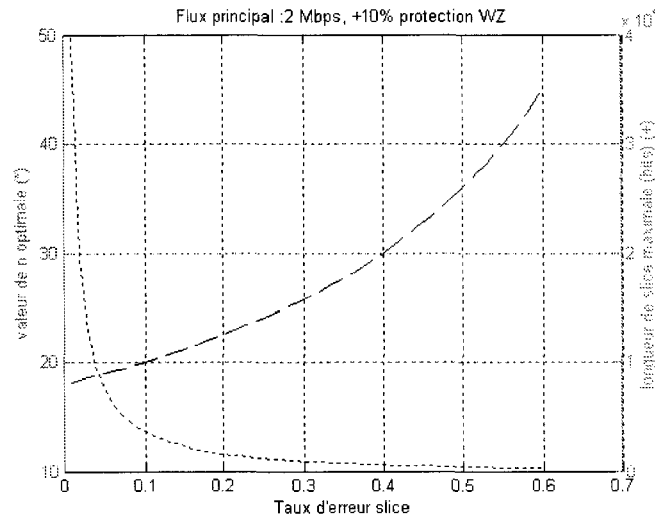


Figure 3.13 : Valeur de n optimale et de la longueur de slice de parité correspondante en fonction du taux de slices en erreur.

La longueur des slices de parité moyenne étant connue, on peut déterminer le nombre de slices de parité que l'on peut adjoindre au flux principal pour chaque description.

3.3.2 Détermination de la description optimale

Nous avons établi, Equations 3.8 et 3.9, l'expression de la distorsion globale de transmission moyenne d'une image du GOP. On admet, pour simplifier, que la probabilité d'erreur sur les paquets des slices I, P et B est la même et que le nombre de slices de parité envoyées avec chaque type d'image est constant. Les probabilités d'affichage des versions fines, basse résolution et en erreur des slices sont alors identiques pour tout type d'image. La distorsion globale peut alors s'écrire :

$$\begin{aligned}
 D_{tr} &= \frac{1}{N} \times \left[\begin{aligned} &(nbI \times D_1^I \times w_i + nbP \times D_1^P \times w_p + nbB \times D_1^B \times w_b) \times P_{wz}(\overline{F} \cap S) \\ &+(nbI \times D_{Err}^I \times w_i + nbP \times D_{Err}^P \times w_p + nbB \times D_{Err}^B \times w_b) \times P_{wz}(F \cap S) \end{aligned} \right] \\
 &= \frac{1}{N} \times \left[\begin{aligned} &A \times P_{wz}(\overline{F} \cap S) \\ &+B \times P_{wz}(F \cap S) \end{aligned} \right] \quad (3.12)
 \end{aligned}$$

Où l'on sait que :

- A est fonction de la distorsion liée à la substitution de la version fine de l'image par la version basse résolution. Dans le cas d'une image prédite, seule la différence de

précision de la dfd de l'image est alors perdue, les vecteurs de mouvement étant identiques.

- B est fonction de la distorsion liée à la perte de la slice avec une éventuelle dissimulation d'erreur. Ces deux cas ont été évalués séparément.

Le Tableau ci-dessous donne, à titre d'exemple, les valeurs moyennes obtenues pour les séquences au format CIF@2Mbps.

	EQM / variance	EQM	PSNR (dB)
Erreur	0,214	370,48	22,44
Dissimulation	0,035	60,20	30,33
8:1 coef. DCT	0,010	17,59	35,68
8:2 coef. DCT	0,005	9,51	38,35
8:4 coef. DCT	0,002	2,99	43,38

Tableau 3.3 : Valeurs moyennes des EQM, EQM normalisées par la variance de la séquence et PSNR pour le format vidéo CIF@2Mbps.

On constate logiquement que, dans l'Equation 3.14, B (distorsion liée à l'échec de la correction) est très supérieur à A (distorsion liée à la substitution par la version basse résolution) quelle que soit la description bas débit utilisée. Minimiser la distorsion globale de transmission revient donc à minimiser $P_{wz}(F \cap S)$, la probabilité d'afficher une slice issue de la dissimulation d'erreur.

Dans le cas où la protection est suffisante ($P_{wz}(F \cap S) = 0$ et $P_{wz}(\bar{F} \cap S) = P(S)$), la distorsion moyenne d'une image peut s'écrire :

$$D_w = \frac{1}{N} \times (nbI \times D_1^I \times w_i + nbP \times D_1^P \times w_p + nbB \times D_1^B \times w_b) \times P(S) \quad (3.13)$$

Le cas optimal est le cas où la protection est nécessaire et suffisante. Le nombre de slices non reçues par image est exactement égal à la capacité correctrice du code de Reed-Solomon, $(n - k)$. Si on néglige la corrélation entre erreurs successives, ce cas correspond à un taux d'erreur slice

$P(S) = \frac{n - k}{n}$. L'expression de la distorsion devient alors:

$$D_w = \frac{1}{N} \times (nbI \times D_1^I \times w_i + nbP \times D_1^P \times w_p + nbB \times D_1^B \times w_b) \times \frac{n - k}{n} \quad (3.14)$$

On peut utiliser ces résultats pour tracer les distorsions issues d'une transmission optimale pour différents taux d'erreur maximaux² dans les cas SLEP (cf. Figure 3.14). La longueur des slices de

² Une démarche d'optimisation similaire donnée récemment dans [RAN06b] est un cas particulier de notre développement où le GOP ne contient pas d'images B, soit $M = 0$ et 100% de bloc codés en inter dans les images P.

parité est déterminée dans notre application par la résolution de la description bas débit choisie. Si le nombre de slices en erreur est inférieur à la capacité correctrice, la qualité de l'image reconstruite sera fonction du nombre de slices substituées qui est le nombre de slices corrompues. Nous rappelons que le cas FEC est un cas particulier du schéma SLEP où la version bas débit est la version fine de la séquence vidéo. Dans le cas du schéma optimal FEC, si le nombre d'erreur est inférieur à $(n - k)$, la qualité reste identique.

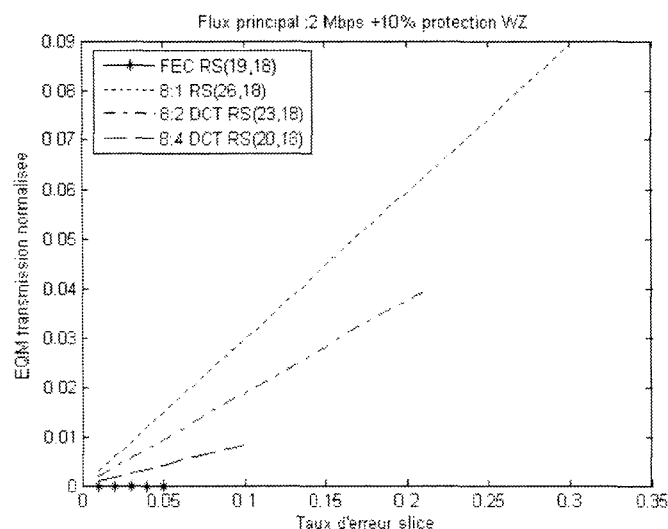


Figure 3.14 : Distorsion optimale de la protection Wyner-Ziv en fonction du nombre de slices en erreur, exemple du flux vidéo au format CIF.

On vérifie que plus la description sur laquelle porte la protection est bas débit, plus le taux d'erreur slice maximal toléré sera élevé. Le nombre moyen de slices de parité ajouté à chaque image ou capacité correctrice est d'autant plus élevé que le débit de l'image est faible. Le nombre moyen de slices de parité est presque doublé pour chaque diminution de résolution spatiale d'un facteur 2. Les schémas WZ de résolutions 8:8 (FEC), 8:4, 8:2 et 8:1 permettent d'obtenir une image reconstruite sans erreur (bien que de résolution spatiale des slices variable) pour des taux d'erreur slice maximaux de 5%, 10%, 21% et 29% respectivement.

Les calculs exposés précédemment sont utilisés pour permettre l'adaptation du modèle analytique aux séquences à analyser et le choix des paramètres de la protection pour chaque description. Dans le paragraphe suivant, nous discutons de l'influence de la résolution flux principal sur la qualité de l'image affichée et la résistance aux erreurs.

3.3.3 Effet de la résolution spatiale de l'image d'origine

Les formats du flux principal (pleine résolution) envisagés sont CIF, SD et HD. Les versions de résolutions spatiales inférieures ont été obtenues par décimation DCT de rapports 8 :4, 8 :2, 8 :1. Si la pleine résolution de départ est un format HD, les tailles d'images obtenues sont alors très proches des versions 16 :9 de formats de diffusions existants. Les tailles d'images réduites accessibles à moindre complexité sont listées dans le tableau ci-dessous :

Rapport de décimation	Taille d'image correspondante		
8 :8	1920 x 1080 (HD)	720 x 486 (NTSC)	352 x 288 (CIF)
8 :4	960 x 540	360 x 243	176 x 144 (QCIF)
8 :2	480 x 270	180 x 121	88 x 72
8 :1	240 x 135	90 x 60	44 x 36

Tableau 3.4 : Récapitulatif des tailles d'image à relier aux résolutions obtenues par décimation DCT.

Le système de transmission proposé dans ce chapitre est basé sur la protection des informations basse résolution du flux MPEG principal. En cas d'erreurs de transmission, le flux de Wyner-Ziv permet de corriger et d'afficher la version basse résolution des zones corrompues de l'image. L'intérêt d'appliquer notre schéma à un flux principal de haute résolution est que, même pour les taux d'erreur les plus élevés, la résolution spatiale minimale protégée (cas de la décimation de rapport 8 :1) est proche du QCIF, donc diffusable.

Dans ce schéma, l'augmentation de la protection fournie par le codage de Reed-Solomon est proportionnelle à la réduction de débit (au niveau slice) réalisée grâce à la décimation DCT. Le Tableau 3.5 présente les réductions de débit obtenues pour les différentes versions basse résolution dans les cas CIF et HD.

	réduction cas CIF	réduction cas HD
16 coef. DCT	49,55%	23,19%
4 coef. DCT	73,10%	46,49%
1 coef. DCT	82,49%	63,56%

Tableau 3.5 : Comparaison des réductions de débit obtenues après décimation pour les formats CIF et HD.

La réduction de débit liée à la décimation des coefficients DCT est nettement moins efficace dans le cas des séquences HD. La réduction de débit dépend du contenu du bloc d'origine, nous avons donc évalué le contenu moyen d'un bloc DCT quantifié pour les différents formats utilisés. Afin de lever toute ambiguïté concernant l'impact du contenu des séquences sur les valeurs calculées, nous avons inclus les résultats des tests obtenus pour des séquences identiques

à des résolutions différentes. Les amplitudes moyennes des coefficients DCT d'un bloc sont reportés dans le Tableau suivant.

Séquences HD (« fairytale »)								Séquences SD (« fairytale »)							
50	2,1	1,2	0,4	0,1	0	0	0	50	2,8	1,9	0,8	0,3	0,1	0	0
3,3	1,2	0,6	0,2	0,1	0	0	0	3,3	1,6	0,9	0,4	0,2	0	0	0
1,6	0,8	0,4	0,1	0	0	0	0	1,8	1,1	0,6	0,2	0,1	0	0	0
0,9	0,4	0,2	0,1	0	0	0	0	1,2	0,6	0,3	0,1	0	0	0	0
0,4	0,2	0	0	0	0	0	0	0,5	0,3	0,1	0	0	0	0	0
0,3	0,1	0	0	0	0	0	0	0,3	0,1	0	0	0	0	0	0
0,1	0	0	0	0	0	0	0	0,1	0	0	0	0	0	0	0
0,1	0	0	0	0	0	0	0	0,1	0	0	0	0	0	0	0
Séquences SD (« 7_12 test »)								Séquences CIF (« 7_12 test »)							
59	3,7	2,4	1,1	0,5	0,2	0,1	0	60	3,7	2,6	1,2	0,6	0,3	0,1	0,1
3,4	1,8	1,1	0,5	0,3	0,1	0	0	3,4	1,9	1,2	0,6	0,4	0,1	0,1	0
1,9	1,2	0,7	0,3	0,1	0	0	0	2	1,3	0,8	0,4	0,2	0,1	0	0
1,3	0,6	0,3	0,1	0	0	0	0	1,5	0,7	0,4	0,2	0,1	0	0	0
0,6	0,3	0,1	0	0	0	0	0	0,8	0,4	0,2	0,1	0	0	0	0
0,2	0,1	0	0	0	0	0	0	0,4	0,2	0,1	0	0	0	0	0
0	0	0	0	0	0	0	0	0,1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tableau 3.6 : Bloc moyen après DCT et quantification pour divers formats et séquences.

On constate que la répartition des coefficients DCT non nuls du bloc varie selon l'ensemble de séquences considéré, ce qui est dû à leur contenu et ne prête pas, dans notre application, à conséquence.

Par contre, le nombre moyen de coefficients DCT non nuls par bloc et leur amplitude moyenne varie en fonction de la résolution. Pour les deux ensembles tests évalués, la version de plus haute résolution contient moins de hautes fréquences que la version de résolution réduite, quantitativement et qualitativement. On peut supposer que ce phénomène est lié au contenu de la scène. Par exemple, le fait d'augmenter la résolution n'apportera pas forcément plus d'information si la scène est peu texturée. L'essentiel des détails existants serait déjà capturé par la résolution spatiale inférieure.

Si le bloc contient moins d'énergie dans les hautes fréquences, leur suppression provoquera une réduction de débit plus faible. La conséquence pour notre système sera une protection moins forte, donc des performances atténuées pour la transmission d'un flux HD par rapport à celles d'un flux SD. Afin d'améliorer les performances du système de transmission, le pourcentage du débit consacré à la protection doit donc être augmenté. La solution classique consistant à réduire le débit d'information et augmenter le débit de la protection peut être utilisée.

3.4 Conclusion

Dans ce chapitre, notre schéma de protection contre les erreurs de transmission basé sur le paradigme de Wyner-Ziv a été présenté. Un flux de données de parité portant sur les informations de basse résolution spatiale est transmis conjointement au flux MPEG principal. Ce flux est construit en appliquant un codage de Wyner-Ziv à une description bas débit obtenue par décimation DCT du flux principal. La qualité de l'image décimée pour une réduction de débit conséquente est satisfaisante et en général visuellement meilleure que dans le cas SNR (absence d'artefacts de codage renforcés). La diminution de résolution par décimation DCT permet en outre au décodeur d'obtenir une version de taille réduite de la séquence à moindre complexité.

Un modèle analytique permettant d'estimer la distorsion de l'image affichée après correction a été exposé. Le modèle utilise les caractéristiques de distorsion et de débit de la séquence qui doivent être calculées en amont et permet d'évaluer la dégradation moyenne de l'image reconstruite après transmission. Pour un taux d'erreur et une longueur de salve d'erreur moyens fixés, les paramètres qui minimisent la distorsion finale du système peuvent être sélectionnés.

Nous avons étudié l'application d'une protection uniforme à toutes les images de la séquence encodée. Or on a vu que la portée d'une erreur est fonction du type d'image qu'elle affecte (cf. paragraphe 3.2.4). Une erreur sur une image I se propagera à toutes les images du GOP alors qu'une erreur sur une image B ne se propagera pas. Surprotéger les images I au détriment des images B paraît donc pertinent. Mais protéger une image B coûte moins cher que de protéger une image I, la taille des slices après compression étant inférieure. Il y a donc un compromis entre distorsion liée au taux d'erreur slice résiduel et distorsion liée à la propagation des erreurs à étudier.

La propagation des erreurs étant prise en compte avec précision dans le modèle, l'effet d'une protection variable des images selon leur mode de codage peut en être déduit. Des tests préliminaires ont montré que l'on peut obtenir une amélioration allant jusqu'à 2,5 dB pour des taux d'erreur élevés en surprotégeant les images I et P, au prix d'une perte de 0,5 à 1 dB pour les taux d'erreur plus faibles. L'utilisation du modèle analytique développé pour optimiser la répartition de la protection RS selon le type d'image constitue une des perspectives de notre travail.

Dans le chapitre suivant, nous mettons en œuvre le modèle analytique et le schéma de protection décrit. Nous nous attacherons à valider les résultats du modèle ainsi qu'à illustrer la qualité de l'image obtenue. En particulier, la résistance aux erreurs sera mise à l'épreuve en appliquant au flux transmis des motifs de pertes relevés lors de transmissions sans fil réelles.

4 Résultats analytiques et expérimentaux

Nous avons proposé dans le chapitre précédent un nouveau schéma de protection des données permettant d'obtenir une image décodée dont la qualité varie graduellement en fonction des caractéristiques du canal. Considérons la transmission d'un flux MPEG2 non protégé sur un canal non fiable. Conjointement à ce flux principal, une version de résolution spatiale réduite du flux obtenue par codage de Wyner-Ziv est transmise. En cas d'erreur de transmission, si le décodage du flux de Wyner-Ziv fonctionne, les slices corrompues du flux principal sont remplacées par les slices correspondantes du flux de Wyner-Ziv. Cette substitution entraîne une légère augmentation de la distorsion dans l'image reconstruite due à la diminution de résolution spatiale. Dans le cas où la version de Wyner-Ziv ne peut être décodée, on considère que la slice en erreur est dissimulée.

Nous allons dans ce dernier chapitre évaluer les performances du schéma proposé en termes de résistance / sensibilité aux erreurs. Dans les réseaux de transmissions de paquets, les pertes se produisent le plus souvent par salves, elles peuvent être modélisées et simulées à l'aide d'une chaîne de Markov à deux états (bon / mauvais). La corrélation entre les résultats de simulation et ceux du modèle analytique (cf. Chapitre III, paragraphe 2) sera étudiée. Une confrontation avec des conditions réelles de transmission sera ensuite réalisée afin de confirmer la validité du système proposé. Pour cela, nous appliquerons des motifs d'erreurs enregistrés lors de transmissions réelles sur une liaison du type IEEE 802.11 (WiFi).

Nous comparerons également les résultats obtenus avec ceux d'un codage FEC classique. La protection par FEC est la plus efficace quand le taux d'erreur du canal est connu et fixe. Mais dans le cas contraire, ses performances se dégradent rapidement (cliff effect, cf. Chapitre I, paragraphe 2.2). Or, si on considère la transmission d'informations sur un réseau aux caractéristiques hétérogènes, le taux d'erreur variera en fonction des conditions de transmission. Le taux d'erreur que subira le flux reçu est donc particulièrement imprévisible, rendant l'usage du FEC problématique.

4.1 Description des expériences proposées

Nous présentons dans ce paragraphe le scénario de transmission auquel le schéma de protection exposé au chapitre précédent est appliqué. Clairement, ce schéma s'adresse à des transmissions vidéo par paquets puisqu'il suppose que les slices en erreur sont effacées et que leur position est connue du décodeur. En pratique, les données sont réparties dans des paquets avant émission et

un contrôle d'intégrité est réalisé sur les données grâce à l'en-tête des paquets à la réception. L'étape de mise en paquet des données doit donc être détaillée. Finalement, les paramètres du système doivent être adaptés au type d'erreur rencontré donc au canal de transmission utilisé. Les motifs de pertes mesurés lors de transmissions WiFi seront présentés et serviront de base pour la détermination des paramètres utilisés pour les simulations et le modèle analytique.

4.1.1 Contexte général

Considérons le cas d'une transmission vidéo temps réel sur un réseau à commutation de paquets comme Internet. On peut distinguer principalement deux scénarios d'application potentiels:

- Le cas d'une transmission unicast d'une vidéo éventuellement pré-encodée disponible en diverses qualités stockées sur le serveur et transmise avec la protection qui lui est associée. C'est le cas par exemple des applications de vidéo à la demande. Le serveur peut alors estimer la qualité d'image et la protection à envoyer en fonction des caractéristiques du récepteur et du canal lors de l'établissement de la connexion.
- Le cas d'une transmission multicast (diffusion) qui peut être considéré comme une collection de transmissions unicast. L'état de la liaison (débit et qualité) est d'autant plus incertain que le nombre de destinataires est élevé. De plus, le serveur ne peut pas adapter le format de la vidéo émise à chaque format d'affichage des multiples récepteurs. Afin de ne pas surcharger inutilement le réseau, le serveur enverra de préférence la résolution la plus élevée. La réduction de taille d'image est ainsi laissée au terminal ou à un transcodeur situé dans un nœud intermédiaire (passerelle d'entrée d'un LAN).

Le schéma de protection systématique que nous avons développé peut être utilisé avantageusement avec ces deux modes de transmission. Notre système permet de découpler le débit alloué à la protection et la capacité correctrice. Le serveur peut ainsi adjoindre au flux vidéo transmis un flux de protection de taille constante, la capacité correctrice variera en fonction de la résolution spatiale protégée. Augmenter la résistance aux erreurs n'implique donc pas d'augmenter le volume global de parité transmis, le flux vidéo principal n'a donc pas à être modifié. L'optimisation de la chaîne de transmission s'en voit simplifiée.

De plus, les flux de protection sont des descriptions d'un flux vidéo générées par codage de Wyner-Ziv. Afin d'être décodées, les descriptions de Wyner-Ziv nécessitent une information adjacente ici constituée par la version corrompue du flux vidéo d'origine. Or on peut considérer qu'une version de résolution spatiale réduite d'un flux vidéo est incluse dans la version de résolution supérieure. Une description basse résolution peut donc être obtenue à partir de ce dernier à l'encodeur comme au décodeur, aux erreurs de transmission près. Tout flux de

protection obtenu par codage de Wyner-Ziv peut donc être adjoint à tout flux vidéo de résolution supérieure. L'architecture du serveur est simplifiée, le nombre de flux (vidéo et protection) à stocker diminue grâce à la compatibilité entre le flux vidéo et le flux de protection de résolutions spatiales différentes.

Pour un débit de la protection fixé, on peut donc adjoindre au flux vidéo principal différents flux de protection de capacité correctrice variable. Le choix du taux de protection à appliquer au flux vidéo à transmettre dépend de la qualité de la liaison. Or les liaisons physiques formant le trajet des paquets peuvent être filaires et /ou sans fil et évoluer dans le temps. Lors d'une transmission sur un réseau filaire, les erreurs seront principalement dues aux congestions et délais de transmission excessifs. Lors d'une transmission sans fil, les erreurs sont essentiellement dues aux bruits et évanouissements. Pour ces raisons, dans les deux cas, les erreurs se produisent par salves et le taux d'erreur est très variable [SKO01]. Pour le réseau Internet, des tests de transmission pour diverses tailles de paquet et à divers débit ont donné des valeurs du taux d'erreur de 0% à 60% et des longueurs de salve d'erreur de 2 à 20 paquets pour les débits les plus élevés [HOR99]. On constate néanmoins que les longueurs de salve les plus courtes, un paquet isolé ou 2 paquets consécutifs, sont beaucoup plus fréquentes [PER02].

Afin d'évaluer les performances d'un schéma de transmission, plusieurs approches sont possibles :

- La simulation de la transmission permet de tester le comportement du système à partir des paramètres significatifs de la transmission, déterminés par des mesures préalables. Un modèle adéquat permet ensuite de reproduire les phénomènes identifiés et mesurés. Dans notre cas, on peut modéliser les salves d'erreur par un modèle de Gilbert-Elliott (cf. Chapitre III paragraphe 2.1). Deux paramètres sont alors à déterminer pour caractériser le canal à simuler : le taux d'erreur paquet moyen et la longueur moyenne d'une salve d'erreur.
- L'usage d'un modèle de canal au lieu d'un canal de transmission réel introduit cependant une approximation. Les résultats obtenus par les simulations doivent idéalement être validés par comparaison avec des cas de transmission réelle. Cependant, la mise en place d'un matériel de transmission et de mesure dédié n'est pas toujours possible. L'enregistrement des motifs de pertes issus d'une transmission réelle constitue alors une bonne alternative. Les expériences peuvent ainsi être effectuées hors connexion à condition que toutes les contraintes de la transmission d'origine soient scrupuleusement respectées.

Dans le cas d'une transmission vidéo, l'importance d'un élément du flux transmis dépend des caractéristiques de la séquence (mouvement, texture) et du mode de codage (inter ou intra). Les

conséquences d'un paquet perdu sur la qualité d'image décodée dépendent également des autres pertes subies au sein du même bloc issu du codeur de canal (les n slices du codeur RS). L'état du canal et les paramètres du codeur de canal sont significatifs. Ces phénomènes sont pris en compte dans le modèle analytique développé au chapitre précédent. Pour obtenir des résultats moyens représentatifs, le nombre d'échantillons utilisés doit être suffisant pour que les phénomènes attendus se produisent. Habituellement, un grand nombre d'expériences est réalisé puis moyenné, mais une alternative est d'utiliser un flux de données très long.

Nous présentons dans les paragraphes suivants les détails significatifs concernant les expériences utilisées pour tester notre système de transmission. En particulier, la répartition des données en paquets constitue une étape clef de l'adaptation au réseau et ses contraintes seront explicitées. Finalement, le scénario de transmission réel utilisé est décrit.

4.1.2 Mise en paquets

La mise en paquets doit être conçue en tenant, si possible, compte de la structure des données à transmettre [CLA90]. Chaque paquet doit contenir des données indépendamment décodables de sorte à ce que l'effet d'un paquet perdu n'excède pas le contenu de ce paquet. Pour un flux MPEG2, les en-têtes doivent se situer en début de paquet afin de faciliter la resynchronisation du décodeur en cas d'erreurs. Il est conseillé de répéter les en-têtes de séquence, de GOP voire d'image périodiquement ou de mettre en place leur retransmission. Nous considérons par la suite que la répétition de ces en-têtes est suffisante pour assurer leur bonne réception et ne considérons que les cas de pertes de slices.

Dans le cas du standard MPEG2, une slice est définie par un nombre fixe de macroblocs. Sa taille après codage entropique est variable selon le contenu de l'image mais surtout selon le mode de codage de l'image. Conformément au schéma de protection utilisé, les slices de parité ont la taille de la slice image la plus longue de la version bas débit servant à les générer. Le Tableau 4.1 donne la taille moyenne des slices de parité, qui correspond à la moyenne de la longueur maximale des slices de la version bas débit utilisée pour générer le flux WZ, et la taille moyenne des slices du flux MPEG2 principal.

	slice I	slice P	slice B
flux MPEG	1213,01	738,59	417,38
parité 8:8 (FEC)	1473,63	965,99	580,05
parité 8:4	907,93	428,13	270,66
parité 8:2	437,59	184,63	147,80
parité 8:1	194,85	98,99	105,99

Tableau 4.1 : Longueurs de slice moyennes (en octets) du flux principal et de la parité selon le mode de codage de l'image et la version de résolution spatiale réduite utilisée.

Idéalement, chaque paquet doit contenir exactement une slice. Mais leur taille est très variable et, bien que la taille d'un paquet RTP soit théoriquement variable, il est fréquent qu'elle soit déterminée par le matériel utilisé au niveau de la couche de liaison. Dans le cas où une taille de paquet fixe est imposée, les slices dont la taille excède celle du paquet seront fragmentées. Au récepteur, lors du réassemblage des paquets formant une slice, toute slice corrompue sera intégralement effacée. En conséquence, le taux d'erreur slice sera le plus souvent supérieur au taux d'erreur paquet. Si, en revanche, sa taille est inférieure à celle du paquet, le reste du paquet sera rempli de zéros, ce qui constitue une mauvaise utilisation des ressources. De plus, le surcoût de débit lié à l'ajout des en-têtes de paquets rend favorables les plus grandes tailles de paquet. Il faut donc trouver un compromis entre résistance aux erreurs et efficacité de la transmission [BAS99].

On s'assurera qu'un paquet contient au plus une slice afin que la perte d'un paquet ne cause pas la perte de plusieurs slices. En particulier les slices de parité étant naturellement plus courtes que les slices images du flux principal, permettre la transmission de plusieurs slices dans un même paquet nuirait à la résistance du schéma. Il est donc raisonnable d'utiliser une taille de paquet de l'ordre de la taille d'une slice de parité, afin d'utiliser efficacement la bande passante allouée. On préférera se baser sur la taille d'une slice de parité B, qui est la plus courte et, pour notre structure de GOP, la plus fréquente (100 à 300 octets selon le flux de protection utilisé). La figure suivante illustre la mise en paquet de 300 octets des slices images et parité (diminution de résolution de rapport 8:2) pour un GOP de 12 images de structure IBBP...

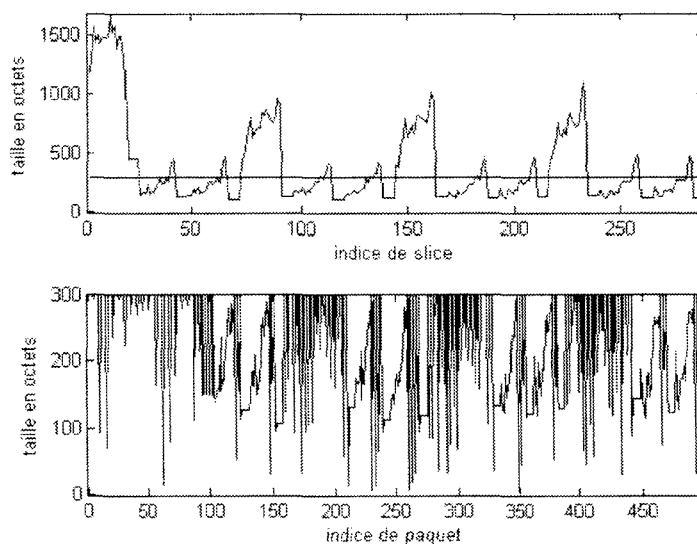


Figure 4.1 : Illustration de la mise en paquets de taille fixe (300 octets) des slices image et parité pour un GOP de la séquence encodée.

On notera que considérer une taille de paquet fixe et remplir la fin du paquet de zéros le cas échéant revient à majorer la probabilité d'erreur paquet puisqu'elle dépend de la longueur du paquet. De plus, bien que ce remplissage par des zéros nuise à l'efficacité de la transmission, mettre au plus une slice par paquet transmis améliore le brassage des données pour un débit de la source constant. Le nombre de paquets transmis par image est plus élevé, mais la corrélation des erreurs portant sur les slices de deux images successives diminue (cf. Annexe II).

La fragmentation des slices en paquets peut être évitée si l'encodeur autorise des slices d'un nombre de macroblocs variable [BAS99] mais cette fonctionnalité n'est généralement pas implémentée dans les codecs MPEG2 actuels. La taille de paquet est le plus souvent déterminée par le matériel utilisé lors des expériences. Nous présentons dans le paragraphe suivant les mesures qui seront utilisées pour la validation expérimentale de notre système de protection et les contraintes associées.

4.1.3 Cas d'une transmission réelle

Dans le cas d'une transmission sans fil, les congestions, les évanouissements et interférences vont influencer sur le débit de la liaison. Si les conditions de propagation sont mauvaises, le taux d'erreur paquet sera élevé. La durée de vie de ces phénomènes fait que les pertes de paquets se produisent essentiellement par salves.

Le projet IST-WCAM (Wireless Cameras and Audio-visual seaMless networking) est un projet de recherche et développement initié par la société de l'information de la commission européenne qui a eu pour objectif l'étude, le développement et la validation d'un système de transmission audiovisuel sans fil sécurisé cohérent [WCA04]. Afin de mieux comprendre les interactions entre les différents composants constituant une transmission sans fil, un ensemble de mesures a été effectué lors du festival d'Annecy 2005. Des transmissions de données à des débits, tailles de paquets et conditions de propagation différents ont été mises en places. Les motifs de pertes de paquets relevés ont été enregistrés de sorte à pouvoir en étudier les caractéristiques ultérieurement.

La plateforme de test est composée d'une interface client / serveur fonctionnant sous Windows XP et de cartes 802.11g (cf. Figure 4.2), les protocoles de transport TCP/IP et UDP/IP ont été implémentés. Dans la mesure où TCP est basé sur la retransmission des paquets perdus, il induit un délai de transmission problématique pour la communication de données temps réel. UDP est mieux adapté à la transmission d'un flux vidéo, mais ne fournit aucune garantie sur la délivrance des paquets. Il est donc nécessaire d'utiliser des techniques de contrôle d'erreur.

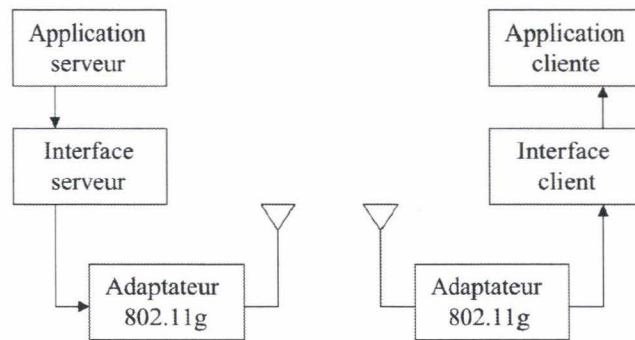


Figure 4.2 : Plateforme de test utilisée par le projet WCAM [CHI05].

Grâce aux informations transmises via des en-têtes propriétaires, la localisation des paquets perdus peut être enregistrée au niveau de l'application cliente [CHI05]. Les motifs de perte de paquets ainsi obtenus peuvent être utilisés hors-ligne pour simuler la transmission de diverses séquences dans des conditions identiques et évaluer les performances de schémas de codage robustes aux erreurs et / ou de dissimulation d'erreur.

Lors des expériences, les données issues du codeur vidéo étaient transmises à un débit de 500 ou 2000 kbps, dans des paquets de différentes tailles. Les scénarios de transmission incluent les situations de mobilité / staticité, en intérieur / extérieur pour différents trajets du récepteur.

Les données que nous avons transmises sont issues d'un codeur MPEG2. Le débit du flux vidéo compressé auquel nous avons appliqué les motifs de perte est de 2Mbps. Un débit de 500 Kbps ne permet pas d'obtenir une qualité d'image suffisante pour le format CIF considéré (352 x 288 pixels). D'autre part, une taille d'image inférieure du flux principal ne serait pas intéressante dans le cadre de la scalabilité spatiale. Afin de limiter l'approximation réalisée sur le taux d'erreur slice après transmission et d'utiliser efficacement la bande passante disponible, il est préférable d'utiliser la taille de paquet la plus petite, ici 300 octets (cf. paragraphe précédent). Une taille de 1200 octets permettrait néanmoins d'améliorer les performances du système en termes de résistance aux erreurs (peu de slices fragmentées) au prix d'une efficacité réduite de transmission (remplissage des paquets contenant les slices les plus courtes par des zéros).

Les caractéristiques des motifs de pertes utilisés sont reportées dans le tableau 4.2. On notera qu'une salve d'erreur d'ordre k débute et se termine par un paquet perdu, le nombre de paquets reçus consécutifs dans l'intervalle étant au plus de $k-1$ [FER05]. La longueur de salve d'erreur donnée dans le tableau est d'ordre un, il s'agit de la longueur moyenne d'une suite ininterrompue de paquets en erreur. Il faut prendre en compte que ces salves peuvent n'être séparées que par un faible nombre de paquets reçus, le nombre de paquets en erreur pour une image peut donc être beaucoup plus élevé que cette valeur moyenne.



Nom	Lieu	Taille de paquet (octets)	description	Taux d'erreur paquet	Longueur moyenne d'une salve
Data 7	BonLieu	300 et 1200	Intérieur : trajet forum, sortie par la porte principale et retour – point d'accès au rez de chaussé	$1,5 \cdot 10^{-3}$	1,8
Data 8	BonLieu	300 et 1200	Intérieur, trajet forum, sortie par la porte de derrière et retour – rez de chaussé	$1,22 \cdot 10^{-2}$	1,34
Data 9	BonLieu	300 et 1200	Intérieur : trajet escaliers jusqu'au hall d'exposition, sortie de gauche, retour puis sortie par la porte principale, retour, sortie de droite, retour, escaliers – point d'accès au premier étage	$1,6 \cdot 10^{-3}$	1,69
Data 10	BonLieu	300 et 1200	Intérieur : statique – point d'accès au premier étage, station de base au deuxième	$3 \cdot 10^{-4}$	1,2
Data 11	BonLieu	300 et 1200	Intérieur : statique – point d'accès premier étage, station de base au café « le bistrot »	$9,4 \cdot 10^{-3}$	1,1

Tableau 4.2 : Caractéristiques des motifs d'erreurs de transmission relevés lors de la campagne de mesure WCAM.

Globalement, le taux d'erreur paquet relevé n'excède pas 2% mais les slices image pouvant correspondre à plusieurs paquets, le taux d'erreur slice sera supérieur. La longueur de salve moyenne n'excède pas 1,8 paquets, mais le nombre moyen d'erreurs slice consécutives (ordre 1) risque d'être inférieur à cause du regroupement des paquets. Pour l'application du modèle analytique, des taux d'erreur slice moyens allant de 1% à 20% et une longueur de salve de 1,2 slices ont été utilisés comme paramètres du canal de Gilbert. Les résultats sont donnés et interprétés dans le paragraphe suivant. La fin du chapitre concerne les résultats obtenus en appliquant des motifs de pertes issus de mesures préalables. La corrélation entre les résultats analytiques et expérimentaux est également étudiée.

4.2 Résultats du modèle analytique

Le modèle analytique nous permet d'obtenir une estimation de la distorsion moyenne finale de l'image à partir des paramètres suivants:

- la qualité du canal de transmission,
- le volume global alloué à la protection du flux principal,

- la capacité correctrice du codage de Reed-Solomon mis en oeuvre,
- la résolution spatiale garantie par la protection.

Les expériences ont été menées sur un ensemble de séquences vidéo test de caractéristiques variées éditées par le groupe d'expert VQEG (Video Quality Expert Group, [VQEG06]). Elles sont au format CIF et encodées à 2Mbps en progressif. Un format d'image entrelacé peut également être utilisé à condition d'adapter les paramètres du schéma de protection à la taille des trames. Les images sont partitionnées en slices de 22 macroblocs (une ligne de macroblocs). Nous comparons les performances du FEC avec celles de notre schéma SLEP utilisant trois descriptions de résolutions spatiales réduites différentes pour un débit total constant : 8:1 (seul coefficient DC est conservé), 8:2 (les 2x2 premiers coefficients DCT) et 8:4 (les 4x4 premiers coefficients DCT). On notera que, dans les schémas traditionnels de transmission par paquets, la protection est réalisée par l'application d'un FEC à travers les slices / paquets d'un flux vidéo. Ce procédé constitue un cas particulier du schéma SLEP où la description WZ est de résolution identique à celle du flux principal. Les performances seront évaluées en termes de résistance aux erreurs de transmission pour différents taux d'erreur slice. Les effets sur la qualité finale de la vidéo affichée de la longueur d'une salve d'erreur et d'une augmentation du débit consacré à la protection sont ensuite abordés.

4.2.1 Analyse de la résistance aux erreurs du schéma

Nous utilisons l'EQM comme mesure de distorsion car c'est actuellement le critère le plus utilisé bien qu'il traduise mal la qualité de l'image perçue par l'observateur dans le cas d'une diminution de résolution spatiale. L'EQM de transmission de chaque image est donc calculée puis moyennée pour l'ensemble de la séquence décodée. L'EQM affichée dans les figures est normalisée par la variance de l'image. La Figure 4.3 présente les résultats du modèle analytique obtenus pour les séquences au format CIF encodé à 2Mbps et un taux de protection de 10% du débit total. Pour les calculs, une salve d'erreur moyenne de 1,2 paquets et des taux d'erreur paquet moyens allant jusqu'à 30% ont été utilisés.

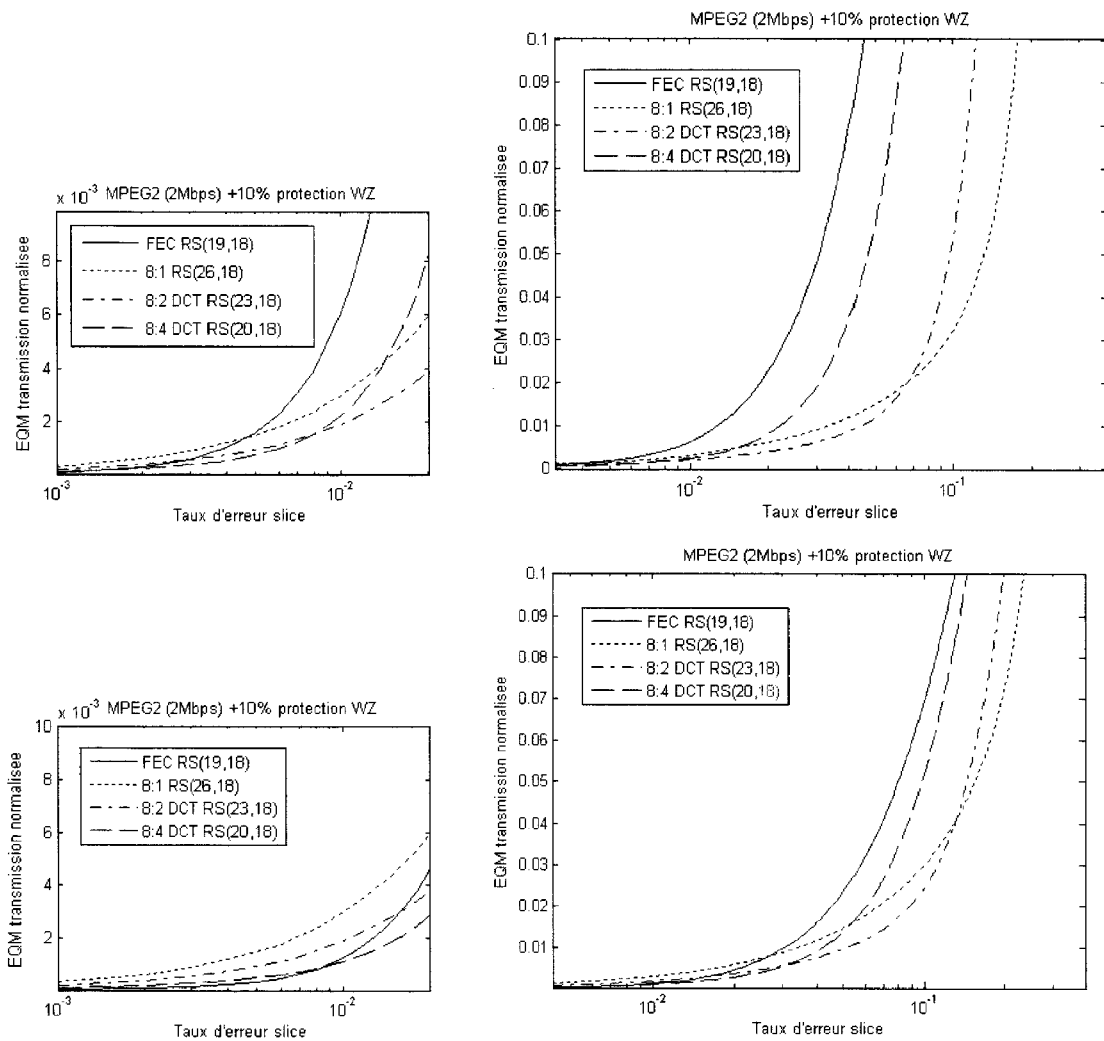


Figure 4.3 : Résultats du schéma SLEP appliqué à des descriptions de résolutions spatiales variables donnés sous forme d'EQM de transmission moyenne, sans (en haut) et avec dissimulation d'erreur (en bas).

Le taux de slices affichées en erreur / dissimulées est également fourni afin de mettre en évidence les performances du schéma concernant la résistance aux erreurs (cf. Figure 4.4). Il permet de distinguer la part de la distorsion due aux pertes de slice de celle due aux substitutions de slice et de mieux comprendre les variations numériques de l'EQM.

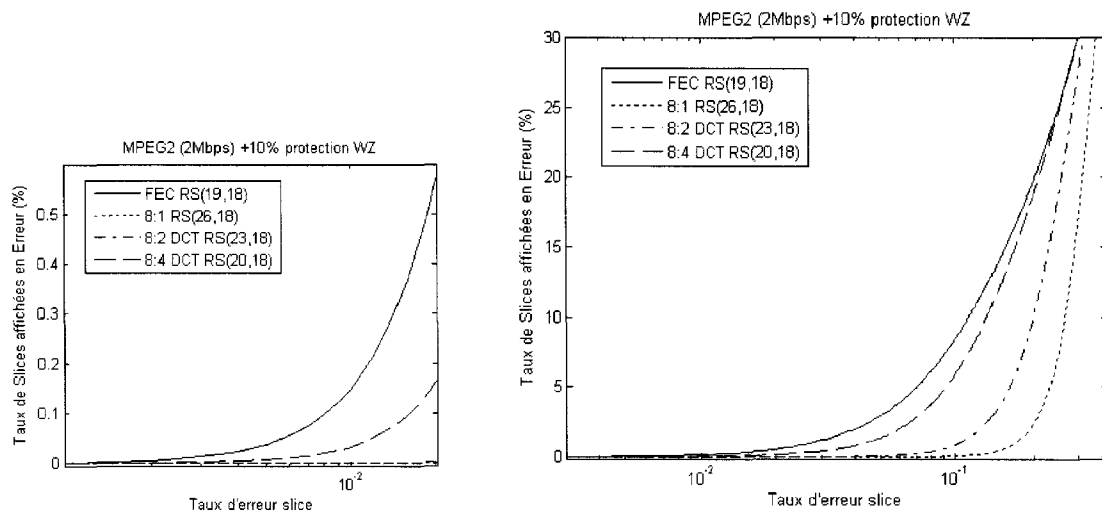


Figure 4.4 : Résultats du schéma SLEP appliqué à des descriptions de résolutions spatiales variables donnés sous forme de taux de slices affichées en erreur moyen.

Dans le cas général, on observe, comme prévu, que plus le taux d'erreur slice (ou paquet) est élevé, plus la description permettant d'obtenir la distorsion minimale est bas débit. L'intersection entre deux courbes correspond au taux d'erreur pour lequel l'échec de la protection de la version la plus fine rend le gain de qualité (lié à une plus haute résolution spatiale) insuffisant par rapport au préjudice qu'elle cause en termes de protection. Pour chaque description servant à générer la parité, on peut délimiter deux zones :

- La première est celle où la capacité correctrice est suffisante et où la correction des erreurs fonctionne. La distorsion est alors liée à la variation de qualité entre la version fine et la version utilisée pour générer la parité. Elle augmente lentement avec la probabilité que les slices soient perdues mais corrigées.
- La seconde est celle où la capacité correctrice du code est dépassée, si bien que la correction des erreurs échoue. La distorsion est alors liée à la perte du contenu des slices en erreur. Elle augmente rapidement avec la probabilité que les slices soient perdues et non corrigées et peut être fortement réduite grâce aux techniques de dissimulation d'erreur.

La protection apportée par le FEC (décimation de rapport 8:8) permet de reconstruire la version fine des slices. Mais la correction ne fonctionne que pour les taux d'erreur les plus faibles (inférieurs à 5%). Si on diminue la résolution spatiale de la description servant de base au codage de Wyner-Ziv, la résistance aux erreurs augmente. En effet, une description de résolution spatiale plus faible sera de débit plus bas et la longueur maximale d'une slice après encodage

entropique est donc plus courte. On peut transmettre un plus grand nombre de slices de parité pour un débit de protection égal. La description la plus grossière (décimation de rapport 8 :1) permet d'accéder à la meilleure résistance aux erreurs, mais la qualité de l'image reconstruite pour des taux d'erreur faibles diminue. Cette solution est donc adaptée aux pires, ou aux plus imprévisibles, conditions de transmission.

Dans le Chapitre III paragraphe 3.2, nous avons évalué les performances des différents schémas dans le cas d'erreurs de transmission non corrélés et d'une protection optimale (juste suffisante). Les taux de perte de slice maximaux permettant une reconstruction sans erreur obtenus dans ce cas étaient respectivement de 5%, 10%, 21% et 29% pour les schémas FEC, 8 :4, 8 :2, 8 :1 (cf. Figure 3.14). L'application d'erreurs corrélés (longueur moyenne d'une salve d'erreur de 1,2 paquets) n'a pas permis d'obtenir de tels résultats. L'effet de la corrélation des erreurs sur les performances des schémas est détaillé dans le paragraphe suivant.

4.2.2 Effet de la longueur des salves d'erreur

Il est bien connu que les salves d'erreur sont plus difficiles à corriger que les erreurs isolées. Le modèle de canal de Gilbert-Elliott utilisé (cf. Annexe II) permet de simuler des erreurs paquets corrélés en spécifiant la longueur moyenne d'une salve d'erreur (L_b paquets) en plus du taux d'erreur moyen (P_e). Il est ainsi possible d'estimer la probabilité notée $P(m,n)$ que m paquets soient perdus sur n transmis.

Lorsque les erreurs sont corrélés, on impose que les erreurs se produisent par salves de longueur moyenne L_b . Analytiquement si on augmente la longueur moyenne d'une salve d'erreur (L_b), quand le canal passe dans l'état « Bad », la probabilité qu'il y reste (P_{BB}) augmente comme le montre l'équation suivante (cf. Annexe II) :

$$P_{BB} = 1 - P_{BG} = 1 - \frac{1}{L_b} \quad (4.1)$$

Pour un taux d'erreur moyen constant, si le nombre moyen d'erreurs par salve augmente, le nombre de salves d'erreur moyen par bloc de n paquets (la fréquence des salves d'erreur) diminue en conséquence.

La figure 4.5 illustre les variations de probabilité d'erreur paquet pour les valeurs $L_b = 2$ et $L_b = 5$ paquets et un taux d'erreur moyen de 30%. Le cas d'erreurs indépendantes, correspondant au cas d'un canal symétrique lors d'une transmission binaire (Binary Symetric channel, BSC), est inclus dans le modèle comme un cas particulier où :

$$L_b = \frac{1}{1 - P_e} \quad (4.2)$$

Dans le cas présent, il correspond à une longueur de salve d'erreur moyenne :

$$L_b = \frac{1}{0,7} \approx 1,43, \text{ pour un taux d'erreur moyen de } 30\% \text{ et } 1,11 \text{ pour } 10\%.$$

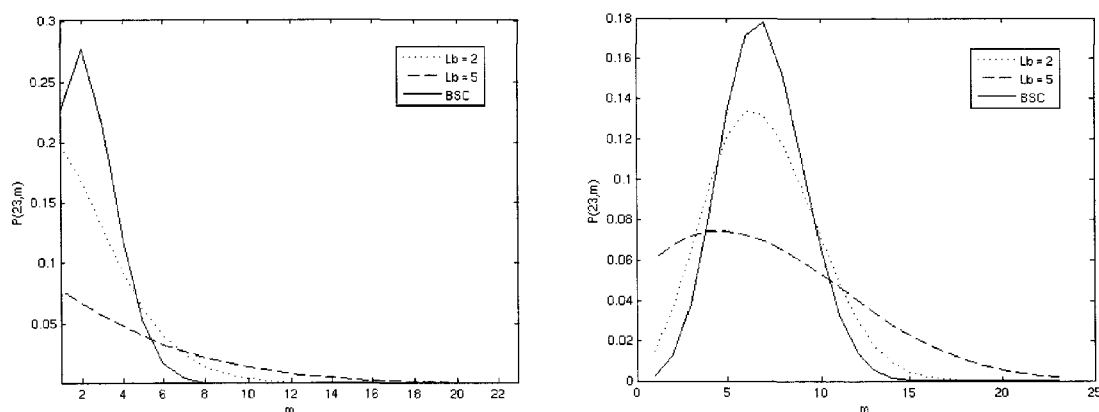


Figure 4.5 : Variations de $P(m,23)$ pour diverses longueurs de salve d'erreur et un taux d'erreur paquets moyen de 10 et de 30%.

On constate bien que pour un taux d'erreur de 30%, si les erreurs sont indépendantes, le cas le plus probable est le cas où 7 paquets sont reçus en erreur sur les 23 paquets transmis ($0,3 \times 23 = 6,9$).

Si la longueur moyenne d'une salve d'erreur est de 2 paquets, la probabilité que 6 ou 8 paquets soient perdus (donc que 3 ou 4 salves de 2 paquets se produisent) augmente au détriment de la probabilité d'obtenir 7 paquets en erreur. En effet, la probabilité d'observer une (des) erreur(s) isolée(s) parmi les erreurs affectant le bloc de 23 paquets est plus faible que celle d'observer des salves de deux erreurs consécutives.

Si la longueur moyenne d'une salve d'erreur est de 5 paquets, la probabilité que les paquets en erreur soient isolés est encore plus faible. Le cas le plus probable est celui d'obtenir exactement une salve de 5 paquets en erreur. La probabilité d'obtenir plusieurs (2 ou 3) salves de 5 erreurs par bloc (donc jusqu'à 15 paquets en erreur sur 23) augmente par rapport aux cas d'erreurs moins corrélées. Les salves sont plus longues donc leur fréquence d'apparition moyenne diminue, ce qui explique les variations plus douces de la courbe.

Concernant le modèle de Gilbert-Elliott (cf. Annexe II), à taux d'erreur moyen égal, l'augmentation de la longueur d'une salve d'erreur se traduit par une diminution de la probabilité de passer de l'état bon à l'état mauvais et de l'état mauvais à l'état bon. Les deux états du canal

auront donc une durée moyenne plus longue, les transitions étant plus rares. La densité de probabilité du nombre d'erreurs s'uniformise et le nombre d'erreurs le plus probable augmente.

Or le schéma développé permet d'augmenter la capacité correctrice, au prix d'une légère distorsion sur les slices corrigées. La répartition des erreurs au sein du flux est donc lourde de conséquences. Nous cherchons à évaluer l'effet d'une augmentation de la longueur moyenne d'une salve d'erreur sur l'efficacité de la correction du système proposé. Le nombre d'erreurs par bloc de paquets transmis a une incidence sur la capacité correctrice du codeur de Reed-Solomon.

Les valeurs $L_b = 2$ et $L_b = 5$ paquets ont donc été appliquées au modèle analytique développé, les résultats sont représentés dans la figure suivante :

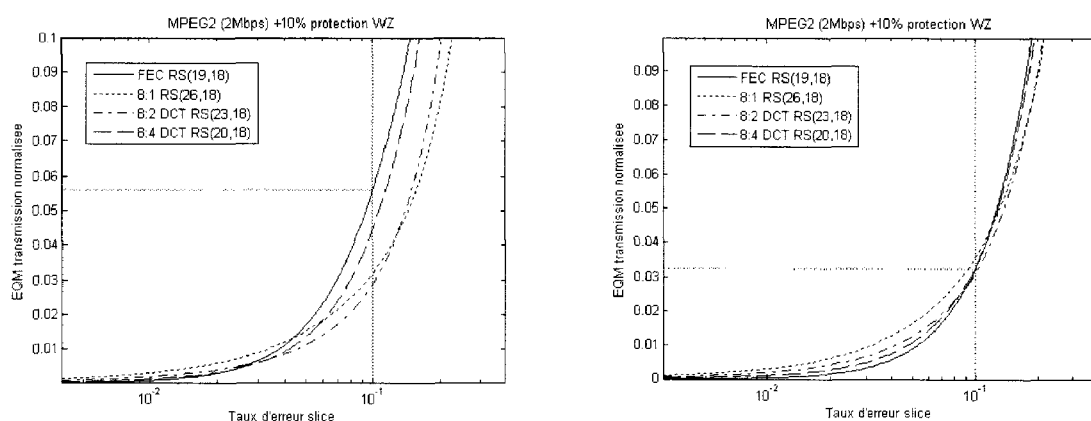


Figure 4.6 : Effet d'une augmentation de la longueur moyenne d'une salve d'erreur de 2 paquets (à gauche) à 5 paquets (à droite) sur la distorsion moyenne de la séquence décodée (dissimulation d'erreur incluse).

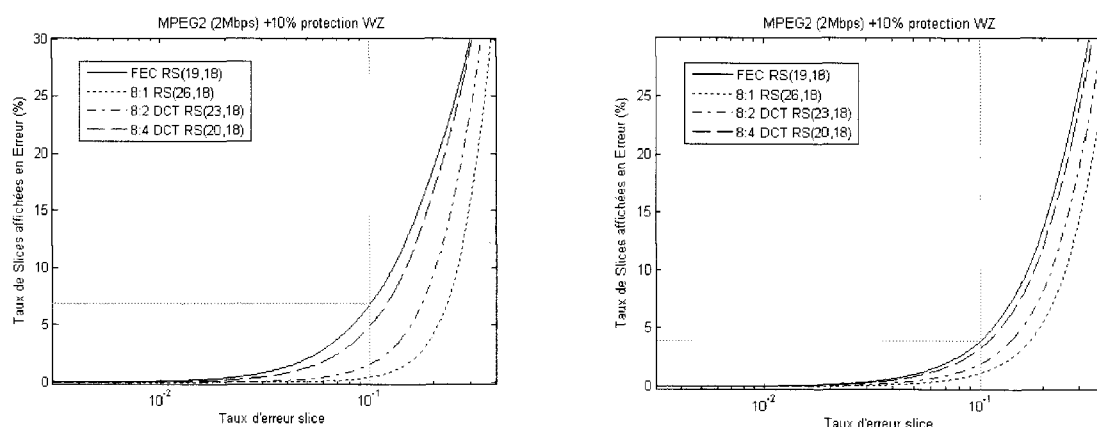


Figure 4.7 : Effet d'une augmentation de la longueur moyenne d'une salve d'erreur de 2 paquets (à gauche) à 5 paquets (à droite) sur le taux de slices affichées en erreur.

Lorsque la longueur moyenne d'une salve d'erreur augmente, à taux d'erreur moyen égal, l'écart entre les différentes solutions est réduit, tous les schémas se rapprochent du cas FEC. En effet, la probabilité d'obtenir peu d'erreurs par bloc diminue au profit de la probabilité d'obtenir, même plus rarement, un grand nombre d'erreurs par bloc (cf. Figure 4.5). Cette répartition des erreurs est nettement plus préjudiciable au décodage RS qui échoue plus fréquemment. En particulier, la diminution de résolution ne permet plus forcément de garantir une capacité correctrice suffisante, comme dans le cas FEC.

Notre schéma permet toujours d'améliorer la protection en diminuant le taux de slices affichées en erreur par rapport au schéma FEC classique, bien que ses performances soient globalement dégradées. Cependant, le débit de protection considéré (10%) est faible, une augmentation de ce taux de protection permettrait évidemment d'améliorer les performances de l'ensemble des schémas considérés. L'effet d'une augmentation du débit de la protection est étudié dans le paragraphe suivant.

4.2.3 Effet d'une augmentation du débit alloué à la protection

Nous avons également évalué l'effet d'une augmentation du débit consacré à la parité, sans modification du débit consacré aux données image. Les résultats obtenus reportés Figure 4.9 et 4.10 illustrent l'évolution de la résistance aux erreurs du schéma quand le taux de protection passe de 10% à 20%.

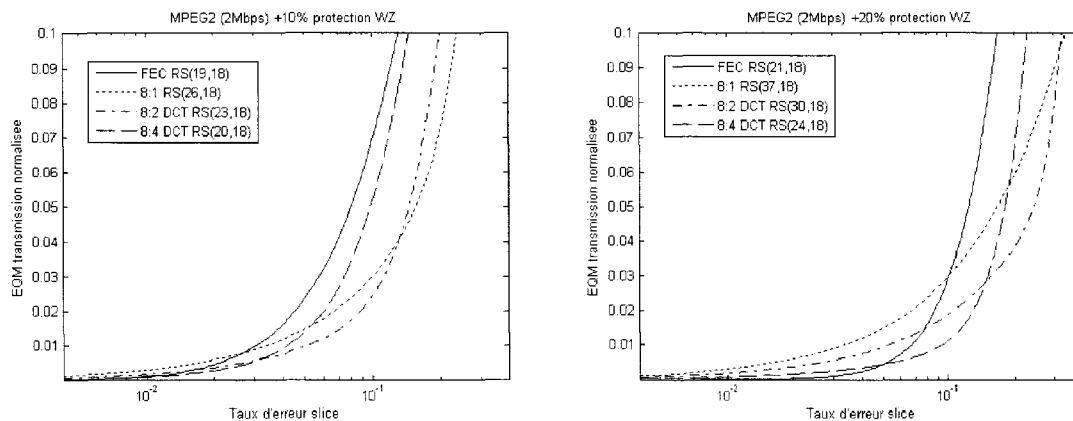


Figure 4.8 : Comparaison de la distorsion moyenne de la séquence décodée pour des taux de protection de 10% (à gauche) et 20% (à droite).

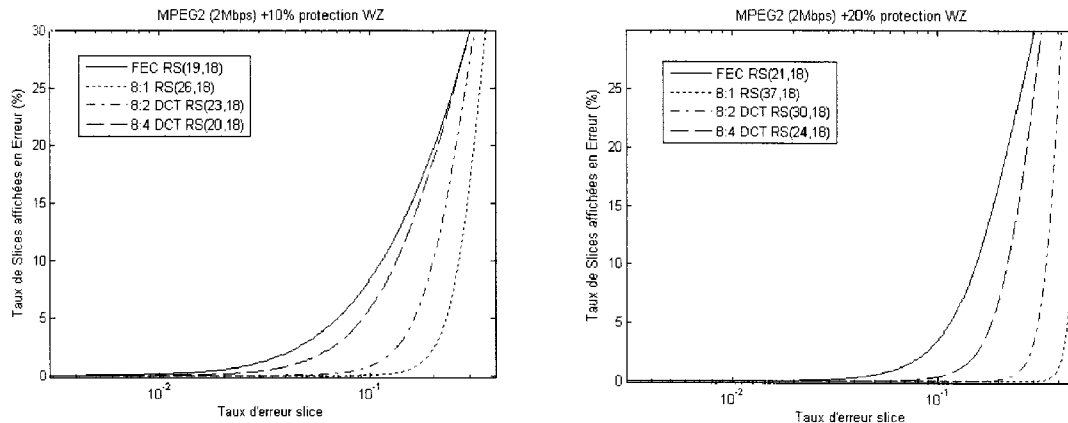


Figure 4.9 : Comparaison du taux de slice affichées en erreur moyen de la séquence décodée pour des taux de protection de 10% (à gauche) et 20% (à droite).

On constate qu'en plus d'une amélioration globale des performances, l'écart entre les différentes courbes se creuse. En effet, plus la description utilisée pour générer le flux de Wyner-Ziv est grossière, plus les slices de parité associées sont courtes. Pour un même volume de données, le nombre de slices de parité transmises est alors plus élevé et la capacité correctrice améliorée. Si le volume dédié à la protection augmente, l'écart entre les valeurs prises par le paramètre n dans les différentes configurations du système augmente également. La protection apportée est d'autant plus importante, ce qui favorise la protection par un flux de WZ.

Remarque : Effet d'une diminution de débit du codage source

On notera que, pour libérer de la bande passante et augmenter le volume de la parité, le procédé classique consiste à diminuer le débit du flux principal. Or si on augmente le taux de compression du codage source, les coefficients DCT du flux vidéo seront plus grossièrement quantifiés, le nombre moyen de coefficients DCT non nuls par bloc et leur amplitude après quantification diminuera. La décimation forcera donc à zéro un nombre de coefficients plus faible et d'amplitude moindre et la réduction de débit sera moins conséquente.

Nous avons vérifié que pour une diminution de 10% du débit issu du codage source, qui permet donc de doubler le débit de la protection, la perte d'efficacité de la réduction de débit n'est pas significative. La réduction de débit du flux principal permet surtout de consacrer une plus grande partie de la bande passante totale à la protection.

On notera d'autre part que si le codage source est plus grossier, la variation de qualité entre la résolution fine et les versions de remplacement (erreur, dissimulation ou substitution) sera un peu moins élevée. La qualité de l'image décodée, pour des taux d'erreur faible diminuera

légèrement car le codage de source est plus sévère. Mais la réduction de débit du flux principal permettant de consacrer une part de la bande passante plus importante à la protection, la résistance aux erreurs s'en verra nettement améliorée et la qualité de la séquence vidéo décodée augmentera substantiellement pour des taux d'erreur plus élevés.

4.3 Résultats expérimentaux

L'objectif de ce paragraphe est la validation du système de protection d'un flux vidéo grâce à la transmission d'un flux supplémentaire obtenu par codage de Wyner-Ziv d'une version basse résolution spatiale du flux vidéo principal. Nous avons, dans le paragraphe précédent, analysé les performances de ce schéma en fonction des caractéristiques du canal de transmission et des paramètres du système. Les résultats du modèle analytique nous ont permis d'estimer les performances du schéma de protection proposé en fonction de la résolution spatiale de la description WZ utilisée et des conditions de transmission.

Pour valider les résultats obtenus, nous appliquons le système proposé à un flux vidéo transmis sur un canal simulé et comparons les performances des simulations à celles estimées par le modèle. Finalement, l'efficacité de notre système de protection est évaluée dans le cas d'une transmission sur un canal sans fil.

Dans les expériences décrites dans la suite de ce chapitre, le flux principal est codé selon la norme MPEG2. Le débit de la protection a été fixé à 10% du débit du flux principal, car les taux d'erreur paquet moyens ciblés sont faibles (cf. Tableau 4.2). Dans le cas d'erreurs non corrélées, cette valeur assure une protection efficace pour des taux d'erreur slice de 5% (cas FEC) à 30% (cas 8:1) comme nous l'avons vu au Chapitre III, paragraphe 3.2 (Figure 3.14).

La dissimulation d'erreur est appliquée aux slices perdues non corrigées en copiant la slice correspondante de l'image de référence précédente. L'EQM est utilisée comme mesure de distorsion. La séquence transmise (7-12 test, [VQE06]) possède 2640 images de 18 slices chacune. La quantité de données générées à laquelle les conditions de transmission sont appliquées permet d'obtenir des valeurs moyennes représentatives.

4.3.1 Résultats de simulation

Les paramètres de la protection appliqués dans cette section correspondent à la modélisation du paragraphe précédent. Le flux principal et le flux de protection sont transmis dans des paquets

de taille variable de sorte à ce que chaque paquet contienne exactement une slice. Les erreurs sont générées à l'aide du modèle de canal de Gilbert-Elliott avec une longueur de salve d'erreur moyenne de 1,2 et un taux d'erreur variant de 1 à 20%.

Nous détaillons les résultats obtenus pour 1% des paquets /slices en erreur, l'ensemble des résultats est reporté Annexe V. Puis nous nous intéressons à la corrélation entre les résultats de simulation et ceux issus du modèle analytique.

4.3.1.1 Taux d'erreur résiduel et distorsion

Les Figures 4.10 et 4.11 illustrent respectivement les résultats de la transmission simulée en termes d'erreurs slices résiduelles et de distorsion associée pour un taux d'erreur moyen de 1% à titre d'exemple.

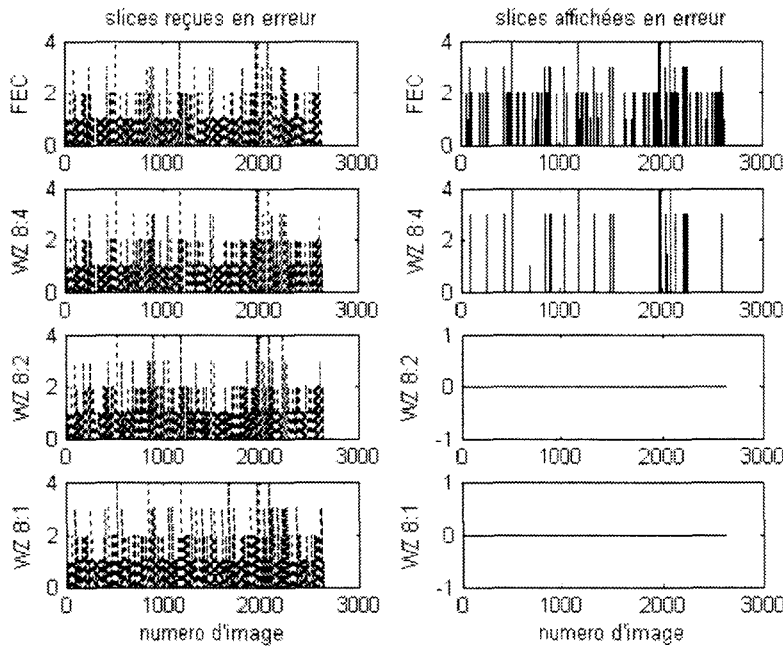


Figure 4.10 : Nombre de slices reçues et affichées en erreur en fonction du numéro d'image pour la simulation d'une transmission sur un canal de Gilbert-Elliott de taux d'erreur moyen 1% et de salve d'erreur d'une longueur moyenne de 1,2 slices.

On constate, comme le modèle analytique l'avait prévu, que l'usage d'un flux de protection basé sur une description de résolution spatiale plus faible permet une meilleure résistance aux erreurs. À la réception, le nombre de slices en erreur par image varie de 1 à 4, ces erreurs touchent les slices image ou les slices de parité. Le nombre de slices en erreur dans l'image reconstruite varie en fonction de la capacité correctrice du schéma considéré.

La capacité correctrice du schéma FEC est d'une slice, la correction échoue donc dans la plupart des cas. La capacité correctrice du schéma 8 :4 est de 2 slices et n'est donc toujours pas suffisante pour permettre la reconstruction de toutes les slices perdues. Les schémas 8 :2 et 8 :1 ont respectivement des capacités correctrices de 4 et 8 slices, ils ont conduit ici à un taux d'erreur slice résiduel nul.

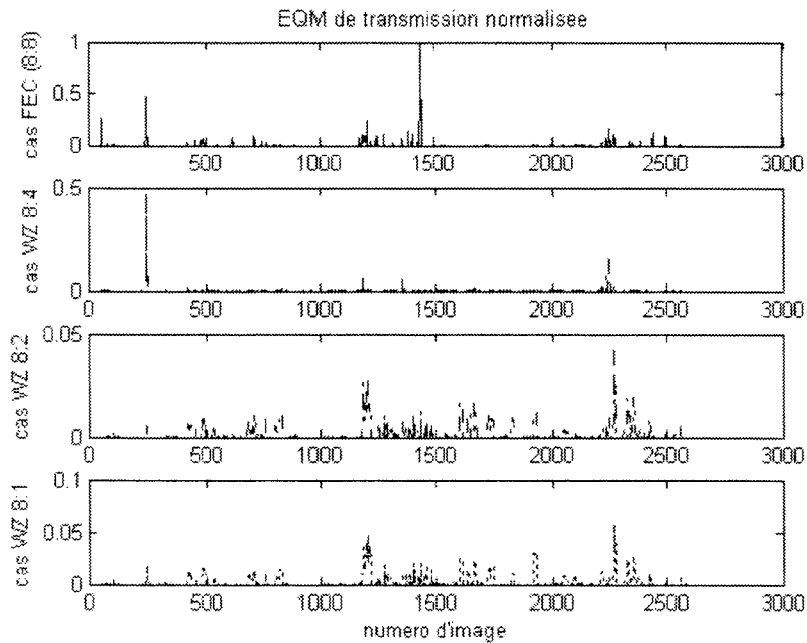


Figure 4.11 : EQM normalisée en fonction du numéro d'image associée à la simulation d'une transmission sur un canal de Gilbert-Elliott de taux d'erreur moyen 1% et de salve d'erreur d'une longueur moyenne de 1,2 slices.

La Figure 4.11 met en évidence le gain associé à l'usage d'une protection portant sur une description de résolution spatiale inférieure. En effet, pour le cas FEC, la protection n'est pas efficace et les distorsions liées à la transmission sont fréquentes et de grande amplitude. Le schéma 8 :4 permet de corriger une partie des erreurs de transmission et de diminuer significativement la distorsion globale de la séquence reconstruite. Mais la capacité correctrice étant insuffisante, certaines images présentent toujours des distorsions inacceptables.

Les schémas 8 :2 et 8 :1 assurent une reconstruction sans erreur des slices basse résolution qui sont substituées aux slices reçues en erreur. Cette substitution entraîne une diminution de la qualité de la slice reconstruite associée à la diminution de résolution spatiale. À l'évidence, les performances du schéma 8 :2 sont meilleures que celles du schéma 8 :1 puisque sa capacité correctrice est suffisante et la résolution spatiale protégée supérieure.

4.3.1.2 Effet du mode de codage de l'image / du mouvement sur la distorsion

L'observation des Figures 4.10 et 4.11 révèle une grande variabilité de l'EQM associée à la substitution d'une slice en erreur par une slice corrigée ou dissimulée. Ces disparités sont liées à deux facteurs : le mode de codage de l'image et la présence / l'absence de mouvement. Ces phénomènes sont illustrés Figure 4.12.

- La substitution par une slice corrigée d'une slice image codée en intra induit une diminution de résolution de l'image affichée. Dans le cas d'une image codée en inter, l'information de mouvement est intacte, seule l'erreur de prédiction est altérée. Or l'erreur de prédiction concerne en particulier les contours et détails fins de l'image, sa perte se traduit donc par un léger flou. Les valeurs d'EQM associées à ces modifications sont faibles que la séquence soit animée ou non.
- La dissimulation d'erreur remplace une slice perdue non restaurée par une slice de l'image de référence précédente. En cas de mouvement, les deux slices ont des contenus très différents et la distorsion associée au remplacement est inacceptable. Les distorsions les plus amples sont associées aux zones les plus animées de la séquence, que l'image soit codée en intra ou en inter.

La Figure 4.12 donne des exemples d'images décodées après correction / dissimulation des erreurs de transmission par les schémas FEC, WZ 8 :4, WZ 8 :2, WZ 8 :1. Il s'agit de la dernière image P du GOP, ses slices 4 et 5 (tee-shirt sombre 59) sont perdues. D'autre part, l'image I de ce GOP contient elle aussi deux erreurs de transmission aux slices 7 et 8 (tee-shirt blanc 87).

La capacité correctrice du schéma de protection FEC n'est que d'une slice par image, la correction échoue donc dans ce cas. La dissimulation d'erreur est utilisée pour reconstruire les slices perdues de l'image I et de l'image P mais la séquence étant fortement animée, le résultat n'est pas satisfaisant. L'image P subit donc les erreurs de l'image I précédente qui se propagent (bien qu'atténuées) et les erreurs dues à la dissimulation d'erreur sur l'image P elle-même.

Les schémas de protection par un flux de WZ basé sur une description basse résolution du flux vidéo ont une capacité correctrice suffisante. Ils permettent d'afficher une image correcte bien que les slices reconstruites soient de résolution spatiale réduite. Dans une image codée en intra, la diminution de résolution peut être gênante (cas 8:1) mais elle n'est pas perceptible dans le cas 8:4. Les slices 7 et 8 sont corrompues par les substitutions de slices dans l'image I en début de GOP, cette substitution entraîne une erreur lors du décodage des images suivantes du GOP puisque l'image de référence a un contenu différent à l'encodeur (pleine résolution) et au décodeur (basse résolution).

L'estimation de mouvement peut être réalisée sans perte puisque les informations de mouvement de la description WZ sont rigoureusement identiques à celles du flux pleine résolution. Les slices 4 et 5 sont affectées d'un léger flou dû à la diminution de résolution de l'erreur de prédiction. On remarque aussi la diminution de résolution des blocs codés en intra de ces slices P. Ces phénomènes sont également présents dans les cas WZ 8 :2 et 8 :4 mais ils sont moins visibles voire négligeables (cas 8 :4).

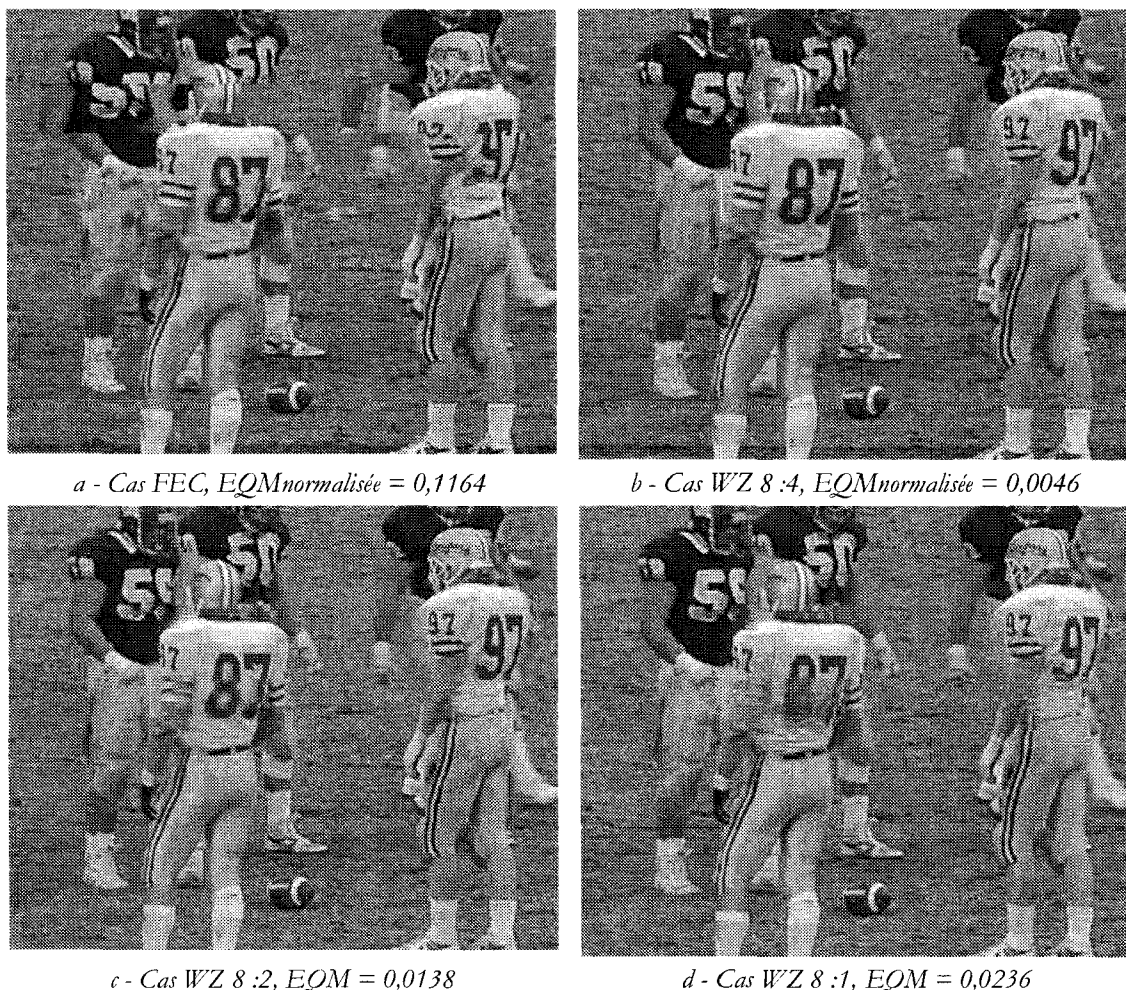


Figure 4.12 : Illustration de la qualité visuelle associée aux schémas de protection pour une image P contenant 2 slices en erreur.

L'évaluation des erreurs slices résiduelles et de la distorsion de l'image reconstruite confirme la supériorité des performances obtenues par le schéma de protection SLEP basé sur une scalabilité fréquentielle par rapport au schéma FEC. La comparaison subjective montre clairement le bénéfice associé à l'utilisation du schéma SLEP en cas de mouvement dans la séquence à protéger.

4.3.1.3 Évaluation de la validité du modèle analytique

Nous avons développé un modèle analytique permettant d'estimer la distorsion due à la transmission à partir des paramètres du canal de Gilbert-Elliott. Ce modèle ne prend pas en compte certains phénomènes qui atténuent la propagation des erreurs de transmission [FAR99]. On citera par exemple la saturation, qui a une valeur erronée supérieure à 255 attribuera la valeur 255 pour un codage sur 8 bits de la valeur du pixel. Ce seuillage diminue en pratique l'EQM associée à l'erreur de transmission. Ces phénomènes ont des conséquences sur la distorsion finale qu'il nous faut quantifier pour valider notre modèle.

Pour chaque réalisation du canal testée, les performances en termes de résistance aux erreurs et de qualité décodée ont été évaluées. Les mêmes paramètres de canal (taux d'erreur moyen et longueur moyenne de salve d'erreur) ont été appliqués au modèle analytique et les performances estimées relevées. Nous avons ainsi pu estimer la précision de l'analyse. Les résultats de ces diverses réalisations sont présentés Figure 4.13 et détaillés dans l'Annexe V.

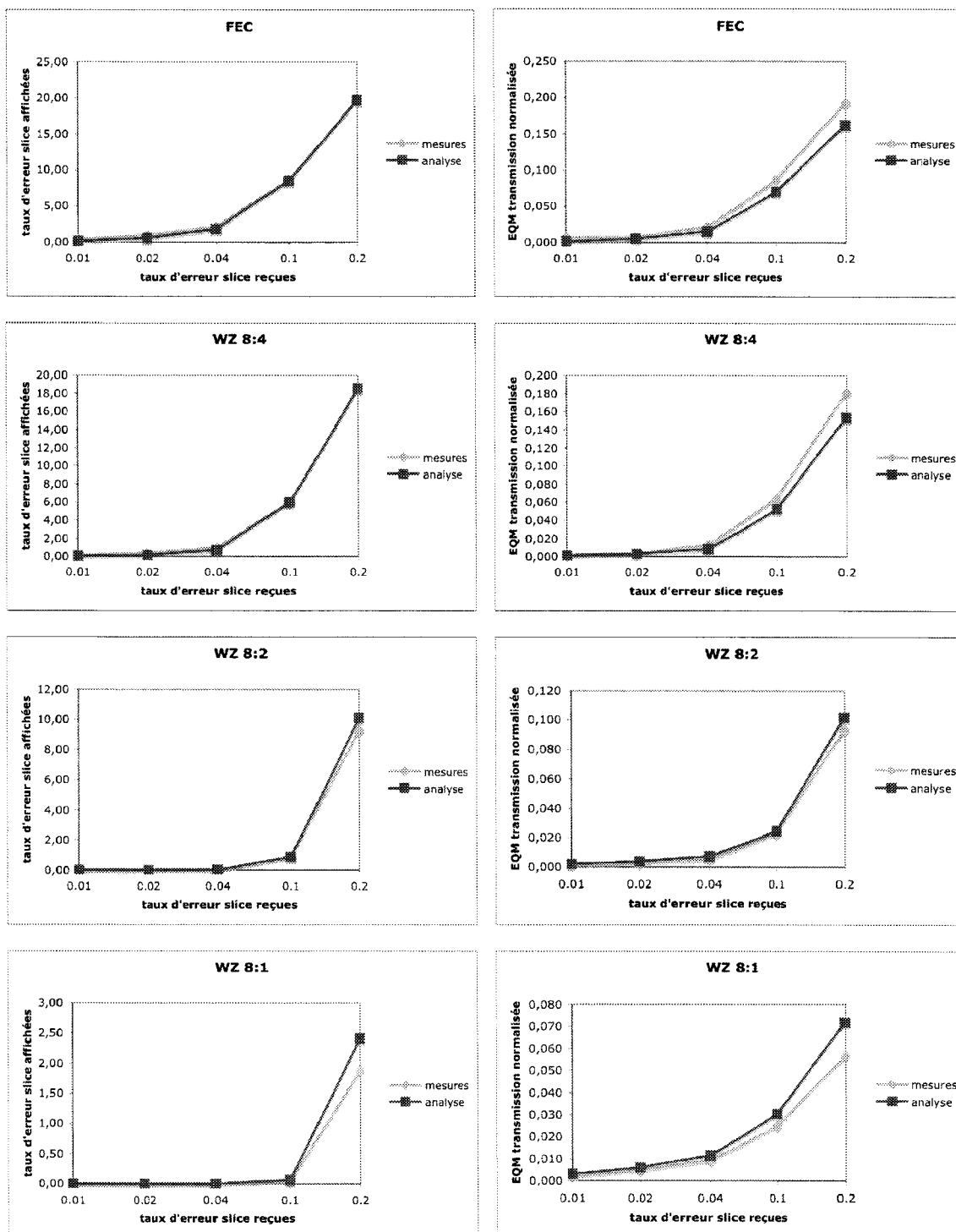


Figure 4.13 : Comparaison entre valeurs analytiques et résultats de simulations en termes de taux de slices affichées en erreur (à gauche) et d'EQM (à droite) pour les différents schémas de protection.

Le taux d'erreur résiduel et la distorsion donnés par la démarche analytique correspondent à ceux mesurés après simulation. Les résultats de simulation sont légèrement supérieurs à l'estimation du modèle, en particulier quand le taux d'erreur slice est le plus élevé. La différence peut s'expliquer par plusieurs facteurs :

- La répartition des erreurs (ou plutôt salves d'erreurs) simulées qui n'est pas complètement uniforme. Si les erreurs non corrigées touchent plus fortement les images P et B, la propagation des erreurs est moins importante (cas WZ 8 :1, pour 20% des slices reçues en erreurs par exemple). La distorsion mesurée est donc plus faible que celle estimée par le modèle qui considère une probabilité d'erreur résiduelle équivalente sur tous les types d'images. De la même façon, les erreurs résiduelles ne sont pas exactement uniformément réparties sur les différentes sections de la séquence test. Si un passage vidéo fortement animé est moins corrompu que le reste de la séquence après correction, l'EQM mesurée sera plus faible que la valeur moyenne donnée par le modèle (cf. Figure 4.14). L'EQM de transmission est donc fortement dépendante de la section de séquence touchée.
- Les phénomènes qui atténuent la propagation des erreurs, assimilables à un filtrage spatial équivalent [FAR99] qui ne sont pas pris en compte par notre analyse.

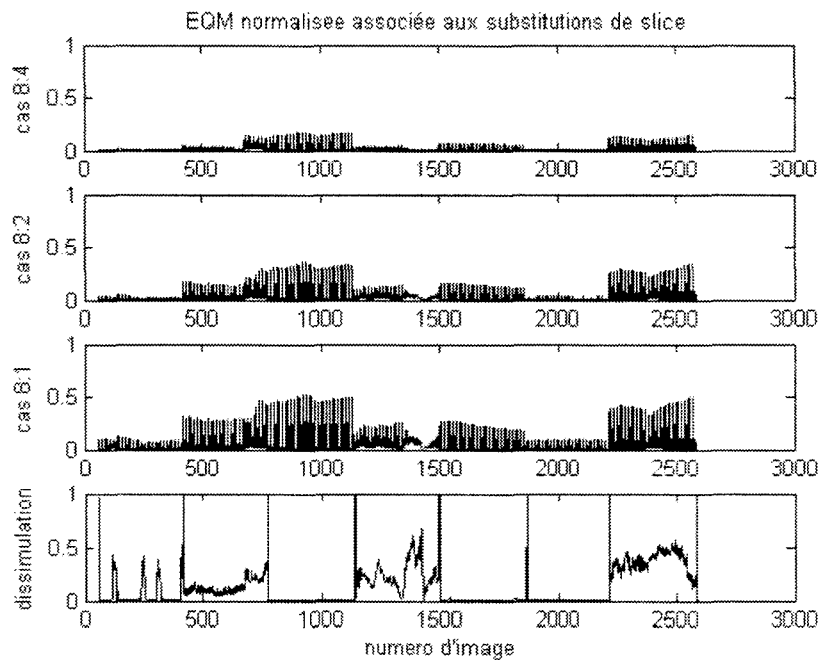


Figure 4.14 : Variation de l'EQM de transcodage et de dissimulation en fonction du numéro d'image.

Après comparaison des valeurs estimées par le modèle analytique et des valeurs mesurées après simulation, le modèle analytique est validé. Pour un canal de Gilbert-Elliott spécifié, les valeurs obtenues par chaque simulation et l'analyse sont fortement corrélées. Il nous reste à vérifier l'efficacité du système en lui appliquant des erreurs correspondant à une transmission réelle.

4.3.2 Résultats de transmission

Dans ce paragraphe, le schéma proposé est utilisé pour améliorer une transmission vidéo sur un réseau local sans fil (Wireless Local Area Network, WLAN) basé sur la technologie 802.11. La modélisation des canaux peut seulement approximer le comportement de canaux sans fil [8]. Nous avons appliqué des motifs de pertes issus de mesures préalables afin de tester l'efficacité de notre schéma SLEP.

L'ensemble des résultats pour les différents scénarii est reporté Annexe VI. Les résultats des motifs de pertes 10 et 11 mettent en évidence la résistance aux erreurs du schéma SLEP avec un taux d'erreurs slice résiduel faible à nul pour les schémas WZ. La distorsion résultante pour ces motifs est toujours nettement en faveur des schémas SLEP, avec un minimum atteint par le schéma 8 :4. Les ensembles de données 7, 8 et 9 présentent des motifs d'erreurs dont les salves ne sont pas uniformément distribuées. De plus, l'estimation du taux d'erreur slice à partir du taux d'erreur paquet dans le cas d'erreurs corrélées est un problème complexe. Ces phénomènes ne sont pas pris en compte par le modèle analytique développé qui n'est donc plus adéquat pour l'estimation de la distorsion finale. L'usage de mesures concrètes pour l'étude des performances de notre système de transmission est donc justifié.

Nous détaillons ci-après les résultats du motif référencé data 8, le pire des cas en termes de taux d'erreur paquet moyen. Ce motif résulte d'une transmission en intérieur avec le point d'accès qui se rapproche puis s'éloigne de la station de base. Le taux d'erreur paquet moyen est de 1,2% et la longueur de salve d'erreur moyenne (ordre 1) est 1,3 paquets. Comme le nombre de slices de parité adjoint à chaque flux dépend des paramètres de la protection, la position des erreurs à la réception varie d'un flux à l'autre. Le Tableau 4.3 donne les taux d'erreurs slice et longueurs de salves d'erreur calculés après application des motifs d'erreur paquet aux différents flux transmis. Limiter à un le nombre de slices de parité maximal par paquet agit comme un brassage en augmentant la distance (en nombre de paquets) entre deux images successives quand le nombre de slices de parité augmente. La résistance aux erreurs est donc renforcée au prix d'une baisse d'efficacité de la transmission (remplissage de zéros et délai).

	taux d'erreur	longueur de salve
FEC	1,64%	2,04
WZ 8:4	1,46%	1,60
WZ 8:2	0,85%	1,41
WZ 8:1	0,95%	1,33

Tableau 4.3 : Caractéristiques des erreurs slice à l'entrée du décodeur après application du motif d'erreur référencé data 8 pour les divers schémas.

Le taux de slices reçues en erreur moyen est modifié par rapport au taux de paquets reçus en erreur à cause de la fragmentation des slices les plus longues lors de la mise en paquets. Le taux d'erreur slice des images I augmente jusqu'à 3% alors que celui des images B ne dépasse pas 1,4% à cause de leurs longueurs respectives comme le montre le Tableau 4.4. La longueur moyenne d'une salve d'erreur est à l'origine de 1,3 paquets. La concentration de ces erreurs se traduit par une moyenne de plus de 5 slices en erreur par image corrompue, valeur à comparer aux capacités correctives des différents schémas (respectivement de 1, 2, 4 et 8 slices en erreur).

		Taux de slices reçues en erreur	Nombre moyen de slices en erreur par image (qd erreur)
FEC	I	3,10%	8,13
	P	1,93%	5,26
	B	1,40%	4,73
WZ 8:4	I	2,56%	6,65
	P	1,61%	4,84
	B	1,41%	4,82
WZ 8:2	I	1,87%	7,92
	P	0,70%	3,53
	B	1,00%	5,56
WZ 8:1	I	0,99%	5,70
	P	1,27%	6,06
	B	1,14%	6,27

Tableau 4.4 : Caractéristiques des erreurs slices associées au motif de pertes référencé data 8 pour les différents types d'image.

La Figure 4.15 présente les résultats en termes de taux d'erreur slice avant et après correction.

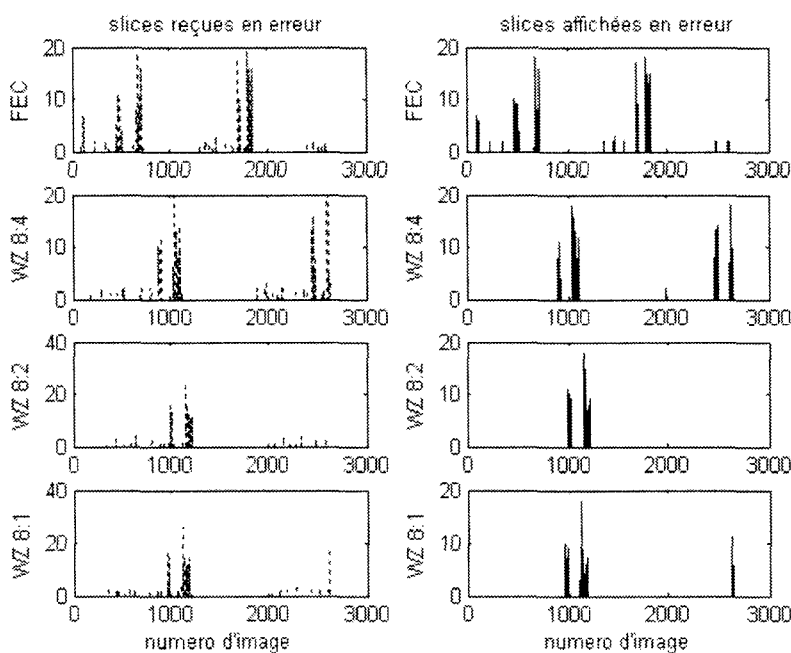


Figure 4.15 : Taux d'erreur slice avant / après correction pour les différents schémas de protection et le motif de pertes référencé data 8.

Comme on peut l'observer sur la Figure 4.15, le schéma FEC affiche plus de slices en erreur que les schémas utilisant un flux de protection WZ car la protection par codage RS est plus faible. Nous vérifions que les schémas SLEP, qui utilisent des descriptions de résolutions spatiales inférieures, ont une résistance aux erreurs supérieure avec une correction très efficace. Cependant, aucun des schémas évalués n'est capable de garantir une reconstruction sans erreur face aux pires salves d'erreur. Dans la mesure où le nombre de slices image perdues excède le nombre de slices de parité transmises, la correction échoue. On remarque que pour certaines images, les 18 slices source sont perdues et que dans ce voisinage 24% des slices sont perdues (plus de 300 slices sur 75 images). Dans ces conditions, une protection plus forte ne garantirait pas la reconstruction sans erreurs des slices perdues car les slices de parité étant transmises immédiatement après les slices images, elles seraient également corrompues. Un brassage supplémentaire d'ordre très élevé (plusieurs images) des paquets pourrait résoudre le problème mais au prix d'un délai fortement accru (quelques secondes).

La Figure 4.16 donne l'EQM de transmission résultant de la substitution des slices en erreur par leur version basse résolution spatiale (correction) ou par la slice correspondante de l'image de référence précédente (dissimulation). Le Tableau 4.5 donne les résultats moyens obtenus pour le motif de perte data 8. Les valeurs analytiques correspondantes sont fournies à titre indicatif.

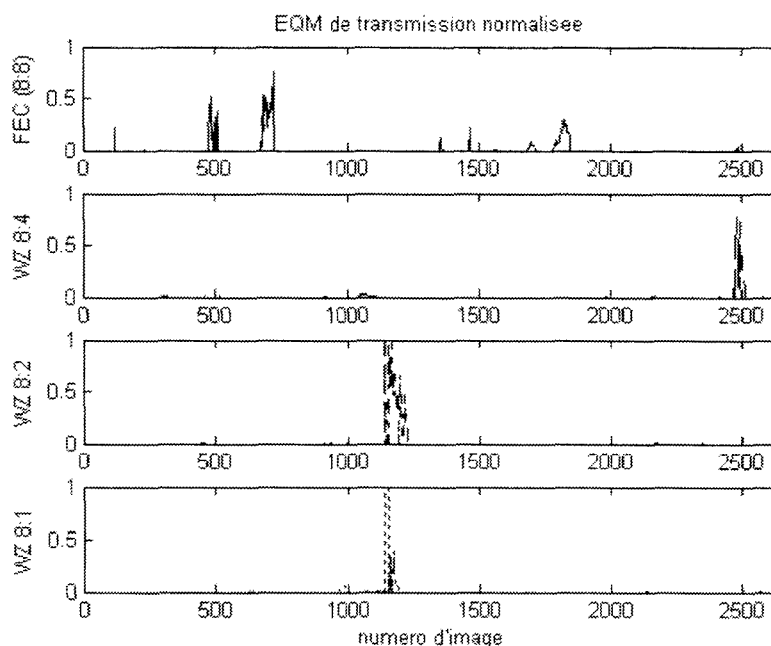


Figure 4.16 : Distorsion globale liée à la transmission pour les différents schémas évalués.

La Figure 4.16 valide l'efficacité de notre schéma et l'amélioration de qualité de la séquence décodée dans le cas d'une protection par les flux WZ. Dans les cas 8 :1 et 8 :2, la distorsion de transmission est essentiellement due aux substitutions, à la diminution de résolution spatiale et non à la dissimulation d'erreurs non corrigées. Le schéma SLEP ne protège qu'une basse résolution spatiale ce qui permet d'augmenter la résistance aux erreurs. La qualité de la séquence décodée est plus constante que celle obtenue par le schéma FEC améliorant donc la qualité d'expérience de l'observateur.

Le taux d'erreur slice résiduel après décodage RS reste important, mais essentiellement dû à une salve d'erreur principale dans le cas des schémas SLEP. L'impact visuel de ces erreurs est grand mais limité à un intervalle de temps de 2 secondes, le reste de la séquence étant de bonne qualité visuelle. Pendant ces deux secondes, la correction est impossible par les schémas de protection usuels. Une solution consisterait à retransmettre les slices perdues après ces deux secondes (quand l'état du canal redevient favorable) ce qui implique la présence d'une mémoire tampon suffisante au récepteur.

Resultats généraux :

Les figures 4.17 et 4.18 regroupent les résultats expérimentaux obtenus après application des différents motifs de pertes en termes de taux d'erreurs slice résiduel et d'EQM de transmission.

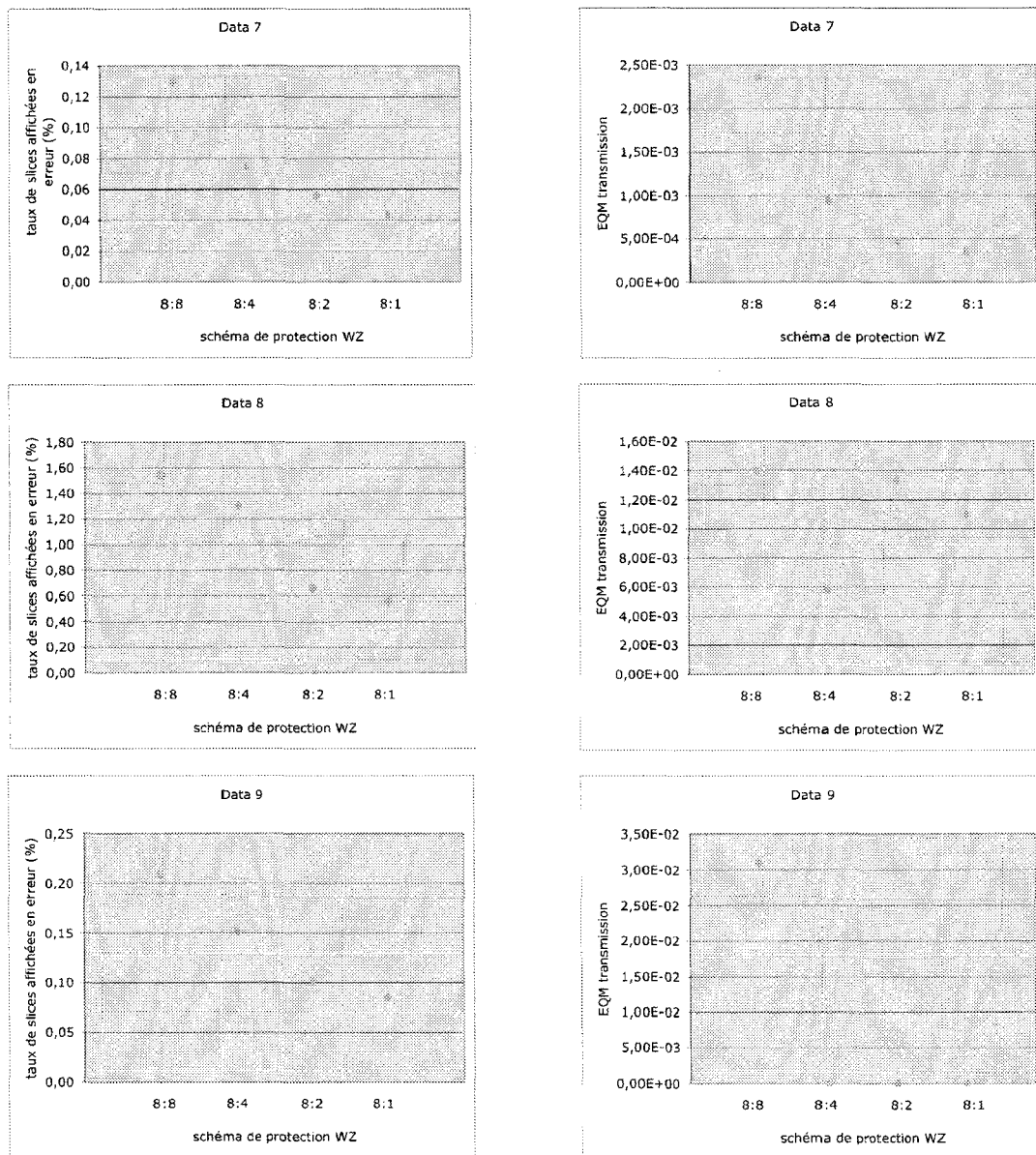


Figure 4.17 : Résultats expérimentaux obtenus pour les motifs de pertes data 7 à data 9.

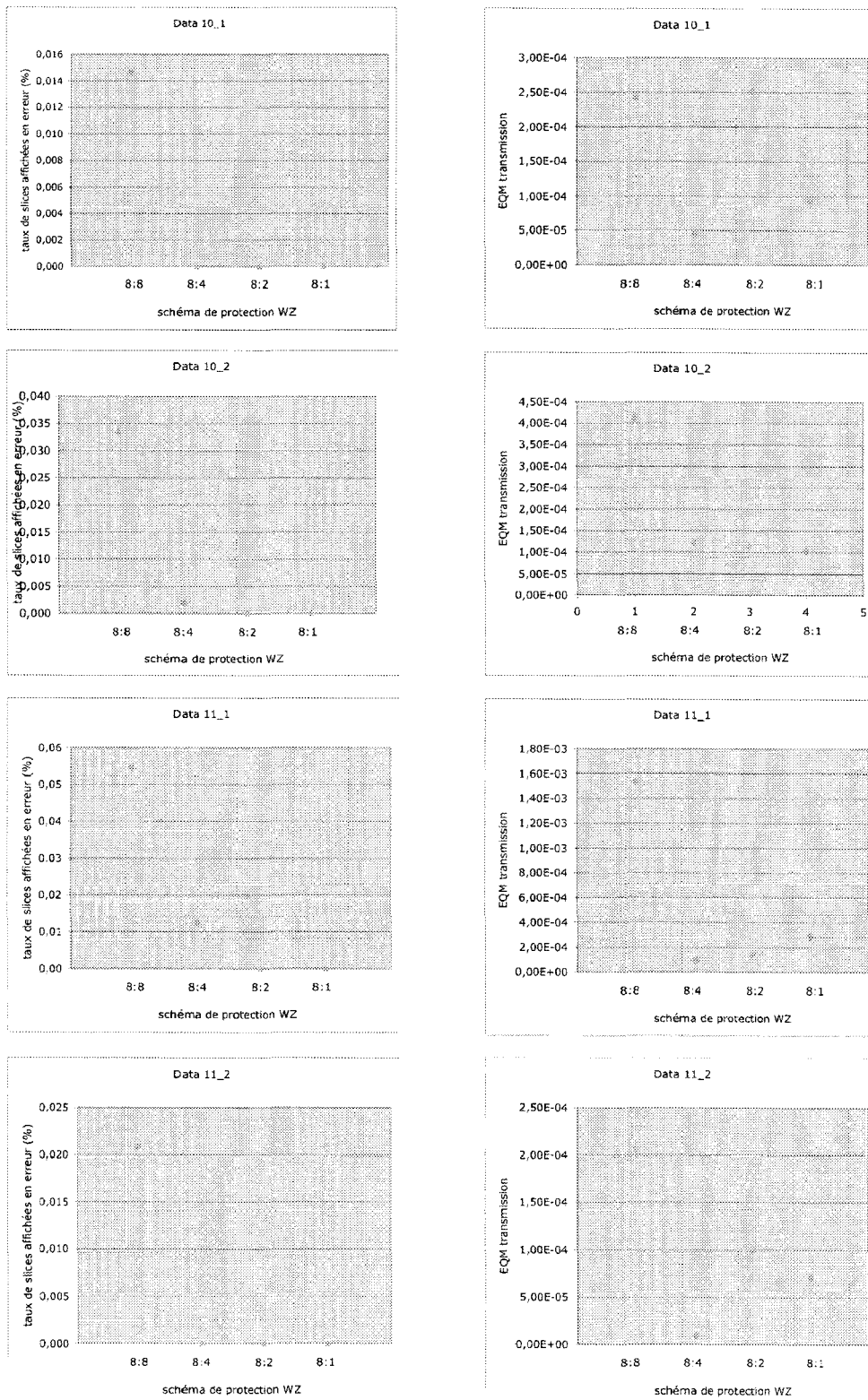


Figure 4.18 : Résultats expérimentaux et analytiques obtenus pour les motifs de pertes data 10 et data 11.

On remarque que les salves d'erreurs réelles ont une distribution non-uniforme, elles sont corrélées et n'affectent que certaines sections de la séquence, en particulier quand le point d'accès est mobile (motifs data 7, 8 et 9). Or l'EQM de transcodage et l'EQM liée à la dissimulation varient fortement en fonction du contenu des images et de leur mode de codage (cf. Figure 4.14). D'autre part, on notera que le modèle de canal utilisé pour la partie analytique ne correspond pas à ce cas de figure, l'estimation des performances voit sa précision diminuer.

Pour tous les motifs de pertes, on observe que le taux d'erreur slice résiduel et la distorsion obtenus par les schémas de protection SLEP sont inférieurs à ceux obtenus par le schéma FEC. La description optimale varie en fonction du motif d'erreur considéré. Pour les motifs data 7, 8 et 9, la concentration des erreurs est telle qu'aucun des schémas proposés ne peut garantir la correction, le schéma WZ 8 :1 donne les meilleurs résultats en termes de taux d'erreur slice résiduel et d'EQM. Pour le motif data 11_1, la capacité correctrice du flux WZ 8 :2 suffit à garantir une reconstruction sans erreur, les salves d'erreur étant réparties de façon plus uniforme. La capacité correctrice du flux WZ 8 :4 est à peine dépassée et le gain de qualité associée à la résolution spatiale 8 :4 fait que l'EQM globale du cas 8 :2 est supérieure à celle du cas 8 :4 (compromis qualité / correction). Pour les motifs data 10_1, 10_2 et 11_2, le taux d'erreur d'entrée est plus faible, un flux WZ 8:4 donne une résistance aux erreurs suffisante tout en préservant la qualité des slices perdues reconstruites.

4.4 Conclusion

Dans ce chapitre, le modèle analytique permettant d'estimer la distorsion liée à la transmission a été mis en oeuvre. Des simulations ont permis d'en valider la précision pour le type de canal modélisé, un canal de Gilbert-Elliott. Finalement, des motifs de pertes correspondant à la transmission d'un flux de paquets de taille fixe sur une liaison sans fil ont été appliqués afin d'évaluer l'efficacité du système en conditions réelles, en incluant l'étape de mise en paquet.

Les résultats analytiques et expérimentaux démontrent clairement le gain apporté par un schéma de protection SLEP par rapport à un schéma FEC classique en termes de résistance aux erreurs. Diminuer la résolution spatiale protégée d'un facteur deux horizontalement et verticalement permet de doubler la capacité correctrice du codage de Reed-Solomon pour un débit global identique. Nous avons fait varier le taux d'erreur moyen, la longueur moyenne d'une salve d'erreur et le pourcentage de débit consacré à la protection. Dans tous les cas évalués, l'utilisation d'un flux WZ pour protéger le flux principal permet de diminuer le taux d'erreur slice résiduel et offre de meilleures performances qu'un schéma de protection FEC classique.

Le schéma SLEP permet de diminuer le taux de slices à dissimuler, au prix d'une augmentation du taux de slices substituées. Or la distorsion induite par la résolution spatiale réduite des slices reconstruites est, particulièrement en cas de mouvement, faible par rapport à celle due à la dissimulation des slices perdues et la qualité d'image obtenue par le schéma SLEP est supérieure. Les différents flux de protection WZ permettent de réaliser un compromis entre qualité (résolution spatiale) et résistance aux erreurs (capacité correctrice) sans modifier le codage source du flux principal. Même dans les pires conditions de transmission, ou les plus imprévisibles en termes de répartition des erreurs, la protection du flux principal par un flux WZ permet d'améliorer significativement les performances.

Conclusion et perspectives

Nous avons proposé dans cette thèse un système de protection d'un flux vidéo comprimé contre les erreurs de transmission dans le cadre d'une communication par paquets. Ce système utilise la scalabilité fréquentielle et le codage de Wyner-Ziv pour améliorer la qualité de la vidéo diffusée. Le schéma est indépendant du codec sous-jacent, il est ici appliqué au cas d'un codeur MPEG2, mais peut être adapté au standard H.264. Il peut également être utilisé pour tout type de transmission. On notera cependant qu'une détection des erreurs sur les slices reçues est requise pour que le décodage de Reed-Solomon soit à effacement et la capacité correctrice maximale. Les réseaux basés sur une transmission par paquets (IP, DVB, ATM, etc.) incluent un contrôle d'intégrité des données, notre système y est donc particulièrement adapté.

Dans le premier chapitre, nous avons présenté les principales techniques de codage résistant aux erreurs rencontrées dans la littérature. Parmi celles-ci, les techniques exploitant les caractéristiques de la source lors de la conception du codeur canal donnent de meilleurs résultats en accordant une priorité variable aux éléments du flux selon leur importance visuelle. Le codage scalable en est un exemple, il permet d'adapter au mieux la transmission des signaux aux diverses ressources disponibles (réseaux ou terminaux). Mais cette technique souffre encore de plus faibles performances en compression et ne constitue pas à ce jour une solution efficace. Les recherches dans ce domaine étant prolifiques, nous avons cherché comment réaliser cette protection hiérarchique sans sacrifier l'efficacité de codage.

Le chapitre II retrace l'historique des principaux aboutissements du codage source avec information adjacente, application des théorèmes de Slépián-Wolf et Wyner-Ziv. Le principe général est de considérer l'information séparant la source de l'information adjacente qui lui est corrélée comme un bruit de transmission. Le codage de Wyner-Ziv pratique se base sur l'emploi d'un codage canal comme codage de source, il offre en conséquence des propriétés intéressantes de résistance aux pertes d'information. En particulier, le schéma SLEP (« Systematic Lossy Error Protection ») qui émule la scalabilité SNR, sans en reproduire les inconvénients en termes de complexité et inefficacité de codage, a retenu notre attention.

Notre schéma SLEP à scalabilité spatiale est explicité dans le chapitre III. La complexité habituellement liée au filtrage spatial est évitée en effectuant la diminution de résolution dans le domaine fréquentiel, procédé référencé comme scalabilité fréquentielle. Les effets de la diminution de résolution spatiale sur le débit et la qualité d'image ont été étudiés et la supériorité de notre approche par rapport à une approche SNR démontrée. Nous avons proposé un modèle analytique qui prend en compte les paramètres du codeur Reed-Solomon, les caractéristiques du canal de transmission conjointement avec celles de la séquence compressée. Ce modèle permet d'estimer la probabilité d'erreur slice résiduelle et la distorsion finale due à la transmission, il est utilisé pour l'optimisation du système.

Finalement, le chapitre IV regroupe les résultats de l'analyse, des simulations et des expériences réalisées. La pertinence du modèle analytique a été validée à l'aide des simulations. L'étude comparative des performances de schémas de transmission a montré le gain apporté par l'utilisation d'une description de WZ comme protection contre les pertes de paquets par rapport à une protection par FEC classique. Le système permet un compromis entre résistance aux erreurs et résolution minimale reconstruite sans erreur. La distorsion due à la dissimulation d'erreur est remplacée par une distorsion associée à la scalabilité fréquentielle nettement moins préjudiciable. Réduire la résolution spatiale garantie d'un facteur deux permet de doubler la capacité correctrice du schéma, menant à une qualité d'image diffusée supérieure, quelles que soient les conditions de transmission.

Les intérêts de ce système sont multiples. Tout d'abord, la simplicité des composants mis en oeuvre rend possible son implémentation dans un système temps réel. On notera que le seul bloc fonctionnel supplémentaire par rapport à un codec classique est la décimation dans le domaine DCT. Or cette méthode simple de transcodage constitue un premier pas vers la diminution de taille d'image qui peut se révéler nécessaire pour adresser les diverses résolutions des terminaux de diffusion. D'autre part, le système utilise une scalabilité spatiale qui permet d'adapter la résolution spatiale diffusée aux conditions du canal, mais le flux principal est monocouche. La compatibilité avec les systèmes existants est donc parfaitement préservée, ce qui en favoriserait le déploiement. On notera également que le système développé est indépendant du codec et du mode de transmission. Il sera donc possible de l'adapter facilement aux futurs techniques de codage basées DCT et réseaux de communication basés sur une transmission par paquets. Enfin, la résistance aux erreurs du schéma peut être adaptée aux caractéristiques d'erreur rencontrées sans que le codage source n'ait à être modifié, simplifiant l'optimisation de la chaîne de transmission.

Perspectives :

L'indépendance du système au codec sous-jacent autorise l'emploi de techniques de codage sophistiquées pour la génération du flux principal qui n'ont pas été considérées ici (contrôle du débit, codage résistant aux erreurs). Certaines de ces fonctionnalités peuvent être exploitées par notre système de protection afin d'améliorer les performances globales. Par exemple, si un canal de retour est disponible, on peut imaginer varier le type de protection en fonction des conditions de transmission. L'adaptation du système au standard H.264, qui permet entre autres d'obtenir des facteurs de compression nettement supérieurs, est également envisagée.

Une autre perspective de travail serait l'amélioration de la granularité de notre système. En effet, les rapports de résolution envisagés sont au nombre de 4 (8 :8, 8 :4, 8 :2 et 8 :1), donnant accès à quatre niveaux de protection. Premièrement, ces différents niveaux peuvent être combinés afin d'obtenir un compromis plus fin entre résolution spatiale et résistance aux erreurs (protection par de multiples flux WZ). Ensuite, la scalabilité fréquentielle peut être complétée par une scalabilité SNR et/ou temporelle. L'implémentation de la scalabilité temporelle peut être réalisée simplement en changeant la description protégée sur le critère du mode de codage de l'image (I, P, B). Le modèle analytique développé qui calcule séparément la distorsion moyenne des différents types d'image, pourrait alors être utilisé pour l'optimisation d'un schéma SLEP à deux niveaux de scalabilité spatiale et temporelle.

Enfin, on pourrait développer au sein du décodeur une méthode de dissimulation des erreurs spécifique à notre système de protection. Cette méthode de dissimulation d'erreur permettrait non seulement d'estimer les informations perdues mais surtout de compléter les informations partielles disponibles (information de basse résolution spatiale). En effet, si la protection est correctement paramétrée, on peut obtenir un taux d'erreur résiduel quasi nul. Les slices non reçues sont reconstruites avec une résolution spatiale inférieure mais l'information de mouvement reste intégralement disponible. Ces éléments pourraient être exploités conjointement à une technique de récupération des informations de texture afin de reconstituer une image de qualité supérieure.

Conclusion

Annexes

Annexe I : transmission par paquet

L'acheminement des informations à travers le réseau s'effectue suivant des règles de communication strictes appelées protocoles. Ces protocoles tiennent compte de tous les détails de la communication depuis les caractéristiques de la ligne physique jusqu'à la représentation de la communication. Une manière de simplifier le problème consiste à séparer l'ensemble des fonctionnalités en couches protocole. Le modèle d'interconnexion des systèmes ouverts (« Open System Interconnection », OSI, cf. [SCH00] et Figure A.1), défini par l'ISO en 1985, donne les caractéristiques physiques et logicielles nécessaires à l'interconnexion en réseau des systèmes ouverts. Bien que l'aspect protocole de communication ne constitue pas le sujet de cette étude, les outils fournis par les différentes couches du protocole doivent être pris en compte lors de l'optimisation des paramètres d'un schéma de vidéotransmission et nécessitent donc un examen préalable.

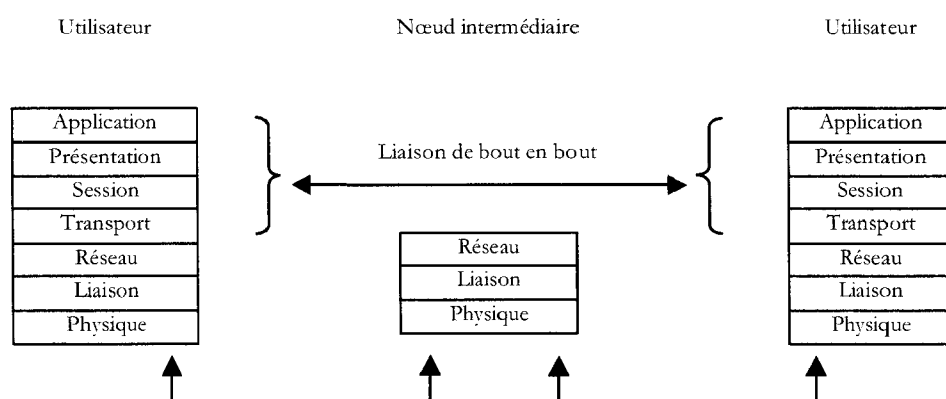


Figure A.1 : Architecture OSI.

Le réseau Internet est une collection de réseaux hétérogènes (cf. Figure A.2), dont les caractéristiques varient selon la localisation et le temps. Il définit un ensemble de protocoles, indépendants du support physique sous-jacent, ainsi qu'une méthode d'interconnexion. Le développement d'Internet a donné lieu à au modèle TCP/IP (Transmission Control Protocol/ Internet Protocol) du nom des 2 protocoles les plus connus. Ils s'appuient sur la transmission de paquets de données de taille variable numérotés et identifiés. Les différents paquets issus d'un même flux suivent des chemins indépendants en fonction du trafic sur les différentes lignes du réseau et sont ré-assemblés à destination. L'hétérogénéité des réseaux de distribution filaires ou mobiles s'appuyant sur le protocole réseau IP et la diversité des systèmes hôtes ne permet pas de garantie de qualité de service (QoS). On qualifie pour cela les services Internet de « best effort » [CIV00].

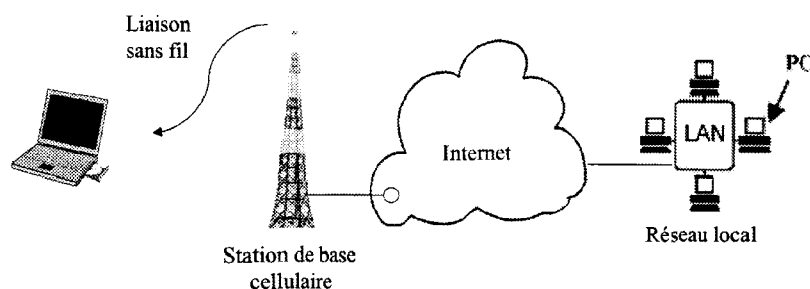


Figure A.2 : Réseau de communication mixte.

La couche protocole de transport TCP ne garantit la délivrance des paquets dans l'ordre d'envoi que grâce à la retransmission des paquets perdus et n'est donc pas adapté au cas d'une transmission temps-réel. Pour pallier les faiblesses du couple TCP/IP, L'IETF (Internet Engineering Task Force) a développé les protocoles associés UDP (User Datagram Protocol) et RTP (Real-time Transport Protocol) [RIV02]. UDP permet l'envoi des paquets en mode datagram, c'est-à-dire sans contrainte sur le chemin physique emprunté. RTP assure le bon séquençage des flux audio et vidéo et prévoit même des en-têtes dédiées aux flux MPEG [RFC98].

Les fonctionnalités fournies par les couches protocoles RTP/UDP intervenant au-dessus de la couche IP, pertinentes pour notre étude incluent :

- La mise en paquet des informations (couche application, protocole RTP),
- L'attribution d'un numéro de séquence aux paquets émis / reçus (protocole RTP),
- Le contrôle d'intégrité des informations reçues (couche transport, protocole UDP),
- Les fragmentations / réassemblages éventuels de paquets (couche réseau, protocole IP).

Ces services sont rendus possibles grâce aux informations supplémentaires contenues dans l'en-tête ajoutée au paquet par chaque couche protocole traversée. Elles représentent 40 octets de surcoût par paquet dans le cas d'une transmission utilisant la suite de protocoles RTP/UDP/IP (cf. Figure A.3 et [RFC96]). Ces fonctionnalités permettent au récepteur d'identifier et de localiser les paquets corrompus par des erreurs binaires et les paquets perdus pendant la transmission. Tout paquet passé à l'application peut donc être considéré comme sans erreur.

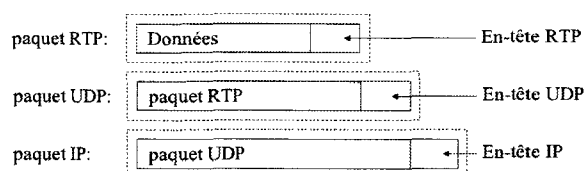


Figure A.3 : Constitution du paquet IP par encapsulations successives des données [RFC98].

Cependant il reste un grand nombre de problèmes à résoudre. Comme stipulé plus haut, le chemin emprunté par les paquets sur le réseau Internet est une collection de nœuds reliés par des liaisons de caractéristiques variables. Les congestions réseau restent à considérer et peuvent causer des pertes de paquets subites et importantes [WUD00]. De plus, en termes de débits, la vitesse de transmission globale sera celle du tronçon le plus lent et une surexploitation d'un des liens favorisera le phénomène de congestion. Il est donc primordial de veiller à une utilisation efficace de la bande passante disponible en réduisant le débit de la source audiovisuelle autant que possible.

Annexe II : Définition du canal de Gilbert-Elliott

Le paquet IP est transmis sur le médium de transmission physique utilisé. Au récepteur, les paquets corrompus sont supprimés (on ne considère pas ici l'usage des techniques de retransmission). Certains paquets n'arrivent pas (ou trop tard) à destination à cause de congestions (filaire) ou d'évanouissements (sans fil) qui sont des phénomènes potentiellement durables. La modélisation du canal doit rendre compte de l'alternance de ces états (bon et mauvais).

Comme ces pertes sont corrélées, elles peuvent être modélisées par un canal de Gilbert [GIL60], constitué d'une chaîne de Markov à deux états (cf. Figure A.4). Le canal est caractérisé par ses probabilités de transition de l'état « Good » (G) à l'état « Bad » (B) et inversement : P_{GB} et P_{BG} .

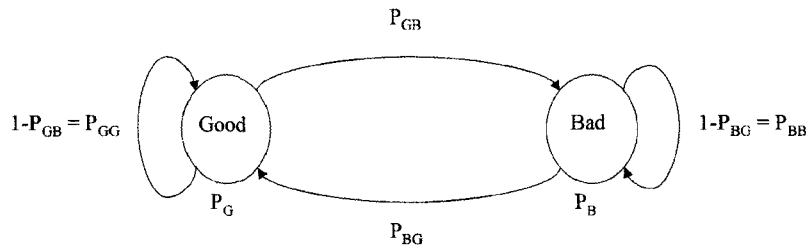


Figure A.4 : Canal de Gilbert.

Avec : $P_B = 1-h$, la probabilité d'erreur symbole dans l'état « Bad »,
 $P_G = 1-k$, la probabilité d'erreur symbole dans l'état « Good ».

On définit : $P(B)$, la fraction du temps passée dans l'état « Bad »,
 $P(G)$, la fraction du temps passée dans l'état « Good ».

Or, la probabilité d'être dans l'état B, c'est la probabilité que de l'état G, le canal soit passé à l'état B ou que le canal ait été initialement dans l'état B et y soit resté, ce que l'on peut écrire :

$$P(B) = P(B).P_{BB} + P(G).P_{GB}$$

On en déduit :

$$P(B) = P(B).P_{BB} + [1 - P(B)].P_{GB}$$

$$P_{GB} = P(B).[1 - P_{BB} + P_{GB}] = P(B).(P_{BG} + P_{GB})$$

$$P(B) = \frac{P_{GB}}{(P_{BG} + P_{GB})}$$

On obtient de la même façon l'expression de $P(G)$ en fonction des probabilités de transition du canal :

$$P(G) = \frac{P_{BG}}{(P_{BG} + P_{GB})}$$

Cas simplifié :

On peut appliquer ce modèle à la transmission de symboles de m bits. Dans ce cas, on considère que si le canal est dans l'état « G », les symboles émis seront reçus sans erreur. Si le canal est dans l'état « B », le bruit est prédominant. La probabilité que le symbole reçu soit le symbole envoyé (donc que chacun de ses bits soit reçu sans erreur) est d'autant plus faible que :

- la probabilité d'erreur binaire dans l'état B est haute,
- le nombre de bits constituant chaque symbole est élevé.

Dans le cas d'une transmission de symboles, on peut donc approximer les probabilités d'erreur symbole en choisissant $h = 0$ et $k = 1$ [YEE95]. Autrement dit dans l'état G les symboles seront reçus sans erreur ($P_G = 0$) alors que dans l'état B ils seront systématiquement perdus ($P_B = 1$). En conséquence, on peut exprimer les caractéristiques du canal sous forme de deux paramètres plus intuitifs :

- taux d'erreur symbole moyen : $P_e = P_B.P(B) + P_G.P(G) = P(B)$,
- longueur moyenne d'une salve d'erreur : $L_b = \frac{1}{P_{BG}}$.

On peut définir la matrice des probabilités de transition:

$$M = \begin{bmatrix} P_{GG} & P_{GB} \\ P_{BG} & P_{BB} \end{bmatrix} = \begin{bmatrix} 1 - P_{GB} & P_{GB} \\ P_{BG} & 1 - P_{BG} \end{bmatrix}.$$

On définit $\rho = 1 - P_{GB} - P_{BG}$, la corrélation entre deux symboles consécutifs. En utilisant les expressions de ρ et P_e , on peut écrire :

$$P_{BB} = P_e + (1 - P_e)\rho$$

$$P_{GG} = (1 - P_e) + P_e\rho$$

Et la matrice de probabilités de transition devient :

$$M = \begin{bmatrix} 1 - P_e(1 - \rho) & P_e(1 - \rho) \\ (1 - P_e)(1 - \rho) & 1 - (1 - P_e)(1 - \rho) \end{bmatrix}$$

Probabilité d'échec du décodage de Reed-Solomon

Le décodeur RS opère sur des symboles. Le taux d'erreur symbole d'entrée vu par le décodeur RS est le taux d'erreur paquet/slice. On s'intéresse à la probabilité que i paquets soient en erreur à la réception de n paquets, notée $P(i, n)$. La détermination de $P(m, n)$ [ELL63] [HOR99] est reportée ci-après.

On définit la probabilité d'occurrence d'un gap (intervalle entre un paquet en erreur et le paquet en erreur suivant, cf. Figure A.5) de longueur exacte v :

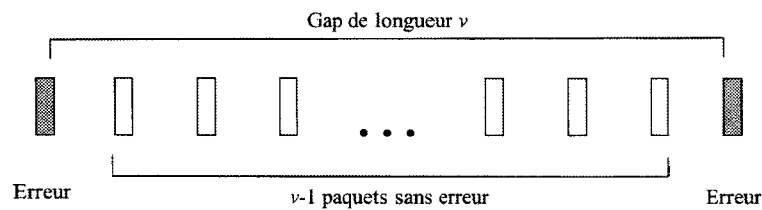


Figure A.5 : Illustration d'un gap de longueur v .

$$\begin{aligned}
 g(v) &= P(0^{v-1}/1) = 1 - P_{BG} & v = 1 \\
 &= P_{BG} \times (1 - P_{GB})^{v-2} P_{GB} & v > 1
 \end{aligned}$$

On définit la probabilité d'occurrence d'un gap de longueur supérieure à $v - 1$:

$$\begin{aligned}
 G(v) &= P(0^{v-1}/1) = 1 & v = 1 \\
 &= P_{BG} \times (1 - P_{GB})^{v-2} & v > 1
 \end{aligned}$$

On définit la probabilité de perdre $m - 1$ paquets parmi les $n - 1$ paquets suivants un paquet perdu:

$$\begin{aligned}
 R(m, n) &= G(n) & m = 1 \\
 &= \sum_{v=1}^{n-m+1} g(v) \times R(m-1, n-v) & 2 \leq m \leq n
 \end{aligned}$$

On définit la probabilité de perdre m paquets parmi les n paquets transmis:

$$\begin{aligned}
 P(m, n) &= 1 - \sum_{m=1}^n P(m, n) & m = 0 \\
 &= \sum_{v=1}^{n-m+1} P(B) \times G(v) \times R(m, n-v+1) & 1 \leq m \leq n
 \end{aligned}$$

Annexe III : Détermination de la distorsion moyenne d'une image du GOP

On cherche à évaluer l'effet de la propagation des erreurs sur la distorsion moyenne d'une image du GOP. Nous prendrons l'exemple du GOP le plus couramment utilisé (cf. Figure A.6).

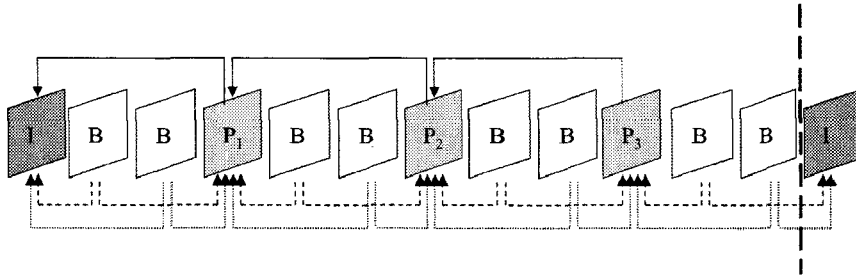


Figure A.6 : Illustration de la structure du GOP $N = 12$, $M = 3$.

Dans un premier temps, pour simplifier l'explication, on suppose par la suite qu'une slice encodée en mode inter ne contient aucun macrobloc codé en intra et que les blocs d'une image B sont tous interpolés.

On note :

D_{img}^I l'EQM de transmission d'une image d'une image codée en intra (image I) reconstruite,

D_{img}^P l'EQM de transmission d'une image résultant d'une prédiction (image P) reconstruite,

D_{img}^B l'EQM de transmission d'une image résultant d'une interpolation (image B) reconstruite.

D_{DFD}^P l'EQM de transmission de la partie transmise de l'image P (image différentielle),

D_{DFD}^B l'EQM de transmission de la partie transmise de l'image B (image différentielle).

Une image I est codée sans appel à une référence, son EQM peut s'écrire :

$$D_{img}^{I^0} = D_{img}^I$$

Une image P est codée par prédiction par rapport à une référence qui est soit une image I, soit une image P. Son EQM résulte des erreurs de transmission de l'image différentielle et de celles de son image de référence :

$$D_{img}^{P_1} = D_{DFD}^{P_1} + D_{img}^I$$

$$D_{img}^{P_2} = D_{DFD}^{P_2} + D_{img}^{P_1} = D_{DFD}^{P_2} + D_{DFD}^{P_1} + D_{img}^I$$

$$D_{img}^{P_3} = D_{DFD}^{P_3} + D_{img}^{P_2} = D_{DFD}^{P_3} + D_{DFD}^{P_2} + D_{DFD}^{P_1} + D_{img}^I$$

Les images B sont codées par interpolation de 2 images voisines qui sont de type I ou P. Leur EQM résulte des erreurs de transmission de l'image différentielle et de ses images de référence :

$$D_{img}^{B1} = D_{DFD}^{B1} + \frac{1}{2}(D_{img}^I + D_{img}^{P1}) = D_{DFD}^{B1} + \frac{1}{2}(2D_{img}^I + D_{DFD}^{P1})$$

$$D_{img}^{B2} = D_{DFD}^{B2} + \frac{1}{2}(D_{img}^{P1} + D_{img}^{P2}) = D_{DFD}^{B2} + \frac{1}{2}(2D_{img}^I + 2D_{DFD}^{P1} + D_{DFD}^{P2})$$

$$D_{img}^{B3} = D_{DFD}^{B3} + \frac{1}{2}(D_{img}^{P2} + D_{img}^{P3}) = D_{DFD}^{B3} + \frac{1}{2}(2D_{img}^I + 2D_{DFD}^{P1} + 2D_{DFD}^{P2} + D_{DFD}^{P3})$$

$$D_{img}^{B4} = D_{DFD}^{B4} + \frac{1}{2}(D_{img}^{P3} + D_{img}^I) = D_{DFD}^{B4} + \frac{1}{2}(2D_{img}^I + D_{DFD}^{P1} + D_{DFD}^{P2} + D_{DFD}^{P3})$$

L'EQM de transmission moyenne d'une image du GOP est donnée par l'expression suivante :

$$\begin{aligned} D_{img} &= \frac{1}{12}(D_{img}^I + 2D_{img}^{B1} + D_{img}^{P1} + 2D_{img}^{B2} + D_{img}^{P2} + 2D_{img}^{B3} + D_{img}^{P3} + 2D_{img}^{B4}) \\ &= \frac{1}{12}(12D_{img}^I + 9D_{DFD}^{P1} + 6D_{DFD}^{P2} + 3D_{DFD}^{P3} + 2D_{DFD}^{B1} + 2D_{DFD}^{B2} + 2D_{DFD}^{B3} + 2D_{DFD}^{B4}) \end{aligned}$$

Il est intéressant d'observer le nombre d'occurrence de l'EQM de chaque image. On peut affecter un facteur de pondération moyen à l'EQM moyenne calculée pour chaque type d'image.

Pour plus de clarté, on note :

nbI le nombre d'images I par GOP, $nbI = 1$,

nbB_c le nombre d'image B consécutives entre deux images P ou I et P : $nbB_c = M - nbI = M - 1$,

nbP le nombre d'image P dans un GOP : $nbP = \frac{N - nbI - nbB_c}{M} = \frac{N - M}{M}$,

nbB le nombre d'image B dans un GOP, $nbB = N - (nbP + nbI) = N - \frac{N}{M}$,

On peut écrire que :

- Les erreurs sur une image I affecteront toutes les images du GOP : poids $w_{I,th} = N$.
- Les erreurs sur une image B n'affecteront qu'elles-mêmes : poids $w_{B,th} = 1$.
- Les erreurs sur la $i^{ème}$ image P (cf. Figure A.7) affecteront :
 - Elle-même : poids 1
 - Les images B qui la précèdent immédiatement par prédiction bidirectionnelle (sens inverse) : poids $\frac{1}{2} \times nbB_c$
 - Les images B suivantes du reste du GOP par prédiction bidirectionnelle (sens direct) : poids $\frac{1}{2} \times (nbB - nbB_c \times i)$.

- Les images P suivantes par prédiction (sens direct) : poids $nbP - i$
- Les images B suivantes par prédiction bidirectionnelle (sens inverse) des images P suivantes contaminées (sauf les deux dernières du GOP prédites à partir de l'image I suivante) : poids $\frac{1}{2} \times (nbB - nbB_c \times (i + 1))$

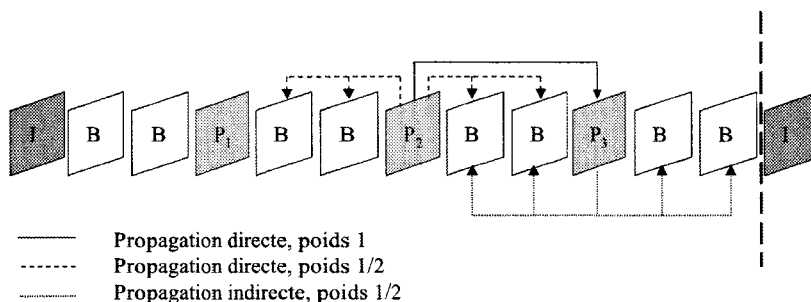


Figure A.7 : Illustration de la propagation d'une erreur sur une image P.

La somme de ces termes donne, pour la $i^{\text{ème}}$ image P un poids total :

$$w_{p^i} = 1 + nbP + nbB - i \times (1 + nbB_c)$$

A partir de la valeur moyenne de $i \in [1, nbP]$: $i_{\text{moy}} = \frac{1 + nbP}{2}$, on obtient le poids moyen de l'erreur sur une image P :

$$w_p = 1 + nbP + nbB - \frac{1 + nbP}{2} \times (1 + nbB_c)$$

que l'on peut exprimer ensuite en fonction de M et N :

$$w_p = 1 + \frac{N - M}{M} + N - \frac{N}{M} - \frac{M + N - M}{2M} \times (1 + M - 1) = \frac{N}{2}$$

En pratique, les macroblocs des images codées en inter peuvent être codés en intra. Le codeur effectue ainsi un rafraîchissement périodique des slices par insertion de macroblocs codés en intra dans les images codées en inter. On note :

- β_0 le taux de blocs non intra dans une image P,
- β_1 le taux de blocs non intra dans une image B.

NB : le cas théorique précédente supposait $\beta_0 = \beta_1 = 1$.

On peut, à partir des poids théoriques obtenus précédemment, dériver les valeurs suivantes :

- Les erreurs sur une image I affecteront :
 - Elles-mêmes

- Les autres images du GOP excepté les macroblocs codés en intra: poids $w_i = 1 + nbP \times \beta_0 + nbB \times \beta_1$.
- Les erreurs sur une image B n'affecteront qu'elles-mêmes : poids $w_B = 1$.
- Les erreurs sur la $i^{\text{ème}}$ image P affecteront :
 - Elle-même : poids 1
 - Les blocs non codés en intra des images B qui la précèdent immédiatement par prédiction bidirectionnelle (sens inverse) : poids $\frac{1}{2} \times nbB_c \times \beta_1$
 - Les blocs non intra des images B suivantes du reste du GOP par prédiction bidirectionnelle (sens direct) : poids $\frac{1}{2} \times (nbB - nbB_c \times i) \times \beta_1$
 - Les blocs non intra des images P suivantes par prédiction (sens direct) : poids $(nbP - i) \times \beta_0$
 - Les blocs non intra des images B suivantes par prédiction bidirectionnelle (sens inverse) des images P suivantes contaminées (sauf les deux dernières du GOP prédites à partir de l'image I suivante) : poids $\frac{1}{2} \times (nbB - nbB_c \times (i + 1)) \times \beta_0 \times \beta_1$

La somme de ces termes donne, pour la $i^{\text{ème}}$ image P un poids total :

$$w_{p_i} = 1 + \frac{nbB_c}{2} (1 - i) \times \beta_1 + \frac{nbB}{2} \times \beta_1 + (nbP - i) \times \beta_0 + \frac{(nbB - nbB_c(i + 1))}{2} \times \beta_0 \times \beta_1$$

On calcule la valeur moyenne de $i \in [1, nbP]$: $i_{moy} = \frac{1 + nbP}{2}$, et on l'injecte pour obtenir le poids moyen de l'erreur sur une image P, et on l'exprime ensuite en fonction de M et N :

$$w_p = 1 + \frac{nbB_c \times (1 - nbP) + 2nbB}{4} \times \beta_1 + \left(\frac{nbP - 1}{2}\right) \times \beta_0 + \frac{2 \times nbB - nbB_c \times (3 + nbP)}{4} \times \beta_0 \times \beta_1$$

$$w_p = 1 + \frac{(M - 1) \times (2M - N) + 2MN - 2N}{4M} \times \beta_1 + \frac{N - 2M}{2M} \times \beta_0 + \frac{2MN - 2N - (M - 1) \times (2M + N)}{4M} \times \beta_0 \times \beta_1$$

Annexe IV : Détail des séquences utilisées

Nous présentons dans cette annexe les séquences utilisées pour l'évaluation des performances du système développé.

- « 7_12 test »

concaténation de 7 séquences de 12 secondes à 30 images/sec

cif : 352 x 288

ntsc : 720 x 486

	Difficulté de codage	Psnr (cif@2Mbps)	Psnr (ntsc@5Mbps)
Conference	difficile	33,8 dB	34,1 dB
MobileCalendar	très difficile	23,1 dB	24 dB
OldBoat	facile	37,2 dB	35,3 dB
Football	moyen	35,2 dB	36,1 dB
Map	facile	38,5 dB	38,26 dB
Speaker	facile	42,1 dB	41,5 dB
Flowers	difficile	33,1 dB	32,9 dB
Moyenne		30,4 dB	31,1 dB

- « Fairytale »

5 séquences de 10 secondes à 25 images/sec

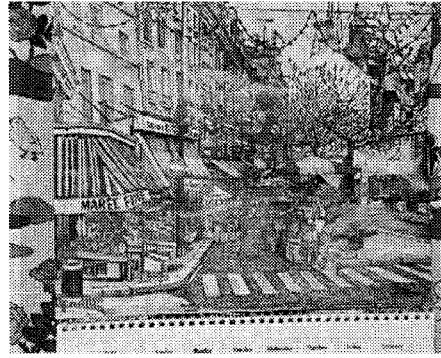
HD : 1920 x 1080 entrelacé

	Difficulté de codage	Psnr (18 Mbps)	Psnr (20Mbps)
CrowdRun	difficile	28,1 dB	28,61 dB
ParkJoy	difficile	25,9 dB	26,3 dB
DucksTakeOff	difficile	28,85 dB	29,26 dB
IntoTree	facile	33,86 dB	34,22 dB
OldTownCross	facile	36,04 dB	36,3 dB
Moyenne		27,9 dB	28,61 dB

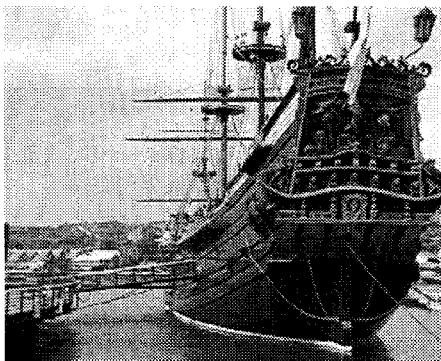
7_12 test



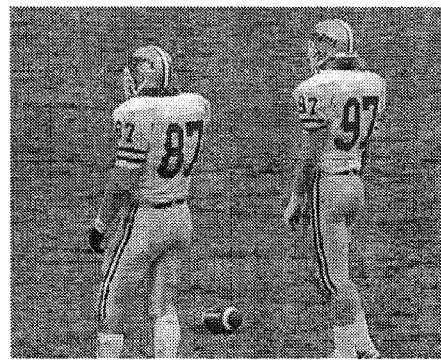
Conference



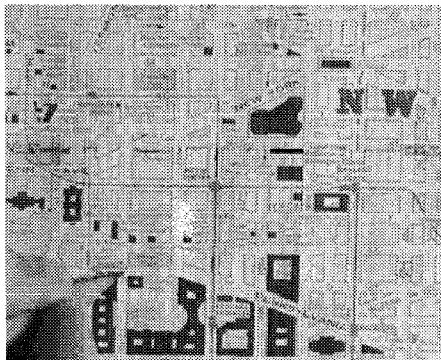
MobileCalendar



OldBoat



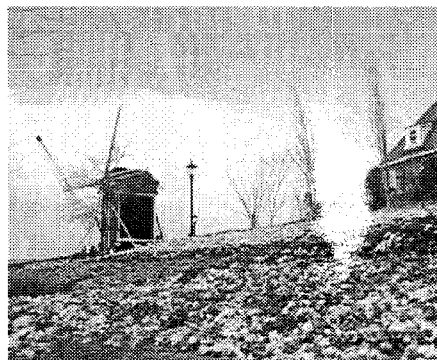
Football



Map



Speaker

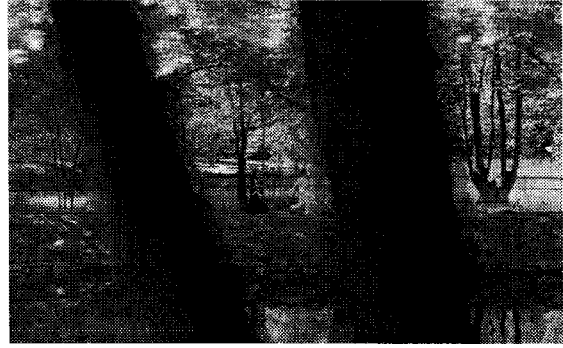


Flowers

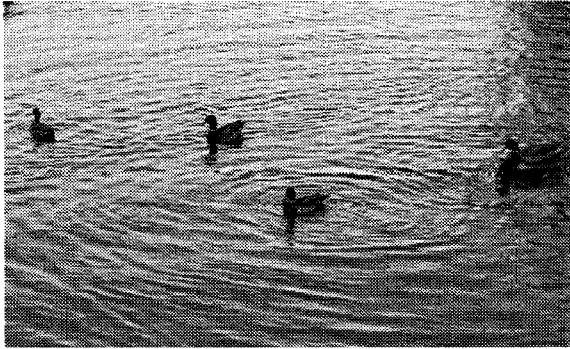
Fairytale



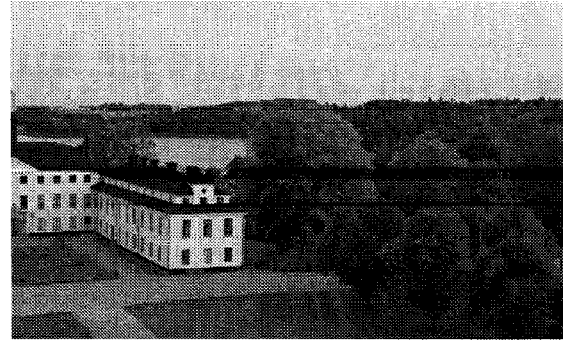
CrowdRun



ParkJoy



DucksTakeOff



IntoTree



OldTownCross

Annexe V : Résultats de simulations

Rappels :

Erreurs générées par un modèle de canal de Gilbert-Elliott

Séquence encodée à 2Mbps par un codeur MPEG2

10% du débit sont consacrés à la protection

paramètres RS :

- FEC (19,18), corrige jusqu'à 1 slice en erreur sur les 19 reçues,
- 8 :4 (20,18), corrige jusqu'à 2 slices en erreur sur les 20 reçues,
- 8 :2 (23,18), corrige jusqu'à 4 slices en erreur sur les 23 reçues,
- 8 :1 (26,18), corrige jusqu'à 8 slices en erreur sur les 26 reçues,

Paquets supposés de taille variable

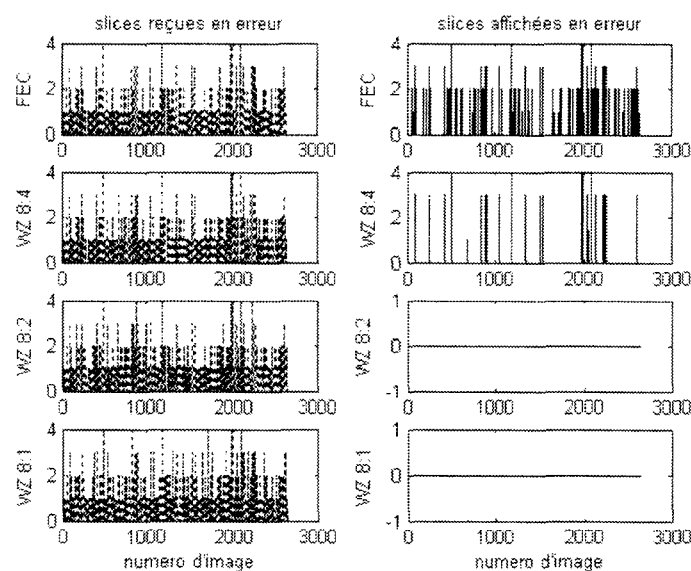
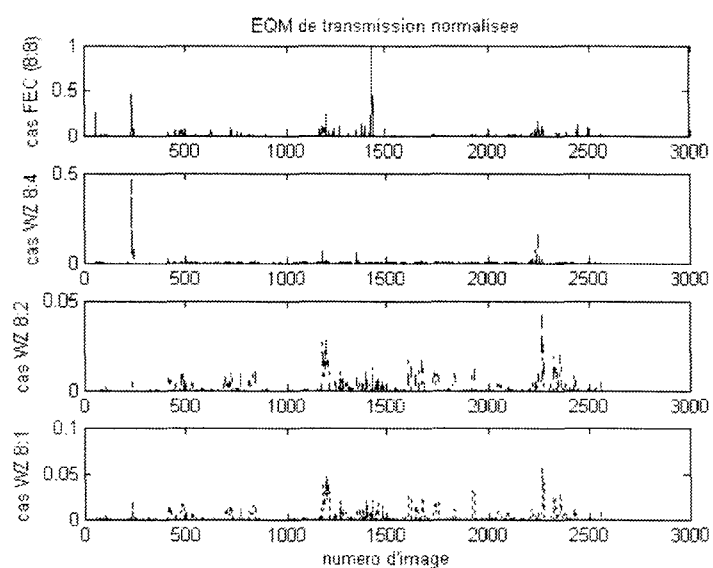
Une slice par paquet exactement

Dissimulation d'erreur incluse : remplacement de la slice indisponible par la slice correspondante de l'image de référence (I ou P) précédente.

La mesure de distorsion est l'EQM.

La variance moyenne de la séquence codée est de 1735. La variance de chaque image sert de référence à l'EQM normalisée.

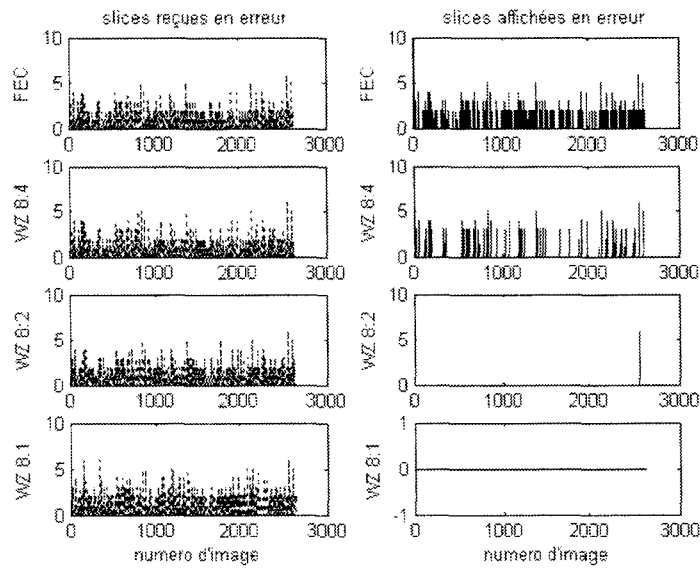
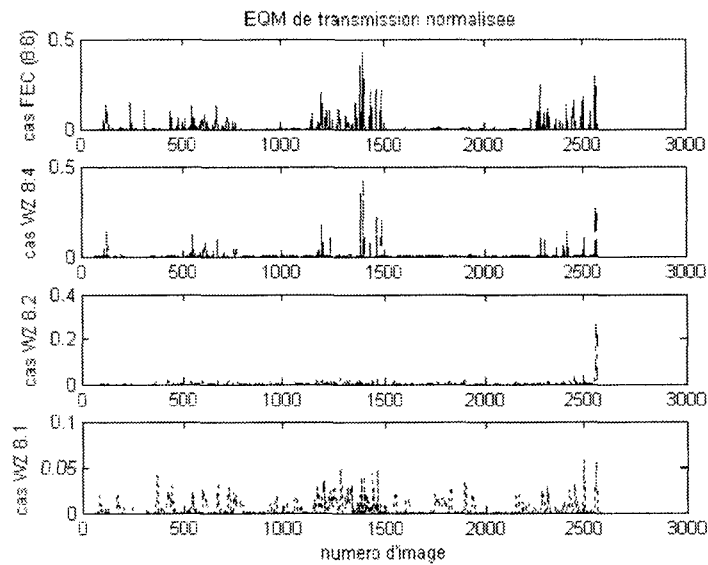
1% d'erreur



taux de slices affichées en erreur	mesures (%)	analyse (%)
WZ 8:8 (FEC)	4,34E-01	1,74E-01
WZ 8:4	1,47E-01	4,36E-02
WZ 8:2	0,00E+00	5,17E-04
WZ 8:1	0,00E+00	4,91E-06

eqm de transmission normalisée	mesures	analyse
WZ 8:8 (FEC)	7,22E-03	1,43E-03
WZ 8:4	1,88E-03	1,23E-03
WZ 8:2	1,38E-03	2,03E-03
WZ 8:1	2,33E-03	3,15E-03

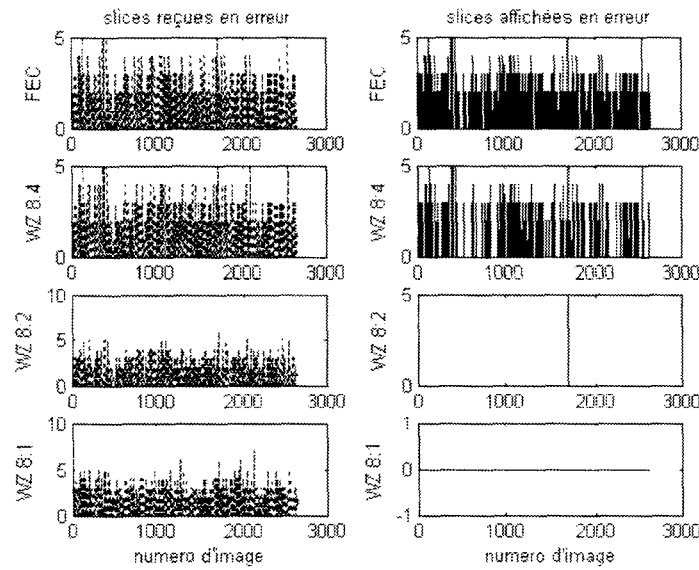
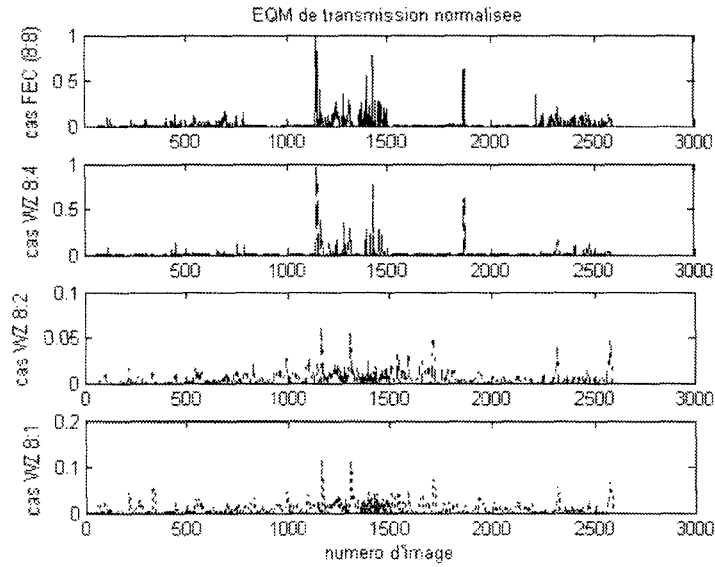
2% d'erreur



taux de slices affichées en erreur	mesures (%)	analyse (%)
WZ 8:8 (FEC)	1,02E+00	5,98E-01
WZ 8:4	3,98E-01	1,79E-01
WZ 8:2	1,26E-02	3,53E-03
WZ 8:1	0,00E+00	5,44E-05

eqm de transmission normalisée	mesures	analyse
WZ 8:8 (FEC)	7,71E-03	4,92E-03
WZ 8:4	3,58E-03	3,06E-03
WZ 8:2	3,11E-03	3,95E-03
WZ 8:1	4,80E-03	6,04E-03

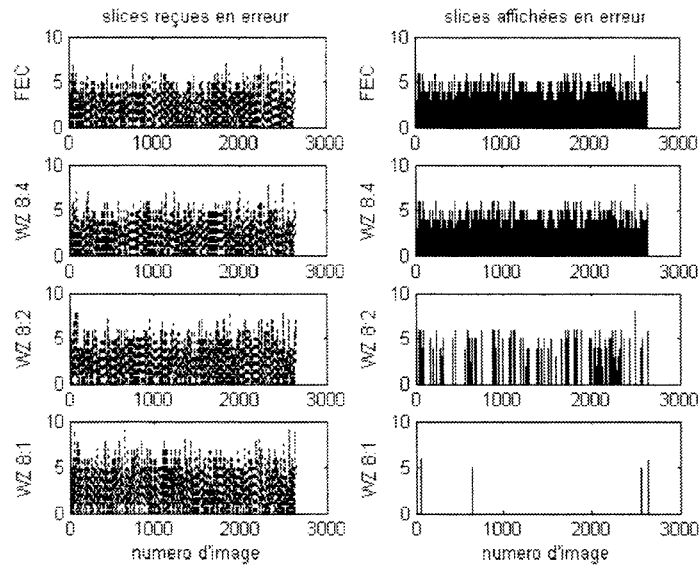
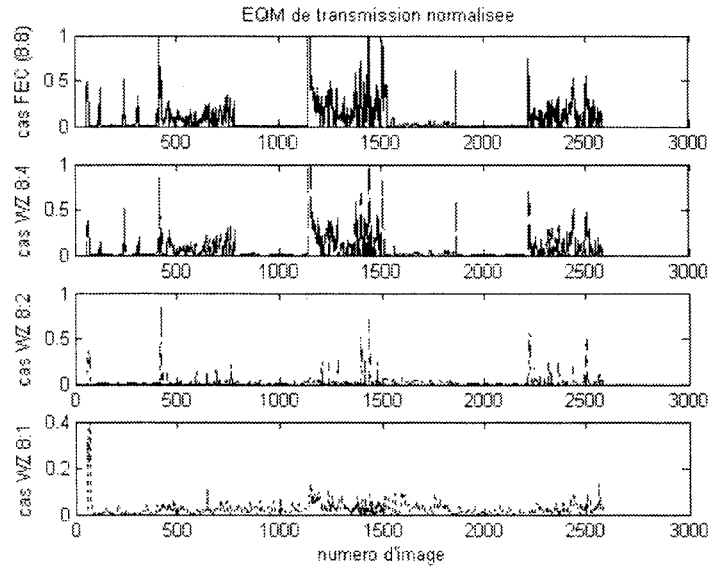
4% d'erreur



taux de slices affichées en erreur	mesures (%)	analyse (%)
WZ 8:8 (FEC)	2,17E+00	1,83E+00
WZ 8:4	9,74E-01	7,09E-01
WZ 8:2	1,05E-02	2,30E-02
WZ 8:1	0,00E+00	5,41E-04

eqm de transmission normalisée	mesures	analyse
WZ 8:8 (FEC)	2,15E-02	1,51E-02
WZ 8:4	1,32E-02	8,42E-03
WZ 8:2	5,20E-03	7,30E-03
WZ 8:1	9,36E-03	1,15E-02

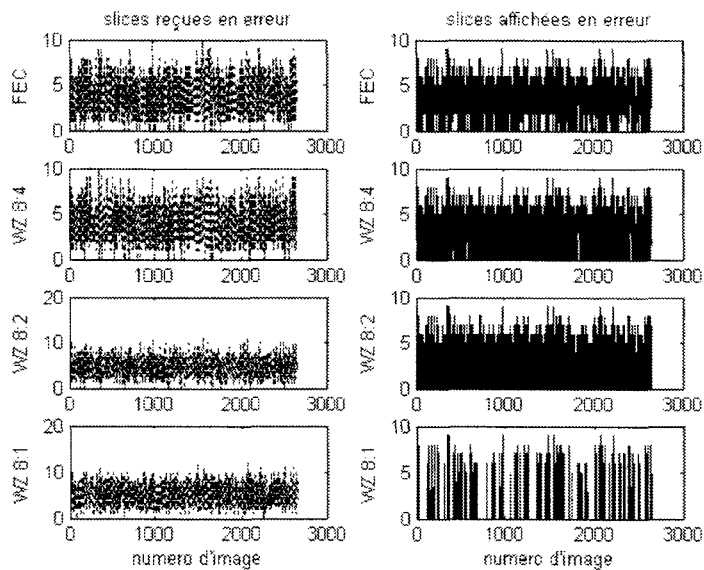
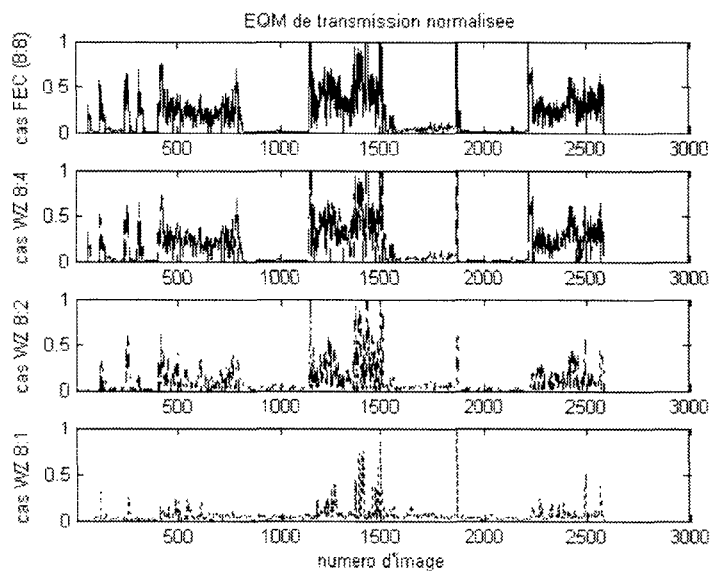
10% d'erreur



taux de slices affichées en erreur	mesures (%)	analyse (%)
WZ 8:8 (FEC)	8,67E+00	8,54E+00
WZ 8:4	5,99E+00	5,95E+00
WZ 8:2	8,97E-01	8,95E-01
WZ 8:1	4,63E-02	6,59E-02

egm de transmission normalisée	mesures	analyse
WZ 8:8 (FEC)	8,69E-02	7,02E-02
WZ 8:4	6,53E-02	5,23E-02
WZ 8:2	2,34E-02	2,45E-02
WZ 8:1	2,48E-02	3,03E-02

20% d'erreur



taux de slices affichées en erreur	mesures (%)	analyse (%)
WZ 8:8 (FEC)	1,97E+01	1,97E+01
WZ 8:4	1,85E+01	1,85E+01
WZ 8:2	9,25E+00	1,01E+01
WZ 8:1	1,87E+00	2,41E+00

eqm de transmission normalisée	mesures	analyse
WZ 8:8 (FEC)	1,93E-01	1,62E-01
WZ 8:4	1,81E-01	1,53E-01
WZ 8:2	9,32E-02	1,01E-01
WZ 8:1	5,65E-02	7,18E-02

Annexe VI : Résultats d'expériences

Rappels :

Erreurs issues de la campagne de mesure WCAM

Séquence encodée à 2Mbps par un codeur MPEG2

10% du débit sont consacrés à la protection

paramètres RS :

- FEC (19,18), corrige jusqu'à 1 slice en erreur sur les 19 reçues,
- 8 :4 (20,18), corrige jusqu'à 2 slices en erreur sur les 20 reçues,
- 8 :2 (23,18), corrige jusqu'à 4 slices en erreur sur les 23 reçues,
- 8 :1 (26,18), corrige jusqu'à 8 slices en erreur sur les 26 reçues,

paquets de taille fixée à 300 octets

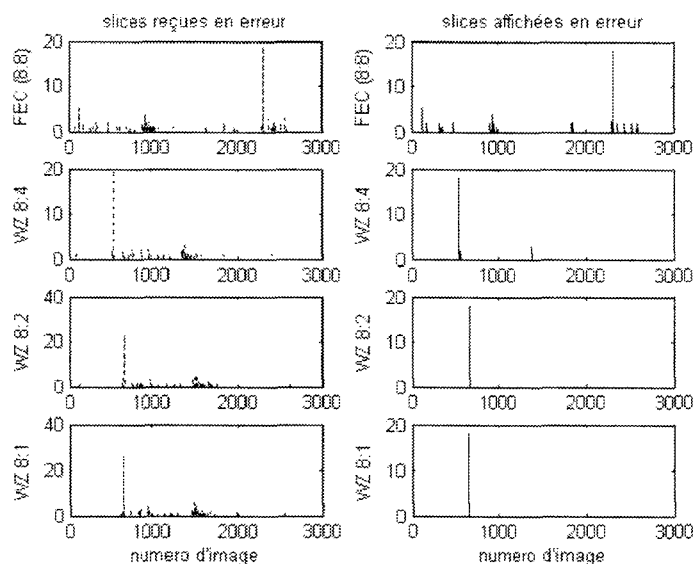
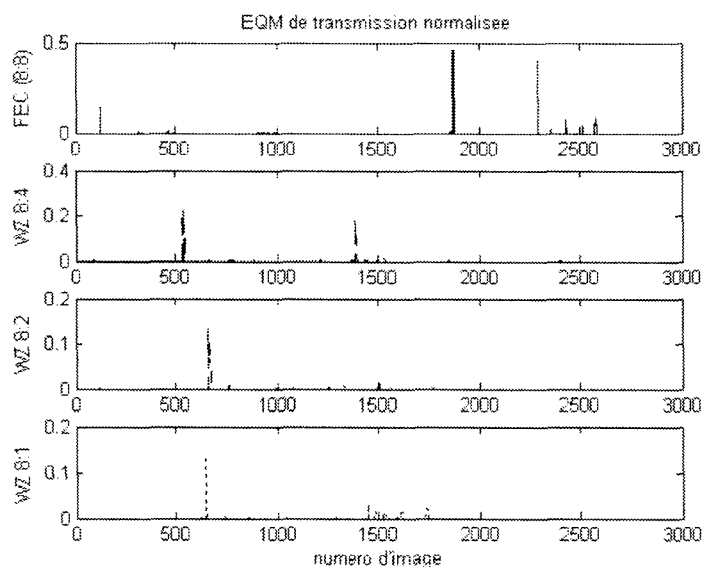
maximum 1 slice par paquet

Dissimulation d'erreur incluse : remplacement de la slice indisponible par la slice correspondante de l'image de référence (I ou P) précédente.

La mesure de distorsion est l'EQM.

La variance moyenne de la séquence codée est de 1735. La variance de chaque image sert de référence à l'EQM normalisée.

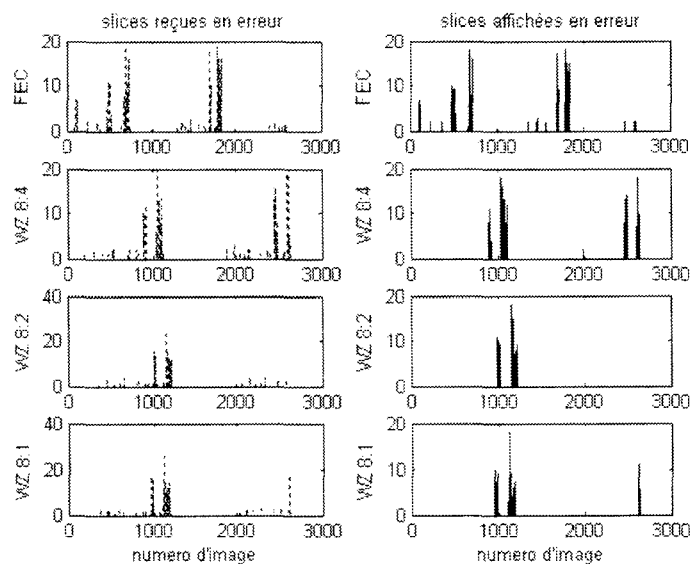
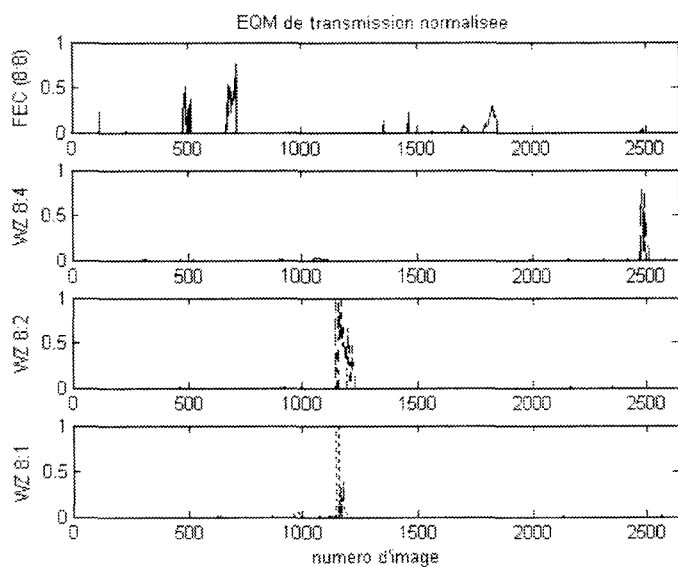
Data 7



taux de slices affichées en erreur	mesures (%)	analyse (%)
WZ 8:8 (FEC)	1,31E-01	7,23E-03
WZ 8:4	7,58E-02	1,88E-03
WZ 8:2	5,68E-02	1,55E-04
WZ 8:1	4,42E-02	7,63E-06

eqm de transmission normalisée	mesures	analyse
WZ 8:8 (FEC)	2,38E-03	5,94E-05
WZ 8:4	9,61E-04	2,60E-04
WZ 8:2	4,92E-04	3,17E-04
WZ 8:1	3,67E-04	4,06E-04

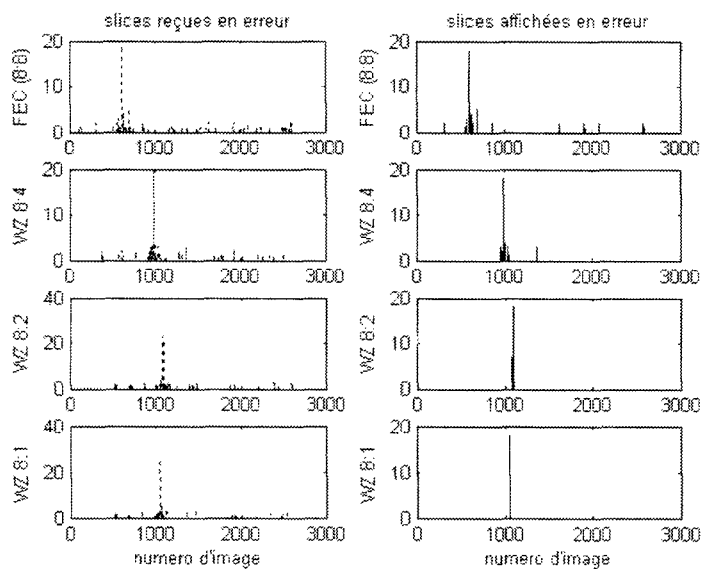
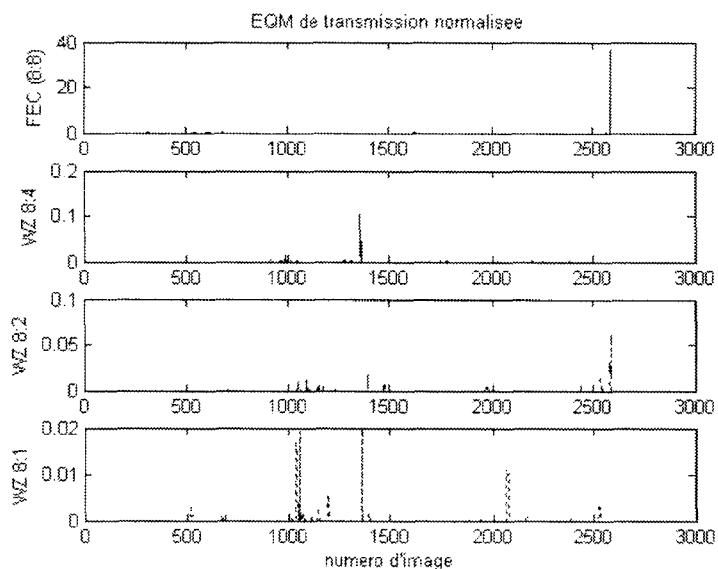
Data 8



taux de slices affichées en erreur	mesures (%)	analyse (%)
WZ 8:8 (FEC)	1,55E+00	2,48E-01
WZ 8:4	1,31E+00	1,05E-01
WZ 8:2	6,71E-01	1,08E-03
WZ 8:1	5,75E-01	2,05E-05

eqm de transmission normalisée	mesures	analyse
WZ 8:8 (FEC)	1,40E-02	2,04E-03
WZ 8:4	5,86E-03	2,00E-03
WZ 8:2	1,34E-02	1,61E-03
WZ 8:1	1,11E-02	2,81E-03

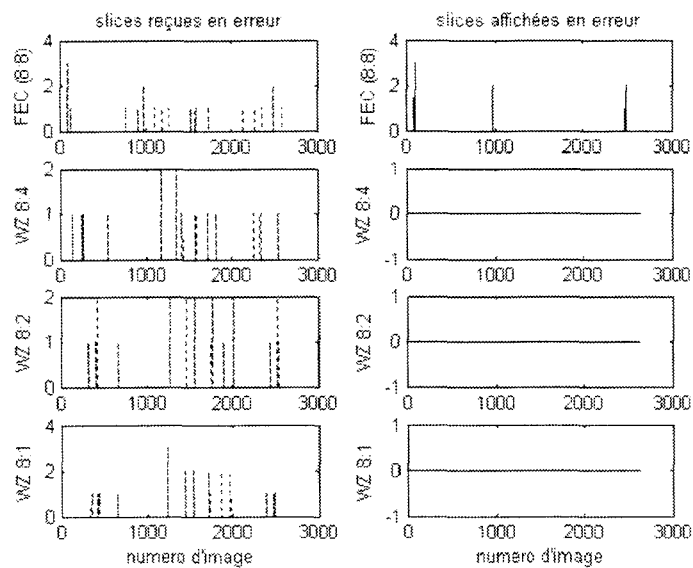
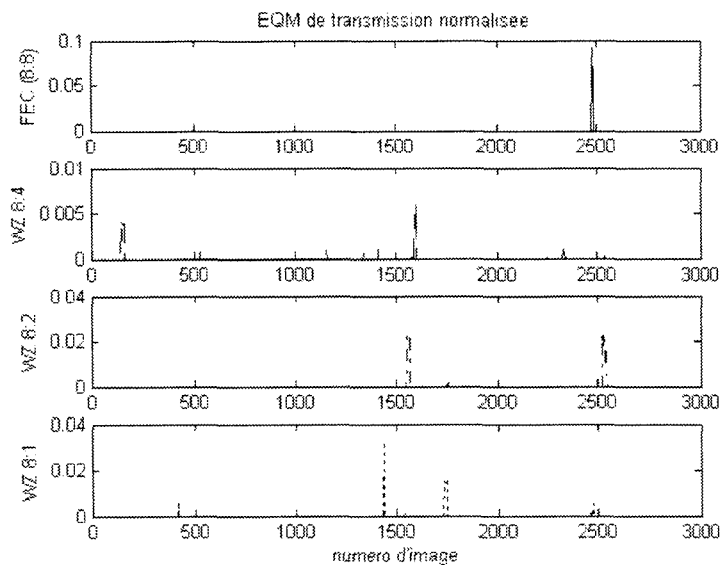
Data 9



taux de slices affichées en erreur	mesures (%)	analyse (%)
WZ 8:8 (FEC)	2,10E-01	1,18E-02
WZ 8:4	1,54E-01	2,92E-03
WZ 8:2	1,03E-01	3,97E-04
WZ 8:1	8,63E-02	9,67E-05

eqm de transmission normalisée	mesures	analyse
WZ 8:8 (FEC)	3,12E-02	9,67E-05
WZ 8:4	1,36E-04	2,25E-04
WZ 8:2	1,94E-04	3,97E-04
WZ 8:1	1,81E-04	6,56E-04

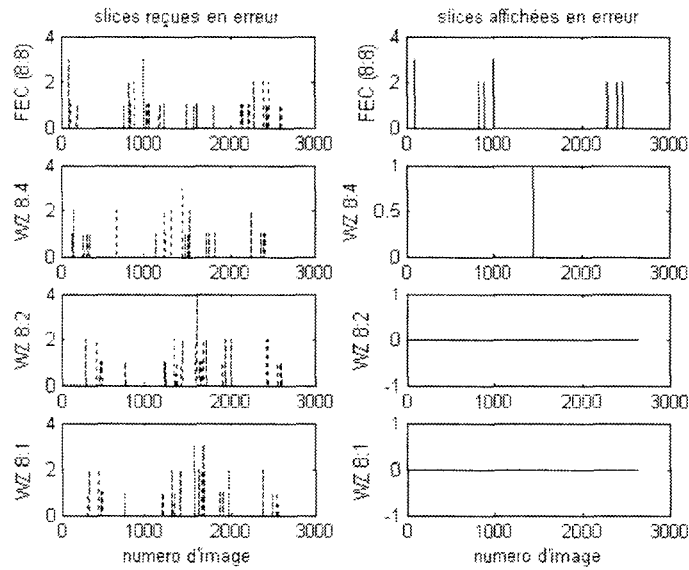
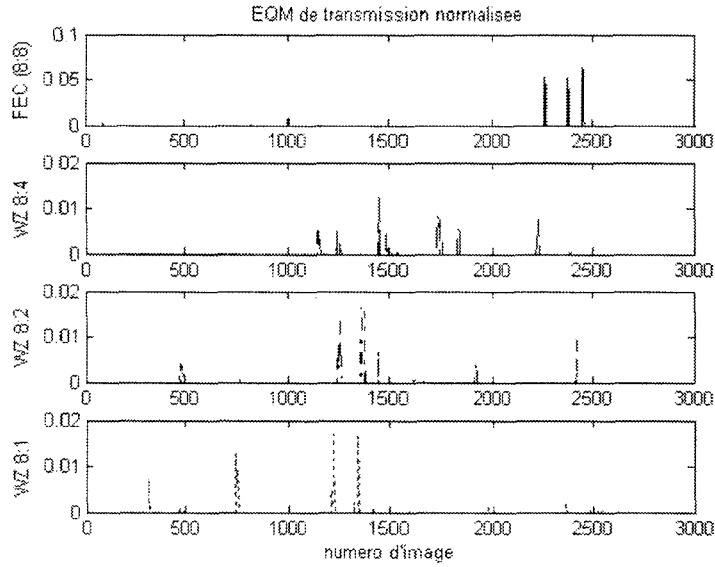
Data 10_1



taux de slices affichées en erreur	mesures (%)	analyse (%)
WZ 8:8 (FEC)	1,47E-02	5,89E-04
WZ 8:4	0,00E+00	1,81E-05
WZ 8:2	0,00E+00	4,49E-08
WZ 8:1	0,00E+00	6,00E-11

eqm de transmission normalisée	mesures	analyse
WZ 8:8 (FEC)	2,44E-04	4,84E-06
WZ 8:4	4,82E-05	3,72E-05
WZ 8:2	2,53E-04	6,71E-05
WZ 8:1	9,39E-05	1,12E-04

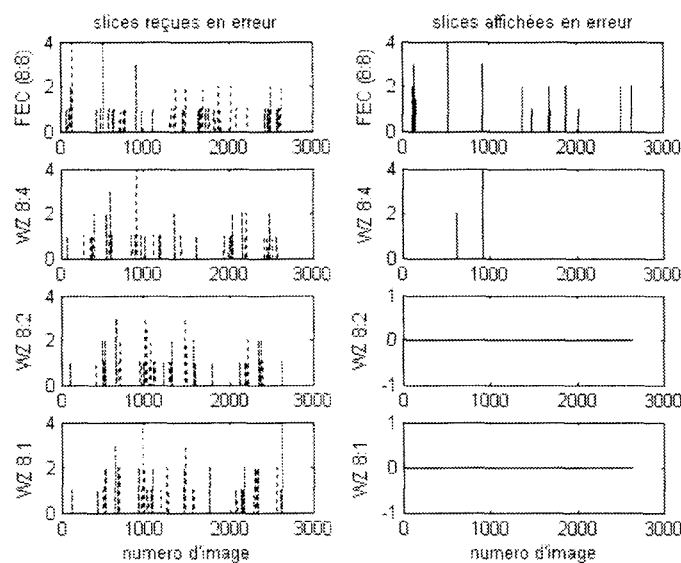
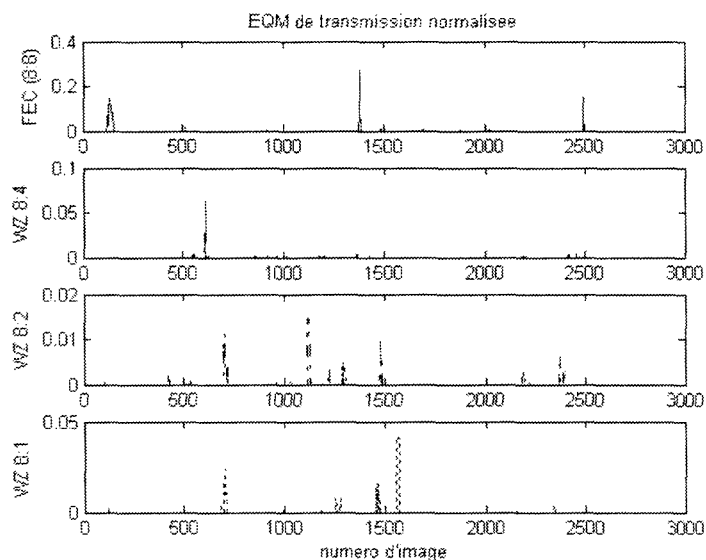
Data 10__2



taux de slices affichées en erreur	mesures (%)	analyse (%)
WZ 8:8 (FEC)	3,37E-02	4,69E-03
WZ 8:4	2,10E-03	2,46E-04
WZ 8:2	0,00E+00	5,23E-08
WZ 8:1	0,00E+00	9,83E-09

eqm de transmission normalisée	mesures	analyse
WZ 8:8 (FEC)	4,16E-04	3,85E-05
WZ 8:4	1,25E-04	1,04E-04
WZ 8:2	1,18E-04	1,62E-04
WZ 8:1	1,04E-04	2,81E-04

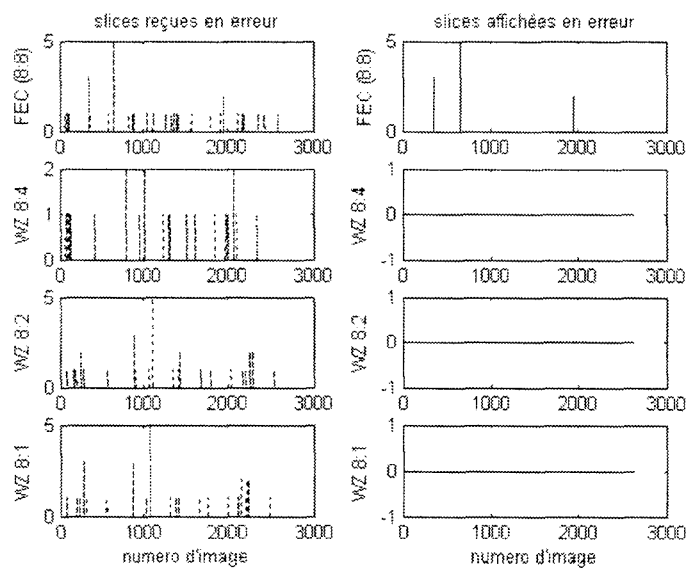
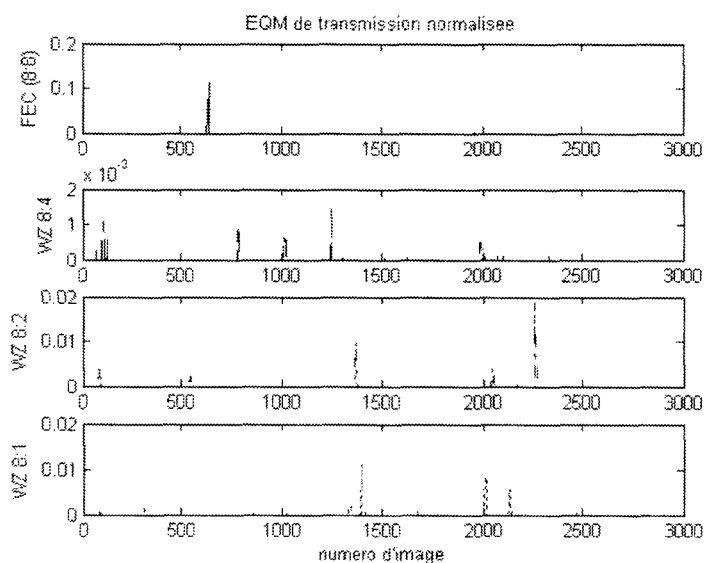
Data 11_1



taux de slices affichées en erreur	mesures (%)	analyse (%)
WZ 8:8 (FEC)	5,47E-02	1,06E-03
WZ 8:4	1,26E-02	3,32E-05
WZ 8:2	0,00E+00	1,12E-06
WZ 8:1	0,00E+00	4,47E-10

eqm de transmission normalisée	mesures	analyse
WZ 8:8 (FEC)	1,54E-03	8,68E-06
WZ 8:4	1,05E-04	3,73E-05
WZ 8:2	1,53E-04	7,50E-05
WZ 8:1	2,98E-04	1,31E-04

Data 11__2



taux de slices affichées en erreur	mesures (%)	analyse (%)
WZ 8:8 (FEC)	2,10E-02	1,33E-03
WZ 8:4	0,00E+00	5,21E-05
WZ 8:2	0,00E+00	9,20E-06
WZ 8:1	0,00E+00	3,83E-07

eqm de transmission normalisée	mesures	analyse
WZ 8:8 (FEC)	2,07E-04	2,83E-05
WZ 8:4	1,13E-05	6,10E-05
WZ 8:2	1,01E-04	2,61E-04
WZ 8:1	7,21E-05	1,40E-04

Liste des figures

Figure 1.1 : Schéma bloc d'un système de transmission typique.....	6
Figure 1.2 : Vecteurs de base de la DCT.....	10
Figure 1.3 : Illustration de l'effet de la quantification sur la qualité de l'image décodée et le taux de compression (CR).	10
Figure 1.4 : Schéma bloc d'un codeur MPEG.	11
Figure 1.5 : Effet de la propagation d'une erreur binaire liée à l'usage de techniques de compression sur une image fixe.....	13
Figure 1.6 : Illustration de la propagation d'une erreur liée à l'usage de techniques de prédiction temporelle.....	13
Figure 1.7 : Illustration du principe d'estimation de la valeur du symbole reçu.....	16
Figure 1.8 : Modes d'application du FEC : (a) aux données binaires et (b) à travers des paquets de données.....	17
Figure 1.9 : Insertion de marqueurs de synchronisation dans le flux MPEG2.	20
Figure 1.10 : Illustration du décodage des codes à longueur variable réversibles.	21
Figure 1.11: Réorganisation des bits du flux codé par EREC.	21
Figure 1.12 : Schéma bloc d'un codec à descriptions multiples.	23
Figure 1.13 : Illustration des zones de décision pour une 4/16 QAM pour a) les bits de haute priorité et b) les bits de basse priorité.	25
Figure 1.14 : Schéma bloc d'un codeur par couches avec transport hiérarchique.....	27
Figure 1.15 : Schéma bloc d'une transmission en mode data partitionning.	28
Figure 1.16 : Illustration de la dissimulation par interpolation spatiale à l'aide d'un filtre médian : image originale, corrompue et après dissimulation.	30
Figure 1.17 : Illustration de la récupération du mode de codage d'un bloc en MPEG2.	32
Figure 1.18 : Illustration du "Block Matching Algorithm".....	33
Figure 2.1 : Scénario de transcodage pour les architectures de réseau sans fil.....	37
Figure 2.2 : Schéma bloc du codage / décodage séparé de deux sources corrélées X et Y. ...	40
Figure 2.3 : Schéma bloc du codage / décodage conjoint de deux sources corrélées X et Y. ...	40
Figure 2.4 : Schéma bloc du codage séparé / décodage conjoint de deux sources corrélées X et Y.....	40
Figure 2.5 : Différents cas de codage de sources corrélées [SLE73].....	41
Figure 2.6 : Cas du codage de la source X avec information adjacente Y.....	42

Figure 2.7 : Représentation de l'ensemble des débits réalisables pour la représentation de deux sources corrélées X et Y d'après le théorème de Slepian-Wolf [SLE73].	43
Figure 2.8 : Codage de la source X avec information adjacente Y disponible uniquement au décodeur.	43
Figure 2.9 : Partitionnement des symboles en sous-ensembles.	46
Figure 2.10 : Décodage de la valeur de X grâce à l'information adjacente Y.	47
Figure 2.11 : Regroupement des symboles de la source en sous-ensembles.	47
Figure 2.12 : Analogie entre codage de canal (en haut) et codage de source avec information adjacente (en bas).	48
Figure 2.13 : Quantification scalaire emboîtée.	49
Figure 2.14 : Illustration du principe de codage / décodage de Wyner-Ziv, exemple simple d'un quantificateur "emboîté" scalaire à 4 sous-ensembles.	50
Figure 2.15 : Codeur intra images / décodeur inter images basé sur la séparation des images paires et impaires.	52
Figure 2.16 : Evaluation des performances du système proposé pour la séquence Foreman [AAR04].	53
Figure 2.17 : Diagramme en bloc de l'encodeur et du décodeur PRISM [PUR02].	54
Figure 2.18 : Comparaison de l'algorithme PRISM et du codec H. 263+ pour la séquence Football [PUR02].	54
Figure 2.19 : Codeur de Slepian-Wolf par couche.	56
Figure 2.20 : Codeur de Wyner-Ziv par couches [XUQ04].	57
Figure 2.21 : Comparaison des performances du codeur h. 26L FGS et du codeur WZ par couches, utilisant des flux H26L de différents débits comme couche de base / information adjacente.	57
Figure 2.22 : Observation de l'évolution du PSNR dans le cas d'1% de macroblocs en erreur dans le cas d'un codeur WZ par couches et d'un codeur h. 26L FGS.	58
Figure 2.23 : Encodeur vidéo résistant à la propagation temporelle des erreurs [SEH04].	60
Figure 2.24 : Comparaison du PSNR de l'approche proposée par [SEH04] avec la limite théorique et l'insertion d'images codées en intra.	60
Figure 2.25 : Illustration du codage source canal systématique.	61
Figure 2.26 : Schéma de protection systématique d'un flux vidéo grâce au paradigme de Wyner-Ziv [AAR03b].	62
Figure 2.27 : Codeur de Wyner-Ziv basé sur une quantification scalaire et des turbo-codes [AAR03b].	62
Figure 2.28 : Protection d'un flux vidéo par de multiples flux de Wyner-Ziv.	63
Figure 2.29 : Application des codes de Reed-Solomon à travers les slices d'une image.	64
Figure 2.30 : Schéma de protection systématique avec pertes d'un flux vidéo MPEG utilisant les codes de Reed-Solomon.	64
Figure 2.31 : Comparaison des performances du schéma pour une protection de qualité grossière, plus fine ou mixte [RAN05].	65

Figure 3.1 : Système proposé, protection contre les erreurs systématique basée sur une scalabilité spatiale.....	68
Figure 3.2 : Codeur (a) et décodeur (b) à scalabilité spatiale [WAN02b].	69
Figure 3.3 : Représentation des vecteurs de base de la DCT 1D.....	70
Figure 3.4 : Schéma de principe de la diminution de résolution spatiale par décimation des coefficients DCT [RAO91].....	71
Figure 3.5 : Réduction de la résolution spatiale d'une image par masquage fréquentiel	72
Figure 3.6 : Sérialisation des coefficients DCT après décimation.	72
Figure 3.7 : Schéma de protection avec décimation DCT.....	73
Figure 3.8 : Illustration de l'effet de la requantification (image extraite de la séquence Foreman).	75
Figure 3.9 : Illustration de l'effet d'une décimation DCT.....	76
Figure 3.10 : Canal de Gilbert.....	80
Figure 3.11 : Probabilités d'affichage des slices.....	81
Figure 3.12 : Illustration de la structure du GOP $N = 12$, $M = 3$	83
Figure 3.13 : Valeur de n optimale et de la longueur de slice de parité correspondante en fonction du taux de slices en erreur.....	87
Figure 3.14 : Distorsion optimale de la protection Wyner-Ziv en fonction du nombre de slices en erreur, exemple du flux vidéo au format CIF.....	89
Figure 4.1 : Illustration de la mise en paquets de taille fixe (300 octets) des slices image et parité pour un GOP de la séquence encodée.....	97
Figure 4.2 : Plateforme de test utilisée par le projet WCAM [CHI05].	99
Figure 4.3 : Résultats du schéma SLEP appliqué à des descriptions de résolutions spatiales variables donnés sous forme d'EQM de transmission moyenne, sans (en haut) et avec dissimulation d'erreur (en bas).	102
Figure 4.4 : Résultats du schéma SLEP appliqué à des descriptions de résolutions spatiales variables donnés sous forme de taux de slices affichées en erreur moyen.	103
Figure 4.5 : Variations de $P(m,23)$ pour diverses longueurs de salve d'erreur et un taux d'erreur paquets moyen de 10 et de 30%.	105
Figure 4.6 : Effet d'une augmentation de la longueur moyenne d'une salve d'erreur de 2 paquets (à gauche) à 5 paquets (à droite) sur la distorsion moyenne de la séquence décodée (dissimulation d'erreur incluse).....	106
Figure 4.7 : Effet d'une augmentation de la longueur moyenne d'une salve d'erreur de 2 paquets (à gauche) à 5 paquets (à droite) sur le taux de slices affichées en erreur.	106
Figure 4.8 : Comparaison de la distorsion moyenne de la séquence décodée pour des taux de protection de 10% (à gauche) et 20% (à droite).	107
Figure 4.9 : Comparaison du taux de slice affichées en erreur moyen de la séquence décodée pour des taux de protection de 10% (à gauche) et 20% (à droite).	108

Figure 4.10 : Nombre de slices reçues et affichées en erreur en fonction du numéro d'image pour la simulation d'une transmission sur un canal de Gilbert-Elliott de taux d'erreur moyen 1% et de salve d'erreur d'une longueur moyenne de 1,2 slices.	110
Figure 4.11 : EQM normalisée en fonction du numéro d'image associée à la simulation d'une transmission sur un canal de Gilbert-Elliott de taux d'erreur moyen 1% et de salve d'erreur d'une longueur moyenne de 1,2 slices.	111
Figure 4.12 : Illustration de la qualité visuelle associée aux schémas de protection pour une image P contenant 2 slices en erreur.	113
Figure 4.13 : Comparaison entre valeurs analytiques et résultats de simulations en termes de taux de slices affichées en erreur (à gauche) et d'EQM (à droite) pour les différents schémas de protection.	115
Figure 4.14 : Variation de l'EQM de transcodage et de dissimulation en fonction du numéro d'image.	116
Figure 4.15 : Taux d'erreur slice avant / après correction pour les différents schémas de protection et le motif de pertes référencé data 8.	119
Figure 4.16 : Distorsion globale liée à la transmission pour les différents schémas évalués.	120
Figure 4.17 : Résultats expérimentaux obtenus pour les motifs de pertes data 7 à data 9.	121
Figure 4.18 : Résultats expérimentaux et analytiques obtenus pour les motifs de pertes data 10 et data 11.	122
Figure A.1 : Architecture OSI.	131
Figure A.2 : Réseau de communication mixte.	132
Figure A.3 : Constitution du paquet IP par encapsulations successives des données [RFC98].	133
Figure A.4 : Canal de Gilbert.	135
Figure A.5 : Illustration d'un gap de longueur ν	137
Figure A.6 : Illustration de la structure du GOP $N = 12, M = 3$	139
Figure A.7 : Illustration de la propagation d'une erreur sur une image P.	141

Liste des tableaux

Tableau 1.1 : Caractéristiques des applications de communication vidéo principales [WAN02].	7
Tableau 1.2 : Quelques outils de codage robuste aux erreurs disponibles dans les standards actuels de compression vidéo.	19
Tableau 3.1 : Caractéristiques des images basses résolution obtenues par requantification (SNR) et masquage fréquentiel (spatial).	74
Tableau 3.2 : Réduction de débit moyenne associée à la décimation DCT pour une séquence au format CIF@2Mbps.	86
Tableau 3.3 : Valeurs moyennes des EQM, EQM normalisées par la variance de la séquence et PSNR pour le format vidéo CIF@2Mbps.....	88
Tableau 3.4 : Récapitulatif des tailles d'image à relier aux résolutions obtenues par décimation DCT.	90
Tableau 3.5 : Comparaison des réductions de débit obtenues après décimation pour les formats CIF et HD.	90
Tableau 3.6 : Bloc moyen après DCT et quantification pour divers formats et séquences.....	91
Tableau 4.1 : Longueurs de slice moyennes (en octets) du flux principal et de la parité selon le mode de codage de l'image et la version de résolution spatiale réduite utilisée.	96
Tableau 4.2 : Caractéristiques des motifs d'erreurs de transmission relevés lors de la campagne de mesure WCAM.....	100
Tableau 4.3 : Caractéristiques des erreurs slice à l'entrée du décodeur après application du motif d'erreur référencé data 8 pour les divers schémas.	118
Tableau 4.4 : Caractéristiques des erreurs slices associées au motif de pertes référencé data 8 pour les différents types d'image.....	118

Liste des tableaux

Références

- [AAR02] A. Aaron, B. Girod, « Compression with side information using turbo codes », Data Compression Conference (DCC'02), pp. 252-261, Snowbird, USA, Jan. 2002.
- [AAR02b] A. Aaron, B. Girod, « Wyner-Ziv coding of motion video », Asilomar Conference on Signals, Systems and Computers, Pacific Grove, USA, Nov. 2002.
- [AAR03] A. Aaron, E. Setton, B. Girod, « Towards practical Wyner-Ziv coding of video », Proc. IEEE International Conference on Image Processing (ICIP'03), Barcelona, Spain, Sep. 2003.
- [AAR03b] A. Aaron, S. Rane, D. Rebollo-Monedero, B. Girod, « Systematic lossy forward error protection for video waveforms », Proc. IEEE International Conference on Image Processing (ICIP'03), Barcelona, Spain, Sep. 2003, Invited Paper.
- [AAR04] A. Aaron, S. Rane, E. Setton, B. Girod, « Transform-domain Wyner-Ziv codec for video », Proc. Visual Communications and Image Processing, (VCIP'04), San Jose, CA, Jan. 2004.
- [AAR04b] A. Aaron, S. Rane, B. Girod, « Wyner-Ziv video coding with hash-based motion compensation at the receiver », Proc. IEEE International Conference on Image Processing, (ICIP'04), Singapore, Oct. 2004.
- [BAO96] J. Bao, H. Sun, T. C. Poon, « HDTV down conversion decoder », IEEE Trans. On Consumer electronics, Vol. 42, No. 3, pp. 402-410, Aug. 1996.
- [BAS99] A. Basso, G. L. Cash, M. R. Civanlar, « Real-time MPEG2 delivery based on RTP : implementation issues », Signal processing, Image Communication Vol. 15, pp. 165-178, 1999.

-
- [BHA96] V. Bhaskaran, K. Konstantinides, « Image and video compression standards – algorithms and architectures », Kluwer academic publishers, 1996.
- [CHE04] S. Cheng, Z. Xiong, « Successive refinement for the Wyner-Ziv problem and layered code design », Proc. Data Compression Conference (DCC'04), Snowbird, UT, Mar. 2004.
- [CHE95] C. T. Chen, « Error Detection and concealment with an unsupervised MPEG2 video decoder », Journal of visual communication and image representation, Vol. 6, No. 3, pp. 265-279, Sep. 1995.
- [CHI05] T. K. Chiew, P. Ferre, D. Agrafiotis, A. Molina, A . R. Nix, D. R. Bull, « Cross-layer WLAN measurement and link analysis for low latency error resilient wireless video transmission », Proc. of IEEE International Conference on Consumer Electronics (ICCE'05), Las Vegas, USA, Jan. 2005.
- [CIV00] M. R. Civanlar, « Compressed video over networks », Chap. on Internet video, Signal processing and communication series, Reibman and Amy R., Sep. 2000.
- [CIV01] M. R. Civanlar, G. L. Cash, « A practical system for MPEG2-based video on demand over ATM packet networks and the WWW », Signal processing : Image communications, Vol. 8, pp. 221-27, 1996.
- [CLA90] D.D. Clark, D.L. Tennenhouse, « Architectural considerations for a new generation of protocols », Proc. of the ACM Sigcomm'90, pp. 200-208, 1990.
- [ELL63] E. O. Elliott, « Estimates of error rates for codes on burst noise channels », Bell Syst. Tech. J. 42, pp. 1977-1997, Sep. 1963.
- [FAR99] N. Farber, K. Stuhlmuller, B. Girod, « Analysis of error propagation in hybrid video coding with application to error resilience », Proc. International Conference on Image Processing (ICIP'99), Vol. 2, pp. 550-554, 1999.
- [FER05] P. Ferre, D. Agrafiotis, T.K. Chiew, A. Doufexi, A. Nix and D. Bull « Packet Loss Modelling for H.264 Video Transmission over IEEE 802.11g Wireless LANs » Proc. of International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Montreux, Switzerland, Apr. 2005.

- [GIL60] N. Gilbert, « Capacity of a burst noise channel », Bell Syst. Tech. J. 39, pp. 1253-1265, Sep. 1960.
- [GIR05] B. Girod, A. Aaron, S. Rane, D. Rebollo-Monedero, « Distributed video coding », Proceedings of the IEEE, Special Issue on Video Coding and Delivery, Vol. 93, no. 1, pp. 71-83, Jan. 2005, Invited paper.
- [GIR99] B. Girod, K. Stuhlmüller, M. Link and U. Horn. « Packet Loss Resilient Internet Video Streaming ». SPIE Visual Communications and Image Processing (VCIP'99), January 1999, San Jose, CA.
- [GOU05] C. Goudemand : « Amélioration des transmissions vidéo MPEG-2 sur lignes ADSL basée sur une optimisation conjointe source/canal de la qualité vidéo », Thèse de doctorat, Université de Valenciennes et du Hainaut Cambresis, Dec. 2005.
- [HOR99] U. Horn, K. Stuhlmüller, M. Link, and B. Girod, « Robust Internet video transmission based on scalable coding and unequal error protection », Image Communication, vol. 15, no. 1-2, pp. 77-94, Sep. 1999.
- [HUH06] H.R. Hu, K.R. Rao, « Digital image quality and perceptual coding », CRC press book, Taylor and Francis group, 2006.
- [ISO04] ISO/IEC JTC1/SC29/WG11, « Report of Ad hoc Group on Scalable Video Coding », MPEG04/M10758, Seattle, July 2004.
- [ISO92] ITU-T, ISO/IEC JTC 1/SC 29, « Digital compression and coding of continuous-tone still images: requirements and guidelines », ITU-T Recommendation T.81, ISO/IEC 10918- 1 (JPEG) (Committee Draft), 1992.
- [ISO93] ISO/IEC JTC1/SC29/WG11, test model 5, Doc. N0400, Draft revision 2, Apr. 1993.
- [ISO94] ISO/IEC, 13818-2, « Information Technology—Generic Coding of Moving Pictures and Associated Audio, Part 2: Video », Recommendation ITU H.262, Nov. 1994.
- [ITU96] IUT-T Recommandations H.223 : « Multiplexing protocol for low bitrate multimedia communications », 1996.

-
- [ITU98] ITU recommendation H.263, « Video codec for low bitrate communication », Geneva, Switzerland, 1998.
- [JUN03] J. Jung, M. Antonini, M. Barlaud, « Optimal decoder for block-transform based video coders », IEEE Trans. on Multimedia, Vol. 5, pp. 145-160, Jun. 2003.
- [KAS01] A. K. Kastaggelos, Y. Eisenberg, F. Zhai, R. Berry, T. N. Pappas, « Advances in efficient resource allocation for packet based real-time video transmission », Proceedings of the IEEE, pp. 135-147, Vol. 93, No. 1, Jan. 2001.
- [LAM93] W. M. Lam, A. R. Reibman, B. Liu, « Recovery of lost or erroneous received motion vectors », Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'93), Vol. 5, pp. 417-420, 1993.
- [LIN83] S. Lin, D. J. Costello, « Error Control Coding : fundamentals and applications », Englewood cliffs, Prentice Hall, 1983.
- [LIN84] S. Lin and D. J. Costello Jr., « Automatic-repeat-request error-control schemes », IEEE Communication Mag., Vol. 22, pp. 5-17, Dec. 1984.
- [NET95] A. N. Netravali, B. G. Haskell, « Digital pictures: Representation, Compression and standards », 2nd edition, Plenum Press, 1995.
- [PER02] C. S. Perkins, L. Gharai, T. Lehman, and A. Mankin. « Experiments with delivery of HDTV over IP networks », Proc. of the 12th International Packet Video Workshop, Pittsburgh, Apr. 2002.
- [PRA00] S. S. Pradhan, K. Ramchandran, « Distributed source coding : symmetric rates and applications to sensor networks », Proc. of IEEE Data Compression Conference (DCC'2000), pp. 363-372, Snowbird, Utah, USA, Mar. 2000.
- [PRA01] S. S. Pradhan, R. Puri, K. Ramchandran, « (n,k) source-channel erasure codes: Can parity bits also refine quality? » Proc. Conf. Inform. Sciences and Syst., 2001.
- [PRA01b] S. S. Pradhan, K. Ramchandran, « Enhancing analog image transmission systems using digital side information: a new wavelet-based image coding paradigm », Proc. of IEEE Data Compression Conference (DCC'2001), pp.63-72, Los Alamitos, CA, USA.

-
- [PRA02] S. S. Pradhan, J. Kusuma, K. Ramchandran, « Distributed compression in a dense microsensor network », *IEEE Signal Processing Mag.*, pp. 51-60, Mar. 2002.
- [PRA03] S. S. Pradhan, J. Chou, K. Ramchandran, « Duality between source coding and channel coding and its extension to the side information case », *IEEE Trans. on Information Theory*, Vol. 49, No. 5, May 2003.
- [PRA99] S. S. Pradhan, K. Ramchandran, « Distributed source coding using syndromes (discus) : Design and construction », *Proc. of IEEE Data Compression Conference (DCC'99)*, pp. 158–167, Snowbird, USA, Mar. 1999.
- [PRO95] J. G. Proakis, *Digital Communications*, 3rd ed., New York, McGraw-Hill, 1995.
- [PUR02] R. Puri, K. Ramchandran, « PRISM: A new robust video coding architecture based on distributed compression principles », *Proc. Conference on Communication, Control and Computing*, Allerton, IL, Oct. 2002.
- [RAM05] M. Ramon, F.-X. Coudoux, M.-G. Gazalet, « Efficient priority encoding transmission scheme Based on Wyner-Ziv coding and spatial scalability », *Proc. SPIE Visual Communications and Image Processing (VCIP'05)*, Beijing, China, Jul. 2005.
- [RAM06] M. Ramon, F.-X. Coudoux, M.-G. Gazalet, M. Gharbi, P. Corlay, « Systematic Lossy Error Protection of Video based on Reduced Spatial Resolution », *International conference on electronics, circuits and Systems (ICECS'06)*, Nice, France, Dec. 2006.
- [RAN04] S. Rane, A. Aaron, B. Girod, « Systematic lossy forward error protection for error resilient digital video broadcasting », *Proc. SPIE Visual Communications and Image Processing (VCIP'04)*, San Jose, CA, Jan. 2004.
- [RAN04b] S. Rane, B. Girod, « Analysis of error-resilient video transmission based on systematic source-channel coding », *Proc. Picture Coding Symposium (PCS'04)*, San Francisco, CA, Dec. 2004.
- [RAN05] S. Rane, A. Aaron, B. Girod, « Error-resilient video transmission using multiple embedded Wyner-Ziv descriptions », *Proc. IEEE International Conference on Image Processing (ICIP'05)*, Genoa, Italy, Sep. 2005.

-
- [RAN05b] S. Rane, B. Girod, « Systematic lossy error protection versus layered coding with unequal error protection », Proc. SPIE Conference on Image and Video Communications and Processing (IVCP'05), San Jose, CA., Jan. 2005.
- [RAN06] S. Rane, B. Girod, « Systematic Lossy Error Protection based on H.264/AVC Redundant Slices », Proc. SPIE Visual Communications and Image Processing, (VCIP'06), San Jose, CA. Jan. 2006.
- [RAN06b] S. Rane, P. Baccichet, B. Girod, « Modeling and Optimization of a Systematic Lossy Error Protection System based on H.264/AVC Redundant Slices », Proc. Picture Coding Symposium (PCS'06), Beijing, China, Apr. 2006.
- [RAO91] K. R. Rao, P. Yip, « Discrete Cosine Transform, algorithms, advantages, applications », Academic press inc., Chap. 7 pp.143, 1991.
- [RED96] D. W. Redmill, N. G. Kingsbury, « The EREC : An error resilient technique for coding Variable length blocks of data », IEEE Trans. Image Processing, Vol. 5, No. 5, pp 565-574, Jul. 1996.
- [RFC96] IETF RFC 1889, H. Schulztrine, S. Casner, R. Frederick, V. Jacobson, « RTP : a transport protocol for real-time applications », Jan. 1996.
- [RFC98] D. Hoffman, G. Fernando, V. Goyal, M.R. Civanlar, « RTP payload format for MPEG1 / MPEG2 video », IETF RFC2250, Jan. 1998.
- [RFC02] IETF RFC 3453, « The use of forward error correction in reliable multicast », network working group, Dec. 2002.
- [RIB97] J. Ribas-Corbera, D.L. Neuhoff, « On the optimal block size for block-based, motion compensated video coders », SPIE Proc. of Visual Communications and Image Processing (VCIP'97), Vol. 3024, pp. 1132-1143, Feb. 1997.
- [RIV02] E. Rivier, « Communication audiovisuelle », Chap. 14, méthodes de codage de canal, Springer Ed. Nov. 2002.
- [RIZ97] L. Rizzo, « Effective erasure codes for reliable computer communication protocols », ACM SIGCOMM'97 Computer Communication Review, Vol. 27, No. 2, pp.24-36, Apr. 1997.

- [SCH00] H. Schulzrinne, « Compressed video over networks », Chap. 3, IP networks, Signal processing and communication series, Reibman and Amy R., Sep. 2000.
- [SEH03] A. Sehgal, N. Ahuja, « Robust predictive coding and the Wyner-Ziv problem », Proc. IEEE Data Compression Conference (DCC'03), Snowbird, UT, Mar. 2003, pp. 103-112.
- [SEH04] A. Sehgal, A. Jagmohan, N. Ahuja, « Wyner-Ziv coding of video: an error-resilient compression framework », IEEE Trans. on Multimedia, Vol. 6, No. 2, pp. 249-258, 2004.
- [SEH04b] A. Sehgal, A. Jagmohan, N. Ahuja, « Scalable video coding using Wyner-Ziv codes », Proc. of Picture Coding Symposium (PCS'04), San Francisco, California, USA, Dec. 2004.
- [SHA48] E. Shannon, « A mathematical theory of communication », Bell System Tech. J. 27, pp. 379-423, 623-656, 1948.
- [SKO98] A.N. Skodras, C.A. Christopoulos, « On the Down-Scaling of Compressed Pictures », Proc. of the IEEE International Workshop on Intelligent Signal Processing and Communication Systems (ISPACS'98), pp. 363-367, Melbourne, Australia, Nov. 4-7, 1998.
- [SLE73] D. Slepian, J. K. Wolf, « Noiseless coding of correlated information sources », IEEE Trans. on Information Theory, Vol. 19, pp. 471-480, Jul. 1973.
- [SMI94] B. C. Smith, « Implementation techniques for Continuous Media Systems and applications », PhD thesis, university of California, Berkeley, 1994.
- [STA03] V. Stankovic, R. Hamzaoui, Zixiang Xiong, «Product code error protection of packetized multimedia bitstreams », Proc. of International Conference on Image Communication (ICIP'03), pp. 77-80, Vol. 1, Sep. 2003.
- [STU00] K. Stuhmuller, N. Farber, M. Link, B. Girod, « Analysis of video transmissions over lossy channels », IEEE journal on selected areas in communications, Vol. 18, No. 6, pp. 996-1011, Jun. 2000.
- [SUJ00] J. K. Su, J. J. Eggers, B. Girod, « Illustration of the duality between channel coding and rate distortion with side information », Proc. 34th Asilomar Conf. Signals, Systems, & Computers, Pacific Grove, CA, USA, Oct. 29-Nov. 1 2000.

- [SUN92] H. Sun, K. Challapali, J. Zdepski, « Error concealment in digital simulcast AD-HDTV decoder », IEEE Trans. Consumer Electron., Vol. 38, pp. 108-117, Aug. 1992
- [SUN94] H. Sun, J. Zdepski, « Error concealment strategy for picture header loss in MPEG compressed video », SPIE conf. High Speed Networking and Multimedia Computing, pp. 142-52, San Jose, CA, Feb. 1994.
- [SUN95] H. Sun, W. Kwok, « Concealment of damaged blocks transform coded images using projection onto convex sets », IEEE Trans. Image processing, Vol. 4, pp. 470-477, Apr. 1995.
- [TAU02] D. S. Taubman, M. Marcellin « Jpeg2000: Image Compression Fundamentals, Standards, and Practice », Kluwer Academic Publishers, Jan. 2002.
- [VIL99] J. D. Villasenor, Y. Q. Zhang, J. Wen, « Robust Video coding algorithms and systems », Proceedings of the IEEE, Vol. 87, pp. 1724-33, Oct. 1999.
- [VIT71] A. J. Viterbi, « Convolutional Codes and Their Performance in Communication Systems », IEEE Trans. on Communications Technology, Vol. 19, No. 5, pp. 751-772, Oct. 1971.
- [VQE00] VQEG, video quality expert group, « Final report from the VQEG on the validation of objective models of video quality assessment », Mar. 2000, <ftp://ftp.its.bldrdoc.gov>.
- [VQE06] VQEG Multimedia Group TEST PLAN Draft Version 1.15 Sep. 28, 2006.
- [WAN02] Y. Wang, J. Ostermann, Y. Q. Zhang, « Video processing and communications», Chap. 14 : Error control in video communications, Prentice Hall, 2002.
- [WAN02b] Y. Wang, J. Ostermann, Y. Q. Zhang, « Video processing and communications », Chap. 11 : Scalable video coding, Prentice Hall, 2002.
- [WAN05] Y. Wang, A. R. Reibman, S. Lin, « Multiple Description Coding for video delivery », Proceedings of the IEEE, Vol. 93, No. 1, Jan. 2005.

- [WAN93] Y. Wang, Q. F. Zhu, L. Shaw , « Maximally Smooth image Recovery in transform coding », IEEE Trans. Communication, Vol. 41, pp. 1544-1551, Oct. 1993.
- [WAN98] Y. Wan, Q. F. Zhu, « Error control and concealment for video communication : a review », Proceedings of the IEEE, Vol. 86, No.5, May 1998.
- [WCA04] Site officiel du projet « Wireless Cameras and Audio-visual seaMless networking » <http://www.ist-wcam.org/>.
- [WEN98] J. Wen, J. Villasenor, « Reversible Variable Length Codes for robust image and video transmsission », Data Compression Conference (DCC'98), Snowbird, UT, pp. 471-80, Apr. 1998.
- [WUD00] D. Wu, Y. T. Hou, Y. Q. Zhang, « Transporting Real-Time Video over the Internet : challenges and approaches », Proceedings of the IEEE, Vol. 88, No. 12, pp. 1855-1875, Dec. 2000.
- [WYN76] A. D. Wyner, J. Ziv, «The rate-distortion function for source coding with side information at the decoder », IEEE Trans. on Information Theory, Vol. 22, pp.1-10, Jan. 1976.
- [XIO03] Z. Xiong, A. Liveris, S. Cheng, Z. Liu, « Nested Quantization and Slepian-Wolf Coding: A Wyner-Ziv Coding Paradigm for iid Sources », Proc. IEEE Workshop on Statistical Signal Processing, St. Louis, MO, 2003.
- [XIO04] Z. Xiong, A. Liveris, S. Cheng, « Distributed source coding for sensor networks », IEEE Signal Processing Magazine, Vol. 21, pp. 80-94, Sep. 2004.
- [XUQ04] Q. Xu, Z. Xiong, « Layered Wyner-Ziv video coding », in Visual Communications and Image Processing (VCIP'04), Vol. 5308 of Proceedings of SPIE, pp. 83-91, San Jose, California, USA, Jan. 2004.
- [XUQ04b] Q. Xu, « Layered Wyner-Ziv coding for noisy channels », A Thesis, Submitted to the Office of Graduate Studies of Texas A&M University, Aug. 2004.
- [XUQ06] Q. Xu, V. Stankovic, Z. Xiong, « Layered Wyner-Ziv video coding for transmission over unreliable channels », Signal Processing: Special Issue on Distributed Source Coding, Vol. 86, pp. 3212-3225, Apr. 2006.

-
- [YEE95] J. R. Yee, E. J. Weldon, « Evaluation of error correcting codes on a Gilbert channel », IEEE Trans. on Communications, Vol. 43, No. 8, pp.2316-2323, Aug. 1995.
- [ZAM02] R. Zamir, S. Shamai (Shitz), U. Erez, « Nested Linear /Lattice Codes for Structured Multiterminal Binning », IEEE Trans. on Information Theory, Vol. 48, pp. 1250-1276, Jun. 2002.
- [ZAM96] R. Zamir, « The rate loss in the Wyner-Ziv problem », IEEE Trans. on Information Theory, Vol. 42, No. 6, pp. 2073-2084, Nov. 1996.
- [ZHU00] Q. F. Zhu, Y. Wang, « Compressed Video Over Networks », Chap. 6, Error concealment for video communications, Signal processing and communication series, Reibman and Amy R., Sep. 2000.
- [ZHU05] X. Zhu, S. Rane, B. Girod, « Systematic lossy error protection for video transmission over wireless ad hoc networks », Proc. SPIE Visual Communications and Image Processing (VCIP'05), Beijing, China, Jul. 2005.

Schéma de protection systématique à scalabilité fréquentielle d'un flux vidéo compressé basé sur le codage de Wyner-Ziv

Les fournisseurs d'accès Internet proposent aujourd'hui des offres « quadruple play » (services Internet, TVIP éventuellement en Haute Définition, vidéo à la demande et services mobiles) qui reflètent la convergence actuelle des différents réseaux de diffusion. Les transmissions vidéo numériques doivent s'adapter à des ressources de plus en plus hétérogènes en termes de réseau et de récepteur. Typiquement, les flux vidéo sont transmis sous forme de paquets dont la perte a des conséquences extrêmement dommageables sur la qualité vidéo reconstruite. Il est donc nécessaire de développer des nouvelles solutions de transmission flexibles et résistantes aux erreurs.

Nous présentons dans le cadre de cette thèse un système original de protection systématique d'une transmission vidéo basé sur une scalabilité fréquentielle. La scalabilité fréquentielle permet d'accéder à une version basse résolution spatiale de la vidéo à moindre complexité. Elle présente également l'avantage de faciliter l'adaptation du format d'affichage de la séquence vidéo à l'hétérogénéité des récepteurs. Le système proposé adjoint une description basse résolution de la séquence, obtenue par codage de Wyner-Ziv, au flux principal. Au décodeur, ce flux est utilisé pour reconstruire sans erreur une version de résolution spatiale inférieure des zones de l'image corrompues. En diminuant la résolution spatiale du flux adjacent, on peut augmenter la capacité correctrice du système proposé sans augmenter le débit dédié à la protection. La qualité de la séquence vidéo diffusée évolue graduellement en fonction des conditions de transmission, sans nécessiter un codage par couches. De plus notre système est parfaitement compatible avec les systèmes de diffusion existants.

Un modèle analytique a été développé afin d'estimer la distorsion due à la transmission et les paramètres optimaux du système proposé (capacité correctrice et résolution protégée). Des motifs de pertes de paquets corrélées issus de simulations et de mesures réelles lors d'une transmission 802.11g (WiFi) ont été utilisés. Les performances de la protection ont été évaluées en termes de résistance aux erreurs et qualité vidéo reconstruite. Finalement, les résultats analytiques et expérimentaux du système mettent en évidence la supériorité de la solution proposée par rapport à une protection classique.

Mots clés : Protection inégale contre les erreurs, scalabilité fréquentielle, codage systématique, codage de Wyner-Ziv.

Systematic frequency scalability protection scheme of a compressed video stream based on Wyner-Ziv coding

Internet providers currently propose « quadruple play » services (Internet services, IPTV optionally in high definition, video on demand, mobile services) which illustrate nowadays convergence between different broadcasting networks. Digital video transmissions have to adapt increasingly heterogeneous resources concerning network and reception devices. Typically, video streams are formatted into packets which loss has extremely damaging consequences on reconstructed video quality. It is necessary to find a new flexible error resilient transmission solution.

We present an original systematic error protection scheme for video transmission based on frequency scalability. Frequency scalability allows to reduce a sequence spatial resolution with low complexity. It also eases the adaptation of the video sequence display format to heterogeneous receivers. In this scheme, a reduced spatial resolution description of the sequence obtained using Wyner-Ziv coding is transmitted jointly with the main video stream. At the decoder side, the side stream is used to ensure reconstruction of the low spatial resolution version, which is substituted to the corrupted main stream parts. By lowering the side stream resolution, a higher corrective capacity can be achieved at the same protection bit rate. The video sequence quality degrades gracefully with worsening error conditions, without requiring layered coding. Moreover the proposed system is fully backward compatible.

An analytical model has been developed to estimate the transmission distortion, which allowed us to derive the optimal system parameters (corrective capacity and protected resolution). Correlated loss patterns from simulations and 802.11g transmission measurements have been used. Performances of our scheme have been evaluated upon robustness to packet loss and reconstructed video quality. Finally, analytical and experimental results put in relief the superiority of the proposed solution compared to a classical protection.

Keywords : Unequal error protection, frequency scalability, Systematic coding, Wyner-Ziv coding

